THE EUROPEAN UNIVERSITY INSTITUTE

Department of Economics

ESTIMATION ERROR IN UNOBSERVED COMPONENT MODELS

Christophe PLANAS

Thesis submitted for assessment with a view of obtaining the Degree of Doctor of the European University Institute

Florence, January 1995

THE EUROPEAN UNIVERSITY INSTITUTE

Department of Economics

ESTIMATION ERROR IN UNOBSERVED COMPONENT MODELS

Christophe PLANAS

The thesis committee consists of:

Prof. Luc Bauwens, C.O.R.E.
Prof. Andrew Harvey, L.S.E.
Prof. Agustín Maravall, E.U.I. (Supervisor)
Prof. Grayham Mizon, E.U.I.
Prof. Alain Monfort, C.R.E.S.T.

Acknowledgements

I would like to thank my supervisor, Agustin Maravall, for the interest with which he has followed this work and for his very useful comments and suggestions. I would also like to express my gratitude to my family and to my girlfriend, Anna, for their constant support.

Contents

Chapter 1

Introduction

Economists often concentrate their interest on some particular underlying parts of economic variables. For instance, long-run paths or cyclical fluctuations are the subject of macroeconomists' attention. Such views fit into the Unobserved Components modelling framework, in which a time series is assumed to be made up of several components that are not directly observed. Corresponding to the common use of unobserved components as a tool for economic analysis, over these last twenty years a significant body of the statistical literature has been devoted to the study of Unobserved Component models. However, these models have encountered a problem in that they are not identified. Every analyst wishing to decompose a time series into unobserved components must make an arbitrary assumption about the components. This dissertation discusses the identification problem in Unobserved Components models.

In recent developments in economic research, unobserved components appear in a variety of problems. Firstly, these issues commonly arise when a variable which is supposed to play an important economic role is not directly observed. This concerns for example the modelling of monetary authority credibility (Weber (1992)); the estimation of the persistence of economic shocks (Cochrane (1988)); or the estimation of the underground economy (Aigner, Schneider and Ghosh (1988)). Unobserved components are also relevant when agents are believed to react differently to distinct movements in a variable in which they are interested. In particular, the permanent and the transitory evolutions of the same economic variables have been seen to play different roles in the Friedman (1957) Permanent Income Hypothesis or in Lucas' (1976) analysis of the agents' reactions to the changes in the price level. The Business Cycle literature focusses on some underlying movements in time series, in particular on the trend evolution and on the cyclical fluctuations of the economy (see for example Beveridge and Nelson (1981), Nelson and Plosser (1982), Watson (1986), and Harvey (1985), among others). Outside the arena of economic research, unobserved components play an important role in short-term policy making and in the monitoring of the economy. Typically, governments use seasonally adjusted series and, on occasion, trends, to set short-term macroeconomic policy. The standard decomposition involved here is into the seasonally adjusted series plus a seasonal component, where the seasonally adjusted series may be split into a trend and an irregular component.

The wide use of unobserved components leads to a practical need for estimators. Methods for estimating the components may be classified in two main groups: empirical methods using ad-hoc filters and model-based approaches. The first group began to be developed in the beginning of the century, and are at present widely used in applications. Two examples are the Hodrick-Prescott filter for detrending time series, and the X11 filter, or some variant thereof, for seasonally adjusting time series. These procedures allow for evolving components, in opposition to the regression approach which assumes deterministic components. However, a common feature of ad-hoc filters is that they do not properly take into account the stochastic properties of the series under analysis and, as a consequence, the use of these methods presents some inconveniences. For instance, the estimation error and the components forecasts are not available, and no rigorous diagnosis of the decomposition adequacy is provided. Model-based approaches which consider linear stochastic processes for the components can bring an answer to these problems. We shall consider a class of models which has become very popular in the statistical analysis of time series: the class of AutoRegressive Integrated Moving Average (ARIMA) models. These models, developed by Box and Jenkins (1970), have been shown to be a relatively flexible class, and also useful in applications. It is thus natural to adopt them to describe the behavior of the series and its components. Chapter 2 underlines this choice through the discussion of some historical considerations about the components and of the limitations displayed by ad hoc filters.

However, even if it has been found that it is convenient to parametrize the models for the components using ARIMA models, all the parameters cannot be identified. This identification problem is in fact general in the unobserved components analysis. In the vast majority of cases, statisticians have solved it using some *a priori* considerations for postulating "desirable" models. Chapter 3 formally discusses the identification problem in unobserved components models and presents the most popular identification criteria in use. It will be seen that the identification problem turns out to be simply a noise repartition problem between the components.

Chapter 4 presents the component estimation in the model-based approaches. Two estimation algorithms can be used: the Kalman Filter and the Wiener-Kolmogorov filter. Since it is better suited for analytical discussion, we shall use the Wiener-Kolmogorov filter. The derivation of the estimators will be detailed, and the stochastic properties of the estimators will be discussed. In particular, their similarities and their discrepancies with the original components models will be analyzed. Some new properties of the estimators concerning the structure of the estimators cross-covariances and their relationship with the estimation error will be derived.

Presenting the estimation errors and analysing the estimators' properties, Chapter 4 will point out the fact that both of them depend on the identifying assumptions. This is a crucial point which has to be underlined: any identifying assumption has some particular consequences for both the stochastic properties of the component estimators and also the precision of the estimation. Given that any identifying assumption is arbitrary and cannot be justified by any compelling reason, it seems sensible to look for the model for the components that can be estimated with the maximum precision. Every analyst would benefit by dealing with a signal that can be estimated accurately. The monitoring of the economy and short-term macroeconomic policy might also be improved if they were to be conducted signals less obscured by the estimation error. Chapter 5 discusses how the precision of the component estimation is related to the set of the identifying assumptions. Some functions of the noise repartition will be derived. These functions will appear to be simply second-order polynomials in the noise repartition, with coefficients easily available. They will be used in Chapter 6 to derive some simple rules for selecting the best estimated decomposition. It will be shown that the thus-obtained decompositions possess some other attractive properties. Some optimal properties of a popular identification criterion, the canonical decomposition, will arise. The main tool for this analysis is provided by the ARIMA models framework, but our results are also valid for any linear stochastic processes.

The discussion is illustrated with some important types of ARIMA models in Chapter 7, and extended in Chapter 8 to the components' growth rates. Finally, some actual time series are analysed in chapter 9.

Chapter 2

The Development of the ARIMA-Model-Based Approach for Unobserved Component Analysis: a Brief History

2.1 Introduction

In this second chapter, we review some of the most important steps in the process leading to the development of the ARIMA-model-based decomposition of time series. Within this framework, two different fundamental notions are brought together: the idea of unobserved components and a probabilistic theory based on parametric models. The idea that time series are made up of different unobserved components has a very long history; this will be summarized in section 1. A detailed review of the early developments can be found in Nerlove, Grether and Carvalho (1979). May be the first model-based approach to the components analysis was the regression approach, where a deterministic function for the component is assumed and estimated by least-squares. For example, a trend was modelled as a polynomial in time. However, the deterministic assumption was soon felt restrictive, and some more flexible procedures have been developed. Having discarded the possibility of deterministic behaviors, a prominent feature of the early analysis of the components is that it was conducted in the absence of an explicit and suitable model for the series under analysis. Thus the filters constructed for the estimation of the components were mostly ad hoc, based on empirical considerations. In this dissertation, we will refer to the component estimation procedures which do not explicitly assume a model for the component under analysis as "empirical methods". Being constructed independently of the stochastic properties of the series analyzed, empirical filters display some important limitations. The principle of empirical filters, their advantages and defects will be discussed in section 2.

The development of a probabilistic theory during the twentieth century supplied an answer to the criticisms of the empirical methods in use. An important step in time series modeling was the generalization of ARIMA models by Box and Jenkins (1970). These models are of interest because they provide a simple way to model seasonal time series with relatively few parameters. We devote section 3 to the description of these models. If they were initially designed for the statistical study of observed time series, the Box-Jenkins ARIMA models have provided the basis of a model-based approach for the analysis of the components. Characterizing the components in this way has allowed analysts to avoid the problems encountered with empirical methods. This is the approach that we will consider in this dissertation, and the general specification for ARIMA-model-based decomposition of time series will be presented in section 4. Emphasis will be put on the fact that if it is possible to represent the components in terms of ARIMA models, the decomposition of a time series into unobserved components is in general not unique.

2.2 The Idea of Unobserved Components

Nerlove and al. (1979) locates in astronomy, during the seventeenth and eighteenth centuries, the origin of the perception of a series of observations as the output of different processes. Famous mathematicians such as Laplace, Euler and Lagrange used this representation of series when studying the planetary orbits. The success of the idea of unobserved components in astronomy inspired some meteorologists like B. Ballot (1847) who started looking for periodicities in atmospheric events. B. Ballot is also considered as the first to have built a procedure for seasonal adjustment: a good description of his work is contained in Nerlove and al. (1979, p.354-360). Economics imported the idea of unobserved components during the nineteenth century. Among the precursors was J. Gilbart ((1852), quoted in Kuznets (1933)) who noted a seasonal pattern in the circulation of banknotes in the UK.

The first to explicitly state the unobserved components hypothesis was Persons (1919, 1923). According to Persons, time series are composed of:

- a long-term tendency or secular trend;
- a wavelike or cyclical movement;
- a seasonal movement;
- a residual variation.

Persons' publications were followed by an increase in the interest for unobserved components analysis. However, the statisticians soon faced several difficulties, mainly related to the characterization of the components. As Macaulay (1931) noted, trends and cycles were evolving over time and were "not necessarily representable throughout [their] length by any simple mathematical equation" (p. 38, quoted in Bell and Hillmer (1984)). Fitting specific trend functions such as a linear deterministic trend was thus excluded. Macauley (1931), Joy and Thomas (1928) among others used moving averages trend estimates. Such approaches preserve the time-varying nature of the trend: as noted in Kendall (1976), a moving average filter for trend estimation amounts to approximate the trend by local polynomials. In current usage in macroeconomics, a popular empirical filter for removing trend is the Hodrick-Prescott filter (1980).

Empirical filters were also used for the seasonal adjustment of time series. King (1924), for example, considered the median of successive sets of data points. There was a general agreement on the idea that the seasonal components of time series also change over time. Here again, the development of moving average filters was related to the difficulty of precisely characterizing the seasonal component. During the sixties a moving average filter, the X-11 procedure (see Shiskin and al. (1967)), became probably the most commonly used tool for the seasonal adjustment of economic time series. Data producing agencies and institutions related to policy making have applied it widely, and continue to use it intensively to this day, perhaps in its variant X-11 ARIMA (see Dagum (1980)). Given the popularity of empirical filters, we now turn to a discussion of their advantages and inconveniences. The important point is that the problems encountered with these filters arise from their ad hoc nature, i.e. from the absence of a statistical model for the components.

2.3 Empirical Moving Average Filters

A moving average filter of length $2M + 1$ can be expressed as:

$$C(B) = \sum_{i=-M}^{M} c_i B^i,$$

where throughout the dissertation B stands for the backshift operator such that for a time series z_t and any integer k, $B^k z_t = z_{t-k}$. Symmetric filters are often employed because they induce no phase-shift in the estimation of the components.

The principle of moving average filters can be easily understood in the frequency domain, where the interpretation can be drawn in terms of bandpass filters. For example, assume that the filter $C(B)$ is designed to detrend a time series. The trend component represents the long term variation of a time series, or in other words movements with low frequencies. A filter designed to eliminate a trend component will thus have a zero gain in the low frequency region, and a gain of 1 for the other frequencies. For seasonal adjustment purposes, the filters will cancel the variability associated with the seasonal frequencies simply by displaying a zero gain in these regions and a close to one gain in the other frequencies. Of course, the original series must be adequate, and effectively possess the features that the filter is supposed to remove.

Moving average filters have been developed precisely in order to present this bandpass structure. This class of filters are not model dependent, and in practice they are implemented independently of the stochastic properties of the series under analysis. This is very convenient when a statistical agency has to routinely adjust hundreds of time series, since the same filter can be applied to every series. The success of X11 for the seasonal adjustment of

- time series can be partly explained in this way. Clearly, the positive aspect of ad hoc filters is a practical consideration which is less relevant for academic research where emphasis is placed on methodological issues concerning just a few series.

Cleveland and Tiao (1976) and Burridge and Wallis (1984), have shown that the X-11 weights are the solution of a particular model, and in this way an underlying stochastic mechanism can be implicitly assumed as generating the filter. Hence, applying the same filter to each series considered can be seen as assuming that the same model applies to all series; this is excessively restrictive. It is however true that many series have stochastic structures that are relatively similar, and not too far away from models generating X-11 weights; this also explains the success of X-11.

A further problem of ad hoc filters is that they do not provide any statistics for checking the adequacy of the estimation results. At the extreme, one can extract with X-11 a seasonal component from a white noise variable (see Maravall (1993b)). Moreover, neither the estimation error on the component nor the component forecasts are available. Given the relevance that this knowledge may have for policy making, its absence constitutes a serious shortcoming of empirical filters.

Empirical filters have been developed partly because proper statistical models for analyzing time series were not satisfactory, or simply were not available. The components were believed to evolve over time and it was not clear how to specify explicit models to describe them. Without a proper statistical model for the series under analysis, the filters have been built in an ad hoc manner which has induced the limitations that we have described. To overcome them, in the last fifteen years model-based approaches considering linear stochastic processes have been developed. Such approaches were made possible by progresses in applied time series modelling that we briefly summarize now.

2.4 Linear Time Series Modelling

We review some of the most important steps in the development of applied time series analysis relevant to our work.

Yule (1927) first introduced the class of AutoRegressive (AR) processes. He fitted a low order process to the Wolfer's sunspot data by least squares methods. The Moving Averages (MA) processes were introduced by Slutsky (1937). Wold (1938) fitted MA models to data and established the important innovation representation for stationary processes, which is also known as the Wold representation theorem. The differencing of time series was used by Robb (1929) when dealing with the removal of the trend component.

Mann and Wald (1943) discussed inference in time series models. They derived the asymptotic theory for maximum likelihood estimators in AR models. Interested readers can find a good presentation of their work in Anderson (1971). Durbin (1960) and Walker (1962) among others considered mixed ARMA process when studying the properties of maximum likelihood estimates.

Whittle (1952) was probably the first to use high lags in time series models to account for seasonality. He fitted an 8-lag AR process to the Beveridge wheat price series (Beveridge (1921) and (1922)). For the six-month sunspot data, he selected a 22-lag AR model, but with the second to twenty-first lag coefficients set to zero. These representations may be seen as the first step toward seasonal time series modeling.

The contribution of Box and Jenkins (1970) was decisive for applied time series modelling. They set the general class of AutoRegressive Integrated Moving Average models, which consists of a mixed ARMA process for the series made stationary by differencing. For nonseasonal time series, these models are specified as:

$$\phi(B)\Delta^d z_t = \theta(B)a_t,$$

where $\phi(B)$ and $\theta(B)$ are polynomials satisfying the stationary and invertibility conditions respectively, $\Delta = 1 - B$ is the difference operator, d denotes the minimum number of differences required to render the series stationary and a_t is a white noise variable. If the polynomials $\phi(B)$ and $\theta(B)$ are respectively of order p and q, then z_t is said to follow an ARIMA(p,d,q) model.

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

$$\Phi(B^m)\Delta_m^D z_t = \Theta(B^m)\alpha_t,$$

where $\Phi(B^m)$ and $\Theta(B^m)$ are polynomials in B^m, of order respectively P and Q, which satisfy the stationarity and invertibility condition, and $\Delta_m = 1 - B^m$ is the seasonal differencing operator. This nonstationary operator has its roots $e^{i2k\pi/m}$, $k = 0, 1, \cdots, m-1$, evenly spaced on the unit circle. The parameter D represents the minimum number of differences required to make the series stationary. It is usually assumed that the relationship between the same month in successive years is common to all months, so the parameters of the polynomials $\Phi(B^m)$ and $\Theta(B^m)$ are constant.

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

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

It then comes out that the series z_t follow a multiplicative seasonal ARIMA model, specified as:

$$\phi(B)\Phi(B^m)\Delta^d \Delta_m^D z_t = \theta(B)\Theta(B^m)a_t.$$

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

Formulation of an ARIMA model requires first to select the appropriate order of the polynomials and the adequate order of differenciation. For the first choice, Box and Jenkins recommend the use of the autocorrelation and partial autocorrelation functions. The asymptotic theory for sample autocorrelations, derived by Bartlett (1946), serves as a statistical basis for conducting this choice. Unit root tests, as developed by Dickey and Fuller (1976) and Phillips (1987), help in selecting the adequate order of differenciation. Once a model is selected, then estimation of the parameters may be conducted by

Maximum Likelihood or some Least-Squares criterion. Several statistics may be used for checking the model adequacy, in particular the residual autocorrelations and the Box-Pierce-Ljung Q-statistics. A particular type of ARIMA model which has shown to be well adapted for many seasonal series is the $(0,1,1)(0,1,1)_m$ model, also called the "airline model".

The ARIMA models have provided the ground for a model-based analysis of Unobserved Components in time series. Cleveland and Tiao (1976) were, to our knowledge, among the first to consider this possibility, and they used ARIMA models in conjunction with signal extraction theory. That is the approach that will be adopted in this dissertation. Other approches are of course possible. In particular, while the ARIMA-model-based approach deduces the models for the components from the observed series model, the Structural Time Series (STS) (Engle (1978), Harvey and Todd (1983)) models proceed by directly specifying linear stochastic processes for the components. However, STS models are closely related to ARIMA models, and Maravall (1985) showed that they turn out to be a particular case of the ARIMA model specification. In that sense we can say that STS models are "encompassed" by ARIMA models. We believe it is better to deal with the most general linear model. Our choice insures that the analysis will be valid for any linear model.

We now present the general model specification, introducing the notation that will be used throughout the discussion, and we will see how ARIMA models may be used to characterize the components.

2.5 ARIMA-Model-Based Decomposition of Time Series

2.5.1 General Model Specification

The ARIMA-model-based approach assumes that an observed series x_t is made up of Unobserved Components, typically a signal s_t and a nonsignal n_t. The UC and the observed series are assumed to be related according to:

$$x_t = n_t + s_t, \tag{2.1}$$

where the additivity may be obtained after some suitable transformation of the observed series. For example, logs may be taken if the initial relationship is multiplicative. Possible applications of (2.1) will be discussed in the next subsection. Decompositions into more than two components are sometimes considered. For instance, a nonseasonal component n_t can be written as the sum of a trend and an irregular component. For our purposes it will be convenient to deal with the simplest two component model, one being of interest, the other one summing up all other components that may be present.

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

Assumption 1: The Unobserved Components are uncorrelated.

This assumption may appear to be somewhat restrictive, and in fact it is not required in order to obtain estimates of the UC (see Whittle (1963)). Some decompositions used in the literature consider correlated components (see for example Watson (1986)). Probably the most popular example of correlated components is given by the Beveridge-Nelson (1981) decomposition of I(1) series into a temporary and a permanent component. In this procedure both components turn out to be defined as linear combinations of the observed series x_t (see Maravall (1993b)). Since x_t is stochastic, the Beveridge-Nelson decomposition implicitly assumes that the components share the same innovation, which is a strong assumption.

Assuming instead independent components is a simplification which has some intuitive appeal. It is justified by the idea that the evolution of the different components is driven by separate forces. A typical illustration of the applicability of this representation is provided by the monetary control problem. This arises because central banks often rely on Seasonally Adjusted (SA) money demand estimators to take decision about money supply in the next period. In this case, the orthogonality hypothesis amounts to considering the seasonal and long-term evolution of the monetary aggregates as being driven by different causes: the long-term path would be related to the economic fundamentals, while the seasonal variations would be related to events such as holidays timing or the Christmas period. This seems reasonable and thus supports the use of seasonally adjusted series for policy making. In general, the orthogonality hypothesis is standard in practical applications such as the short-term monitoring of the economy.

The next assumptions concern the stochastic structures of the components.

Assumption 2:

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

$$\begin{aligned} \phi_n(B)n_t &= \theta_n(B)a_{nt}, \\ \phi_s(B)s_t &= \theta_s(B)a_{st}, \end{aligned} \tag{2.2}$$

where the variables a_{nt} and a_{st} are normally distributed white noise with variances V_n and V_s. The models are not reducible; that is each pair of polynomials $\{\phi_n(B), \theta_n(B)\}$, $\{\phi_s(B), \theta_s(B)\}$, are prime. Furthermore, the polynomials $\phi_n(B)$, $\phi_s(B)$, $\theta_n(B)$, and $\theta_s(B)$, of order respectively p_n, p_s, q_n, and q_s, may have their roots on or outside the unit circle, but $\phi_n(B)n_t$ and $\phi_s(B)s_t$ are required to be stationary. Also, as implied by assumption 1, the innovations a_{nt} and a_{st} are independent.

The specification of ARIMA models for Unobserved Components can be found in Cleveland and Tiao (1976), Box, Hillmer and Tiao (1981), Pierce (1978), Burman (1980) and Hillmer and Tiao (1982) for the early references.

A restriction must however be considered:

Assumption 3: The AR polynomials $\phi_n(B)$ and $\phi_s(B)$ do not share any common roots, and the MA polynomials $\theta_n(B)$ and $\theta_s(B)$ have no common unit root.

The first part of the assumption implies that the spectra of the UC do not have peaks at the same frequencies. Given that different components are associated with different spectral peaks, the first part of assumption 3 concerning the AR polynomials seems a reasonable feature of the decomposition. The second part of the assumption concerning the MA polynomials is required in order to obtain an invertible overall model, which insures a finite mean squared estimation error, as we shall see in chapter 4. From assumptions 1, 2, and 3, we obtain:

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

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

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

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

$$\theta_x(B)a_t = \phi_s(B)\theta_n(B)a_{nt} + \phi_n(B)\theta_s(B)a_{st}, \tag{2.3}$$

where a_t is a normally distributed white noise with variance V_a. Without loss of generality, we set $V_a = 1$ so that all other variances will be expressed as a fraction of V_a. Since the MA polynomials $\theta_n(B)$ and $\theta_s(B)$ are assumed not to share a common unit root, the MA process $\theta_x(B)a_t$ is invertible; this can be directly seen for (2.3). Equation (2.3) also implies that the order of the MA polynomial is constrained by:

$$q_x \leq max(p_n + q_s, p_s + q_n). \tag{2.4}$$

Equations (2.1), (2.2) and assumptions 1, 2, and 3 constitute a model which will be refered to as model (A). This is the general model that will be discussed in this dissertation. It includes more than the ARIMA-model-based: in fact it is valid for any linear process with normal innovations.

Since our discussion will focus on the characterization of the components, and since the model for observed series can be consistently esitmated, we shall retain the following assumption:

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

In other words, the polynomials $\phi_x(B)$, $\theta_x(B)$, and the innovation variance V_a are known. As briefly discussed in the previous section, the knowledge of the model followed by the observed series is reached after estimation using Box-Jenkins techniques. Since identification, estimation, and diagnostic checking of ARIMA models are now well established procedures, we shall not discuss this stage. Interested readers are refered to Box and Jenkins (1970).

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

$$A_i = V_i \frac{\theta_i(B)\theta_i(F)}{\phi_i(B)\phi_i(F)}, \tag{2.5}$$

where F is the forward operator defined as $F = B^{-1}$. Using the Fourier transform $B = e^{-iw}$ in (2.5), w denoting frequency in radians such that $w \in [-\pi, \pi]$, the equation above also defines the spectra $g_i(w)$. When one or both components are nonstationary, neither the spectra nor the ACGF of the nonstationary components and of the observed series are strictly defined: the presence of a unit root in a AR polynomial $\phi_i(B)$ implies an infinite peak in $g_i(w)$ and thus an infinite variance. However, the definitions of $g_i(w)$ and A_i provided in (2.5) may be extended to cover nonstationary cases, as in Hillmer and Tiao (1982), Bell and Hillmer (1984), and Harvey (1989), who refer to them as pseudo-spectrum and pseudo-ACGF. Since we do not make any assumptions about order of integration of the observed series, we will refer to the functions $g_i(w)$ and A_i, $i = x, n, s$, simply as the spectrum and the ACGF, whether the components are stationary or not in order to simplify the presentation.

2.5.2 Characterization of the Components: some examples

Model (A) is quite general. It embodies many possible applications, the most important of which are possibly detrending of time series, seasonal adjustment, cycle analysis and noise extraction. These applications involve the components discussed by Persons, namely the trend, the seasonal, the cycle and the irregular component. To give some concreteness to our discussion, we briefly present how in practice ARIMA models may be used to characterize these most common components.

Trend Component

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

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

where $0 \leq d \leq 3$, and $\psi_s(B) a_{st}$ is a low order ARMA processes. In the ARIMA-model-based approach, trends are often specified as IMA(2,2) models. Other model specifications used for example by Harrisson and Steven

(1976), by Harvey and Todd (1983), and by Ng and Young (1990), and which are commonly encountered in the STS approach, consider "second order" random walks processes such that:

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

where the drift is itself a random walk:

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

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

The above formulation may be easily interpreted as a stochastic extension of linear deterministic trends. Setting V_μ and V_s to zero, so that $\Delta \mu_t$ is constant, the corresponding deterministic trend function is trivially obtained as a quadratic polynomial of time if $a_{\mu t}$ has a non-zero mean, as a linear function of time otherwise. With the IMA(2,2) modelisation, this is equivalent to having roots of 1 in the MA polynomial. Allowing for stochasticity instead of deterministic expressions enables the trend component to evolve over time, which is an expected feature of the specification. Models for stochastic linear trends have been exhaustively discussed in Maravall (1993a).

Seasonal Component

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

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

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

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

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

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

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

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

Cyclical Component

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

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

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

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

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

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

Irregular Component

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

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

2.5.3 Admissible Decompositions.

Since the AR polynomials are identified directly from the factorization of $\phi_x(B)$, the unknowns of the model consist of the coefficients of the polynomials $\theta_s(B)$, $\theta_n(B)$, and the innovation variances V_s and V_n. The fundamental problem is that, by equations (2.1), (2.2), and assumptions 1-3, these parameters cannot be uniquely deduced: the models for the components are not identified. In model (A), information on the stochastic structure of the components is brought by the observed series and by the overall relationship (2.3). This information, however, is not enough to uniquely determine the models for the components: the equation (2.3) implies a system of $\max(p_s+q_n, p_n+q_s)$

covariance equations while the number of unknowns is $q_n + q_s + 2$. So when $\max(p_s + q_n, p_n + q_n) < q_n + q_s + 2$, and in the absence of an additional assumption, there exists an infinite number of ways to decompose the series x_t. Any decomposition consistent with the overall model for the observed series and insuring non negative components spectra will be called an 'admissible decomposition'. All admissible decompositions are of course observationally equivalent.

We illustrate this important point of underidentification of UC models with the following example:

Example: Trend plus Cycle decomposition.

Consider the following decomposition:

$$\begin{aligned} x_t &= n_t + s_t \quad \text{with}, \\ \Delta s_t &= a_{st}, \\ (1 - \phi B) n_t &= (1 + \theta_n B) a_{nt}, \end{aligned} \tag{2.6}$$

where we constrain the parameter ϕ to be negative and such that $|\phi| > \theta_n$. The assumptions 1-3 are supposed to hold. Equations (2.6) represents a simple model designed to decompose a time series into a trend (s_t) and a cycle component (n_t). The trend is modeled as a random walk and the cycle as a stationary ARMA(1,1), with period 2. It is a particular case of the model used in Stock and Watson (1993) to analyze the business cycle and to forecast recessions. This specification implies that the observed series x_t follows an ARIMA(1,1,2) model specified as:

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

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

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

from where we obtain the following system of covariance equations:

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

Taking $\phi=-.7$, $\theta_n=.3$, $V_n=.129$, and $V_s=5V_n=.645$, the model for the observed series is then obtained as:

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

The series is sinulated on figure 2.1 and the components s_t and n_t on figure 2.2. On subsequent figures, this model will be refered to as "model 1".

Now we change the specification of the components' models and take:

$$\begin{aligned}
\Delta s_t &= (1+.172B)a_{st} \qquad V_s=.470 \\
(1+.7B)n_t &= (1+.434B)a_{nt} \qquad V_n=.220.
\end{aligned}$$

Figure 2.1: Observed Series in Trend plus Cycle example

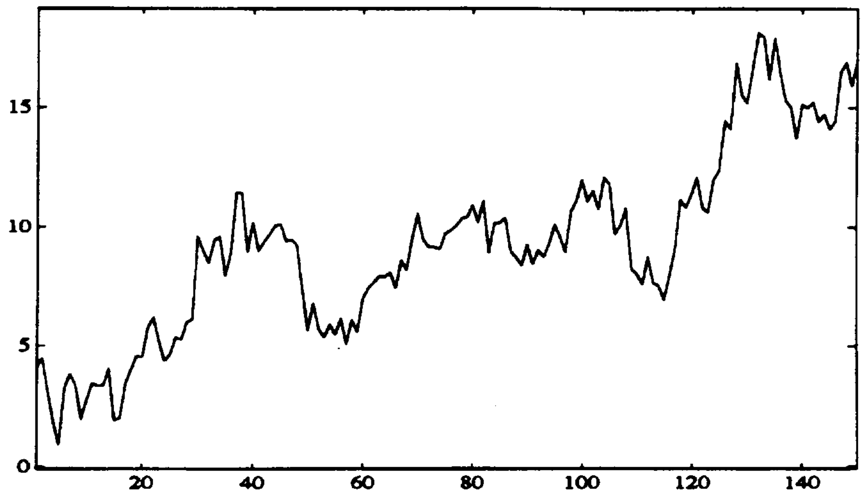
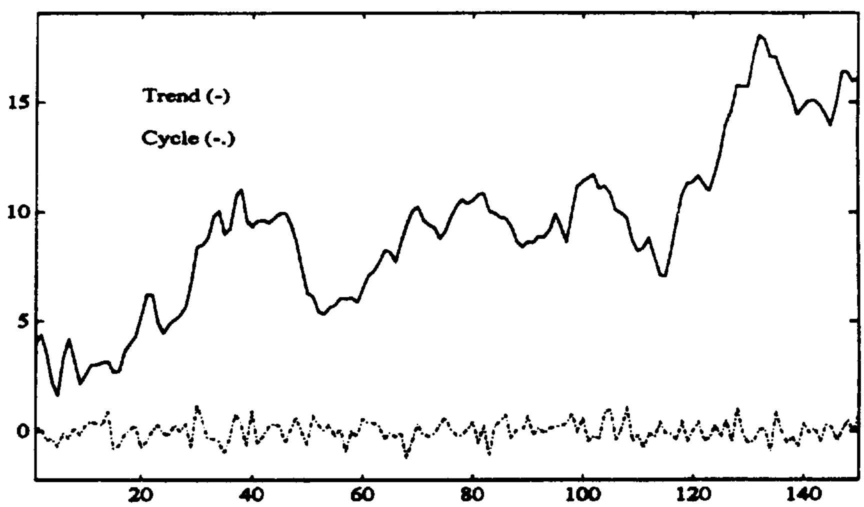


Figure 2.2: Components in Trend plus Cycle example



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

$$
\begin{aligned}
1+\theta_1^2+\theta_2^2 &= (1+.872^2+.120^2)V_s+(1+.576^2+.434^2)V_n \\
-\theta_1(1-\theta_2) &= .977V_s-.576^2V_n \\
-\theta_2 &= .120V_s-.434V_n.
\end{aligned}
$$

It can easily be checked that for these values for θ_s, V_s, θ_n, V_n, the MA parameters θ_1 and θ_2 remain unchanged. This new decomposition, referred to as "model 2", is thus consistent with the overall model that (2.6) had generated. Two model specifications have generated the same observed series. They correspond thus to two admissible decompositions.

What is the difference between the two decompositions ? The spectra of the components for these two decompositions are plotted on figure 2.3 and 2.4. It can be seen that for each component, the spectra obtained from the two model specifications differ only by a constant. This constant can be interpreted as the size of an orthogonal white noise which has been interchanged between the two components. To isolate it, it is convenient to look at the spectra minima. For the first model, the trend spectrum has a minimum at the π frequency equals to: $g_s(\pi)=V_s/4=.161$. In the second case, this minimum becomes: $V_s/8=.081$. Therefore, a white noise variable of variance $V_s/8$ has been removed from the trend component spectrum. This noise has been assigned to the nonsignal component which becomes more stochastic. At the extreme, we could remove all the noise from the trend component and assign it to the 'cycle'. The decomposition that we would obtain is found to

Figure 2.3: Spectra for the cycle component

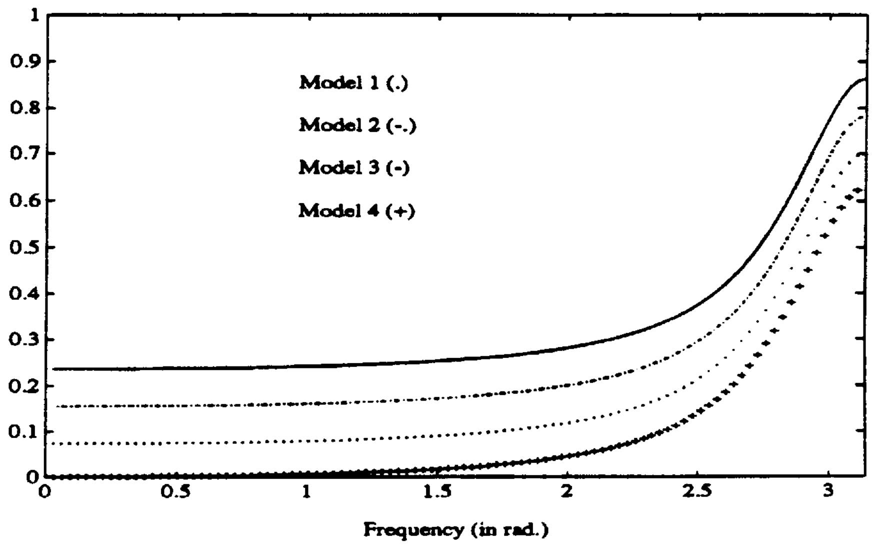
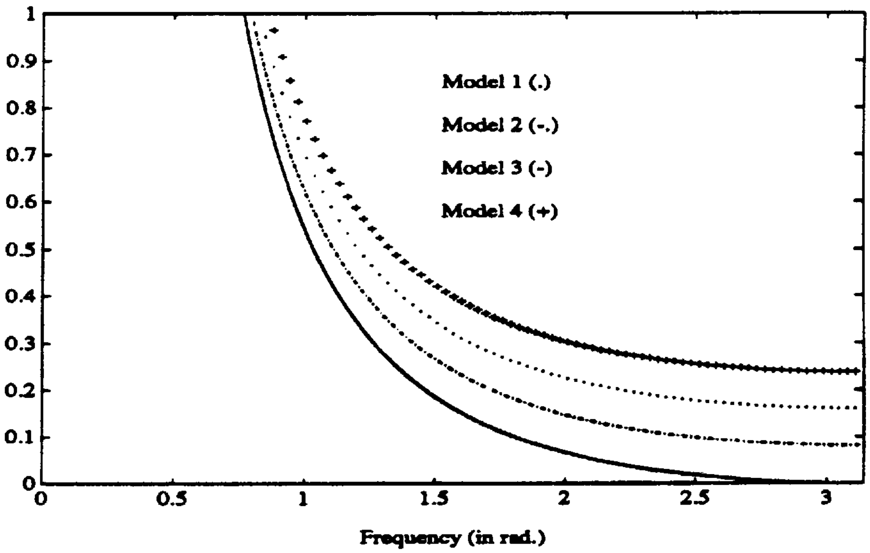


Figure 2.4: Spectra for the trend component



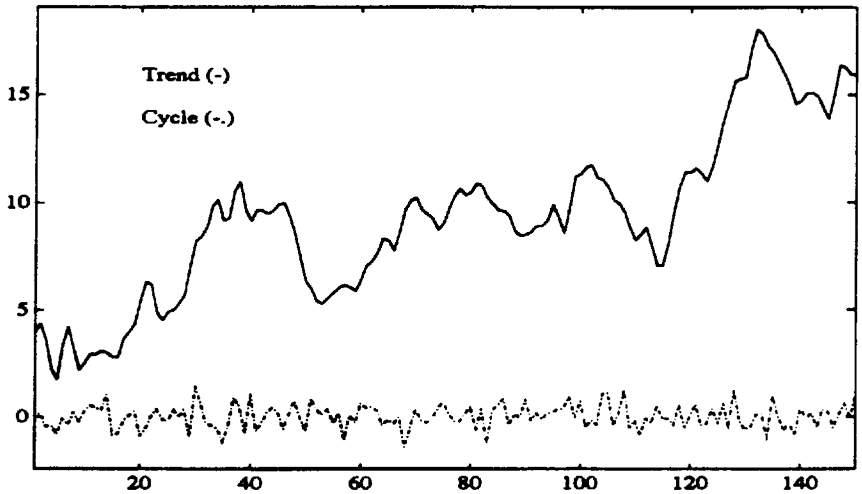
be:

$$\begin{aligned} \Delta s_t &= (1+B)a_{st} \qquad V_s = .161 \\ (1+.7B)n_t &= (1+.496B)a_{nt} \qquad V_n = .306, \end{aligned} \tag{2.7}$$

and the trend is now noninvertible, the zero in the spectrum at the π-frequency meaning that it does not embody any orthogonal noise. The spectrum for these components are also seen on figures 2.3 and 2.4. If we look on figure 2.5 at the plot of the components so-obtained and compare it with the components of the first decomposition, one notices that the trend is smoother and the 'cycle' noisier. Decompositions where one component is noninvertible while the other concentrates all the noise of the model are called 'canonical' (see Hillmer and Tiao (1982)). Alternatively, one may be interested by the cycle analysis. In this case, it is possible to assign all the noise of the model to the trend component, and to remove it from the spectrum of the cycle. We thus obtain a second canonical decomposition, denoted "model 4" on figures 2.3 and 2.4, where the canonical component is now the cycle.

As shown on figures 2.3 and 2.4, the identification problem can be thus seen as the problem of determining the spectrum of the component within the range delimited below by the spectrum of the component free of noise and above by the spectrum of the component concentrating all the noise of the model. The admissible decompositions are thus characterized by a particular noise allocation. The identification problem in Unobserved Components models will be treated more thoroughly in chapter 3.

Figure 2.5: Trend plus Cycle example: canonical trend and noisy cycle



2.6 Presentation of the problem

It is tempting to relate the underidentification of the components with the lack of a precise definition of the components. For example, suppose we are interested in removing the seasonal variations of a time series. The first question that one would immediately ask is: what is seasonality ? Unfortunately there is no generally accepted definition of what constitutes a seasonal component. And even if we accept to define a seasonal component by the spectral peaks at the seasonal frequencies (see for example Granger (1978), Bell and Hillmer (1984)), such a definition is not precise enough to imply a unique model for the seasonal component. Spectral peaks are generated by large roots in AR polynomial; nothing is said about what should be the MA polynomial and the component innovation variance. In the same way, a trend component may be defined by an infinite spectral peak at the low frequencies, but from this definition several models for a trend component may be derived. In other words, unobserved components cannot be precisely defined, and as a consequence they are not identified. Identification problems are also encountered in simultaneous econometric models. But, if in these cases economic theory enables the analysts to overcome the underidentification of the model by setting at zero some parameters, in statistics, as noticed in Maravall (1988b), no such a priori information is available. Any model-based approach must thus consider an arbitrary assumption on the components. This gives rise to two important problems in UC models. Firstly, as recognized by statisticians (see Bell and Hillmer (1984)), it makes difficult to evaluate a signal extraction procedure: it is not possible to compare methods estimating different signals. Secondly, given that one is interested in a signal, which model form should be chosen ? Obviously, it would be desirable to attenuate the arbitrariness inherent in any signal extraction procedure. In this dissertation, we discuss the choice of a UC model specification according to some "optimality" criteria concerning estimation and some other desirable properties. For our purposes, the most attractive feature of the ARIMA-model-based approach is that the arbitrary assumptions on the components models are made explicit, while they are somewhat hidden in empirical methods.

Different model specifications yield different properties and different degree of accuracy of the estimator. Bell and Hillmer (1984) already noticed that "the accuracy of the estimator depends heavily on what is being estimated". Since any unobserved component model specification is an arbitrary choice difficult to rationalize, why to not use the most accurately estimated decomposition ? The estimation error is important for data producing agencies and researchers which are clearly interested in the most precise historical estimators. Similarly, policy making and evaluation would benefit of being conducted on the most precisely estimated concurrent signals. The concern of selecting a decomposition in respect to the estimation accuracy was already underlying in Maravall and Pierce (1986) when they concluded: "why so much emphasis on seasonal adjustment? Perhaps attention should shift to estimation of a smoother signal less affected by revisions (possibly some type of trend)". The "revisions" refers to a type a estimation error affecting concurrent estimates that will be presented in chapter 4. Maravall and Pierce were clearly interested in best estimated models. In Chapter 6, we shall provide an answer to the question they raise. We first need to present a formal discussion of the identification problem in UC models and a brief overview of the most popular identification criteria.

Chapter 3

The Identification Problem in Unobserved Components Models.

3.1 Introduction

We have just seen that in the "Trend plus Cycle" example, the decomposition was not unique. We had a system of 3 equations for 4 unknowns which thus could not be uniquely solved. Each set of parameters consistent with this system and insuring non negative components spectra defined an admissible decomposition. The difference between admissible decompositions could also be interpreted in terms of different noise repartitions. In this chapter, we formally discuss the identification problem for the general model (A) under the two possible perspectives: as a parametric problem and as a noise repartition problem. Some popular identification criteria used in the statistical litterature will be presented. We first need some essential concepts about identifiability.

3.2 Identifiability: Basic Concepts.

The following definitions are mainly extracted from Rothenberg (1971) and Hotta (1989).

Definition 3.1 *A structure S is a set of hypothesis which implies a unique distribution function of the observable Y, say $P[Y/S]$. The set denoted $\mathcal{S}$ of all a priori possible structures is called a model.*

From this definition, a structure is identified if there exists a unique inverse association between S and $P[Y/S]$. Considerig model (A), each structure is formed by a particular set of parameters $S_A = \{\theta_{n1}, \cdots, \theta_{nq_n}, \theta_{s1}, \cdots, \theta_{sq_s},$- $V_n, V_s\}$ lying within the admissible parameter space. Model (A) would thus be identified if the distribution of the observable x_t would be generated by a unique set of parameters S_A.

Definition 3.2 *Two structures S_1 and $S_2 \in \mathcal{S}$ are said to be observationally equivalent if $P[Y/S_1] = P[Y/S_2]$ almost everywhere.*

A model will thus be identified if and only if there does not exist two different structures which are observationally equivalent.

Definition 3.3 *A structure $S \in \mathcal{S}$ is said to be identifiable if there is no other structure in $\mathcal{S}$ which is observationally equivalent. A model $\mathcal{S}$ is identified if all the structures are identified.*

In practice, we will consider a weaker condition and say that a model is identified if almost all the structures (not in a probabilistic sense) are identified.

3.3 The identification problem as a parametric problem: identification by zero-coefficients restrictions.

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

3.3.1 A necessary condition for identifiability.

For the general model (A), the relationship (2.3) provides the following identity:

$$\theta_x(B)\theta_x(F) = \phi_s(B)\phi_s(F)\theta_n(B)\theta_n(F)V_n + \phi_n(B)\phi_n(F)\theta_s(B)\theta_s(F)V_s, \quad (3.1)$$

which implies a set of covariances equations by equating the coefficient the coefficient of B^j. The right hand side of (3.1) contains $q_n + q_s + 2$ unknowns, which are $\theta_{n1}, \cdots, \theta_{nq_n}, V_n, \theta_{s1}, \cdots, \theta_{sq_s}$, and V_s, while the left hand side yields $q_x + 1$ covariances equations. So when $q_n + q_s + 1 > q_x$, the system is underidentified and instead of a unique decomposition, a set of observationally equivalent decompositions is obtained. Using $q_x = max(p_n + q_s, p_s + q_n)$, we can easily deduce that the necessary condition for identification of model (A) is:

$$q_s < p_s \quad \text{or} \quad q_n < p_n. \quad (3.2)$$

We now prove that this condition is also sufficient for the identification of model (A).

3.3.2 A sufficient condition for identifiability.

For certain types of decompositions, Hotta (1989) has shown that condition (3.2) is sufficient to identify the UC models. However, its demonstration did not apply to model (A) which is more general that the specifications that he discussed. Hotta (1989) also considered a seasonal and a nonseasonal component sharing a unit root at the zero frequency. As discussed in Pierce (1979), this hypothesis would imply that no optimal estimators exist, so it is naturally excluded by our model specification. We thus must adapt the Hotta's demonstration to the type of models we are considering. Hotta's methodology is based on the relationship between the spectrum of the observed series and the spectra of the UC. Using the Fourier transform in (3.1), we have:

$$| \theta_x(e^{-i\lambda}) |^2 = | \phi_s(e^{-i\lambda})\theta_n(e^{-i\lambda}) |^2 V_n + | \phi_n(e^{-i\lambda})\theta_s(e^{-i\lambda}) |^2 V_s,$$

where for a polynomial $\eta(e^{i\lambda})$, $| \eta(e^{i\lambda}) |^2 = \eta(e^{i\lambda})\eta(e^{-i\lambda})$. Developing the squared polynomials, we get:

$$\sum_{i=0}^{q_x} \theta_{xi}^+ \cos(i\lambda) = [V_n \sum_{i=0}^{q_n} \theta_{ni}^+ \cos(i\lambda)][\sum_{i=0}^{p_s} \phi_{si}^+ \cos(i\lambda)] +$$

$$+[V_s \sum_{i=0}^{q_s} \theta_{si}^+ \cos(i\lambda)][\sum_{i=0}^{p_n} \phi_{ni}^+ \cos(i\lambda)], \qquad (3.3)$$

where: $\theta_{ki}^+ = \sum_{j=0}^{q_k - i} \theta_{kj}\theta_{ki+j}\delta_i^{**}$, $k = x, n, s$ and $\delta_i^{**} = 1$ if $i = 0$, $\delta_i^{**} = 2$ otherwise. The coefficients ϕ_{si}^+ and ϕ_{ni}^+ are defined similarly.

To prove the sufficiency of condition (3.2), we need a trigonometric property, a lemma and a theorem (Hotta (1989)):

Lemma 3.1 *If* $\sum_{j=0}^{r} a_j \cos(j\lambda) = \sum_{j=0}^{r} b_j \cos(j\lambda)$, $\lambda \in [-\pi, \pi]$, *then* $a_j = b_j$ *for all* j.

To demonstrate this lemma it is enough to see that $\cos(j\lambda), j = 1, \cdots, r$ is a set of linearly independent functions.

Theorem 3.1 *If* $\psi(B)$ *and* $\overline{\psi}(B)$ *are two polynomials of order* b *with roots not inside the unit circle, then:*

$$\sum_{i=0}^{b-j} \psi_i \psi_{i+j} \sigma^2 = \sum_{i=0}^{b-j} \overline{\psi}_i \overline{\psi}_{i+j} \overline{\sigma}^2,$$

for $j = 0, 1, \cdots, b$ *and with* $\psi_0 = \overline{\psi}_0$ *has a unique solution* $\psi_i = \overline{\psi}_i$, $i = 1, \cdots, b$ *and* $\sigma^2 = \overline{\sigma}^2$.

The proof is given in Anderson (1971). This theorem allows us to focus on the identification of the θ_{ni}^+ and θ_{si}^+ since the identification of the MA parameters will follow. The trigonometric property that we will use is:

Property 3.1

$$\cos(k\lambda)\cos(j\lambda) = \frac{1}{2}[\cos((k+j)\lambda) + \cos((k-j)\lambda)].$$

We demonstrate the sufficiency of condition (3.2) for the case where $max(p_n + q_s, q_n + p_s) = q_n + p_s$, the alternative case leading to a similar demonstration. So $q_x = q_n + p_s$ and the necessary condition is: $q_s < p_s$ and/or $q_n < p_n$. We focus on showing that $q_s < p_s$ is sufficient to identify of model (A). We can proceed with $q_s = p_s - 1$ to have the maximum of MA parameters on the signal to identify. We however need a restriction which is: $q_n < q_s$. The opposite case will be discussed later. Also, notice that $q_s = p_s - 1$ and $q_x = p_s + q_n$ imply $p_n \le q_n$.

Using lemma (3.1) and property (3.1) in equation (3.3), we get for $q_s < p_s$ the following system of equations:

$$
\begin{aligned}
2\theta^+_{xp_s+q_n} &= \theta^+_{nq_n}\phi^+_{sp_s}V_n \\
2\theta^+_{xp_s+q_n-1} &= [\theta^+_{nq_n-1}\phi^+_{sp_s} + \theta^+_{nq_n}\phi^+_{sp_s-1}]V_n \\
&\vdots \\
2\theta^+_{xp_s+p_n} &= [\theta^+_{np_n}\phi^+_{sp_s} + \cdots + \theta^+_{nq_n}\phi^+_{sp_s+p_n-q_n}]V_n \\
2\theta^+_{xq_s+p_n} &= [\theta^+_{np_n-1}\phi^+_{sp_s} + \cdots + \theta^+_{nq_n}\phi^+_{sq_s+p_n-q_n}]V_n + \theta^+_{sq_s}\phi^+_{np_n}V_s \\
&\vdots \\
2\theta^+_{xp_s} &= [2\theta^+_{n0}\phi^+_{sp_s} + \theta^+_{n1}\phi^+_{sp_s-1} + \cdots + \theta^+_{nq_n}\phi^+_{sp_s-q_n}]V_n + \\
& \quad [\theta^+_{sq_s}\phi^+_{n1} + \theta^+_{sq_s-1}\phi^+_{n2} + \cdots + +\theta^+_{sp_s-p_n}\phi^+_{np_n}]V_s \\
&\vdots \\
2\theta^+_{xp_n} &= [\theta^+_{nq_n}\phi^+_{sq_n-p_n} + \cdots + 2\theta^+_{np_n}\phi^+_{s0} + \theta^+_{np_n-1}\phi^+_{s1} + \cdots + 2\theta^+_{n0}\phi^+_{sp_n}]V_n + \\
&+ \quad [2\theta^+_{s0}\phi^+_{np_n} + \theta^+_{s1}\phi^+_{np_n-1} + \cdots + +2\theta^+_{sp_n}\phi^+_{n0} + \theta^+_{sp_n+1}\phi^+_{n1} + \cdots]V_s
\end{aligned}
$$

$$
\begin{aligned}
2\theta^+_{x1} &= [\theta^+_{nq_n}(\phi^+_{sq_n-1} + \phi^+_{sq_n+1}) + \cdots + 2\theta^+_{n1}\phi^+_{s0} + 2\theta^+_{n0}\phi^+_{s1}]V_n + \\
&+ [2\theta^+_{s0}\phi^+_{n1} + 2\theta^+_{s1}\phi^+_{n0} + \cdots + (\theta^+_{sp_n-1} + \theta_{sp_n+1})\phi_{np_n}]V_s \\
2\theta^+_{x0} &= [\theta^+_{nq_n}\phi^+_{sq_n} + \cdots + \theta^+_{n1}\phi^+_{s1} + 2\theta^+_{n0}\phi^+_{s0}]V_n + \\
&+ [2\theta^+_{s0}\phi^+_{n0} + \cdots + \theta^+_{sp_n}\phi^+_{np_n}]V_s. \qquad (3.4)
\end{aligned}
$$

In matrix notation: $\Gamma = A\Psi$, where:

$$
\Gamma_{(p_s+q_n+1)\times 1} = \begin{bmatrix} 2\theta^+_{xp_s+q_n} \\ \vdots \\ 2\theta^+_{xp_s+p_n} \\ \vdots \\ 2\theta^+_{x0} \end{bmatrix},
$$

and:

$$\Psi_{(p_s+q_n+1)x1} = \begin{bmatrix} \theta^+_{nq_n} V_n \\ \vdots \\ \theta^+_{np_n} V_n \\ \\ \theta^+_{sq_s} V_s \\ \vdots \\ \theta^+_{s0} V_s \\ \\ \theta^+_{np_n-1} V_n \\ \\ \vdots \\ \theta^+_{n0} V_n \end{bmatrix}.$$

Writing Ψ in this form gives a matrix A of the type:

$$A_{(p_s+q_n+1)\ x\ (p_s+q_n+1)} = \begin{bmatrix} \underset{(q_n-p_n+1)\ x\ (q_n-p_n+1)}{A_{11}} & \underset{(q_n-p_n+1)\ x\ (p_s+p_n)}{0} \\ \\ \underset{(p_s+p_n)\ x\ (q_n-p_n+1)}{A_{21}} & \underset{(p_s+p_n)\ x\ (p_s+p_n)}{A_{22}} \end{bmatrix},$$

with:

$$A_{11} = \begin{bmatrix} \phi^+_{p_s} & 0 & \cdots & 0 \\ \phi^+_{p_s-1} & \phi^+_{p_s} & \cdots & \vdots \\ \vdots & & & 0 \\ \phi^+_{p_s-q_n+p_n} & \cdots & \cdots & \phi^+_{p_s} \end{bmatrix}.$$

To prove the sufficiency of condition (3.2), we have to show that the system (3.4) admits a unique solution, i.e. that A is invertible. We have:

$$
\begin{aligned}
| A | &= | A_{11} || A_{22} - A_{21}A_{11}^{-1}A_{12} |= \\
&= | A_{11} || A_{22} | ,
\end{aligned}
$$

since $A_{12} = 0_{(q_n-p_n+1)x(p_s+p_n)}$. Furthermore $| A_{11} |$ being triangular gives: $| A_{11} |= (\phi^+_{sp_s})^{q_n-p_n+1} = (2\phi_{sp_s})^{q_n-p_n+1}$ which differs from zero if the order of the AR polynomial $\phi_s(B)$ is correctely specified. So A is invertible if A_{22} is. From (3.4), the matrix A_{22}, of order $(p_s + p_n)x(p_s + p_n)$, can be decomposed as:

$$
A_{22} = \begin{bmatrix} \underset{(p_s \text{ x } p_s)}{a_{11}} & \underset{(p_s \text{ x } p_n)}{a_{12}} \\ & \\ \underset{(p_n \text{ x } p_s)}{a_{21}} & \underset{(p_n \text{ x } p_n)}{a_{22}} \end{bmatrix},
$$

where, from (3.4):

$$
a_{11} = \begin{bmatrix} \phi^+_{np_n} & 0 & \cdots & 0 \\ \phi^+_{np_n-1} & \phi^+_{np_n} & \cdots & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & \cdots & 2\phi^+_{np_n} \end{bmatrix}.
$$

The determinant of A_{22} is given by: $| A_{22} |=| a_{11} || a_{22} - a_{21}a_{11}^{-1}a_{12} |$. Since a_{11} is triangular, we have: $| a_{11} |= 2(\phi^+_{np_n})^{p_s}$ Simplifying, we get: $| a_{11} |=$

$2(2\phi_{np_n})^{p_s}$ which is by assumption different from zero. So A_{22} is invertible if a_{22} is. From (3.4), the submatrix a_{22}, of dimension $p_n x p_n$, is given by:

$$a_{22} = \begin{bmatrix} 2\phi^+_{s0} & \phi^+_{s1} & \cdots & 2\phi^+_{sp_n-1} \\ \phi^+_{s1} & 2\phi^+_{s0} & \cdots & 2\phi^+_{sp_n-2} \\ \vdots & & & \vdots \\ \phi^+_{sp_n} + \phi^+_{sp_n+2} & \cdots & 2\phi^+_{s0} + \phi^+_{s2} & 2\phi^+_{s1} \\ \phi^+_{sp_n-1} & \cdots & \cdots & 2\phi^+_{s0} \end{bmatrix}.$$

So the only case where the matrix A is not invertible is: $| a_{22} |= 0$. However, the relationship that it would imply between the parameters of the polynomial $\phi_s(B)$ is so specific that it defines a set of measure zero in the space of the parameters. Hence the matrix A is generally invertible when $q_s < p_s$, and the system (3.4) admits a unique solution. The condition $q_s < p_s$ is thus sufficient to identify model (A). ∎

For sake of simplicity, the demonstration has been conducted under the hypothesis that $p_n < q_s$. It can be easily checked that the opposite case does not change the triangularity property of the matrices A_{11} and a_{11}. So the demonstration remains valid, even if eventually the submatrix a_{22} would come out with different coefficients. So we can conclude that the model (A) is identified almost everywhere under the condition:

$$q_n < p_n \quad or \quad q_s < p_s.$$

It thus possible to restrict the order of the components MA polynomial in order to identify the underlying structure of the model. This has been mostly used in 'structural models':

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

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

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

where u_t and v_t are orthogonal white-noises. Maravall (1985) noted that this model is equivalent to specifying an IMA(2,1) model for the trend component, which makes explicit the identification restrictions in structural models.

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

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

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

Balanced components are more general in the sense that they allow different noise repartitions. The order of the MA polynomial for the remaining component will however be constrained by the overall model for the observed series. Writing (2.4) as:

$$q_x \leq p_x + \max(q_s - p_s, q_n - p_n),$$

the model for the remaining component will be balanced if $q_x \leq p_x$, and with a MA polynomial of higher order than the AR polynomial if $q_x > p_x$. Clearly, in any of these cases the necessary and sufficient condition (3.2) for identification of model (A) is no longer satisfied: an additional condition must be imposed. The Watson's minimax filter and the canonical decomposition are two identification procedures which use some extra considerations in order to obtain identification. With both of them, the identification problem is explicitly handled as a noise repartition problem.

3.4 The identification problem as a noise repartition problem

3.4.1 Writing the decomposition in terms of one single unidentified parameter indexing the noise repartition

As in Watson (1987), we express the models for the Unobserved Components in terms of a single unidentified parameter. We must first make the assumption that the model for one component, let us say s_t, is balanced in order:

Assumption 5: $q_s = p_s$,

so, as previously pointed out, the model for the other component n_t will be balanced if $q_x \leq p_x$, top-heavy $q_x > p_x$. This decomposition is however non identified. Recalling that $g_x(w)$, $g_s(w)$, and $g_n(w)$ denote the spectra of x_t, s_t and n_t at the frequency w, $0 \leq w \leq \pi$, the hypothesis of independence of the components yields the following relationship:

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

As in Burman (1980), we write : $\varepsilon_s = \min_w g_s(w)$ and $\varepsilon_n = \min_w g_n(w)$. The quantity $\varepsilon_s + \varepsilon_n$ can be seen as the variance of a pure noise component embodied in the spectrum of the observed series which can be attributed arbitrarily. We shall denote this variance as V_u, which is also expressed in V_a units. If we remove as much noise as possible from s_t and attribute it to n_t, then we obtain : $g_s^0(w) = g_s(w) - \varepsilon_s$, the spectrum of a noninvertible signal s_t^0, and $g_n^0(w) = g_n(w) + \varepsilon_s$ the spectrum of a nonsignal n_t^0 which concentrates all the noise of the model.

Using $g_s^0(w)$ and $g_s^0(w)$, we now define for $\alpha \in [0,1]$ the unobserved components s_t^α and n_t^α as the processes with spectra :

$$\begin{aligned} g_s^\alpha(w) &= g_s^0(w) + \alpha V_u, \\ g_n^\alpha(w) &= g_n^0(w) - \alpha V_u, \end{aligned} \tag{3.5}$$

so that the spectrum of the observed series can be represented as :

$$g_x(w) = g_s^\alpha(w) + g_n^\alpha(w). \tag{3.6}$$

If we consider the time domain, we have : $A_x = A_s^\alpha + A_n^\alpha$, where in the same way :

$$A_s^\alpha = A_s^0 + \alpha V_u \quad \text{and} \quad A_n = A_n^0 - \alpha V_u. \tag{3.7}$$

All the admissible decompositions are written in term of one single under-identified parameter α which indexes the noise repartition. It is clear that the identification problem arises because we do not know which amount of noise must be assigned to the components : α is not unique, but belongs to the interval $[0,1]$. These notations will enable us to derive some results relating the estimation errors to the unobserved components model specification. At this stage, we use them to present the minimax filter and the canonical specification.

3.4.2 Canonical decomposition

The canonical decomposition was first proposed by Box, Hillmer and Tiao (1978) and Pierce (1978). The approach consists of specifying a component as clean of noise as possible. Hence, a canonical signal has a zero in its spectrum, which corresponds to a unit root in the MA polynomial. That is, a canonical signal is noninvertible. Identification is thus obtained by removing from the signal spectrum its minimum. It corresponds to the component s_t^0 previously defined by the spectrum $g_s^0(w)$. As illustrated in (3.5), an interesting property of canonical decomposition is that the admissible models for a signal can always be written as the sum of the canonical one plus an orthogonal white noise. Furthermore, Hillmer and Tiao (1982) showed that the canonical decomposition minimizes the variance of the signal innovation. This model specification is widely used in the ARIMA-model-based approach. Additional properties of the canonical decomposition will be shown in chapter 6.

A disadvantage of the canonical decomposition is that it is not unique. In (3.5), it is also possible to remove all the noise from the nonsignal n_t and to assign it to the signal s_t, defining a new canonical decomposition with the nonsignal made noninvertible. The problem is that often it is not clear how to justify the choice of a particular canonical specification. Consider for example seasonal adjustment applications where the canonical requirement usually concerns the seasonal component. The underlying idea is that the analyst is interested by what is not seasonal, and wishes to handle a nonseasonal component embodying as much information as possible. This corresponds to the minimum extraction principle (Pierce (1978)). However, the same analyst may also be interested in a nonseasonal component less obscured by noisy movements, and so he may expect an as-smooth-as-possible seasonally adjusted series. The nonseasonal component would thus be taken canonical. Both views can be perfectly justified. Canonical decompositions are not unique and a particular choice may be difficult to rationalize.

3.4.3 Minimax filter

The minimax filter is an identification procedure proposed by Watson (1987) and Findley (1985). It designs the components models in order to maximize the Mean Squared Errors on the components estimators. It is motivated by the idea that, if a filter minimizes the estimation error for a particular unobserved component model specification, the same filter may perform poorly in terms of Mean Squared Error for another possible decomposition. In order to make the estimator "robust" over the range of observationally equivalent decompositions, Watson (1987) built a minimax filter as a Minimum MSE filter applied on the decomposition which maximizes that.

This approach may also be difficult to rationalize. Since all admissible decompositions are observationally equivalent, and only differ by the noise repartition between the components, why to choose the decomposition with maximum MSE on the estimators ? Unobserved components are tools that are designed to capture some phenomenoms. They are expected to help as well as possible the users, and for that to approximate as well as possible the underlying signal.

As far as we know, the Watson's minimax filter was the first approach relating the estimation error to the component model specification. We shall pursue that relationship. Before investigating how the identification of the components influences the accuracy of the estimators, we need to analyze the estimation procedure.

Chapter 4

Minimum Mean Squared Error Estimation of the Components

4.1 Introduction

Estimation of the components is necessary because the components are never observed. In this chapter, we discuss an estimation procedure and present some new properties of the estimators. We will emphasize the fact that the estimator characteristics depend on the assumptions made in order to reach identification.

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

$$\hat{s}_{t/T} = E(s_t/X_T). \tag{4.1}$$

If $t = T$, the conditional expectation (4.1) yields the concurrent estimator of the signal. If $t < T$, (4.1) provides the preliminary estimator of a past realization of the signal, while for $t > T$, (4.1) corresponds to the $t-T$ period-ahead forecast of the signal. In the model-based approach framework, two methods of calculating this expectation can be used, each one having specific advantages. The Kalman filter method (see, for example, Anderson and Moore (1979)) proceeds firstly by writing the model in a state space format, then by setting some initial conditions, and finally by deriving the optimal estimator through recursive computations. Because of its computational tractability, it has been used in many applied works (for a general presentation, see for example Harvey (1989)). The Wiener-Kolmogorov (WK) filter has the benefit of being particularly suited for analytical discussion, and provides as an output a clear and precise information about the structure of the estimators. Its computational tractability has been improved in a decisive way by the T. Wilson algorithm presented in Burman (1980). The theoretical nature of the topic developed in this dissertation motivates our choice of focussing on the WK filter. The results, however, will also apply to estimators obtained with the Kalman filter.

4.2 The Wiener-Kolmogorov filter

For completeness, we present the construction of the WK filter. For stationary time series, this filter is derived as follow (from Whittle (1963)). We must first take the assumption that an infinite set of observations on the process x_t is available. We will denote $\hat{s}_t$ the estimator obtained for an infinite sample, so that $\hat{s}_t = \hat{s}_{t/\infty}$. This assumption will be relaxed later.

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

$$\hat{s}_t = \sum_{k=-\infty}^{\infty} \nu_k x_{t-k},$$

$$= \nu(B)x_t,$$

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

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

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

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

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

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

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

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

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

$$\nu(e^{-iw}) = g_s(w)/g_x(w). \qquad (4.2)$$

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

4.3 Optimal estimators

4.3.1 MMSE Estimators

The WK filter was expressed in (4.2) as the ratio of the spectrum of the signal to the spectrum of the observed series. An appealing feature of ARIMA models is that they provide a convenient way to parametrize the spectrum of time series. Applying the expression (4.2) to model (A), and under the hypothesis of independence of the components, the estimators can be obtained as (see, for example, Hillmer and Tiao (1982)):

$$\begin{aligned}\hat{s}_t &= \frac{g_s(w)}{g_x(w)} x_t = \\ &= V_s \frac{\theta_s(e^{-iw})\theta_s(e^{iw})\phi_n(e^{-iw})\phi_n(e^{iw})}{\theta_x(e^{-iw})\theta_x(e^{iw})} x_t.\end{aligned}$$

and for the nonsignal estimator :

$$\hat{n}_t = \frac{g_n(w)}{g_x(w)} x_t =$$

$$= V_n \frac{\theta_n(e^{-iw})\theta_n(e^{iw})\phi_s(e^{-iw})\phi_s(e^{iw})}{\theta_x(e^{-iw})\theta_x(e^{iw})} x_t, \tag{4.3}$$

The hypothesis of invertibility of the polynomial $\theta_x(B)$ insures the convergence of the filter. It is a symmetric filter, which depends on the polynomials of the models for the observed series and those for the components. This dependence allows the WK filter to adapt itself to the series under analysis, in contrast to the empirical filters which assume that roughly the same model holds for every series.

From (4.3), it is clear that the MMSE estimators are available only when all the polynomials of model (A) have been determined. Estimating the components requires the practitioners to first select an admissible decomposition, that is to make an arbitrary assumption about the stochastic structure of the components. Different assumptions made on the models for the component will imply different properties of the estimators through the squared polynomial $\theta_i(B)\theta_i(F)$ and the innovation variance V_i, where $i = n, s$. However, there is not a strict correspondence between the stochastic properties of the components and those of the estimators. Indeed, some discrepancies do exist. This point has been often discussed in the statistical literature (see for example Bell and Hillmer (1984)). It can be easily understood from a study of the models followed by the estimators.

4.3.2 The distribution of the estimators

Considering the estimator $\hat{s}_t$, we can write:

$$\hat{s}_t = V_s \frac{\theta_s(B)\theta_s(F)\phi_n(F)}{\phi_s(B)\theta_x(F)} a_t. \tag{4.4}$$

Comparing the model for the estimator (4.4) and the model (2.2) for the theoretical signal, it is easily seen that they share the same AR and MA polynomials in B (see, for example, Maravall (1993b)). So, if a component is nonstationary, the component and its estimator will share the same stationarity inducing transformation. Moreover, as a consequence of the validity of the WK filter in nonstationary cases, a nonstationary component and its estimator are cointegrated with cointegrating vector (1,-1).

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

4.4 Covariance between estimators.

Another important discrepancy between the properties of the component and those of the estimators is that, if the components are assumed independent, the Minimum Mean Squared Errors estimators will always be covariated. This is a consequence of the estimation procedure which orthogonally projects both components on a space of dimension one defined by the observed series. The existence of covariances between the estimators even if the theoretical components were assumed independent has been subject of attention in the literature (see for example Nerlove (1964), Granger (1978), Garcia-Ferrer and Del Hoyo (1992)). A similar result has been discussed by Harvey and Koopman (1992) concerning the estimators of the 'pseudo-residuals' a_{nt} and a_{st}. Since the covariances between the components' estimators are described as an undesirable feature of the estimation, selecting the admissible decomposition which minimizes the lag-0 covariance between the estimators seems to be a reasonable identification criterion. The covariances between the estimators are easily obtained from the following lemma:

Lemma 4.1 *If $C(\hat{n},\hat{s})$ denotes the cross-covariance generating function between $\hat{n}_t$ and $\hat{s}_t$, then $C(\hat{n},\hat{s})$ is equal to the ACGF of the model:*

$$\theta_x(B)z_t = \theta_s(B)\theta_n(B)b_t, \tag{4.5}$$

where b_t is a white noise with variance V_nV_s.

<u>Proof:</u> From (4.3), we have:

$$C(\hat{n},\hat{s}) = A_nA_s/A_x. \tag{4.6}$$

Developing and simplifying, we get:

$$C(\hat{n},\hat{s}) = V_nV_s\frac{\theta_n(B)\theta_n(F)\theta_s(B)\theta_s(F)}{\theta_x(B)\theta_x(F)}, \tag{4.7}$$

which is the ACGF of the model (4.5). ∎

So $C(\hat{n},\hat{s})$ can be seen as the ACGF of an ARMA process with AR polynomial $\theta_x(B)$, MA polynomial $\theta_n(B)\theta_s(B)$, and with innovation variance V_nV_s.

The dependence of the covariances between the estimators to the admissible decompositions occurs thus through the MA polynomial and the innovation variance of model (4.5).

Using lemma 4.1, we derive the following properties of the MMSE estimators.

Property 4.1 *The lag-0 covariance between the estimators is always positive.*

This is obvious since, from lemma 1, the lag-0 covariance between the estimators is equal to the variance of the process z_t: $cov[\hat{n}_t, \hat{s}_t] = var[z_t] > 0$. Since the estimators must sum to the observed series, the existence of a cross-spectrum between the estimators indicates that part of the component spectrum is lost by the estimation procedure. In particular, the positive sign of the lag-0 covariance suggests an "underestimation" of the component, in the sense that the estimator will be more stable than the component. This point will be more discussed in the next section.

Property 4.2 *The covariances between the components estimators are symmetric, finite and converge to zero.*

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

Two points make property 4.2 interesting. Firstly, the covariances between the estimators are finite even when the components are nonstationary with unit roots at different frequencies. Secondly, this result holds independently of the components' order of integration. An interesting consequence is:

Lemma 4.2 *When the observed series x_t is nonstationary, the estimators $\hat{s}_t$ and $\hat{n}_t$ are uncorrelated whatever the selected admissible decomposition is.*

Proof: The correlations between the estimators are given by the ratio of the covariances to the product of the standard deviations of the estimators. As shown in property 4.2, the covariances are finite, while x_t being nonstationary implies that at least one component will be nonstationary, with an infinite standard deviation. So the ratio of covariances to standard deviations will tend to zero. The result is clearly valid for all admissible decompositions. ∎

So, when the observed series is nonstationary, the estimation procedure preserves the property of the theoretical components in terms of zero cross-correlations. Given that most of the economic time series encountered in practice are nonstationary, our result has a wide relevance. Care must thus be taken with saying that "whereas the theoretical components are uncorrelated, the estimators will be correlated in general" (Garcia-Ferrer and del Hoyo (1992)). That can be true only if the components are stationary. Notice the following paradox: when x_t is nonstationary, the correlations between the estimators are zero, while:

Property 4.3 *The correlations between the stationary transformation of the components are different from* 0.

Proof: We denote by $\delta_s(B)$ and $\delta_n(B)$ the polynomials inducing stationarity of s_t and n_t. If, for example, only s_t is nonstationary, then we have: $\delta_n(B) = 1$. We have seen that a component and its estimator share the same stationarity inducing transformation, so $\delta_s(B)\hat{s}_t$ and $\delta_n(B)\hat{n}_t$ are stationary. Then the cross-covariances generating function of the stationary transformation of the estimators is given by:

$$C(\delta_s(B)\hat{s}_t, \delta_n(B)\hat{n}_t) = \delta_s(B)\delta_n(F)C(\hat{s}_t, \hat{n}_t)$$

From property 4.2, the estimators are covariated with finite covariances whatever the order of integration of the components is. Each term of $C(\delta_s(B)\hat{s}_t,$

$\delta_n(B)\hat{n}_t)$ will thus be finite and non null. The standard deviations of the stationary transformation of the estimators being finite, the cross-correlations will be non null. ∎

We now focus on the errors in the estimators.

4.5 Estimation errors

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

4.5.1 Final Estimation Error: relationship with the estimators cross-covariances.

Lemma 4.3 *For all admissible decompositions, the theoretical estimators cross-covariance generating function is also the ACGF of the final estimation error* *:*

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

Proof: To simplify the expressions, we will write for any polynomial $\eta(B)$: $\eta = \eta(B)$ and $\overline{\eta} = \eta(F)$. Multiplying $C(\hat{n}, \hat{s})$ in (4.7) by $\theta_x\overline{\theta_x}/\theta_x\overline{\theta_x}$, we have:

$$C(\hat{n}, \hat{s}) = V_n V_s \frac{\theta_x\overline{\theta}_x\theta_n\overline{\theta}_n\theta_s\overline{\theta}_s}{\theta_x\overline{\theta}_x\theta_x\overline{\theta}_x}. \qquad (4.8)$$

The relationship (2.3) between the MA polynomials provides:

$$\theta_x\overline{\theta}_x = V_n\theta_n\overline{\theta}_n\phi_s\overline{\phi}_s + V_s\theta_s\overline{\theta}_s\phi_n\overline{\phi}_n. \qquad (4.9)$$

Inserting (4.9) into (4.8), we get:

$$C(\hat{n}, \hat{s}) = V_n V_s \theta_n\theta_s\overline{\theta}_n\overline{\theta}_s \frac{V_n\theta_n\overline{\theta}_n\phi_s\overline{\phi}_s + V_s\theta_s\overline{\theta}_s\phi_n\overline{\phi}_n}{\theta_x\overline{\theta}_x\theta_x\overline{\theta}_x}.$$

Now writing :

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

the CCGF $C(\hat{n}, \hat{s})$ can be seen as the ACGF of a process z_t defined as :

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

It is then enough to show that, for example, $z_t = s_t - \hat{s}_t$. Developing the last expression, we get :

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

Using the models (2.2) for the components:

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

which can be written in terms of the ACGF:

$$\begin{aligned} z_t &= [A_n/A_x]s_t - [A_s/A_x]n_t = \\ &= [1 - A_s/A_x]s_t - [A_s/A_x]n_t = \\ &= s_t - [A_s/A_x]x_t = \\ &= s_t - \hat{s}_t. \end{aligned}$$

So z_t is exactly the final estimation error on the signal. Working on n_t instead of s_t would lead to $z_t = \hat{n}_t - n_t$ which has the same ACGF than $s_t - \hat{s}_t$. ∎

An immediate consequence of lemmas 4.1 and 4.3 is that:

Lemma 4.4 *The final estimation error $s_t - \hat{s}_t = \hat{n}_t - n_t$ can be seen as the output of an ARMA process given by:*

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

where b_t is a normally distributed independent white noise with variance V_nV_s.

This last result is identical to the one obtained by Pierce (1979) using another demonstration.

The MA polynomial of the process for the final estimation error corresponds to the product of the MA polynomials of the models for the two components, while the AR polynomial is given by the MA polynomial of the model for the observed series. From assumption 3, the observed series' process is invertible, so the final estimation error follows a stationary model, whose variance gives the MSE of the estimation.

The choice of a particular decomposition will affect the estimation error through the MA polynomials $\theta_s(B)$, $\theta_n(B)$, and the innovation variances V_s and V_n. Lemma 4.3 provides the following general result:

Corollary 4.1 *The admissible decomposition minimizing the final estimation error of the components also minimizes the covariances between the estimators.*

Corollary 4.1 suggests that one particular admissible decomposition will have some attractive features, since it will minimize both the variance of the final estimation error of the components and the lag-0 covariance between the estimators. Conversely, a consequence of lemma 4.3 is that the Watson minimax filter will present the inconvenience of maximizing the lag-0 covariance beween the estimators.

Another interesting consequence of lemma 4.3 is that it enables to approximate the Mean Squared Errors on the component historical estimator by the covariance between the estimators. Since the latter are available as output of the estimation procedure, this measure would be "empirical". This has two important implications. Firstly, recall that when the component estimators are obtained from ad hoc filters such as X-11 for seasonally adjusting series, no estimation errors are available. This is a serious limitation of ad-hoc filters, and the need to know the precision of the estimators has been repeatedly pointed out by advisory committees (see for example Hibbert Committee (1988)). Lemma 4.3 suggests that it is possible to use the covariance between the estimators as approximation of the MSE of the estimators when the filtering procedure yields reasonable results. Further studies are however necessary to determine the behavior of this statistic.

Secondly, the final estimation error in the ARIMA-model-based approach was up to this point only theoretically obtained by deriving the variance of the model given in lemma 4.4 from the models for the components. By comparing this measure with the lag-0 covariance computed on the estimators empirically obtained, a specification test for model-based components analysis may be constructed. The asymptotic properties of this test are still to be investigated. By now, these two points remain left as open issues for future research.

4.5.2 Frequency domain analysis of the final estimation error.

A look at the way the filter works on the series, in the frequency domain, helps in understanding the differences between the final estimator and the component. From (4.3), the spectrum $g_{\hat{s}}(w)$ of the MMSE estimator $\hat{s}_t$ is given by:

$$
\begin{aligned}
g_{\hat{s}}(w) &= \frac{g_s^2(w)}{g_x(w)} = \\
&= \frac{1}{1+\frac{g_n(w)}{g_s(w)}} g_s(w) \qquad (4.10)
\end{aligned}
$$

When the relative contribution of the signal is high at a particular frequency w^*, $g_n(w^*)/g_s(w^*) \simeq 0$, and most of the observed series spectrum is used for the signal estimation. The gain of the filter for this frequency is close to 1, and we get: $g_{\hat{s}}(w^*) \simeq g_s(w^*)$. Conversely, when the relative contribution is low at a particular frequency, the WK filter just ignores it for the signal estimation. For example, suppose that either the signal's spectrum has a zero at the frequency w^* or the nonsignal's spectrum admits an infinite peak at the frequency w^*, so that we have: $g_n(w^*)/g_s(w^*) \rightarrow \infty$. Then the spectrum of the signal estimator will display a zero: $g_{\hat{s}}(w^*) \simeq 0$, and the estimator will follow a noninvertible model. This conclusion was already obtained from a direct observation of the model for the estimator. Furthermore, it is easily deduced from (4.10) that $g_{\hat{s}}(w) \leq g_s(w)$ for every frequency, so the signal is always underestimated (see for example Burman (1980)). The frequency domain analysis of the final estimation error shows that the error is mainly related to the frequencies where the stochastic variability of the signal is relatively low. As discussed in Maravall (1993a), the underestimation of the signal will be particularly large for a stable signal, and particularly small for an unstable signal.

4.5.3 Revision error

The hypothesis of having an infinite realization of the series x_t was needed because the WK filter of (4.3) goes to $-\infty$ to ∞. Since the filter is convergent, it can be safely truncated at some point. However, at the beginning or at the end of a sample, the computation of the estimator requires unknown past or future realizations of x_t. We will focus on the distortion induced by the lack of future observations. Near the end of a finite sample, optimal preliminary estimates can be computed by replacing unknown future realizations by their forecasts (Cleveland and Tiao (1976)). The forecast errors imply that the preliminary estimates will be contaminated by an additional error, termed 'revision error'. As new observations become available, forecasts are updated and eventually replaced by the observed values, and the preliminary estimator is revised. The total revision error in the concurrent estimate of s_t, that is the estimate of s_t computed at time t, is given by: $\hat{s}_t - E(\hat{s}_t/X_t)$. To simplify the presentation, we will denote by $E_t\hat{s}_t$ the expectation of the estimate of s_t conditional on the information available at time t, so that: $E_t\hat{s}_t = E(\hat{s}_t/X_t)$. Writing:

$$\begin{aligned} \hat{s}_t &= V_s \frac{\theta_s(B)}{\phi_s(B)} \frac{\theta_s(F)\phi_n(F)}{\theta_x(F)} a_t, \\ &= \xi_s(B) a_t, \end{aligned}$$

with $\xi_s(B) = \cdots + \xi_{s-1}B + \xi_{s0} + \xi_{s1}F + \cdots$, the total revision error can be obtained as:

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

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

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

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

$$V[\hat{s}_t - E_{t+k}\hat{s}_t] = \sum_{i=k+1}^{\infty} \xi_{si}^2. \tag{4.11}$$

The stationarity of the revisions can be easily understood by noticing that the polynomial $\xi_s(B,F)$ is convergent in F. A consequence is that $lim_{k\rightarrow\infty} V[\hat{s}_t - E_{t+k}\hat{s}_t] = 0$: in practice, the revisions become negligible after some number of periods. A convenient algorithm for computing the ψ-weights is developed in Maravall (1994). This makes straightforward the computation of the revision error in (4.11). Adding final estimation error and total revision error, we obtain the *total estimation error* in the concurrent estimates as $s_t - E_t\hat{s}_t$.

Chapter 5

Estimation Errors and Identification of Unobserved Component Models

5.1 Introduction

We have seen that model-based approaches offer some convenient answers to the limitations of empirical filters, since they provide the estimation error, they enable forecasting, and the estimation takes into account the stochastic properties of the series under analysis. However, they cannot avoid an important identification problem. Within the ARIMA-model-based framework, the hypothesis made to characterize the components are explicitly stated. We have also seen that the identification problem turns out to be a noise repartition problem, and that it is difficult to provide a motivation for a particular choice. Analyzing the components' estimators, it appeared that their properties and their precision depend on the selected models for the components. In this chapter, we relate the accuracy of the estimators to the range of admissible decompositions. The results will be used in the next chapter to derive some simple rules for selecting the most precisely estimated decomposition. Some important properties of the canonical decomposition will appear as a by-product of our analaysis.

When dealing with the decomposition of a time series into unobserved components, there are mostly two types of estimators of interest: the final or historical, and the concurrent estimator. Although the first assumes an infinite sample, in practical applications with the usual series length, the convergence of the filter insures that it corresponds to estimators for the central years of the series. For this reason, they are typically of interest to data-producing agencies. The second, the concurrent estimator, are employed by analysts involved in the short-term monitoring of the economy. As discussed in subsection 4.5.3, concurrent estimators are affected by an additional error, the revision error, which is due to the lack of future observations. In applied work and in economic policy making, this additional error may have some important implications. For the monetary policy case, Maravall and Pierce (1983) have studied the misleading effects of the revision error. For both concurrent estimators as well as for historical estimators, the importance of minimizing the estimation error can be understood.

We discuss the relationship between the estimation error and noise repartition for these two cases of applied interest. Then, for sake of generality, we will extend our results to the estimator of s_t computed at any time $t+k$, $k = \cdots, -1, 0, 1, \cdots$, that is to any preliminary estimator ($k > 0$) and to the forecasts of the components ($k < 0$).

We first need to present some notation that will be used in the remainder of the dissertation. We recall that s_t^α represents a signal concentrating a fraction $\alpha \in [0,1]$ of the pure noise part of the observed series. The formal definitions of the decomposition of x_t into s_t^α and n_t^α have been given in chapter 3, equation (3.5). We shall denote by $\hat{s}_t^\alpha$ the optimal estimator of the signal s_t^α, so $\hat{s}_t^0$ is the estimator of the canonical signal with no noise. It is obtained as $\hat{s}_t^0 = \nu_s^0(B)x_t$, where $\nu_s^0(B) = \sum_{i=-\infty}^{\infty} \nu_{si}^0 B^i$ represents the corresponding WK filter. This estimator will be written in terms on the innovations a_t on the observed series as $\hat{s}_t^0 = \xi_s^0(B)a_t$, where:

$$\begin{aligned}\xi_s^0(B) &= \nu_s^0(B)\theta_x(B)/\phi_x(B) = \\ &= \cdots + \xi_{-1}^0 B + \xi_0^0 + \xi_1^0 F + \cdots.\end{aligned}$$

Similar notations are used to define $\nu_n^0(B)$ and $\xi_n^0(B)$ as the WK filters estimating the nonsignal n_t^0 embodying all the noise as a function of the observed series and of the innovations a_t respectively. Finally, we will denote by $h(B) = \sum_{i=-\infty}^{\infty} h_i B^i$ the filter obtained as $h(B) = \phi_x(B)\phi_x(F)/\theta_x(B)\theta_x(F)$. This filter corresponds to the "inverse ACGF" of the model for x_t (see Cleveland (1976)).

5.2 Final estimation error on historical estimators

We write the final estimation error on the signal s_t^α as: $e_t^\alpha = s_t^\alpha - \hat{s}_t^\alpha$. We recall that the variance of the final error in $\hat{s}_t^\alpha$ also gives the variance of the error in $\hat{n}_t^\alpha$.

Lemma 5.1 *The final estimation error e_t^α (and hence the lag-0 covariance between the estimators) is related to the noise repartition α, $\alpha \in [0,1]$, according to:*

$$var[e_t^\alpha] = cov[\hat{n}_t^\alpha, \hat{s}_t^\alpha] =$$

$$= \quad var[e_t^0] + \alpha V_u(1 - 2\nu_{s0}^0) - \alpha^2 h_0 V_u^2, \qquad (5.1)$$

where e_t^0 is the error in the canonical signal estimator.

Proof: Using lemmas 4.1, 4.3 and the expression 3.7 we have:

$$\begin{aligned} ACGF(e_t^\alpha) &= C(\hat{n}_t^\alpha, \hat{s}_t^\alpha) = \\ &= \frac{A_s^\alpha A_n^\alpha}{A_x} = \\ &= \frac{[A_s^0 + \alpha V_u][A_n^0 - \alpha V_u]}{A_x} = \\ &= \frac{A_s^0 A_n^0}{A_x} + \alpha V_u(1 - 2\frac{A_s^0}{A_x}) - \alpha^2 V_u^2 \frac{1}{A_x}, \end{aligned}$$

where use has been made of $A_x = A_s^0 + A_n^0$. The result (5.1) is then immediate given that $A_s^0 A_n^0 / A_x = ACGF(e_t^0)$, that A_s^0 / A_x corresponds to the WK filter $\nu_s^0(B)$, and that $1/A_x = h(B)$. ∎

Lemma 5.1 gives the relationship between noise repartition and final estimation error. As a function of α, the variance of the final estimation error is a second-order polynomial, with known coefficients $V(e_t^0)$, h_0, ν_{s0}^0 and V_u. The first three coefficients can be easily obtained as variances of the following processes:

- for $V(e_t^0)$: $\theta_x(B) z_t = \theta_n^0(B) \theta_s^0(B) b_t^0$, $V(b_t^0) = V_n^0 V_s^0$;
- for h_0: $\theta_x(B) z_t = \phi_x(B) a_t$, $V_a = 1$;
- for ν_{s0}^0: $\theta_x(B) z_t = \theta_s^0(B) \phi_n(B) c_t^0$, $V(c_t^0) = V_s^0$,

where $\{\theta_s^0(B), V_s^0\}$ and $\{\theta_n^0(B), V_n^0\}$ are the MA polynomials and the innovations variances respectively associated with the canonical signal s_t^0 and with a nonsignal component n_t^0 concentrating all the noise. It is interesting to notice that these three models share the polynomial $\theta_x(B)$ as the AR polynomial. The parameter h_0 corresponds to the variance of the "inverse" or "dual" model (see Cleveland (1972)). As discussed in chapter 3, the fourth coefficient V_u is simply the variance of the pure noise part of the observed series. All these coefficients are thus positive, and they can be derived from the model for the observed series in a straightforward manner (see Chapter 7). The expression (5.1) thus constitutes an easy way to obtain the estimation error variance over the range of the admissible decompositions.

5.3 Revision error and total estimation error on concurrent estimators

The revision error on the estimator of s_t^α computed at time t is denoted as: $r_t^\alpha = \hat{s}_t^\alpha - E_t\hat{s}_t^\alpha$. The total estimation error on the concurrent estimator will be written as: $d_t^\alpha = s_t^\alpha - E_t\hat{s}_t^\alpha$. When $\alpha = 0$, r_t^0 and d_t^0 represent the revision error and the total estimation error on the concurrent estimator of a canonical signal respectively. The variances of the estimation errors on the signal and on the nonsignal estimators are of course identical.

Lemma 5.2 *The size of the revision error on the concurrent estimators of the signal depends on the noise repartition through:*

$$var(r_t^\alpha) = var(r_t^0) + 2\alpha(\nu_{s0}^0 - \xi_{s0}^0)V_u + \alpha^2(h_0 - 1)V_u^2, \qquad (5.2)$$

where ξ_{s0}^0 is the central coefficient of the MMSE filter expressing the canonical signal estimator in terms of the innovations a_t on the observed series.

Proof: The estimator of the signal s_t^α can be written as:

$$\begin{aligned}
\hat{s}_t^\alpha &= \frac{A_s^\alpha}{A_x} x_t = \\
&= \frac{A_s^0 + \alpha V_u}{A_x} x_t = \\
&= [\nu_s^0(B) + \alpha V_u h(B)] x_t = \\
&= \sum_{i=-\infty}^{\infty} (\nu_{si}^0 + \alpha V_u h_i) x_{t+i},
\end{aligned}$$

with $\nu_{si}^0 = \nu_{s(-i)}^0$ and $h_i = h_{-i}$ since the WK filter and the polynomial $h(B)$ are symmetric. The concurrent estimate of the signal s_t^α is given by:

$$E_t \hat{s}_t^\alpha = \sum_{i=0}^{\infty} (\nu_{si}^0 + \alpha V_u h_i) x_{t-i} + \sum_{i=1}^{\infty} (\nu_{si}^0 + \alpha V_u h_i) E_t x_{t+i}.$$

So the revision in the concurrent estimate is :

$$\begin{aligned}
r_t^\alpha &= \hat{s}_t^\alpha - E_t \hat{s}_t^\alpha = \\
&= \sum_{i=1}^{\infty} (\nu_{si}^0 + \alpha V_u h_i)(x_{t+i} - E_t x_{t+i}) = \\
&= \sum_{i=1}^{\infty} (\nu_{si}^0 + \alpha V_u h_i) e_t(i), \qquad (5.3)
\end{aligned}$$

where $e_t(i)$ is the ith-period ahead forecast error of x_t. Using:

$$\psi(B) = \frac{\theta_x(B)}{\phi_x(B)} = 1 + \sum_{j=1}^{\infty} \psi_i B^i, \tag{5.4}$$

we can write the ith-period ahead forecast error as:

$$e_t(i) = a_{t+i} + \sum_{j=1}^{i-1} \psi_j a_{t+i-j}. \tag{5.5}$$

Inserting (5.5) into (5.3), we get:

$$\begin{aligned} r_t^\alpha &= \sum_{i=1}^{\infty} \nu_{si}^0 (a_{t+i} + \sum_{j=1}^{i-1} \psi_j a_{t+i-j}) + \\ &+ \alpha V_u \sum_{i=1}^{\infty} h_i (a_{t+i} + \sum_{j=1}^{i-1} \psi_j a_{t+i-j}) = \\ &= \sum_{i=1}^{\infty} l_i a_{t+i} + \alpha V_u \sum_{i=1}^{\infty} m_i a_{t+i}, \end{aligned} \tag{5.6}$$

where:

$$\begin{aligned} l_i &= \nu_{si}^0 + \psi_1 \nu_{si+1}^0 + \psi_2 \nu_{si+2}^0 + \cdots, \\ m_i &= h_i + \psi_1 h_{i+1} + \psi_2 h_{i+2} + \cdots. \end{aligned} \tag{5.7}$$

Denoting $l(F) = \sum_{i=1}^{\infty} l_i F^i$ and $m(F) = \sum_{i=1}^{\infty} m_i F^i$, then (5.6) can be written as:

$$r_t^\alpha = \hat{s}_t^\alpha - E_t \hat{s}_t^\alpha = l(F) a_t + \alpha V_u m(F) a_t. \tag{5.8}$$

So the variance of the revision error is given by:

$$var(r_t^\alpha) = var(l(F)a_t) + V_u^2\alpha^2 var(m(F)a_t) + +2\alpha V_u cov(l(F)a_t, m(F)a_t). \tag{5.9}$$

The first term of the right hand side of (5.8) provides the revision error on the concurrent estimate of the canonical signal:

$$var(r_t^0) = l(F)l(B)\mid_{B=F=0}. \tag{5.10}$$

The second term of the right hand side of (5.8) relates the noise repartition to the revision error through the polynomial $m(F)$. It is worth noting that this polynomial is the same as that which drives the revision error in signal plus noise decompositions in Maravall (1986). This result was as expected since our model specification implicitly consider a third component which is an orthogonal white noise. From (5.7), m_i is the coefficient of the term in F^i, $i > 0$, in the polynomial multiplication $h(B)\psi(B)$. Since $h(B)$ can be written as $1/\psi(B)\psi(F)$, we have: $h(B)\psi(B) = 1/\psi(F)$. Observing that $m(0) = 0$ while, by construction, the coefficient of F^0 in the polynomial $1/\psi(F)$ is unity, we get:

$$1 + m(F) = 1/\psi(F). \tag{5.11}$$

Developing $var(m(F)a_t) = m(F)m(B)\mid_{B=F=0}$, we have:

$$\begin{aligned} m(F)m(B)\mid_{B=F=0} &= (1/\psi(F) - 1)(1/\psi(B) - 1)\mid_{B=F=0}= \\ &= 1/\psi(F)\psi(B)\mid_{B=F=0} -2(1/\psi(B))\mid_{B=0} +1, \end{aligned}$$

which leads to:

$$m(F)m(B) \mid_{B=F=0} = h_0 - 1. \tag{5.12}$$

We now turn to the covariance between $l(F)a_t$ and $m(F)a_t$. From (5.7), l_i is the coefficient of the term in F^i, $i > 0$, in the polynomial multiplication $\nu_s^0(B)\psi(B)$ which yields the WK filter expressing the estimators as function of the innovations: $\nu_s^0(B)\psi(B) = A_s^0/\psi(F) = \xi_s^0(B)$. Therefore, we can write:

$$\begin{aligned} \xi_s^0(B) = \nu_s^0(B)\psi(B) &= [\xi_{s0}^0 + \xi_{s-1}^0 B + \cdots + \\ &+ \xi_{s-j}^0 B^j + \cdots + l_1 F + \cdots + l_j F^j + \cdots]. \end{aligned} \tag{5.13}$$

We can now use (5.13) and (5.11) to simplify $cov(l(F)a_t, m(F)a_t) = \sum_{i=1}^{\infty} l_i m_i$ by writing the WK filter $\nu_s^0(B)$ as a function of the coefficients l_i and m_i:

$$\begin{aligned} \nu_s^0(B) &= \xi_s^0(B)/\psi(B) \\ &= [\xi_{s0}^0 + \xi_{s-1}^0 B + \cdots + \xi_{s-j}^0 B^j + \cdots + l_1 F + \cdots + l_j F^j + \cdots] \\ &\quad [1 + m_1 B + m_2 B^2 + \cdots + m_j B^j + \cdots]. \end{aligned}$$

Considering the central term, we get:

$$\nu_0^0 = \xi_{s0}^0 + \sum_{i=1}^{\infty} l_i m_i \tag{5.14}$$

so: $\sum_{i=1}^{\infty} l_i m_i = \nu_{s0}^0 - \xi_{s0}^0$, i.e. the covariance between the revision on the canonical signal and the revision related to the noise repartition is proportional to the difference between the central coefficient of the WK filter $\nu_s^0(B)$

and the central coefficient of the WK filter expressing the estimator in terms of the innovations on the observed series.

Substituting (5.14) and (5.12) into (5.9), the result in Lemma 5.2 is obtained. ∎

Another way to express this result would make use of: $\xi_{s0}^0 = \nu_{s0}^0 + \sum_{i=1}^{\infty} \nu_{si}^0 \psi_i$, so that $cov(l(F)a_t, m(F)a_t)$ may also be expressed using: $\sum_{i=1}^{\infty} l_i m_i = -\sum_{i=1}^{\infty} \nu_{si}^0 \psi_i$.

Lemma 5.2 presents the relationship between the noise repartition and the revision error on the concurrent estimator of the signal. As for the previous final estimator case, the relationship is simply a second order polynomial in α. The constant terms $V(r_t^0)$ and ξ_{s0}^0 can be computed without difficulty using the algorithm in Maravall (1994). Hence, given a model for the observed series, lemma 5.2 provides a way to compute the revision error on the concurrent estimator of the signal over the range of all admissible decompositions.

We now focus on the total estimation error on the concurrent estimator.

Lemma 5.3 *The variance of the Total Estimation Error (TEE) on the concurrent estimate of the signal is related to the noise repartition through:*

$$var(d_t^\alpha) = var(d_t^0) + \alpha V_u(1 - 2\xi_{s0}^0) - \alpha^2 V_u^2. \tag{5.15}$$

where $var(d_t^0)$ *represents the variance of the estimation error on the concurrent estimator of a canonical signal:* $d_t^0 = s_t^0 - E_t \hat{s}_t^0$.

<u>**Proof:**</u> Since the final estimation error and the revision error are orthogonal (Pierce (1980)), the size of the total estimation error is obtained by summing the size of the final estimation error and the size of the revision error. Adding (5.1) to (5.2) yields (5.15). ∎

Alternative algorithms have been proposed in the literature to compute $V(e_t^\alpha)$, $v(r_t^\alpha)$ and $V(d_t^\alpha)$ for a given admissible decomposition. Hillmer (1985) provides an algorithm for computing the mean squared revisions which takes into account the truncation of the sample at both sides. Where the Kalman filter is to be used, a method for obtaining the errors variances as output of the filter has been described in Burridge and Wallis (1985). Both procedures are however computational algorithms which do not rely on analytical results.

We illustrate our results with the following multiplicative ARIMA model:

$$\Delta\Delta_{12}x_t = (1-\theta_1 B)(1-\theta_{12}B^{12})a_t \qquad \begin{aligned} a_t &\sim NIID(0,1) \\ \theta_1 &= .3 \\ \theta_{12} &= .7 \end{aligned}$$

The series x_t was decomposed into : $x_t = n_t + s_t$, where n_t and s_t represents here the nonseasonal and the seasonal components defined as in model (A). Making s_t noninvertible, we obtain the following model for the nonseasonal component :

$$\Delta^2 n_t^0 = (1 - 1.27B + .29B^2)a_{nt} \qquad V_n = .739$$

The corresponding WK filter is :

$$A_n^0/A_x = \frac{V_n \theta_n^0(B)\theta_n^0(F)\phi_s(B)\phi_s(F)}{\theta_x(B)\theta_x(F)}$$

where :

$$\theta_n^0(B) \;=\; (1 - 1.27B + .29B^2)$$

$$\begin{aligned}\phi_s(B) &= 1 + B + ... + B^{11} \\ \theta_x(B) &= (1 - .3B)(1 - .7B^{12})\end{aligned}$$

This filter yields as central coefficient : $\nu_{n0}^0 = .859$. Given that $\nu_{s0}^\alpha + \nu_{n0}^\alpha = 1$, we have: $\nu_{s0}^0 = .141$. The inverted process z_t such that: $\theta_x(B)z_t = \phi_x(B)a_t$, whose ACGF is $h(B)$, has a variance of : $h_0 = 1.811$. The size of the pure noise part of x_t is found to be : $V_u = g_n^0(\pi) = .305$. For this model specification we have : $cov[\hat{n}_t^0, \hat{s}_t^0] = V[s_t^0 - \hat{s}_t^0] = .089$ while attributing all the noise to the seasonal component would lead to a final error variance of .135. On the concurrent estimators, the total error variance is .181 with a canonical seasonal component, and .289 with a canonical nonseasonal component. In general, for any admissible decomposition, the final error variance (also covariance between $\hat{n}_t^\alpha$ and $\hat{s}_t^\alpha$) is given by (from lemma 5.1):

$$V[e_t^\alpha] = .089 + .219\alpha - .168\alpha^2,$$

and for the Total Estimation Error (using lemma 5.3):

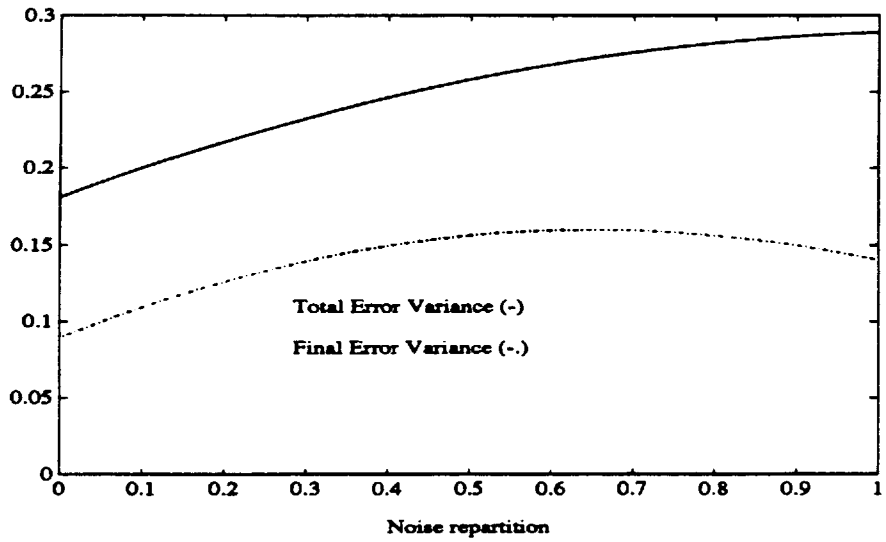
$$V[n_t^\alpha - E_t\hat{n}_t^\alpha] = .181 + .201\alpha - .093\alpha^2,$$

where, for both, $\alpha \in [0, 1]$.

These functions are plotted on figure 5.1.

As expected they are positive for $\alpha \in [0, 1]$. The final error variance, and the lag-0 covariance between the theoretical estimators, vary within the range [.089, .135], and the error on the concurrent estimator within the range [.181, .289] according to the model specification adopted. Both types of errors are minimized for $\alpha = 0$, that is with a canonical seasonal component. The maximum final error is reached around $\alpha = .62$, which is the noise repartition yielded by the Watson minimax filter (Watson (1987)) on the historical

Figure 5.1: Mean Squared Errors in an ARIMA$(0,1,1)_{12}(0,1,1)$ model



estimators. A minimax filter on the concurrent estimates would yield a total estimation error varinace of .289, which corresponds to specifying a canonical nonseasonal component. The revision error corresponds to the area between the two curves, and varies from .091 for an intermediate noise repartition of $\alpha \simeq .10$ to .149 which is obtained on the estimation of a noninvertible trend component.

5.4 Generalization to preliminary estimators and forecasts of the components.

We now generalize our results to any estimator of the signal s_t computed at time $t+k$. Although historical and concurrent estimators are the most interesting in practical applications, for the sake of generality we consider the case of any preliminary estimator or forecast. We write the total estimation error as:

$$\begin{aligned} d^{\alpha}_{t/t+k} &= s^{\alpha}_t - E_{t+k}\hat{s}^{\alpha}_t = \\ &= s^{\alpha}_t - \hat{s}^{\alpha}_t + \hat{s}^{\alpha}_t - E_{t+k}\hat{s}^{\alpha}_t = \\ &= e^{\alpha}_t + r^{\alpha}_{t/t+k}, \end{aligned} \tag{5.16}$$

where e^{α}_t represents the final error and $r^{\alpha}_{t/t+k}$ the revision error on the estimator of s^{α}_t computed at time $t+k$, where k is a signed integer. Clearly, $E_{t+k}\hat{s}^{\alpha}_t$ corresponds to a preliminary estimator when $k > 0$, and to a $-k$-period-ahead forecast when $k < 0$. The WK filter estimating the signal s^{α}_t using finite set of observations $X_{t+k} = [x_1, \cdots, x_{t+k}]$ is denoted $\nu_s^{\alpha,t/t+k}(B)$, so:

$$\hat{s}^{\alpha}_{t/t+k} = E_{t+k}\hat{s}^{\alpha}_t = \nu_s^{\alpha,t/t+k}(B)x_t.$$

We denote $\lambda(B)$ the inverse polynomial:

$$\lambda(B) = \psi(B)^{-1} = \sum_{i=0}^{\infty} \lambda_i B^i, \tag{5.17}$$

where $\lambda_0 = 1$. We also define the coefficient δ_k, $k > 0$, such that: $\delta_k = \sum_{i=0}^{k} \lambda_i^2$. Considering first the case of preliminary estimators ($k > 0$), we get the following result:

Lemma 5.4 *The variance of the error in the preliminary estimator obtained at time $t+k$ of the signal for period t is related to the noise repartition α, $\alpha \in [0,1]$, according to:*

$$var[d^{\alpha}_{t/t+k}] = var[d^{0}_{t/t+k}] + \alpha V_u(1 - 2\nu_{s0}^{0,t/t+k}) - \alpha^2 V_u^2 \delta_k, \tag{5.18}$$

where $k > 0$ and $\nu_{s0}^{0,t/t+k}$ is the coefficient in B^0 of the WK filter estimating a canonical signal s_t^0 at time $t+k$.

__Proof:__ From (5.16) and given the independence of final and revision errors, we have: $var(d^{\alpha}_{t/t+k}) = var(e^{\alpha}_t) + var(r^{\alpha}_{t/t+k})$. The variance of e^{α}_t is given in lemma 5.1. We thus focus on the relationship between the noise repartition α and $r^{\alpha}_{t/t+k}$. Following a reasoning similar to that of Lemma 5.2, we have:

$$\begin{aligned} r^{\alpha}_{t/t+k} &= \hat{s}_t - E_{t+k}\hat{s}_t = \\ &= \sum_{i=-\infty}^{\infty} (\nu^0_{si} + \alpha V_u h_i)(x_{t+i} - E_{t+k}x_{t+i}) = \\ &= \sum_{i=k+1}^{\infty} (\nu^0_{si} + \alpha V_u h_i)(x_{t+i} - E_{t+k}x_{t+i}), \end{aligned} \tag{5.19}$$

since for $i \le k$, $E_{t+k}x_{t+i} = x_{t+i}$. Using: $x_{t+i} = \sum_{j=0}^{\infty} \psi_j a_{t+i-j}$, it is easily obtained that the forecast error on x_{t+i} computed at time $t+k$, $k < i$, is given by:

$$x_{t+i} - E_{t+k}x_{t+i} = \sum_{j=0}^{i-k-1} \psi_j a_{t+i-j}. \tag{5.20}$$

Now inserting (5.20) in (5.19), we get:

$$
\begin{aligned}
r^{\alpha}_{t/t+k} &= \sum_{i=k+1}^{\infty} (\nu^0_{si} + \alpha V_u h_i)(\sum_{j=0}^{i-k-1} \psi_j a_{t+i-j}) = \\
&= \sum_{i=k+1}^{\infty} [(\nu^0_{si} + \psi_1 \nu^0_{si+1} + \cdots)a_{t+i} + \\
&\quad +\alpha V_u (h_i + \psi_1 h_{i+1} + \cdots)a_{t+i}]. \qquad (5.21)
\end{aligned}
$$

Writing:

$$
l_i = \nu^0_{si} + \psi_1 \nu^0_{si+1} + \cdots,
$$

and:

$$
m_i = h_i + \psi_1 h_{i+1} + \cdots, \qquad (5.22)
$$

the variance of the revision error on the nonconcurrent estimates of the signal is given by:

$$
var[r^{\alpha}_{t/t+k}] = var[r^0_{t/t+k}] + \alpha^2 V_u^2 \sum_{i=k+1}^{\infty} m_i^2 + 2\alpha V_u \sum_{i=k+1}^{\infty} l_i m_i. \qquad (5.23)
$$

The first term of the r.h.s. of (5.23) represents the size of the revision error on the nonconcurrent estimator of the canonical signal. It can be computed using: $V(r^0_{t/t+k}) = \sum_{i=k+1}^{\infty}(\xi^0_{si})^2$. The second term of the r.h.s. of (5.23) relates the revision error to the α-squared through the term $V_u^2 \sum_{i=k+1}^{\infty} m_i^2$. As defined in (5.22), m_i can be seen as the term in F^i in the polynomial multiplication $h(B)\psi(B)$ which yields the polynomial $\lambda(F)$ previously defined in (5.17). Hence, for $i \geq 0$, we have $m_i = \lambda_i$ while for $i < 0$, $m_i = 0$ since the polynomial $\lambda(F)$ does not contain any term in B. Summing all the m_i-squared coefficients, we then get:

$$\begin{aligned}\sum_{i=0}^{\infty} m_i^2 &= \lambda(F)\lambda(B) \mid_{B=F=0}= \\ &= \frac{1}{\psi(F)\psi(B)} \mid_{B=F=0}= \\ &= h_0. \end{aligned} \tag{5.24}$$

Now, since for $k > 0$, $\sum_{i=k+1}^{\infty} m_i^2 = \sum_{i=0}^{\infty} m_i^2 - \sum_{i=0}^{k} m_i^2$, we have:

$$\begin{aligned}\sum_{i=k+1}^{\infty} m_i^2 &= h_0 - \sum_{i=0}^{k} m_i^2, \\ &= h_0 - \delta_k, \end{aligned} \tag{5.25}$$

since $m_i = \lambda_i$ for $i > 0$.

The third term of the r.h.s. of (5.23) represents the covariance between the revision error on the canonical signal and the error induced by the noise repartition. It involves the term l_i, which as defined in (5.22) can be seen to be the term in $B^{|i|}$ if $i < 0$, and the term in F^i if $i \geq 0$ in the polynomial multiplication $\nu_s^0(B)\psi(B)$. Thus the nonsymmetric polynomial $l(B,F)$ defined as: $l(B,F) = \sum_{i=-\infty}^{\infty} l_i F^i$, is such that: $l(B,F) = \nu_s^0(B)\psi(B)$ and thus represents the WK filter expressing the final estimator $\hat{s}_t$ in terms of the innovations on the observed series x_t: $l(B,F) = \xi_s^0(B)$. Now since $m(B) = 1/\psi(B)$, we have:

$$l(B,F)m(B) = \nu_s^0(B)\psi(B)/\psi(B) =$$

$$= \nu_s^0(B). \tag{5.26}$$

Then, the sum of all non-zero cross-products $l_i m_i$ yields:

$$\begin{aligned} \sum_{i=0}^{\infty} l_i m_i &= l(B,F)\lambda(B) \mid_{B=F=0}= \\ &= \nu_s^0(B) \mid_{B=F=0}= \\ &= \nu_{s0}^0. \end{aligned} \tag{5.27}$$

Then, since for $k > 0$, $\sum_{i=k+1}^{\infty} l_i m_i = \sum_{i=0}^{\infty} l_i m_i - \sum_{i=0}^{k} l_i m_i$, we have:

$$\begin{aligned} \sum_{i=k+1}^{\infty} l_i m_i &= \nu_{s0}^0 - \sum_{i=0}^{k} l_i m_i = \\ &= \nu_{s0}^0 - \sum_{i=0}^{k} \xi_{si}^0 \lambda_i, \end{aligned} \tag{5.28}$$

where use has being made of $l_i = \xi_{si}^0$ and $m_i = \lambda_i$. We now have to show that $\sum_{i=0}^{k} \xi_{si}^0 \lambda_i = \nu_{s0}^{0,t/t+k}$. Expressing the estimator $\hat{s}_{t/t+k}^0$ as a function of the innovations a_t as: $\hat{s}_{t/t+k}^0 = \xi_s^{0,t/t+k}(B)a_t$, then the filter $\xi_s^{0,t/t+k}(B)$ is defined by:

$$\begin{aligned} \xi_s^{0,t/t+k}(B)a_t &= E_{t+k}\xi_s^0(B)a_t = \\ &= (\cdots + \xi_{s-1}^0 B + \xi_{s0}^0 + \xi_{s1}^0 F + \cdots + \xi_{sk}^0 F^k)a_t. \end{aligned}$$

So the filter $\xi_s^{0,t/t+k}(B)$ corresponds to the filter $\xi_s^0(B)$ for the historical estimator truncated at the term in F^k. The following relationship being still valid:

$$\nu_s^{0,t/t+k}(B)\psi(B) = \xi_s^{0,t/t+k}(B),$$

inverting $\psi(B)$, we have:

$$\begin{aligned}\nu_s^{0,t/t+k}(B) &= \xi_s^{0,t/t+k}(B)\lambda(B) = \\ &= (\cdots + \xi_{s-1}^0 B + \xi_{s0}^0 + \xi_{s1}^0 F + \cdots + \xi_{sk}^0 F^k)\lambda(B).\end{aligned}$$

Looking at the term in B^0 in $\nu_s^{0,t/t+k}(B)$, we directly get:

$$\nu_{s0}^{0,t/t+k} = \sum_{i=0}^{k} \xi_{si}^0 \lambda_i. \tag{5.29}$$

Putting together this last result, (5.25), and (5.28), in (5.23), and adding the variance of the historical error given by lemma 5.1 proves the lemma. ∎

It is interesting to study the behavior of $V(d_{t/t+k}^a)$ when $k \to \infty$. First, trivially from the definition of λ_k, $\lim_{k\to\infty} \delta_k = \lim_{k\to\infty} \sum_{i=0}^{k} \lambda_i^2 = h_0$. Second, since $\lambda(B)\xi_s^0(B) = \nu_s^0(B)$, it is obvious that $\nu_{s0}^0 = \sum_{i=0}^{\infty} \xi_{si}^0 \lambda_i$ which implies that $\lim_{k\to\infty} \nu_{s0}^{0,t/t+k} = \nu_{s0}^0$. Futhermore, since the polynomial $\xi_s^0(B)$ converges in F, $V(r_{t/t+k}^0) = \sum_{i=k+1}^{\infty} l_i^2 = \sum_{i=k+1}^{\infty} (\xi_{si}^0)^2$ must go to zero as k becomes infinite. Replacing δ_k by h_0, ν_{s0}^0 and $V(r_{t/t+k}^0)$ by zero in lemma 5.4 directly yields the variance of the final error given in lemma 5.1. Therefore, as k goes to infinity, the error in the historical estimator is recovered.

The other case of interest is that of the concurrent estimator, for which $k = 0$. Trivially since $\lambda_0 = 1$, $\nu_{s0}^{0,t/t} = \xi_{s0}^0$ and $\delta_0 = 1$. The relationships (5.2) and (5.3) which give the variance of the revision and total error in the concurrent estimator are then immediately obtained.

We now turn our attention to the forecast of the components. The parameter k is now negative, and its absolute value gives the step-ahead forecast horizon.

Lemma 5.5 *The variance of the total estimation error in the forecast obtained at time $t+k$, $k<0$, of the signal for the period t is related to the noise repartition α, $\alpha \in [0,1]$, according to:*

$$var[d^{\alpha}_{t/t+k}] = var[d^{0}_{t/t+k}] + \alpha V_u. \tag{5.30}$$

Proof: This result is immediately obtained by inserting (5.25) and (5.28) in (5.24), and by adding the variance of the historical error given by lemma 5.1. ∎

Lemma 5.5 shows that the error in the component forecast is a linear function of the noise repartition, with a proportional factor of unity. This result follows from the fact that since white noise variables are not predictable, adding a white noise to a component directly increases the forecast error variance by the size of the white noise variance without changing the forecast. Notice that, at the difference of variances of the preliminary estimation errors, the forecast error variance on the signal and that on the nonsignal are not equal.

To summarize, in this section we have derived the relationships between the admissible decompositions and the error in both the historical and concurrent signal estimators, and then we have extended our results to the general case of the estimation of the signal s_t computed at any time $t+k$, $k=\cdots,-1,0,1,\cdots$. An attractive feature of the functions found is that they are particularly simple: they are always polynomials in α (the parameter representing the noise repartition), of order two, and with coefficients which are easily obtainable from the overall model for the observed series. We now use our results to build a procedure for selecting the decomposition minimizing the estimation error. This will lead us to some important features of the canonical assumption.

Chapter 6

Model specification minimizing the estimation error: Properties of the canonical decomposition

6.1 Historical estimators.

Lemma 6.1 *The final estimation error and the lag-0 covariance between the estimators are minimized at:*

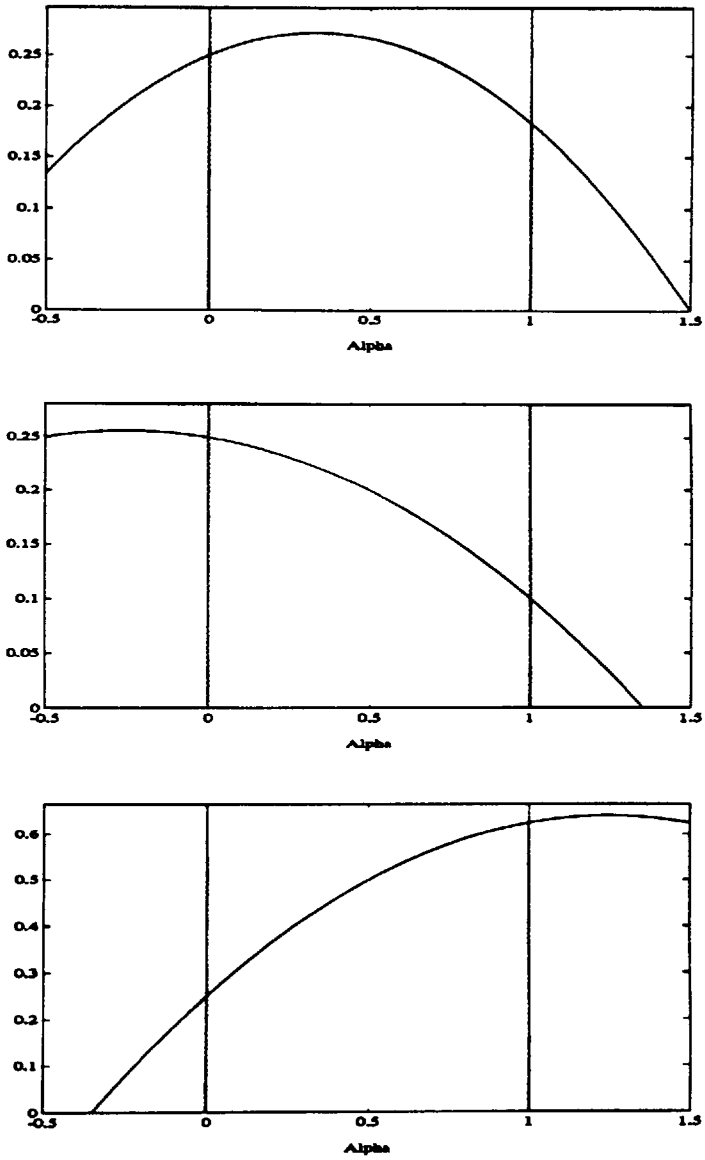
- $\alpha = 0$ *if* $2\nu_{s0}^0 + V_u h_0 < 1$*;*
- $\alpha = 1$ *otherwise.*

Proof: This result is a direct consequence of lemma 5.1. As given in (5.1), the variance of e_t^α is related to the term in α-squared through the coefficient $-h_0$. Since h_0 corresponds to the variance of the "inverse" model, h_0 is always positive and thus $V(e_t^\alpha)$ is a concave function of α over the interval $[0,1]$ (as in figure 5.1). Hence, one of the bounds of $[0,1]$ will necessarily provide the minimum. To locate the noise repartition minimizing the error variance, it is then enough to compare $V(e_t^1)$ and $V(e_t^0)$. Checking the sign of $V(e_t^1) - V(e_t^0) = V_u(1 - 2\nu_{s0}^0 - V_u h_0)$ directly yields the conditions stated in lemma (6.1). Notice that when $2\nu_{s0}^0 + V_u h_0 = 1$, $V(e_t^1) = V(e_t^0)$, and both bounds are solutions to the minimization problem. ∎

Over the range of the admissible decompositions, the variance of the final error may behave according to the three different patterns presented on figure 6.1. Clearly, each pattern is characterized by the position of the maximum, and that position also determines which bound yields the minimum error. The concavity and the symmetry of $V(e_t^\alpha)$ implies that the minimum error always corresponds to the bound that is the farthest away from the value of α maximizing the error. Denoting α_m this value, we have: $\alpha_m = (1 - 2\nu_{s0}^0)/2h_0V_u$. In the first case of figure 6.1, this maximum lies between 0 and 1 and corresponds to the situation where $.5 > \nu_{s0}^0 > .5 - h_0V_u$. The canonical signal will then minimize the error when ν_{s0}^0 is close to $.5 - V_u h_0$. Conversely, when ν_{s0}^0 becomes closer to .5 than to $.5 - V_u h_0$, the minimum switches to the upper bound of the interval $[0,1]$. The possibility that ν_{s0}^0 stands in the middle of $[.5 - h_0V_u]$, or equivalently that $\alpha_m = .5$, does exist, implying that both bounds are solutions to the minimization problem. We shall see in chapter 7 that situations where α_m lies in a close neighbourhood of .5, making similar the error obtained on the bounds of $[0,1]$, may actually be faced.

The second case displayed on figure 6.1 for which $\alpha_m < 0$ occurs when $\nu_{s0}^0 > .5$. The minimum error is then obtained on a signal concentrating

Figure 6.1: Three possible paths for the Final Error Variance



all the noise of the model. The third case of figure 6.1 is symmetric: we have $\alpha_m > 1$, which occurs when $\nu_{s0}^0 < .5 - h_0 V_u$, and the signal must be canonical to minimize the error. Notice that in these last two cases, the other canonical decomposition was maximizing the estimation error, and thus coincides with the Watson minimax filter, which would induce thus a noninvertible component.

Concerning the general aspect of each curve displayed on figure 6.1, it is interesting to notice that a low value of $h_0 V_u^2$ makes the shape of the curves more constant while as ν_{s0}^0 becomes closer to .5, the shape of the curves for α close to zero becomes more flat.

An obvious consequence of lemma 6.1 is a most appealing property of the canonical decompositions:

Corollary 6.1 *A canonical decomposition always minimizes the final estimation error and the lag-0 covariance between the estimators.*

Lemma 6.1 provides a procedure enabling us to identify the canonical decomposition which presents this property. However, it is possible to add a further assumption to model (A) in order to precisely isolate the decomposition solution of the minimization problem. It is also a way to standardize our notations.

Assumption 5.a: The canonical signal s_t^0 is such that: $\nu_{s0}^0 \geq \nu_{n0}^1$.

This assumption can be made without any loss of generality, since up to now s_t and n_t could be interchanged freely. The canonical requirement is now applied to the component which has the largest central coefficient in the WK filter designed to estimate its canonical form. Under assumption 5.a, we get the following result:

Lemma 6.2 *The final estimation error and the lag-0 covariance between the estimators are minimized over the range of all admissible decompositions with a canonical* n_t.

<u>**Proof:**</u> From lemma 6.1, the condition making $\alpha = 1$ solution of the minimization problem was $2\nu_{s0}^0 + V_u h_0 \geq 1$. The demonstration will just consist of showing that $\nu_{s0}^0 \geq \nu_{n0}^1$ set by assumption 5.a implies that this condition is satisfied. Our general model (A) may be respecified as:

$$x_t = s_t^0 + n_t^1 + u_t,$$

where s_t^0 and n_t^1 are the two components in their canonical form and u_t is the white noise present in the series, with variance V_u. Of course, assumption 1 (independence of the components) still holds. Then, using (4.2), the estimator of u_t is obtained as:

$$\begin{aligned} \hat{u}_t &= V_u \frac{\phi_x(B)\phi_x(F)}{\theta_x(B)\theta_x(B)} x_t \\ &= V_u h(B) x_t. \end{aligned}$$

Thus $V_u h_0$ represents the central coefficient of the WK filter estimating u_t. Since s_t^1 represents a signal concentrating all the noise of the model, it can simply be written as: $s_t^1 = s_t^0 + u_t$, and we have:

$$\nu_{s0}^1 = \nu_{s0}^0 + V_u h_0. \tag{6.1}$$

Now, since the estimators of s_t^1 and n_t^1 must sum to the observed series x_t, we have: $\nu_{s0}^1 + \nu_{n0}^1 = 1$. So assumption 5.a may be rewritten as: $\nu_{s0}^0 + \nu_{s0}^1 \geq 1$. Replacing ν_{s0}^1 in this last inequality by the expression (6.1) directly yields:

$$2\nu_{s0}^0 + V_u h_0 \geq 1,$$

which according to lemma 6.1 implies that we are in the case in which $\alpha = 1$ minimizes the estimation error variance, and a canonical n_t is the solution of the minimization problem. ∎

Lemma 6.2 provides a simple procedure to design the Unobserved Components models in order to minimize the component estimation error. Roughly speaking, it states that all the noise of the model must be assigned to the component which is relatively more important. This relative importance is evaluated by comparing the central weight of the WK filters designed to estimate the components in their canonical forms. A very simple way to compute the central weight of the WK filter as the variance of a simple ARMA model has been described in section 5.1.

Several reasons make such a model specification desirable. Firstly, over the range of observationally equivalent decompositions, the model specification selected will provide the most precise estimators. That is, given the model for the observed series, this decomposition will yield the highest coherence between the spectrum of the signal and the spectrum of the estimator. Nerlove (1964) saw such a coherence as a desirable feature when he discussed in the seasonal adjustment context several spectral criteria that adjustment procedures should satisfy. Specifying the unobserved components models as suggested in lemma 6.2 will thus reduce the discrepancies between the model of the signal and the model of its estimator. These discrepancies have been studied in the time domain in section 4.2.2 and for the frequency domain in section 4.4.2. We recall that this very attractive feature of the identification procedure is obtained by minimizing also the covariance between the estimators.

It is interesting to notice that the Watson's minimax filter yields opposite properties. This procedure amounts to conducting a MMSE estimation on the unobserved components model specification which maximizes the final estimation error. Thus, the implicit model specification maximizes the underestimation of the components, the lag-0 covariance between the estimators, and more generally the discrepancies between the model of the theoretical component and its estimator.

We now turn our attention to the error in the concurrent estimator.

6.2 Concurrent estimators.

Lemma 6.3 *The variance of the revision error in the concurrent estimator of the signal s_t^α is maximized at:*

- $\alpha = 0$ *if* $2(\nu_{s0}^0 - \xi_{s0}^0) + V_u(h_0 - 1) \leq 0$;
- $\alpha = 1$ *otherwise.*

Proof: In lemma 5.2, $V(r_t^\alpha)$ is related to the term in α-squared through the coefficient $V_u^2(h_0 - 1)$. Since $h_0 > 1$, $V(r_t^\alpha)$ is a convex function of α. Hence, it will always be maximized at one of the bounds of the finite interval $[0, 1]$. Comparing $V(r_t^1)$ and $V(r_t^0)$ directly yields the conditions stated in lemma 6.3. ∎

We can use lemmas 5.2 and 6.3 to pursue a discussion opened in Bell and Hillmer (1984) about the dependence between the component model specification and the size of the revision in the seasonal adjustment context. The decompositions that they considered were the two canonical decompositions and the model specification implicit in the X-11 filter. Their discussion was centered on the particular model for the observed series:

$$(1 - B)(1 - B^{12})x_t = (1 - \theta_1 B)(1 - \theta_{12}B^{12})a_t,$$

and the series x_t were then decomposed into a seasonal and a nonseasonal component. In this model, we choose the signal s_t^0 as representing a canonical seasonal component. For three set of parameters, they obtained the following results:

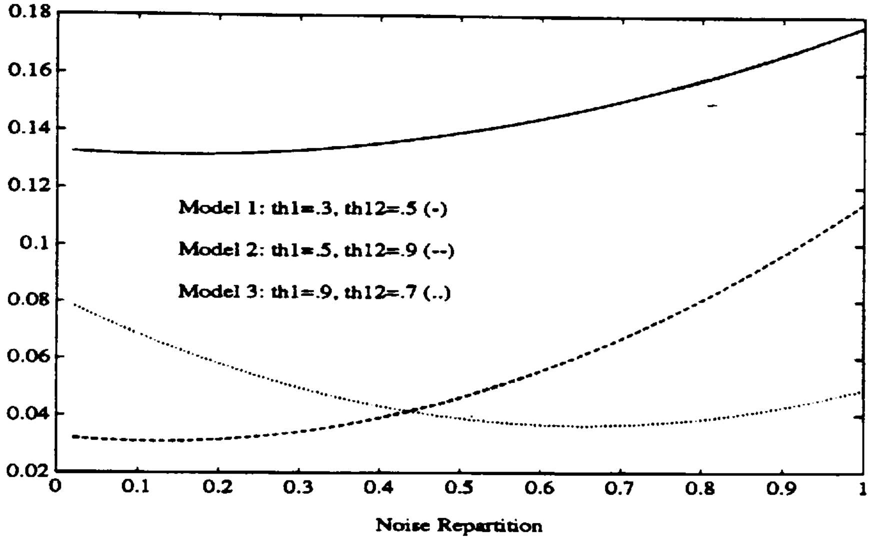
Table 6.1 (reproduced from Bell and Hillmer (1984, p.308))
Mean Squared Revisions.

Parameters:	X-11	Noninvertible Seasonal	Noninvertible Trend
Model 1: $\theta_1 = .3$, θ_{12}=.5	.123	.133	.177
Model 2: $\theta_1 = .5$, θ_{12}=.9	.059	.032	.114
Model 3: $\theta_1 = .9$, θ_{12}=.7	.130	.079	.049

For the decompositions considered, the minimum revisions were obtained with X-11 for the first model, with a noninvertible seasonal component for the second model and with a noninvertible trend for the third model. Figure 6.3 presents the behaviour of the revisions over the range of all admissible decomposition for the three models. The curves have been obtained using lemma 5.2.

For the first model, the minimum revisions are obtained at .1316, corresponding to a seasonal component embodying 16 percent of the pure noise part of the model. Assigning all this noise to the seasonal component and thus specifying a canonical trend maximizes the revisions at .177. This agrees with

Figure 6.2: Mean Squared Revisions in three ARIMA$(0,1,1)_{12}(0,1,1)$ models



lemma 6.3 since for this particular model we found: $\nu_{s0}^{0} = .242$, $\xi_{s0}^{0} = .274$, $V_u = .238$ and $h_0 = 2.051$, so $2(\nu_{s0}^{0} - \xi_{s0}^{0}) + V_u(h_0 - 1) = .184 > 0$. Notice that in this model, X-11 yields the lowest revisions, but it has to be remenbered that it is not a model-based approach, so any comparison is difficult to interpret.

With the second model, the minimum mean squared revisions is reached at .0309 when nearly 12 percent of the pure noise part of the series is assigned to the seasonal component. Conversely, .114 obtained on a noninvertible trend is a maximum over all admissible decompositions. As expected, this result is confirmed by lemma 5.2 since for this model: $\nu_{s0}^{0} = .046$, $\xi_{s0}^{0} = .067$, $V_u = .508$ and $h_0 = 1.403$, so $2(\nu_{s0}^{0} - \xi_{s0}^{0}) + V_u(h_0 - 1) = .162 > 0$. Here, the revision errors yielded by X-11 correspond to the revisions obtained when approximatively 60 percent of the noise is assigned to the seasonal component.

With the third model, a minimum of .0366 is obtained on a seasonal component concentrating nearly 66 percent of the noise of the model. The mean squared revisions of .079 corresponding to a noninvertible seasonal component is a maximum over all admissible decompositions. Calculating the condition stated in lemma 6.3 with, for this model, $\nu_{s0}^{0} = .152$, $\xi_{s0}^{0} = .254$, $V_u = .639$ and $h_0 = 1.246$, yields $-.203 < 0$, which indicates that $\alpha = 0$ is the revision maximizing noise repartition. Here, the X-11 filter yields much higher revisions (.130).

Several points come out of our analysis. Firstly, in the three examples and for the model-based approaches, the maximum mean squared revisions is always obtained at one bound of $[0,1]$. This result is in fact general, and constitutes an obvious consequence of lemma 6.3 that we state as a corollary:

Corollary 6.2 *The variance of the revision error on the concurrent estimator of the components is always maximized with a canonical decomposition.*

Corollary 6.2 generalizes Maravall (1986)'s result to any signal/nonsignal decompositions. Latter, in section 3, we will see that in signal plus noise decompositions, applying the rule provided by lemma 6.3. to determine which model specification maximizes the revision error yields results in agreement with Maravall's findings. Corollary 6.2 points out an unpleasant feature of the canonical decompositions. As we shall see, this unpleasant feature is of secondary importance.

Second, Bell and Hillmer argued that "the magnitude of revisions for a given model depends dramatically on the relevant final adjustment". For the three examples and for the model-based approach, we have evaluated the scale of this dependence. The ranges of variation of the revision obtained with the model-based approach are actually larger than those suggested by

Bell and Hillmer. The best illustration of their statement is provided by the second model, for which the size of the revisions is nearly multiplied by four over the range of all admissible decompositions.

Bell and Hillmer concluded by saying that "no one method gives the lowest mean squared revisions", and thus argued that the use of the revisions to evaluate different seasonal adjustment procedures was inappropriate. They were of course right for the decompositions considered: in the three cases, the specification minimizing the revision variance always corresponded to an intermediate noise repartition. However, Lemma 5.2 gives us the opportunity to build a procedure indicating the model specification minimizing the variance of the total revision on the concurrent estimates:

Lemma 6.4 *The noise repartition $\alpha \in [0,1]$ minimizing the variance of the revisions over the range of all admissible decompositions is given by:*

$$\begin{aligned}
\alpha^* &= -\frac{\nu_{s0}^0 - \xi_{s0}^0}{V_u(h_0 - 1)} \quad \text{if} \quad -\frac{\nu_{s0}^0 - \xi_{s0}^0}{V_u(h_0 - 1)} \in [0,1], \\
\alpha^* &= 0 \quad \text{if} \quad -\frac{\nu_{s0}^0 - \xi_{s0}^0}{V_u(h_0 - 1)} < 0, \\
\alpha^* &= 1 \quad \text{otherwise.}
\end{aligned}$$

Proof: The proof is immediately obtained by minimizing the function $V(r_t^\alpha)$ given in lemma 5.2 with respect to α in the interval $[0,1]$. ∎

Clearly, the specification minimizing the revisions depends on the stochastic properties of the observed series. Notice that there are two cases in lemma 5.4 where this specification corresponds to a canonical decomposition, Hence, one of the two canonical decompositions may also provide the most precise preliminary approximation to the final estimator. However, since the revision error is additional to the final estimation error, it is may be more appropriate to minimize directly the total estimation error on the concurrent estimates. To this we turn next.

Lemma 6.5 *The variance of the Total Estimation Error on the concurrent signal estimator is minimized for $\alpha \in [0,1]$ at:*

- $\alpha = 0$ *if* $2\xi^0_{s0} + V_u < 1$;
- $\alpha = 1$ *otherwise.*

<u>**Proof:**</u> Lemma 5.3 shows that the mean squared total estimation error is a concave function of α. The minimum is thus obtained on one of the bounds of any finite interval. Comparing $V(d^1_t)$ and $V(d^0_t)$ yields the conditions stated in the lemma. ∎

Lemma 6.5 implies the following property of the canonical decompositions:

Corollary 6.3 *The variance of the total estimation error on the concurrent estimates is always minimized with one of the canonical decompositions.*

As in the previous section, we can add without loss of generality an assumption in model (A) in order to precisely isolate the canonical decomposition with the property of Corollary 6.3. It is another possible way to standardize the notations. This assumption, which replaces Assumption 5.a, is stated as:

Assumption 5.b: The signal s^0_t is such that $\xi^0_{s0} > \xi^1_{n0}$,

where ξ^1_{n0} is the central coefficient of the WK filter expressing the estimator of a canonical nonsignal as function of the innovations a_t on the observed series. Under assumption 5.b, we obtain the following lemma equivalent to lemma 6.5:

Lemma 6.6 *Among all admissible decompositions, the one with canonical n_t always minimizes the error in the concurrent estimator.*

Proof: The proof consists of showing that $\xi^0_{s0} > \xi^1_{n0}$ implies that $2\xi^0_{s0}+V_u > 1$. By construction, the coefficients ξ^α_{n0} and ξ^α_{s0} are such that: $\xi^\alpha_{s0} + \xi^\alpha_{n0} = 1$. So $\xi^0_{s0} > \xi^1_{n0}$ is equivalent to:

$$\xi^0_{s0} + \xi^1_{s0} > 1. \tag{6.2}$$

Decomposing the observed series as $x_t = s^0_t + n^1_t + u_t$, where u_t represents a white noise with variance V_u, the estimator of u_t is given by:

$$\begin{aligned} \hat{u}_t &= \xi_u(B)a_t = \\ &= V_u \frac{\phi_x(F)}{\theta_x(F)} a_t, \end{aligned}$$

which implies $\xi_{u0} = V_u$. Since $s^1_t = s^0_t + u_t$, we have under the hypothesis of independence: $\xi^1_s(B) = \xi^0_s(B) + \xi_u(B)$ so: $\xi^1_{s0} = \xi^0_{s0} + V_u$. Inserting this last expression in (6.2) directly yields the condition stated in lemma 6.5 for $\alpha = 1$ to be the solution of the error minimization problem. ∎

Lemma 6.6 provides an easy alternative method of determining the decomposition that can be best estimated: for each canonical component, compare the central coefficient of the WK filters expressing the estimators in terms of the innovations a_t, and keep canonical the component associated with the smallest central coefficient. In other words, the canonical estimator most affected at time t by the innovation on the observed series must be assigned all the noise of the model. We may thus say that the maximum weight must be given to the relatively more important component. A simple look at the location of the maximum of $V(d_T^\alpha)$ indicates that a failure to obey this rule and making instead s_t canonical while assumption 5.b is fulfilled leads to the maximization of the error when $\xi_{s0}^0 > .5$.

The aggregation of the historical estimation error and of the revision error preserves the property of the canonical decompositions that they yield the most precise estimator. However, it is possible that while one canonical decomposition minimizes the error in the historical estimator, the other one minimizes the error in the concurrent estimator. In that case, the aggregation of the two types of errors makes the noise repartition minimizing the error switch to one bound of $[0,1]$ to the opposite bound. This happens when assumptions 5.a and 5.b do not simultaneously hold, for example when $\nu_{s0}^0 > \nu_{n0}^1$ and $\xi_{s0}^0 < \xi_{n0}^1$. Notice that since the revision error variance is always maximized at one bound of $[0,1]$, the switching of solutions means that the decomposition minimizing the final error variance is also the one which maximizes the revision error in the concurrent estimator.

We now check our results on the particular signal plus noise decomposition.

6.3 Example: Signal plus Noise decomposition

The problem of dealing with a time series contaminated by an additional noise is encountered frequently in economics (see for example Pagan (1975)) and in engineering. The typical representation can be presented as follows: the observed series x_t is assumed to be made up of a signal s_t and of a white noise n_t according to:

$$x_t = s_t + n_t.$$

The signal and the observed series are assumed to follow ARIMA models satisfying the assumptions of model (A). Setting the nonsignal as a white noise implies that $\phi_x(B) = \phi_s(B)$. The signal plus noise decomposition is thus a particularly simple case of model (A). It is worthwhile to consider such a model specification for the analysis of the estimation errors since the results can be easily anticipated: it is clear that if we assign all the noise of the model to the signal, no estimation error is made since the signal would be defined as the observed series. So the noise repartition minimizing the estimation errors variances must correspond to the minimum noise extraction. Let us now see if our results are consistent with this expectation. The variance of n_t is such that: $V_n \in [0, V_u]$, where V_u is here the minimum of the spectrum of the observed series. For each value of V_n within this interval, a particular decomposition is obtained. We set: $V_n = V_u$, so that n_t concentrates all the noise of the model and the signal s_t is canonical.

We first deal with the error on the historical estimator. Lemma 6.2 suggests to compare ν_{s0}^0 and ν_{n0}^1. Trivially, removing all the noise from n_t yields: $n_t^1 = 0$, so $\nu_{n0}^1 = 0$. According to lemma 6.2, the signal must concentrate all the noise of the model to minimize the final error variance if: $\nu_{s0}^0 > \nu_{n0}^1 = 0$. This condition must be satisfied since we have seen in subsection 5.1.1 that ν_{s0}^0 is the variance of an ARMA process with AR polynomial $\theta_x(B)$, MA polynomial $\theta_s(B)\phi_n(B)$ and innovation variance V_s^0. Lemma 6.2 yields thus a result consistent with our expectation.

We now focus on the mean squared total revision error. Maravall (1986) showed that the revisions on the concurrent estimator of the signal are maximized when a noninvertible signal is specified. Lemma 6.3 indicates that we should check the sign of $2(\nu_{s0}^0 - \xi_{s0}^0) + V_u(h_0 - 1)$.

Using $\nu_{n0}^0 = V_u h_0$, $\xi_{n0}^0 = V_u$, and the fact that ν_{s0}^0, ν_{n0}^0, and ξ_{s0}^0, ξ_{n0}^0 respectively sum to unity, it is immediately obtained that

$$\nu_{s0}^0 = 1 - V_u h_0.$$

and:

$$\xi_{s0}^0 = 1 - V_u.$$

So:

$$\begin{aligned} 2(\nu_{s0}^0 - \xi_{s0}^0) + V_u(h_0 - 1) &= 2(-V_u h_0 + V_u) + V_u(h_0 - 1) = \\ &= V_u(1 - h_0). \end{aligned}$$

Since, by construction, $h_0 > 1$, we get: $V_u(1 - h_0) < 0$, which implies that the total revision error is maximized on the estimation of a noninvertible signal. This confirms Maravall's findings. Now applying the rule provided by lemma 5.4 to minimize the revision error, we would get as the solution to the minimization problem:

$$\begin{aligned} \alpha^* &= -\frac{\nu_{s0}^0 - \xi_{s0}^0}{V_u(h_0 - 1)} = \\ &= 1. \end{aligned}$$

Our procedure yields thus the expected results: the revision error in the concurrent estimate of the signal is minimized by assigning all the noise to the signal, that is when no noise extraction is performed.

For the total estimation error, lemma 6.6 indicates that it is enough to compare ξ_{s0}^0 and ξ_{n0}^1. Trivially, $\xi_{n0}^1 = 0$, so we have to check if $\xi_{s0}^0 = 1 - V_u$ is positive or not. Since $h_0 > 1$, $1 - V_u > 1 - V_u h_0$ which is positive as we have seen in the final estimator case. Lemma 6.6 thus demonstrates as expected that all the noise must be assigned to the signal in order to obtain the minimum total estimation error.

We thus can conclude that the noise extraction case validates the results presented in section 1 and 2. We now generalize these results to the estimators of s_t computed at any time $t+k$, $k = \cdots, -1, 0, 1, \cdots$.

6.4 Generalization to preliminary estimators and forecast.

About the preliminary estimators, lemma 5.4 has the following implication:

Lemma 6.7 *The variance of the error in the preliminary estimators is minimized at:*

- $\alpha = 0$ *if* $2\nu_{s0}^{0,t/t+k} + \delta_k V_u < 1$;
- $\alpha = 1$ *otherwise.*

<u>**Proof:**</u> This result is a consequence of the concavity of $V(d_{t/t+k}^{\alpha})$. As previously for the lemmas 6.3 and 6.5, the minimum is located on one bound of $[0,1]$. The condition expressed in lemma 6.8 is obtained by comparing $V(d_{t/t+k}^1)$ and $V(d_{t/t+k}^0)$. ∎

The property that a canonical decomposition always minimizes the total estimation error variance is thus true for any preliminary estimators. Writing $\nu_{n0}^{1,t/t+k}$ the central coefficient of the WK filter estimating at time $t+k$ a canonical nonsignal n_t^1, we can as previously add a futher assumption to model (A) in order to standardize the notations.

Assumption 5.c: The signal s_t^0 is such that $\nu_{s0}^{0,t/t+k} > \nu_{n0}^{1,t/t+k}$.

This assumption, which replaces assumption 5.b, allows us to isolate the decomposition minimzing the total estimation error:

Lemma 6.8 *Among all admissible decompositions, the one with canonical n_t always minimizes the error variance in the preliminary estimator for period t obtained at $t+k$, k being a positive integer.*

Proof: We have to show that $\nu_{s0}^{0,t/t+k} > \nu_{n0}^{1,t/t+k}$ implies that $2\nu_{s0}^{0,t/t+k} + \delta_k V_u > 1$. Since $\nu_{s0}^{\alpha,t/t+k} + \nu_{n0}^{\alpha,t/t+k} = 1$, assumption 5.c can be respecified as: $\nu_{s0}^{0,t/t+k} + \nu_{s0}^{1,t/t+k} > 1$. We thus have to show that $\nu_{s0}^{0,t/t+k} + V_u\delta_k = \nu_{s1}^{1,t/t+k}$. In the decomposition $x_t = s_t^0 + n_t^1 + u_t$, the estimator at time $t+k$ of the irregular component u_t is given by:

$$\begin{aligned}
\hat{u}_{t/t+k} &= \xi_u^{t/t+k} a_t = \\
&= E_{t+k}\xi_u(B)a_t = \\
&= E_{t+k}V_u\lambda(F)a_t = \\
&= E_{t+k}V_u(1+\lambda_1 F+\cdots)a_t = \\
&= V_u(1+\lambda_1 F+\cdots+\lambda_k F^k)a_t.
\end{aligned}$$

Since $\nu_u^{t/t+k}(B) = \xi_u^{t+k}\lambda(B)$, it is directly obtained using the last equation above and (5.17) that: $\nu_{u0}^{t/t+k} = V_u\sum_{i=0}^k \lambda_i^2 = V_u\delta_k$. The hypothesis of independence implying: $\hat{s}_t^1 = \hat{s}_t^0 + \hat{u}_t$, we have: $\nu_s^{1,t/t+k}(B) = \nu_{s0}^{0,t/t+k}(B) +$

$\nu_u^{t/t+k}(B)$. It is then true that: $\nu_{s0}^{1,t/t+k} = \nu_{s0}^{0,t/t+k} + V_u\delta_k$. The assumption 5.c is thus equivalent to: $2\nu_{s0}^{0,t/t+k} + \delta_k V_u > 1$, and lemma 6.9 is proved. ∎

Whatever is the set of observations available on x_t and the period for which we are interested in the signal, it is always true that the minimum variance of the total estimation error is obtained by assigning all the noise of the model to the canonical component s_t^0 or n_t^1 whose estimator gives the most weight to the particular realization x_t, the other one being let canonical.

About the forecasts of the components, using lemma 5.5, the following result is immediately obtained:

Lemma 6.9 *Over the range of all admissible decompositions, the variance of the forecast error on the signal is always minimized with the canonical specification.*

Noninvertible components are always best forecasted. This result is not surprising since adding an unpredictable white noise to a variable just increases the forecast error without changing the variable forecast.

Chapter 7

Examples

7.1 Methodology

These results provide the procedure to specify the Unobserved Component model in order to minimize the variance of the components estimation error. For four classes of models for the observed series, we will investigate the type of model specification that such an identification procedure would yield. The estimators on which we shall focus our attention are the two most important ones, namely the historical and the concurrent estimators.

We first present the methodology that shall be followed in each case. The procedure can be split into five main steps:

1. given the model for the observed series, derive the system of covariance equations and identify it with zero coefficient restrictions;

2. compute the minimum of the spectrum of each component; then adding them yield the variance V_u;

3. either derive $V(e_t^0)$, h_0,and ν_{s0}^0 using the models presented in subsection 5.1.1 and check if $2\nu_{s0}^0 + V_u h_0 > 1$ as suggested by lemma 6.1; or compute

the central coefficients ν_{s0}^0 and ν_{n0}^1 of the WK filters designed to estimate the canonical components, compare them and using lemma 6.2 deduce which canonical decomposition minimize the final error variance;

4. provide the corresponding model specification by factorizing the spectrum; an algorithm for factorizing a spectrum can be found in Maravall and Mathis (1994);

5. for concurrent estimators, replace in the last two steps ν_{s0}^0 and ν_{n0}^1 by ξ_{s0}^0 and ξ_{n0}^1 and use either lemma 6.5 instead of 6.1 or lemma 6.6 instead of 6.2.

7.2 A Trend plus Cycle example

We analyze the model discussed in section 2.5.3. To ease the readers understanding, we recall that the model for the observed series was:

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

This series was the sum of a trend and a cycle component, and making the former canonical, we obtained in (2.7):

$$\begin{aligned} \Delta s_t^0 &= (1+B)a_{st}^0 \qquad V_s^0 = .161 \\ (1+.7B)n_t^0 &= (1+.496B)a_{nt}^0 \qquad V_n^0 = .306. \end{aligned}$$

The variance of the pure noise part of this series is given by: $V_u = g_n^0(0) = .237$. The coefficient $V(e_t^0)$, ν_{s0}^0, and h_0 correspond to the variances of the models:

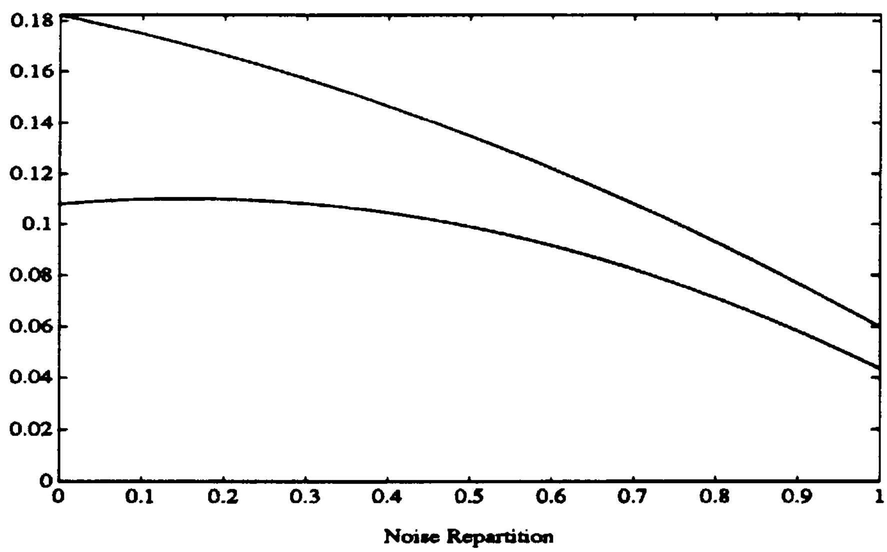
- for $V(e_t^0)$: $\theta(B)z_t = (1+B)(1+.496B)b_t$, $V_b = V_s^0 V_n^0 = .049$;
- for ν_{s0}^0: $\theta(B)z_t = (1+B)(1+.7B)b_t$, $V_b = V_s^0 = .161$;
- for h_0: $\theta(B)z_t = (1-B)(1+.7B)b_t$, $V_b = 1$;

where $\theta(B) = 1 + .404B - .039B^2$. That yields: $V(e_t^0) = .108$, $\nu_{s0}^0 = .440$ and $h_0 = 1.659$. Over the range of all admissible decompositions, the variance of the final estimation error is given by (from lemma 5.1):

$$V(e_t^\alpha) = .108 + .028\alpha - .093\alpha^2.$$

This function is plotted on figure 7.1. It is easily seen that the minimum is

Figure 7.1: Mean Squared Errors in an ARIMA(1,1,2) model



obtained at $\alpha = 1$, in which case $V(e_t^1) = .043$. The decomposition of x_t is thus best estimated when the cycle is canonical while the trend concentrates all the noise of the model. Removing $g_n^0(0)$ from $g_n^0(w)$ and factorizing the resulting spectrum $g_n^1(w)$, the corresponding models for the components are found as:

$$\begin{aligned} \Delta s_t^1 &= (1 - .096B)a_{st}^1 \qquad V_s^1 = .788 \\ (1 + .7B)n_t^1 &= (1 + B)a_{nt}^1 \qquad V_n^1 = .014. \end{aligned}$$

Figure 7.1 also shows that the maximum estimation error variance is reached for an intermediate decomposition where the trend component catches 15,3% of the variance of the noise of the model. This decomposition is the one that the minimax filter would yield.

For concurrent estimation, we need to derive the filter $\xi_s^0(B)$. In the time domain, this filter can be written as:

$$\xi_s^0(B) = V_s^0 \frac{(1+B)(1+F)}{1-B} \frac{1 + .7F}{1 + .404F - .039F^2},$$

which, after calculation, yields $\xi_{s0}^0 = .642$. In the same way, the revision error variance may simply be obtained as $\sum_{i=1}^{\infty}(\xi_{s0}^0)^2$, as (4.11) suggests, and that yields $V(r_t^0) = .075$. So the relationship between error variance on the concurrent estimator and noise repartition is given by:

$$V(d_t^\alpha) = .183 - .067\alpha - .056\alpha^2.$$

This function is also plotted on figure 7.1.

Again, the minimum is reached at .060 on a trend component concentrating all the noise of the model. Concurrent estimation of a canonical trend yields instead a maximum mean squared error of .183. While specifying a canonical cycle minimizes the error variance on both historical and concurrent estimators, in this example, this canonical specification also minimizes the mean squared revision errors.

7.3 Regular ARIMA model

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

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

This simple model, adequate for bi-annual data, has been the subject of a pedagogical discussion in Maravall and Pierce (1987).

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

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

For these models, we have the overall relationship:

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

which provides a system of 3 covariances equations with the four unknowns θ_s, V_s, θ_n, V_n. The system is thus not identified. We overcome this problem by imposing zero-coefficients restrictions on both components. The general expression for factorizing the spectrum of x_t is given by:

$$\frac{1+\theta^2-2\theta cosw}{2-2cos2w} = V_s\frac{1}{2+2cosw} + V_n\frac{1}{2-2cosw}, \tag{7.1}$$

where it is easily obtained that $V_s = (1+\theta)^2/4$ and $V_n = (1-\theta)^2/4$. The amount of noise embodied in the spectra $g_n(w)$ and $g_s(w)$ is given by:

$$\begin{aligned} \epsilon_s &= \min_w g_s(w) = g_s(0) = (1+\theta)^2/16, \\ \epsilon_n &= \min_w g_n(w) = g_n(\pi) = (1-\theta)^2/16, \end{aligned} \tag{7.2}$$

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

$$V_u = \epsilon_s + \epsilon_n = (1+\theta^2)/8. \tag{7.3}$$

Removing ϵ_s from the ACGF of s_t, the central coefficient of the WK filter estimating a canonical seasonal component can be obtained as:

$$\nu_{s0}^0 = [\frac{(1+\theta)^2}{4(1+B)(1+F)} - (1+\theta)^2/16] \cdot [\frac{(1-B^2)(1-F^2)}{(1-\theta B)(1-\theta F)}]_{B=F=0}.$$

Developing, we have:

$$\nu_{s0}^0 = [\frac{(1+\theta)^2(1-B)(1-F)}{4(1-\theta B)(1-\theta F)} - \frac{(1+\theta)^2(1-B^2)(1-F^2)}{16(1-\theta B)(1-\theta F)}]\ |_{B=F=0}=$$

$$= \frac{1+\theta}{2} - \frac{(1+\theta)^2}{8}.$$

Futher simplification of ν_{s0}^0 will not be necessary. We now develop the same analysis to obtain ν_{n0}^1, the central coefficient of the WK filter estimating a canonical trend:

$$\nu_{n0}^1 = [\frac{(1-\theta)^2}{4(1-B)(1-F)} - (1-\theta)^2/16] \cdot [\frac{(1-B^2)(1-F^2)}{(1-\theta B)(1-\theta F)}]_{B=F=0}.$$

Developing, we have:

$$\nu_{n0}^1 = [\frac{(1-\theta)^2(1+B)(1+F)}{4(1-\theta B)(1-\theta F)} + \frac{(1-\theta)^2(1-B^2)(1-F^2)}{16(1-\theta B)(1-\theta F)}] \,|_{B=F=0} =$$

$$= \frac{1-\theta}{2} - \frac{(1-\theta)^2}{8},$$

Comparing the two coefficients, it is easily seen that $\nu_{s0}^0 - \nu_{n0}^1 = \theta/2$ which is positive when $\theta > 0$.

So from lemma 6.2, when θ is positive, the final estimation error and the lag-0 covariance between the estimators of the components are minimized by assigning all the noise of the observed series to the seasonal component s_t, the nonseasonal component n_t being thus a canonical trend. Conversely, θ being negative implies that the MSE and the estimators lag-0 covariance are minimized by considering a canonical seasonal component.

Notice that as θ becomes closer to 1, due to the nearly cancelling AR and MA unit roots of 1, the trend becomes more stable. Our previous analysis indicates that in this case, a canonical trend gives the most accurately estimated decomposition. About the seaonal component, a similar reasoning is valid with θ close to -1. This illustrates the general rule that the most precisely estimated decompositions are the ones where the most stable component is made canonical.

To derive any conclusion concerning the estimation error on the concurrent estimator, we need the coefficient of B^0 in the WK filter estimating the concurrent canonical components. Since we have seen that $\nu_{p0}^{\alpha,t/t} = \xi_{p0}^{\alpha}$, for $p = s,n$, we focus on the central coefficient of the WK expressing the canonical estimators in terms of the innovations a_t. For a canonical seasonal component, we have:

$$\begin{aligned}\xi_{s0}^{0} &= [\frac{(1+\theta)^2}{4}\frac{1}{(1+B)(1+F)} - (1+\theta)^2/16]\cdot[\frac{(1-F^2)}{(1-\theta F)}]_{B=F=0} = \\ &= \frac{1+\theta}{2} - \frac{(1+\theta)^2}{16}.\end{aligned}$$

Proceeding similarly for the canonical signal, we have:

$$\begin{aligned}\xi_{n0}^{1} &= [\frac{(1-\theta)^2}{4}\frac{1}{(1-B)(1-F)} - (1-\theta)^2/16]\cdot[\frac{(1-F^2)}{(1-\theta F)}]_{B=F=0} = \\ &= \frac{1-\theta}{2} - \frac{(1-\theta)^2}{16}.\end{aligned}$$

Comparing ξ_{s0}^{0} and ξ_{n0}^{1}, it follows immediately that $\xi_{s0}^{0} - \xi_{n0}^{1} = 3\theta/4$. So, from lemma 6.4, when θ is positive, making canonical the trend component component yields a total estimation error minimized. Conversely, specifying a canonical model for the seasonal component when θ is negative maximizes the total error variance. In this example, it can be easily checked that the canonical decomposition minimizing the error also maximizes the revision error. However, the final and the concurrent error are always minimized with the same decomposition.

A last point concerns the special case where $\theta = 0$. It is immediately checked that for this value for θ, $\xi_{s0}^0 = \xi_{n0}^1$ and that $\nu_{s0}^0 = \nu_{n0}^1$. So, in that case, both canonical decomposition minimize the error variance on the final and on the concurrent estimators.

7.4 Seasonal ARIMA model

Suppose the observed series x_t follows the model :

$$(1 - B^4)x_t = (1 - \theta B)e_t,$$

$$|\theta| < 1 \qquad e_t \sim NIID(0,1).$$

This model is designed for quaterly time series, and has also been used for illustration in Kohn and Ansley (1986). Suppose we wish to decompose x_t according to : $x_t = s_t + n_t$ where s_t and n_t represent respectively the seasonal and nonseasonal component of the observed series. We then specify the components as:

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

$$(1 - B)n_t = (1 - \theta_n B)a_{nt},$$

where $U(B) = 1 + B + B^2 + B^3$ and $\theta_s(B) = 1 + \theta_{s1}B + \theta_{s2}B^2 + \theta_{s3}B^3$. The polynomial $U(B)$ characterizes the seasonal behavior of the observed series since it implies infinite peaks in the spectrum of the observed series at the frequencies $k\pi/2$, $k = 1, 2$. Similarly, the first difference operator $(1 - B)$ implies an infinite peak in the observed series spectrum at the zero frequency, and thus is associated with the trend behavior of the observed series.

For this decomposition, the overall relationship of the model is given by:

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

which yields a system of 5 equations with 6 unknowns, namely θ_{s1}, θ_{s2}, θ_{s3}, V_s, θ_n and V_n. This system is thus not identified. To reach identification, we impose some zero-coefficient restrictions: $\theta_{s3} = \theta_n = 0$.

The spectrum $g_x(w)$, $w \in [0, \pi]$, of the series x_t may then be written as:

$$\begin{aligned} g_x(w) &= \frac{1 + \theta^2 - 2\theta \cos w}{2 - 2\cos 4w}, \\ &= \frac{V_n}{2 - 2cosw} + \frac{\gamma_0 + \gamma_1 cosw + \gamma_2 cos2w}{4 + 6cosw + 4cos2w + 2cos3w}, \end{aligned}$$

where $\gamma_0 = V_s(1 + \theta_{s1}^2 + \theta_{s2}^2)$, $\gamma_1 = \theta_{s1}(\theta_{s2} + 1)V_s$ and $\gamma_2 = \theta_{s2}V_s$. Using the equality: $2\cos(jw)\cos w = \cos(j - 1)w + \cos(j + 1)w$, we can build the following linear system of equations:

$$\begin{aligned} 1 + \theta^2 &= 4V_n + 2\gamma_0 - \gamma_1 \\ -2\theta &= 6V_n + 2\gamma_1 - 2\gamma_0 - \gamma_2 \\ 0 &= 4V_n + 2\gamma_2 - \gamma_1 \\ 0 &= 2V_n - \gamma_2, \end{aligned}$$

which yields the solution:

$$
\begin{aligned}
V_n &= (1-\theta)^2/16 \\
\gamma_0 &= \theta + 10V_n = \\
&= (10 - 4\theta + 10\theta^2)/16 \\
\gamma_1 &= 8V_n = \\
&= (1-\theta)^2/2 \\
\gamma_2 &= 2V_n = (1-\theta)^2/8.
\end{aligned}
$$

Having identified the spectra of the components for one particular decomposition, we consider the amount of pure noise that they embody. The seasonal component spectrum has a minimum ϵ_s at the zero frequency when $\theta > -.347$, and at a frequency $w^* \in]\pi/2, \pi[$ otherwise. In this latter case, the algebraic expression for ϵ_s is non-trivial, so for the sake of simplication we restrict the analytical derivations to the values of $\theta > -.347$. The graphs provided however consider the complete interval $]-1,1[$ for θ.

For $\theta > -.347$, the minimum of the seasonal component spectrum is given by:

$$
\begin{aligned}
\epsilon_s = \min_w g_s(w) &= g_s(0) = \\
&= (20V_n + \theta)/16 = \\
&= (5 - 6\theta + 5\theta^2)/64.
\end{aligned}
$$

For the nonseasonal component, the amount of noise embodied in the spectrum of n_t is:

$$\min_{w} g_n(w) = g_n(\pi) = (1-\theta)^2/64$$

So, the size of the pure noise part of the observed series corresponds to:

$$V_u = g_n(\pi) + g_s(0) = \frac{3 - 4\theta + 3\theta^2}{32}$$

We need now to calculate ν_{s0}^0. The analytical derivations may be presented by computing ν_{n0}^0 and then using $1 = \nu_{s0}^0 + \nu_{n0}^0$. Writing the WK filter for estimating a nonseasonal component concentrating all the noise of the model, in the time domain we have:

$$\nu_{n0}^0 = [\frac{(1-\theta)^2}{16(2-(B+F))} + \epsilon_s] \cdot [\frac{2-(B^4+F^4)}{1+\theta^2-\theta(B+F)}] \mid_{B=F=0} ,$$

which, after simplification, leads to:

$$\nu_{n0}^0 = \frac{2-\theta-\theta^3}{8} + 2(1+\theta^2)\epsilon_s.$$

Replacing ϵ_s in this last expression by $\epsilon_s = (5 - 6\theta + 5\theta^2)/64$ for the case where $\theta > -.347$, and using $\nu_{s0}^0 + \nu_{n0}^0 = 1$, we get:

$$\nu_{s0}^0 = 1 - \frac{2-\theta-\theta^3}{8} + (1+\theta^2)\frac{5-6\theta+5\theta^2}{32}.$$

Now looking for the central coefficient of the WK filter estimating a canonical nonseasonal component, we have in the time domain:

$$\begin{aligned} \nu_{n0}^1 &= \frac{2-\theta-\theta^3}{8} - 2(1+\theta^2)\epsilon_n = \\ &= \frac{2-\theta-\theta^3}{8} - (1+\theta^2)(1-\theta)^2/32. \end{aligned}$$

Comparing ν_{s0}^0 and ν_{n0}^1, we have:

$$\begin{aligned}\nu_{s0}^0 - \nu_{n0}^1 &= 1 - 2\frac{2-\theta-\theta^3}{8} - (1+\theta^2)\frac{(5-6\theta+5\theta^2)-(1-\theta)^2}{32} = \\ &= \frac{2+2\theta+2\theta^3}{4} - \frac{(1+\theta^2)(4-4\theta+4\theta^2)}{32} = \\ &= \frac{12+20\theta-8\theta^2 3+20\theta^3-4\theta^4}{32}.\end{aligned}$$

The numerator N of this last expression can be factorised as:

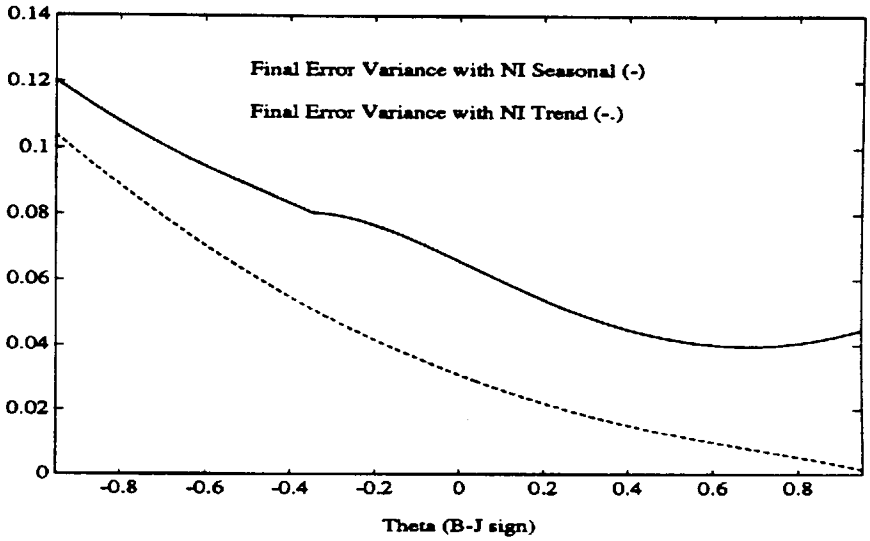
$$N = -4(\theta - 4.827)(\theta + .435)(\theta^2 - .608\theta + 1.43),$$

and it is readily checked that, for $\theta \in]-.347, 1[$, N is always positive. So, when $\theta \in]-.347, 1[$, $\nu_{s0}^0 > \nu_{n0}^1$, thus all the noise of the model must be assigned to the seasonal component in order to obtain the most precise final estimators. We have checked numerically that this result is still valid when $\theta \in]-1, -.347[$. Figure 7.2 displays the final error variance for the two canonical decompositions and for θ varying to -1 to 1. It also shows that the gain in precision obtained by specifying a noninvertible trend increases as θ increases.

To derive any conclusion about the error on the concurrent estimators, we need the central coefficient of the WK filter expressing the estimators in terms of the innovations on the observed series. For the estimator of a nonseasonal component concentrating all the noise of the model, this coefficient is given by:

$$\xi_{n0}^0 = [\frac{(1-\theta)^2}{16(1-B)(1-F)} + \frac{5-6\theta+5\theta^2}{64}][\frac{1-F^4}{1-\theta F}]_{B=F=0} =$$

Figure 7.2: Final Error Variance for Trend and SA Series



$$= \frac{(1-\theta)^2}{16}[1 + (1+\theta) + (1+\theta+\theta^2) +$$

$$+ \frac{1+\theta+\theta^2+\theta^3}{1-\theta}] + \frac{5-6\theta+5\theta^2}{64} =$$

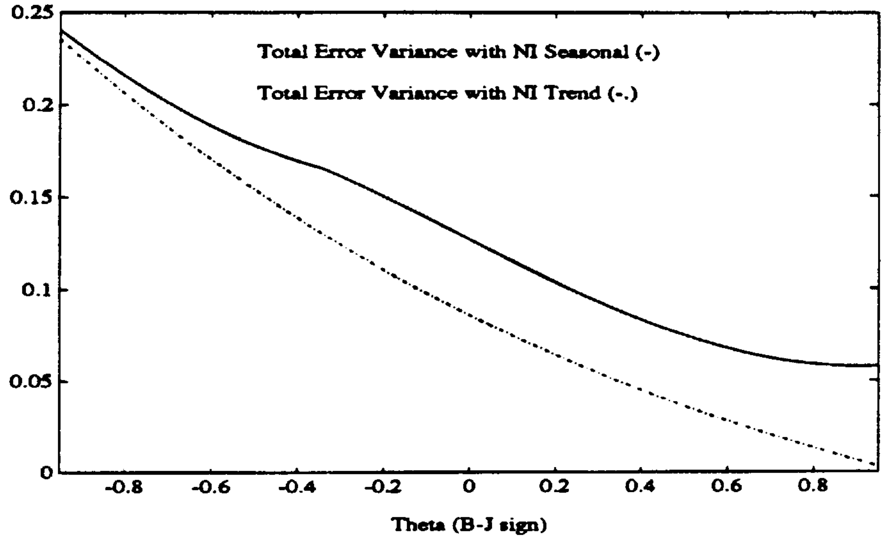
$$= \frac{1-\theta}{4} + \frac{5-6\theta+5\theta^2}{64},$$

and since $\xi_{s0}^0 + \xi_{n0}^0 = 1$, we get:

$$\xi_{s0}^0 = \frac{3+\theta}{4} - \frac{5-6\theta+5\theta^2}{64}.$$

Proceeding in a similar manner, it is easily obtained that the central coefficient of the WK filter estimating a noninvertible nonseasonal in terms of the innovations a_t can be easily obtained as:

Figure 7.3: Total Error Variance for Trend and SA series



$$\xi_{n0}^{1} = \frac{1-\theta}{4} - \frac{(1-\theta)^2}{64}.$$

Now comparing ξ_{s0}^{0} and ξ_{n0}^{1} as suggested in lemma 6.6, we have:

$$\begin{aligned}
\xi_{s0}^{0} - \xi_{n0}^{1} &= \frac{2+2\theta}{4} - \frac{4-4\theta+4\theta^2}{64} = \\
&= \frac{7+9\theta-\theta^2}{16} \\
&= -\frac{(\theta-9.72)(\theta+.72)}{16},
\end{aligned}$$

which is clearly positive for $\theta \in]-.347, 1[$.

Figure 7.3 displays the total error variance in concurrent estimators obtained with the two canonical decompositions. It is seen that specifying a noninvertible trend always minimizes the concurrent estimation error for $\theta \in]-1,1[$. Moreover, the gain in precision increases as θ increases. The ARIMA$(0,1,0)_4(0,0,1)$ models have the property that its seasonally adjusted and seasonal components are always best estimated when the seasonally adjusted series is specified as being noninvertible.

7.5 Airline model

The so-called airline model (see Box-Jenkins (1970), p. 305) constitutes a class of models which has been found to adequately represent a large number of quarterly or monthly economic time series. They are defined by the specification:

$$(1-B)(1-B^s)x_t = (1-\theta_1 B)(1-\theta_s B)a_t,$$

where a_t is a normally distributed white noise and s denotes the number of observations per year. The MA polynomial is constrained to be invertible, and for the observed series to admit a decomposition into a seasonal and a nonseasonal processes, a sufficient condition is that the parameter θ_s is non negative (see Hillmer and Tiao 1982). Then, the observed series x_t is decomposed into a seasonal process s_t with AR polynomial $\phi_s(B) = 1+B+\cdots+B^{s-1}$ and a nonseasonal process n_t with AR polynomial $\phi_n(B) = (1-B)^2$. If we make both processes canonical, then a third component comes out, the irregular component, necessary for the full complement of the decomposition. Because s_t and n_t are free of noise, the irregular component would have its variance maximized. Canonical signal/nonsignal decompositions may then be obtained either by assigning the irregular component to the canonical trend, producing a decomposition into Seasonally Adjusted (SA) series plus (canonical) seasonal component, or by assigning all the noise to the seasonal process, yielding a decomposition into (canonical) trend plus seasonal component.

Figures 7.4 and 7.5 present the canonical specifications which minimizes the error variance on the final estimators and on the concurrent estimators for θ_1 varying within $]-1,1[$ and θ_s within $[0,1[$. On the figures, FEE stands for Final Estimation Error in historical estimators, and TEE for Total Estimation Error in concurrent estimators. Data of two periodicity have been considered: quarterly and monthly.

In both cases, it is interesting to notice that when θ_1 is near to unity, the trend must be made canonical in order to minimize the mean squared final estimation error, while when θ_s is near to its upper bound, then the seasonal component must be held canonical, the remaining component being the seasonally adjusted series. Since $\theta_1 = 1$ corresponds to a deterministic trend component and $\theta_s = 1$ to a deterministic seasonal component, the solution to the error minimization problem seems to be related to the stochastic character of the components. These figures illustrate the result that to minimize the final estimation errors variances, the pure noise part of the observed series must be removed from the relatively more stable component and assigned to the more unstable component. The stochastic nature of the more unstable component is thus "reinforced". This result is in agreement with the analytical solutions obtained in the previous examples, and provides a further interpretation of lemmas 6.2 and 6.6.

On both figures, the continuous line represents the region where $\nu_{s0}^0 = \nu_{n0}^1$, in which case the two canonical decompositions yield the same mean squared

Figure 7.4: Monthly Airline models: specifications minimizing the errors variances

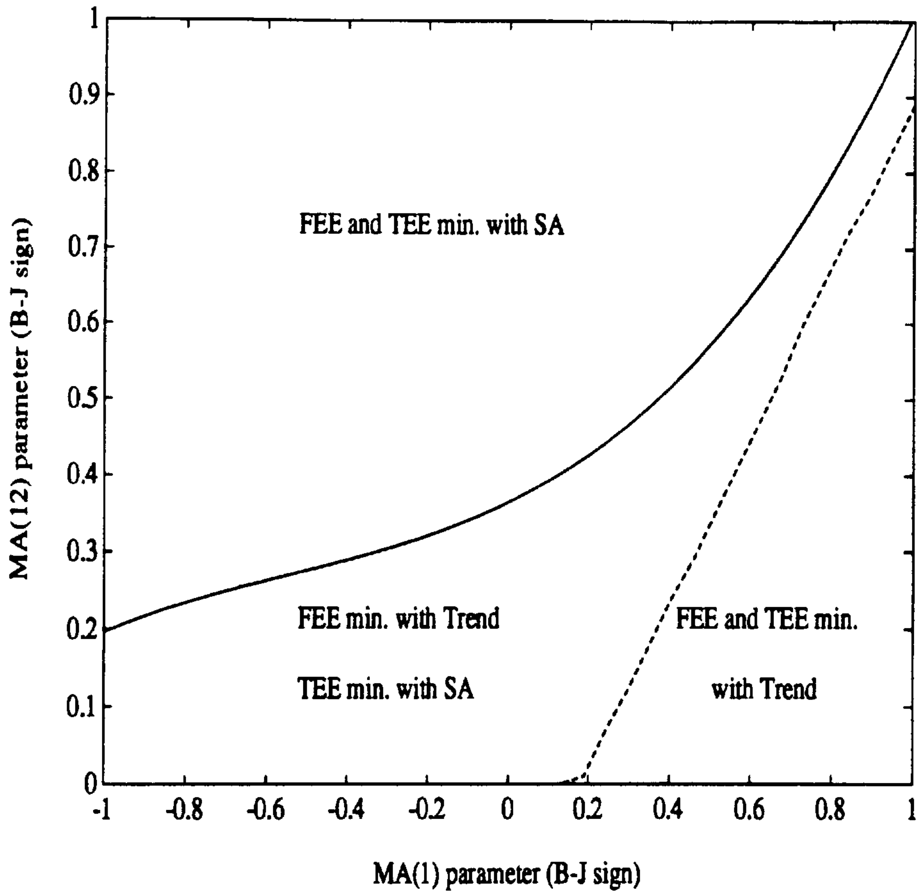
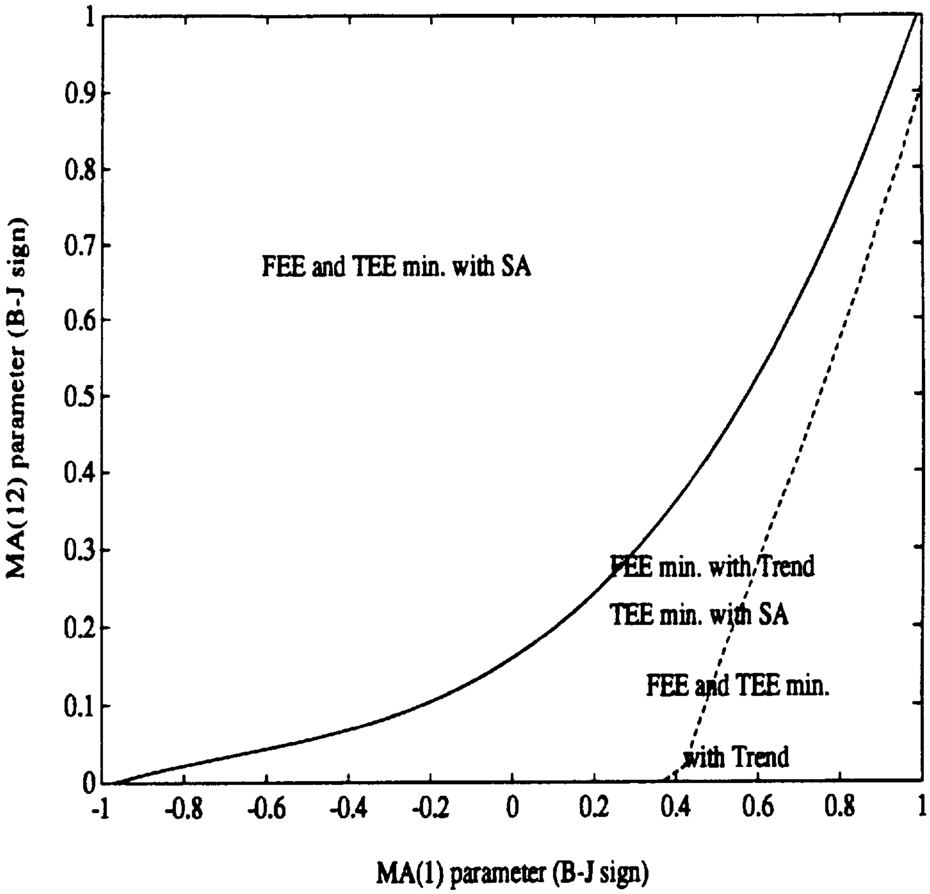


Figure 7.5: Quarterly Airline models: specifications minimizing the errors variances



final estimation error, and both provide components that are estimated the maximum precision. Similarly, the dashed line represents the region where $\xi_{s0}^0 = \xi_{n0}^1$, which implies that the two canonical decompositions yield the same error variance in the concurrent estimators. It is worth noting that there exist cases in which it is irrelevant to select a particular canonical decomposition on the basis on the estimation accuracy.

The space between the two curves represents the region where assumptions 5.a and 5.b do not hold simultaneously: the final estimation error variance is minimized with a canonical trend and the total error variance with a canonical seasonal component. That is, according to lemmas 6.2 and 6.4, we have: $\nu_{s0}^0 > \nu_{n0}^1$ and $\xi_{s0}^0 < \xi_{n0}^1$. The space between the two curves represents thus the region where there is a trade-off between final error and error in concurrent estimator, in the sense that minimizing the final error leads to the maximization of the revision error.

The figures also show the importance of specifying a canonical seasonal in airline type of models for minimizing the estimation errors. This is particularly clear for the case of the quarterly airline model, and in that of the concurrent estimators with monthly observations. This constitutes a comforting feature of seasonally adjusted series, since the ARIMA-model-based approach usually specifies noninvertible models for the seasonal component.

To see how much gain in precision we can expect from our identification procedure, we have computed the estimation errors variance for monthly and quarterly airline models, and for different values of (θ_1, θ_s). The results are displayed in tables 7.1 and 7.2.

A similar analysis can be found in Hillmer (1985) with different sets of parameters and only for noninvertible seasonal components. We consider instead the two canonical decompositions, and we also report, for each of parameters, the maximum estimation error variance. That maximum is associated with the minimax filter.

Our analysis confirms a result already underlined by Hillmer: the estimation errors variances decrease as θ_1 and θ_s increase. For both historical and concurrent estimators, the gain in precision obtained by selecting the best estimated canonical decomposition increases as θ_1 and θ_s increases. This gain may be of significant magnitude: for the historical estimators, when $(\theta_1, \theta_{12}) = (.75, 0)$, then the error variance goes from a maximum of .077 on the canonical seasonal to .019 on a canonical trend, that is smaller by more than a factor 3. Notice that in the vast majority of cases, quarterly airline models yield more precise estimators than the monthly airline models.

Table 7.1a:

Monthly Airline Model: Variance of Final Estimation Error.

θ_1	Model Spec.	$\theta_{12} = 0$	$\theta_{12} = .25$	$\theta_{12} = .5$	$\theta_{12} = .75$
-.75	Canonical Seas.	.410	.504	.436*	.259*
	Max. FEE	.410	.504	.439	.267
	Canonical Trend	.407*	.504	.439	.267
-.50	Canonical Seas.	.308	.377	.327*	.195*
	Max. FEE	.308	.378	.337	.220
	Canonical Trend	.300*	.376*	.337	.220
-.25	Canonical Seas.	.226	.274	.239*	.144*
	Max. FEE	.226	.276	.256	.190
	Canonical Trend	.210*	.271*	.255	.190
0	Canonical Seas.	.164	.197	.173*	.106*
	Max. FEE	.164	.200	.197	.170
	Canonical Trend	.138*	.186*	.191	.168
.25	Canonical Seas.	.121	.143	.129*	.081*
	Max. FEE	.121	.148	.160	.162
	Canonical Trend	.082*	.119*	.139	.146
.50	Canonical Seas.	.096	.113	.106	.070*
	Max. FEE	.096	.122	.145	.168
	Canonical Trend	.042*	.070*	.095*	.118
.75	Canonical Seas.	.077	.118	.116	.076
	Max. FEE	.077	.120	.152	.188
	Canonical Trend	.019*	.036*	.054	.074*

*: Minimum Variance of Final Estimation Error.

Table 7.1b: Monthly Airline Model: Variance of Total Estimation Error.					
θ_1	Model Spec.	$\theta_{12} = 0$	$\theta_{12} = .25$	$\theta_{12} = .5$	$\theta_{12} = .75$
	Canonical Seas.	1.257*	1.151*	.905*	.521*
-.75	Max. FEE	1.261	1.157	.913	.532
	Canonical Trend	1.261	1.157	.913	.532
	Canonical Seas.	.956*	.873*	.685*	.393*
-.50	Max. FEE	.964	.888	.710	.433
	Canonical Trend	.964	.888	.710	.433
	Canonical Seas.	.699*	.641*	.505*	.292*
-.25	Max. FEE	.710	.665	.551	.369
	Canonical Trend	.710	.665	.551	.369
	Canonical Seas.	.491*	.458*	.367*	.215*
0	Max. FEE	.498	.483	.426	.327
	Canonical Trend	.498	.483	.426	.327
	Canonical Seas.	.333	.323*	.269*	.164*
.25	Max. FEE	.333	.337	.324	.292
	Canonical Trend	.326*	.336	.324	.292
	Canonical Seas.	.228	.239	.214*	.139*
.50	Max. FEE	.228	.243	.250	.252
	Canonical Trend	.193 *	.217*	.234	.244
	Canonical Seas.	.149	.205	.207	.143*
.75	Max. FEE	.149	.205	.221	.236
	Canonical Trend	.097*	.120*	.141*	.161

*: Minimum Variance of Total Estimation Error.

Table 7.2a: Quaterly Airline Model: Variance of Final Estimation Error.					
θ_1	Model Spec.	$\theta_{12} = 0$	$\theta_{12} = .25$	$\theta_{12} = .5$	$\theta_{12} = .75$
	Canonical Seas.	.103	.103*	.081*	.045*
-.75	Max. FEE	.103	.107	.088	.056
	Canonical Trend	.102*	.107	.088	.056
	Canonical Seas.	.080	.080*	.064*	.037*
-.50	Max. FEE	.080	.087	.080	.066
	Canonical Trend	.078*	.087	.080	.066
	Canonical Seas.	.062	.064*	.054*	.032*
-.25	Max. FEE	.063	.073	.080	.084
	Canonical Trend	.058*	.073	.080	.084
	Canonical Seas.	.050	.056*	.050*	.031*
0	Max. FEE	.052	.064	.085	.103
	Canonical Trend	.043*	.064	.083	.103
	Canonical Seas.	.047	.059	.056*	.037*
.25	Max. FEE	.047	.071	.097	.125
	Canonical Trend	.033*	.058*	.085	.114
	Canonical Seas.	.048	.073	.071*	.046*
.50	Max. FEE	.048	.082	.115	.150
	Canonical Trend	.029*	.053*	.079	.108
	Canonical Seas.	.053	.092	.091	.060*
.75	Max. FEE	.053	.100	.140	.179
	Canonical Trend	.027*	.046*	.061*	.076

*: Minimum Variance of Final Estimation Error.

Table 7.2b: Quaterly Airline Model: Variance of Total Estimation Error.					
θ_1	Model Spec.	$\theta_{12} = 0$	$\theta_{12} = .25$	$\theta_{12} = .5$	$\theta_{12} = .75$
-.75	Canonical Seas.	.256*	.219*	.164*	.090*
	Max. FEE	.267	.231	.175	.102
	Canonical Trend	.267	.231	.175	.102
-.50	Canonical Seas.	.210*	.180*	.135*	.075*
	Max. FEE	.225	.201	.165	.117
	Canonical Trend	.225	.201	.165	.117
-.25	Canonical Seas.	.172*	.152*	.117*	.066*
	Max. FEE	.190	.184	.170	.148
	Canonical Trend	.190	.184	.170	.148
0	Canonical Seas.	.143*	.135*	.110*	.066*
	Max. FEE	.162	.174	.182	.186
	Canonical Trend	.162	.174	.182	.186
.25	Canonical Seas.	.131*	.135*	.117*	.074*
	Max. FEE	.138	.166	.191	.215
	Canonical Trend	.138	.166	.191	.215
.50	Canonical Seas.	.125	.147*	.137*	.090*
	Max. FEE	.125	.159	.191	.222
	Canonical Trend	.119 *	.154	.187	.218
.75	Canonical Seas.	.122	.167	.166	.113*
	Max. FEE	.122	.169	.202	.229
	Canonical Trend	.102*	.133*	.156*	.171

*: Minimum Variance of Total Estimation Error.

Chapter 8

Extensions to rates of growth

8.1 Introduction

We now extend the previous analysis to the rates of growth of the components. We develop this analysis for its important practical applications, since growth rates are natural tools for short-term policy making. They are involved for example in monetary control or in monitoring the evolution of unemployment.

We assume that the overall relationship $x_t = n_t + s_t$ is expressed in logarithms. We shall consider the following linear approximations to the rate of growth of the signal: $\Delta_d s_t = s_t - s_{t-d}$, where d is an integer between 1 and the number of observation per year m. When $d = 1$, the approximation Δs_t represents the most commonly used rate: for monthly observations, it corresponds to the monthly rate of growth of the signal. When $d = 12$, the approximation refers to the annual rates of growth.

For any value of d, we derive the relationship between model specification and estimation error firstly for the historical estimator, and then the results will be generalized to any preliminary estimator and forecast of the rate of growth. Concurrent estimation error will appear as a particular case. We shall then check that the properties of the canonical decomposition are still valid and we will provide some simple rules to determine the most accurately estimated rates.

8.2 Historical rate of growth estimator: the Final Estimation Error variance.

In the model-based approach, the MMSE estimator of the monthly growth of the signal is obtained simply as:

$$\Delta_d \hat{s}_t = \Delta_d \nu_s(B) x_t,$$

where $\nu_s(B)$ is the WK filter as in (4.3), Writing e_t^R the final estimation error: $e_t^R = \Delta_d s_t - \Delta_d \hat{s}_t$, it is straightforward to check that lemma 4.3 stating the equivalence between ACGF of the final estimation error on $\hat{s}_t$ and the CCGF between the historical estimators $\hat{s}_t$ and $\hat{n}_t$ still holds:

$$ACGF(e_t^R) = CCGF(\Delta_d \hat{s}_t, \Delta_d \hat{n}_t).$$

This implies that properties 4.1, 4.2 concerning the components' estimators are also satified by the estimators of the rates of growth. These properties were specifically the existence of finite and convergent covariances between the estimators. Lemma 4.2 also holds, but a modification is required: the estimators are now uncorrelated if the differenced series $\Delta_d x_t$ is nonstationary.

Denoting by $e_t^{R\alpha}$ the final error on the signal rate of growth estimator associated with the admissible decomposition α, then:

Lemma 8.1 *The variance of $e_t^{R\alpha}$ depends on the noise repartition α, $\alpha \in [0,1]$, according to:*

$$var(e_t^{R\alpha}) = var(e_t^{R0}) + $$
$$+ \; 2\alpha V_u[1 - 2(\nu_{s0}^0 - 2\nu_{sd}^0)] - 2\alpha^2 V_u^2 (h_0 - h_d),$$

where $V(e_t^{R0})$ represents the variance of the final error on the canonical signal rate of growth estimator, ν_{sd}^0 the coefficient of B^d in the WK filter $\nu_s^0(B)$, and h_d the term in B^d in the inverted ACGF $h(B)$.

Proof: The ACGF of the estimation error on the signal rate of growth estimator is given by:

$$ACGF(e_t^{R\alpha}) = (1 - B^d)(1 - F^d) ACGF(s_t^\alpha - \hat{s}_t^\alpha) =$$
$$= (1 - B^d)(1 - F^d) \frac{(A_n^0 - \alpha V_u)(A_s^0 + \alpha V_u)}{A_x} =$$
$$= [2 - (B^d + F^d)][\frac{A_n^0 A_s^0}{A_x} + \alpha V_u (1 - 2\frac{A_s^0}{A_x}) - \alpha^2 V_u^2 \frac{1}{A_x}].$$

The result then follows straightforwardly by noticing that:

$$(1 - B^d)(1 - F^d) A_n^0 A_s^0 / A_x = ACGF(e_t^{R0}),$$

and by taking the terms in B^0 in $(1-B^d)(1-F^d)(1-2A_s^0/A_x) = (1-B^d)(1-F^d)(1-2\nu_s^0(B))$ and in $(1-B^d)(1-F^d)1/A_x = (1-B^d)(1-F^d)h(B)$. ∎

The difference operator introduced to approximate the rate of growth does not change the type of function relating final estimation error variance and noise repartition. It remains a second order polynomial in α, but three new coefficients appear: $V(e_t^{R0})$, ν_{sd}^0 and h_d. Let us consider ρ_d^e, ρ_d^s and ρ_d^h, the lag-d autocorrelation of the models:

- for ρ_d^e: $\theta_x(B)z_t = \theta_n^0(B)\theta_s^0(B)b_t^0, \quad V(b_t^0) = V_n^0 V_s^0$;
- for ρ_d^s: $\theta_x(B)z_t = \theta_s^0(B)\phi_n(B)c_t^0, \quad V(c_t^0) = V_s^0$;
- for ρ_d^h: $\theta_x(B)z_t = \phi_x(B)a_t, \quad V_a = 1$,

where $\{\theta_s^0(B), V_s^0\}$ and $\{\theta_n^0(B), V_n^0\}$ are the MA polynomials and the innovations variances associated with a canonical signal and with a nonsignal component concentrating all the noise, respectively. These models are the ones which generated $V(e_t^0)$, ν_{s0}^0 and h_0 in section 5.1. The coefficient ρ_d^h is thus the lag-d autocorrelation of the "inverse" or "dual" model. Then, we can write the coefficients $V(e_t^{R0})$, ν_{sd}^0 and h_d as: $V(e_t^{R0}) = V(e_t^0 - e_{t-d}^0) = 2(1-\rho_d^e)V(e_t^0)$, $\nu_{sd}^0 = \rho_d^s \nu_{s0}^0$, and $h_d = \rho_d^h h_0$, Given the straightforward availability of these lag-d autocorrelation coefficients, lemma 8.1 provides an easy way to compute the final error variance over the range of all admissible decompositions.

We now focus on the error on the preliminary estimators and the forecasted rates of growth.

8.3 Preliminary estimators and forecasted rates of growth.

We denote by $d_{t/t+k}^{R\alpha}$ the total estimation error in the estimator of $\Delta_d s_t^\alpha$ computed at time $t+k$, $k = \cdots, -1, 0, 1, \cdots$. We have: $d_{t/t+k}^{R\alpha} = \Delta_d s_t^\alpha - E_{t+k}\Delta_d \hat{s}_t^\alpha$. We first consider the case of a preliminary estimator, for which $k \geq -d$.

We modify the definition of the polynomial $\lambda(B)$ and take:

$$\lambda(B) = (1 - B^d)\psi^{-1}(F) = \sum_{i=-d}^{\infty} \lambda_i F^i. \qquad (8.1)$$

We also define the coefficients ζ_{dk}^0 and δ_k, $k > -d$, as: $\zeta_{sd}^{0k} = \sum_{i=-d}^{k}(\xi_{si}^0 - \xi_{si+d}^0)\lambda_i$, and $\delta_{dk} = \sum_{i=-d}^{k} \lambda_i^2$. We then present the following result:

Lemma 8.2 *The Total Estimation Error variance on the preliminary estimates of the signal rate of growth depends on the noise repartition α according to:*

$$var[d^{R\alpha}_{t/t+k}] = var[d^{R0}_{t/t+k}] + 2\alpha V_u(1 - \zeta^0_{dk}) - \delta_{dk}\alpha^2 V_u^2. \qquad (8.2)$$

Proof: The total estimation error on the signal rate of growth can be written as the sum of the final error plus the revision error:

$$\begin{aligned} d^{R\alpha}_{t/t+k} &= \Delta_d s^\alpha_t - E_{t+k}\Delta_d s^\alpha_t = \\ &= \Delta_d s^\alpha_t - \Delta_d \hat{s}^\alpha_t + \Delta_d \hat{s}^\alpha_t - E_{t+k}\Delta_d \hat{s}^\alpha_t = \\ &= e^{R\alpha}_t + r^{R\alpha}_t. \end{aligned}$$

where $r^{R\alpha}_t$ represents the revision error on the estimator : $r^{R\alpha}_t = \Delta_d \hat{s}^\alpha_t - E_{t+k}\Delta_d \hat{s}^\alpha_t$. As the independence between the final and the revision errors on rate of growth estimators still holds, we have:

$$V(d^{R\alpha}_{t/t+k}) = V(e^{R\alpha}_t) + V(r^{R\alpha}_t).$$

Since $V(e^{R\alpha}_t)$ has already been given in lemma 8.1, we focus on $V(r^{R\alpha}_t)$. As the signal estimator was given by:

$$\hat{s}^\alpha_t = \sum_{i=-\infty}^{\infty} (\nu^0_{si} + \alpha V_u h_i) x_{t+i},$$

the estimator of the signal rate of growth can be written as:

$$\Delta_d \hat{s}_t^\alpha = \sum_{i=-\infty}^{\infty} [(\nu_{si}^0 - \nu_{si+d}^0) + \alpha V_u (h_i - h_{i+d})] x_{t+i}.$$

Thus we have:

$$\Delta_d \hat{s}_t^\alpha - E_{t+k} \Delta_d \hat{s}_t^\alpha = \sum_{i=-\infty}^{\infty} [\nu_{si}^0 - \nu_{si-d}^0 + \alpha V_u (h_i - h_{i-d})](x_{t+i} - E_{t+k} x_{t+i}).$$

Using: $x_{t+i} - E_{t+k} x_{t+i} = \sum_{j=0}^{i-k-1} \psi_j a_{t+i-j}$, we get:

$$\begin{aligned}
\Delta_d \hat{s}_t^\alpha - E_{t+k} \Delta_d \hat{s}_t^\alpha &= \sum_{i=k+1}^{\infty} [\nu_{si}^0 - \nu_{si+d}^0 + \alpha V_u (h_i - h_{i+d})] \sum_{j=0}^{i-k-1} \psi_j a_{t+i-j} = \\
&= \sum_{i=k+1}^{\infty} [\nu_{si}^0 - \nu_{si+d}^0 + \psi_1 (\nu_{si+1}^0 - \nu_{si+d+1}^0) + \cdots + \\
&\quad + \alpha V_u (h_i - h_{i+d} + \psi_1 (h_{i+1} - h_{i+d+1}) + \cdots)] a_{t+i}.
\end{aligned}$$

which implies the following expression for the variance of the revision error on the nonconcurrent estimator:

$$\begin{aligned}
var[r_t^{R\alpha}] &= var[r_t^{R0}] + \alpha^2 V_u^2 \sum_{i=k+1}^{\infty} [h_i - h_{i+d} + \psi_1 (h_{i+1} - h_{i+d+1}) + \cdots]^2 + \\
&\quad + 2\alpha V_u \sum_{i=k+1}^{\infty} [\nu_{si}^0 - \nu_{si+d}^0 + \psi_1 (\nu_{si+1}^0 - \nu_{si+d+1}^0) + \cdots] \cdot \\
&\quad \cdot [h_i - h_{i+d} + \psi_1 (h_{i+1} - h_{i+d+1}) + \cdots]. \qquad (8.3)
\end{aligned}$$

Now defining:

$$l_i = \nu_{si}^0 - \nu_{si+d}^0 + \psi_1(\nu_{si+1}^0 - \nu_{si+d+1}^0) + \cdots,$$

and:

$$m_i = h_i - h_{i+d} + \psi_1(h_{i+1} - h_{i+d+1}) + \cdots,$$

equation (8.3) can then be expressed as:

$$var[r_t^{R\alpha}] = var[r_t^{R0}] + \alpha^2 V_u^2 \sum_{i=k+1}^{\infty} m_i^2 + 2\alpha V_u \sum_{i=k+1}^{\infty} l_i m_i. \tag{8.4}$$

The coefficient m_i can be seen as the term in F^i if $i > 0$, and as the term in $B^{|i|}$ otherwise, in the polynomial multiplication: $(1 - B^d)h(B)\psi(B)$. Since $(1 - B^d)h(B)\psi(B) = (1 - B^d)/\psi(F) = \lambda(B)$ which does not contain any term in B^j for $j > d$, the coefficients m_i must be null for $i < -d$, and equal to λ_i when $i \geq -d$. So, we have:

$$\begin{aligned}
\sum_{i=-d}^{\infty} m_i^2 &= \frac{(1 - F^d)}{\psi(B)} \frac{(1 - B^d)}{\psi(F)} \mid_{B=F=0}= \\
&= (1 - F^d)(1 - B^d)h(B) \mid_{B=F=0}= \\
&= 2(h_0 - h_d),
\end{aligned} \tag{8.5}$$

and for $k \geq -d$:

$$\begin{aligned}
\sum_{i=k+1}^{\infty} m_i^2 &= 2(h_0 - h_d) - \sum_{i=-d}^{k} m_i^2 = \\
&= 2(h_0 - h_d) - \delta_{dk}.
\end{aligned} \tag{8.6}$$

In the same way, l_i may be seen as the term in F^i if $i \geq 0$, and as the term in $B^{|i|}$ otherwise, in the polynomial multiplication $(1-B^d)\nu_s^0(B)\psi(B)$. This product yields: $(1-B^d)\nu_s^0(B)\psi(B) = (1-B^d)\xi_s^0(B)$, so for $i > 0$, $l_i = \xi_{si}^0 - \xi_{si+d}^0$. Now multiplying by $m(F) = (1-F^d)/\psi(B)$, we get:

$$(1-B^d)\xi_s^0(B)(1-F^d)/\psi(B) = (1-F^d)(1-B^d)\nu_s^0(B).$$

Looking at the central term, and using $m_i = 0$ for $i < -d$, we have:

$$\sum_{i=-d}^{\infty} l_i m_i = 2(\nu_{s0}^0 - \nu_{sd}^0). \qquad (8.7)$$

In the case where $k \geq -d$, we have:

$$\begin{aligned}
\sum_{i=k+1}^{\infty} l_i m_i &= 2(\nu_{s0}^0 - \nu_{sd}^0) - \sum_{i=-d}^{k} l_i m_i = \\
&= 2(\nu_{s0}^0 - \nu_{sd}^0) - \sum_{i=-d}^{k} (\xi_{si}^0 - \xi_{si+d}^0)\lambda_i = \\
&= 2(\nu_{s0}^0 - \nu_{sd}^0) - \zeta_{sd}^{0k}. \qquad (8.8)
\end{aligned}$$

From (8.4), (8.6), and (8.8), the revision error variance on the preliminary estimator of the signal rate of growth is given by, for $k < -d$:

$$\begin{aligned}
V(r_{t/t+k}^{R\alpha}) &= V(r_{t/t+k}^{R0}) + \alpha V_u[4(\nu_{s0}^0 - \nu_{sd}^0) - 2\zeta_{dk}^0] + \\
&+ \alpha^2 V_u[2(h_0 - h_d) - \delta_{dk}], \qquad (8.9)
\end{aligned}$$

Eventually, adding $V(e_t^{R\alpha})$ given in lemma 8.1 to (8.9) yields the expected result. ∎

Lemma 8.2 provides the general relationship between estimation error of the growth rates and the range of admissible decomposition for estimations computed at any time t+k, $k = \cdots, -1, 0, 1, \cdots$. We can check that it is consistent with the first case of a historical estimator presented above. When k goes to infinity, it is obvious that δ_{dk} tends towards $2(h_0 - h_d)$. On the other hand, we have: $\lim_{k\to\infty} \zeta_{sd}^{0k} = \lim_{k\to\infty}(\xi_{si}^0 - \xi_{si+d}^0)\lambda_i = 2(\nu_{s0}^0 - \nu_{sd}^0)$, since we must obtain the central term of $(1 - B^d)(1 - F^d)\nu_s^0(B)$. . Furthermore, the convergence in F of the polynomial $(1 - B^d)\xi_s^0(B)$ ensures that $V(r_{t/t+k}^{R0})$ tends to zero as k tends to infinity. Taking $k \to \infty$ in equation 8.11 thus yields $V(r_{t/t+k}^{R\alpha}) \to 0$, and the historical estimation case is recovered.

Lemma 8.2 also generates the error variance for concurrent estimators. This particular case is obtained simply by setting k to zero in (8.9), which leads to replace $V(r_{t/t+k}^{R0})$, λ_{dk}, and ζ_{dk}^0, by $V(r_{t/t}^{R0})$, λ_{d0}, and ζ_{dk}^0.

We now focus on the forecasted rates. When $k < -d$, at time $t + k$, the variable s_{t-d}^α is forecasted. Thus, we shall say that the estimation at time $t + k$, $k < -d$, provides a forecasted signal growth rate.

Lemma 8.3 *The variance of the total estimation error on the forecast of the signal rate of growth is related to the noise repartition according to:*

$$var[d_{t/t+k}^{R\alpha}] = var[d_{t/t+k}^{R0}] + 2\alpha V_u. \tag{8.10}$$

where $k < -d$.

-

Proof: The result is immediate using (8.5) and (8.7) in (8.4). ∎

We now discuss the decompositions providing the best estimated growth rates.

8.4 Properties of the canonical decomposition.

8.4.1 Historical estimators

Lemma 8.4 *Over the range of all admissible decomposition, the mean squared error on the historical estimator of the signal growth rate is minimized at:*

- $\alpha = 0$ *if* $2(\nu_{s0}^0 - \nu_{sd}^0) + V_u(h_0 - h_d) < 1$*;*
- $\alpha = 1$ *otherwise.*

<u>**Proof:**</u> In lemma 8.1, the coefficient of α^2 in the expression for $V(e_t^{R\alpha})$ is $-2(h_0 - h_d)$. Since it can be written as: $-2(1 - \rho_d^h)h_0$, where ρ_d^h represents the lag-d autocorrelation of the "inverse" model with variance h_0, we have: $-2(1 - \rho_d^h)h_0 < 0$. So $V(e_t^{R\alpha})$ is a concave function of α over [0,1]. The minimum is thus reached at the boundaries of [0,1]. Comparing $V(e_t^{R1})$ and $V(e_t^{R0})$ directly yields the expected result. ∎

A straightforward consequence of lemma 8.1 is the following property of the canonical decompositions:

Corollary 8.1 *Over the range of all admissible decompositions, a canonical decomposition always minimizes the error variance in the estimator of the signal growth rate.*

The property of canonical decompositions of always providing the best estimated decomposition remains valid when growth rates are being estimated. This result is true irrespective of the order the differencing d that is used to approximate the growth rate. We can understand more clearly which canonical decomposition is best estimated by considering, without loss of generality, the following assumption:

Assumption 6.a: The signal is such that: $\nu_{s0}^0 - \nu_{sd}^0 > \nu_{n0}^1 - \nu_{nd}^1$.

We can then restate lemma 8.4 as:

Lemma 8.5 *Under assumption 6.a, the error variance on the growth rate final estimator is always minimized by specifying a canonical non-signal* n_t.

Proof: We have to show that assumption 6.a implies that the condition $2(\nu_{s0}^0 - \nu_{sd}^0) + V_u(h_0 - h_d) > 1$ given in lemma 8.4 for having $\alpha = 1$ as the solution of the minimization problem is satisfied. Since $\nu_{s0}^\alpha + \nu_{n0}^\alpha = 1$ while, for $j > 0$, $\nu_{sj}^\alpha + \nu_{nj}^\alpha = 0$, assumption 6.a may be rewritten:

$$\nu_{s0}^0 - \nu_{sd}^0 + \nu_{s0}^1 - \nu_{sd}^1 > 1.$$

Now, the WK filter for estimating a pure noise u_t with variance V_u was given as: $\nu_u(B) = V_u h(B)$, and we have already seen that $\nu_s^1(B) = \nu_s^0(B) + \nu_u(B)$. So, as ν_{s0}^1 was found to be equal to $\nu_{s0}^0 + V_u h_0$, we have: $\nu_{sd}^1 = \nu_{sd}^0 + V_u h_d$. Inserting these two results in the last inequality proves the lemma. ∎

The terms $\nu_{s0}^0 - \nu_{sd}^0$ and $\nu_{n0}^1 - \nu_{nd}^1$ represent the central coefficients of the WK filters designed to estimate the canonical components growth rates. Hence, to obtain the best estimated decomposition, it is enough to compare the central coefficients of the WK filters estimating the growth rates of the two canonical components and to assign all the noise of the model to, roughly speaking, the most important component. This result is similar to that derived previously for the levels of the components.

We now turn our attention to the other types of estimators.

8.4.2 Preliminary estimators and forecasts

Lemma 8.6 *For $\alpha \in [0,1]$, the total error variance in the preliminary estimator of the signal growth rate is minimized at:*

- $\alpha = 0$ *if* $\zeta_{sd}^{0k} + (1/2)\delta_{dk}V_u < 1$;
- $\alpha = 1$ *otherwise,*

for $k > -d$.

Proof: The result follows immediately from the fact that δ_{dk} is defined as a sum of squared terms, so $\delta_{dk} > 0$ and $V(d_{t/t}^{R\alpha})$ as presented in lemma 8.2 is a concave function of α. ∎

This lemma suggests another property of the canonical decompositions:

Corollary 8.2 *Over the range of all admissible decompositions, a canonical decomposition always minimizes the total error variance in the preliminary signal growth rate estimator.*

It would be also straightforward to show that canonical decompositions have the unpleasant property of maximizing the revision error in the preliminary estimators of the rates. Hence, it is true that the properties of the canonical decompositions still hold when rates of growth are being considered. There exists however the possibility that, while a canonical specification minimizes the error variance in the preliminary estimator of the level of the component, the other canonical decomposition minimizes the error in growth rate estimator.

Predicting the evolution of a series is of particular interest. For example, forecasted growth rates play an important role in short-term economic policy making. Governments and central banks need to estimate the future paths of of macroeconomic series. Lemma 8.3 indicates how best to forecast growth rates:

Lemma 8.7 *The forecast error variance on growth rates is minimized on canonical components.*

<u>**Proof:**</u> This result follows straightforwardly from lemma 8.3. ∎

As consequence, the most natural tool for monitoring the underlying evolution of economic series is the forecasted trend growth rate. This result has been already emphasized by Box, Pierce and Newbold (1987), who however noticed that, despite this, in practice others estimators are used. For example, practitioners typically consider the forecasted growth rate of the seasonally adjusted series. Such estimators have the inconvenience of being the least precise.

8.5 An Example: A Quarterly Airline Model.

We illustrate these last results with the following example. Let a quarterly series follow the model (in logs):

$$\Delta\Delta_4 x_t = (1 - .3B)(1 - .7B^4)a_t,$$

where a_t is a normally distributed variable with variance $V_a = 1$. This series is decomposed into a seasonal s_t and a nonseasonal component n_t according to:

$$\Delta^2 n_t = (1 + \theta_{n1}B + \theta_{n2}B^2)a_{nt},$$

$$U(B)s_t = (1 + \theta_{s1}B + \theta_{s2}B^2 + \theta_{s3}B^3)a_{nt},$$

where $U(B) = 1 + B + B^2 + B^3$. Making the seasonal component canonical, we obtain, using the procedure described in the previous chapter:

$$\Delta^2 n_t^0 = (1 - 1.269B + .276B^2)a_{nt}^0,$$

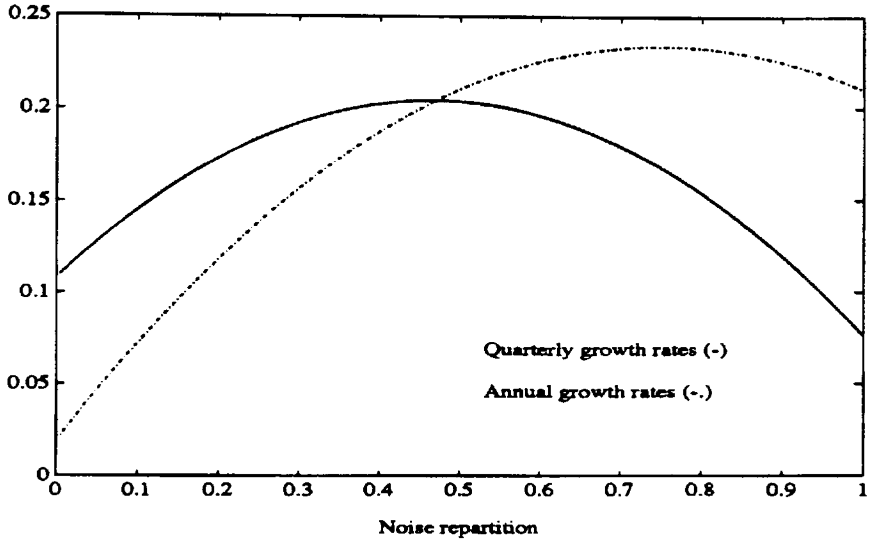
$$U(B)s_t^0 = (1 + .028B - .502B^2 - .526B^3)a_{st}^0, \qquad (8.11)$$

with $V_n^0 = .778$ and $V_s^0 = .009$.

We consider quaterly and annual rates of growth, and we first focus on historical estimators. For a canonical s_t, the coefficients of the WK filter are obtained as: $\nu_{s0}^0 = .119$, $\nu_{s1}^0 = -.039$, and $\nu_{s4}^0 = .095$, while, for a canonical trend: $\nu_{n0}^1 = .333$, $\nu_{n1}^0 = .228$ and $\nu_{n4}^0 = -.008$. Comparing the differences, we get for the monthly growth rates: $\nu_{s0}^0 - \nu_{s1}^0 = .158$, greater than $\nu_{n0}^1 - \nu_{n1}^0 = .105$. Applying lemma 8.5 to the monthly rates of growths estimators point the decomposition where the trend is canonical as being the most accurately estimated. For annual rates of growth, we have: $\nu_{s0}^0 - \nu_{s4}^0 = .025$, less than $\nu_{n0}^1 - \nu_{n4}^0 = .341$. Lemma 8.5 then indicates that, for annual growth rates, the decomposition with canonical seasonal provides the most accurate estimator. Quarterly and annual growth rates yield in this example opposite results.

It is also of interest to relate the error variance to the rate of growth estimator over the range of all admissible decompositions. The pure noise part of the model for x_t has a variance of .302. The variance of the inverted process is $h_0 = 1.817$, and the lag-1 and lag-4 autocorrelations are: $\rho_1^h = -.344$ and $\rho_1^h = -.344$. The final estimation error variance for the decomposition in (8.11) is given by: $V(e_t^0) = .043$. The lag-1 and lag-4 autocorrelations coefficients of the model followed by e_t^0 (see section 8.3) are found to be: $\rho_1^e = -.257$, and $\rho_4^e = .783$.

Figure 8.1: Mean Squared Errors in Growth Rates Final Estimators



Using lemma 8.1, we obtain for the monthly rates of growth and for $\alpha \in [0,1]$:

$$V[e_t^{R\alpha}] = .108 + .413\alpha - .445\alpha^2.$$

For the annual rates of growth, we have instead:

$$V[e_t^{R\alpha}] = .019 + .575\alpha - .384\alpha^2.$$

These two functions are plotted in figure 8.1.

As expected from lemma 8.5, for quartely growth rates, the minimum error variance is obtained at .076 when the trend component is specified noninvertible. The maximum, around .20, corresponds to the case where nearly 45% of the noise of the model is attributed to the seasonal component.

It is the specification that the minimax filter would yield. For annual growth rates, the minimum is obtained at .019 on a canonical seasonal component. The minimax filter yields here an error variance maximized at .24 which corresponds to a seasonal component concentrating around 75% of the noise of the model. The magnitude of the variations of the mean squared estimation error over the range of admissible decompositions emphasizes the interest of identifying the most precisely estimated decomposition.

We turn to the error in the concurrent estimators. The total estimation error for the decomposition (8.11) has a variance of $V(d_t^0) = .088$, lag-1 and lag-4 autocorrelations of -.194 and .743, respectively. From the polynomials $\xi_s^0(B)$ and $\lambda(B)$, computing the coefficients ζ_{sd}^{0k} and δ_{dk} for $d = 1, 4$ and $k = 0$, we get: $\zeta_{s1}^{00} = .535$, $\zeta_{s4}^{00} = .031$, $\delta_{10} = 3.891$ and $\delta_{40} = 3.270$. So, using lemma 8.2, the relationship between noise repartition and mean squared error in the concurrent estimator of the quaterly growth rates is given by:

$$V[e_t^{R\alpha}] = .210 + .281\alpha - .355\alpha^2.$$

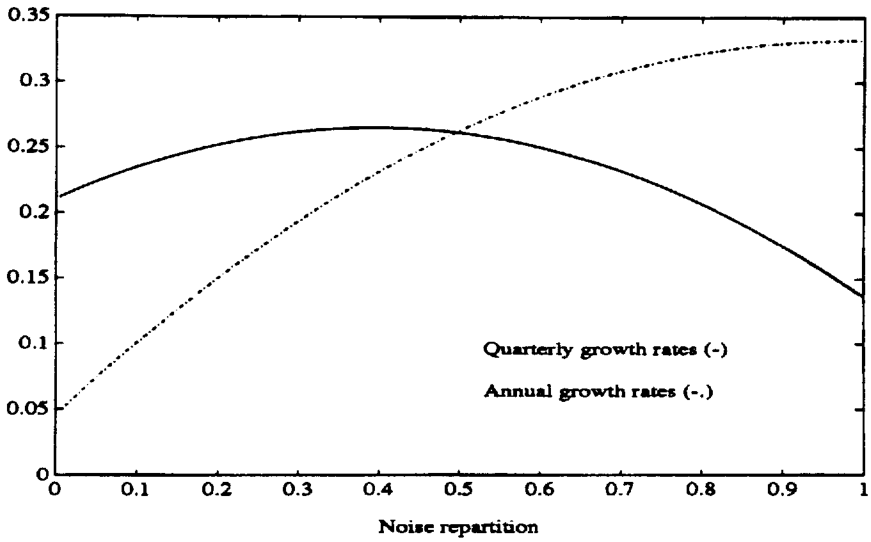
For the annual rates of growth, we have instead:

$$V[e_t^{R\alpha}] = .045 + .586\alpha - .298\alpha^2.$$

The two functions are plotted in figure 8.2.

It is seen on figure 8.2 that, for this quarterly airline model, the components growth rates most accurately estimated remain the same when concurrent estimation is considered. The quarterly growth rates is still better estimated with a canonical trend (.136), while for annual rates of growth, a canonical seasonal still yields a minimum error variance at .045. Notice that for annual growth rates, the decomposition that would yield the minimax filter corresponds to a noninvertible trend. The maximum error variance

Figure 8.2: Mean Squared Errors in Concurrent Growth Rates Estimators



reached is of .343. Selecting instead a canonical seasonal would reduce this error by the proportion of 85%.

Chapter 9

Applications.

9.1 Introduction

The estimation errors associated with different seasonal/nonseasonal decompositions have been analysed for four monthly macroeconomic series : the US exports (1972-1 1989-7), the French money supply (M1, 1978-1 1991-7), the Italian money supply (M1, 1971-1 1991-6) and the Japan exports series (1972-1 1992-7). These series have been collected from the IMF series provided by Datastream. All of them have been modeled in logs. For each series, five types of decomposition are considered: a Seasonally Adjusted (SA) series concentrating all the noise of the model, a Trend representing a canonical nonseasonal component, and a nonseasonal component estimated using a minimax filter designed for historical estimators and concurrent estimates. We also consider a nonseasonal component with an MA polynomial identified by a zero-coefficient restriction; we think that this decomposition is also of interest since it is used in the popular Structural Time Series model approach.

A C program has been written to obtain the corresponding estimation errors, and some checks have been performed using the SEATS software (see

Maravall and Gómez (1992)).

9.2 US Export Series

For the US export series (1972-1 1989-7), an airline model was fitted. The sample size is $T = 211$, and the parameters of the model was obtained as :

Table 9.1:				
Model fitted: $\Delta\Delta_{12}x_t \;=\; (1-\theta_1 B)(1-\theta_{12}B^{12})a_t$				
Parameters:	θ_1	θ_{12}	V_a	Q_{22}
Estimators	.398	.817	$2.39\ 10^{-3}$	26.9
	(.063)	(.045)		

Four outliers have been detected at time t=38,70,72, 121, and with t-value -2.55, -3.30, 2.87, and 3.07, respectively.

The US export series, the Trend and the SA series, i.e. the two competing models for the minimization of the mean squared error in the final and concurrent estimators, are plotted on figures 9.1 and 9.2.

Denoting s_t and n_t the seasonal and the nonseasonal components of x_t, and making both of them canonical, we obtain: $\nu^0_{s0} = .085$, $\nu^1_{n0} = .280$, $\xi^0_{s0} = .114$, and $\xi^1_{n0} = .483$. According to lemmas 6.2 and 6.6, since $\nu^0_{s0} < \nu^1_{n0}$ and $\xi^0_{s0} < \xi^1_{n0}$, the decomposition where the seasonal component is canonical will minimize the estimation error variance for both historical and concurrent estimators.

For intermediate decompositions, we apply lemmas 5.1 and 5.2. In the model for the US exports series, the irregular component has a maximun variance of $V_u = .403$ (in V_a units), while the variance of the "inverted pro-

Figure 9.1: US Exports Series (1972-1 1989-7)

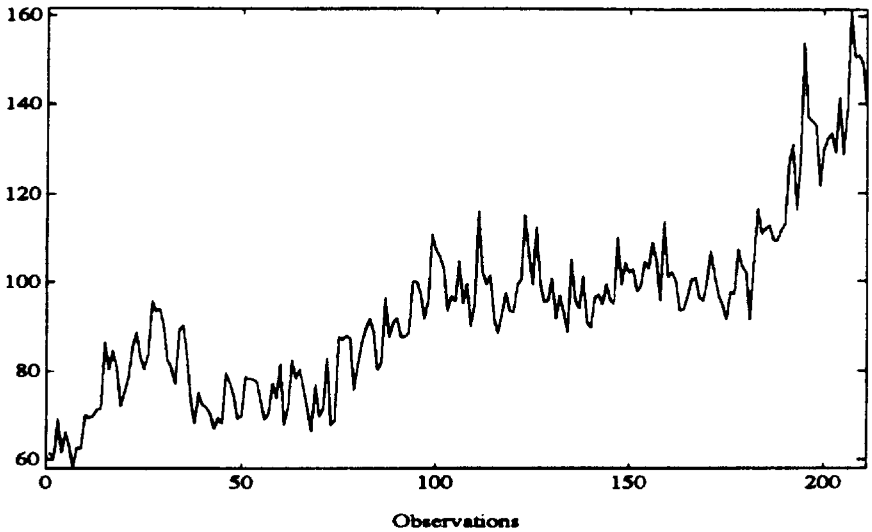


Figure 9.2: XUS: SA Series / Trend

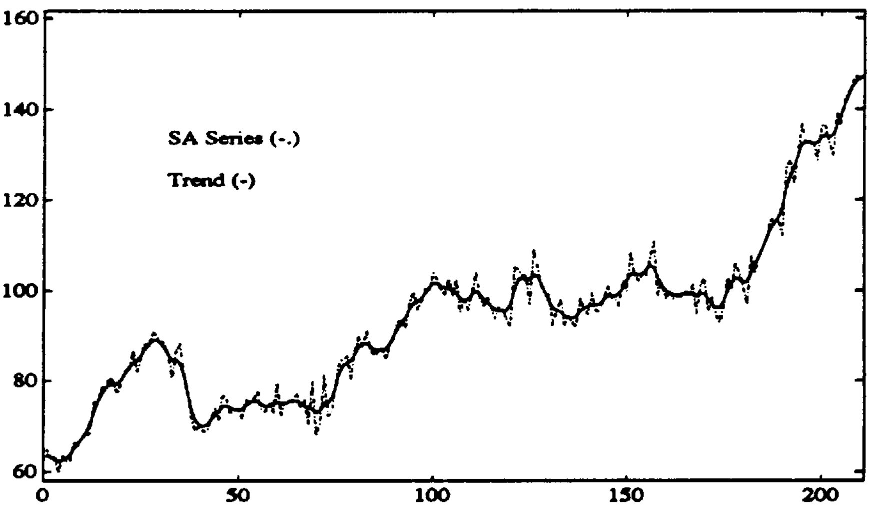
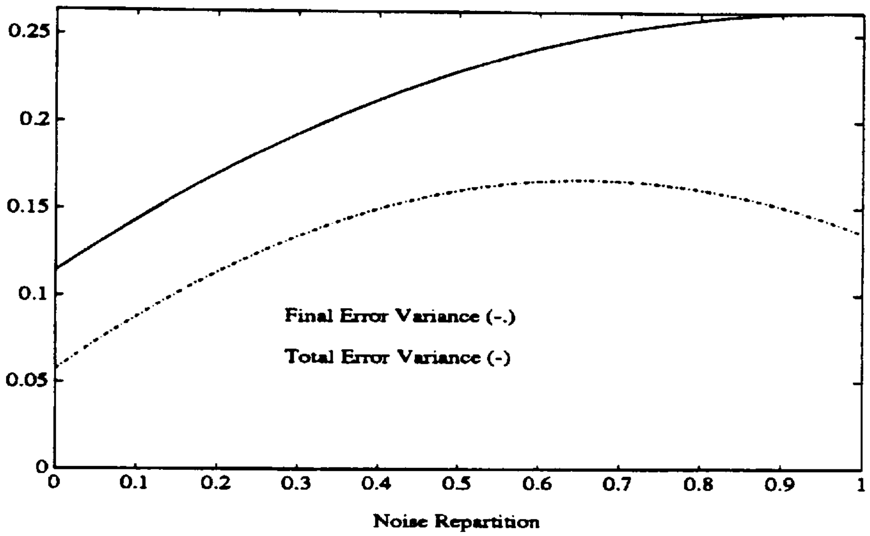


Figure 9.3: XUS Series: Mean Squared Errors



cess" is $h_0 = 1.576$. The mean squared error in historical and concurrent estimators are found to be: $V(e_t^0) = .057$, $V(d_t^0) = .114$. Then, the relationship between the error variance in the historical estimator and the noise repartition, $\alpha \in [0,1]$, is given by:

$$V(e_t^\alpha) = .057 + .334\alpha - .256\alpha^2,$$

and for the total error in concurrent estimators:

$$V(d_t^\alpha) = .114 + .311\alpha - .162\alpha^2.$$

These functions are plotted on figure 9.3, and the estimation errors associated with five different decomposition on table 9.2.

The final error variance and lag-0 covariance between the estimators vary within the range [.057, .172], and, for the total error variance in concurrent estimators, within the range [.114, .275]. The lower bound of both intervals is reached by specifying a canonical seasonal component. This is in agreement with the discussion of the airline models in section 5.3 since the parameter θ_{12} is close enough to the noninvertibility region. The maximum MSE and TEE are obtained using the minimax filter for respectively historical and concurrent estimation. Notice that for concurrent estimation, the minimax filter corresponds to specifying a noninvertible trend component, for which the revision error are maximized at .132. As expected, specifying a IMA(2,1) model for the nonseasonal component yields intermediate estimation errors. In this example, the gain in precision obtained by specifying a canonical seasonal component instead of a canonical trend is substantial: the final error variance is reduced by 66% while the total error variance in concurrent estimators is reduced by 60%.

Table 9.2:

XUS: Models for the nonseasonal component:

SA Series/Canonical seasonal.

Model specification: $\Delta^2 n_t = (1 - .1.382B + .392B^2)a_{nt}$

$V_n = .837$

Final Error var.	Revision error var.	Total estimation error var.
.057	.057	.114

Canonical Trend.

Model specification: $\Delta^2 n_t = (1 + .017B - .983B^2)a_{nt}$

$V_n = .076$

Final Error var.	Revision error var.	Total estimation error var.
.142	.132	.275

Restricting the order of the trend MA polynomial:

Model specification: $\Delta^2 n_t = (1 - .98B)a_{nt}$

$V_n = .302$

Final Error var.	Revision error var.	Total estimation error var.
.166	.103	.269

Minimax filter on historical estimator.

Model specification: $\Delta^2 n_t = (1 - 1.131B + .145B^2)a_{nt}$

$V_n = .416$

Final Error var.	Revision error var.	Total estimation error var.
.172	.085	.256

Minimax filter on concurrent estimator.

Model specification: $\Delta^2 n_t = (1 + .017B - .983B^2)a_{nt}$

$V_n = .076$

Final Error var.	Revision error var.	Total estimation error var.
.142	.132	.275

9.3 French Money Supply (M1)

For the French money supply series (1978-1 1991-7), whose sample size is $T = 163$, an $(2,1,1)\mathbf{x}(0,1,1)_{12}$ model was fitted. The estimated parameters were :

Table 9.3:						
Model fitted: $(1-\phi_1 B-\phi_2 B^2)\Delta\Delta_{12}x_t = (1-\theta_1 B)(1-\theta_{12}B^{12})a_t$						
Parameters:	ϕ_1	ϕ_2	θ_1	θ_{12}	V_a	Q_{20}
Estimators	-.690	-.484	-.203	.497	.170 10^{-3}	26.86
	(.162)	(.085)	.186)	(.081)		

A single outlier appears at time $t = 99$ and with a t-value of 3.10. The AR polynomial have conjugate roots at a frequency close to $2\pi/3$, which are thus associated with the seasonal behaviour of the series.

The observed series and the two competing models for minimizing the estimation error variance are plotted on figures 9.4 and 9.5.

If s_t and n_t represent the seasonal and the nonseasonal part of x_t, making both of them canonical yields: $\nu_{s0}^0 = .447$, $\nu_{n0}^1 = .243$, $\xi_{s0}^0 = .445$, $\xi_{n0}^1 = .400$. Since $\nu_{s0}^0 > \nu_{n0}^1$ and $\xi_{s0}^0 > \xi_{n0}^1$, according to lemmas 6.2 and 6.6, the decomposition where the trend is set canonical will yield the most accurate estimation.

In order to derive the expression for the mean squared error over the range of the admissible decompositions, we compute the coefficients $V(e_t^0)$ and $V(d_t^0)$. This yields: $V(e_t^0) = .111$ and $V(d_t^0) = .220$. In the model for the French M1 series, an irregular component would have a maximum variance

Figure 9.4: French M1 Series (1978-1 1991-7)

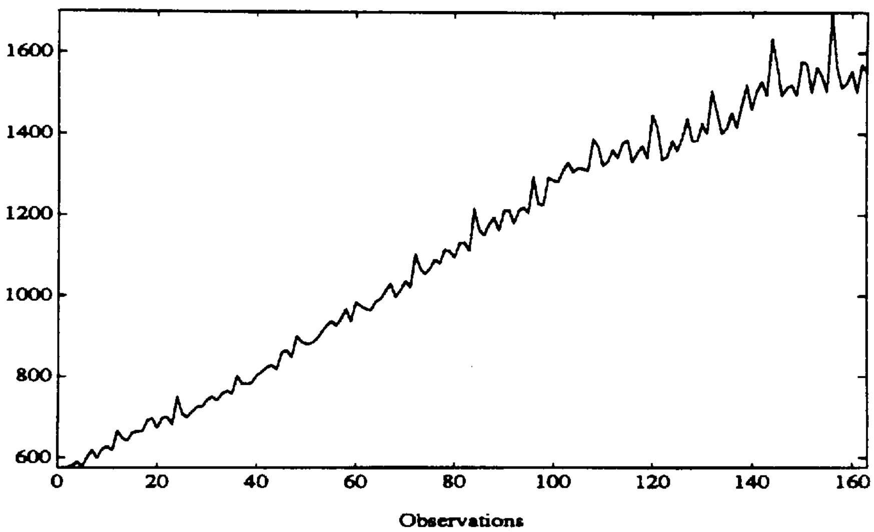


Figure 9.5: M1F: SA Series / Trend

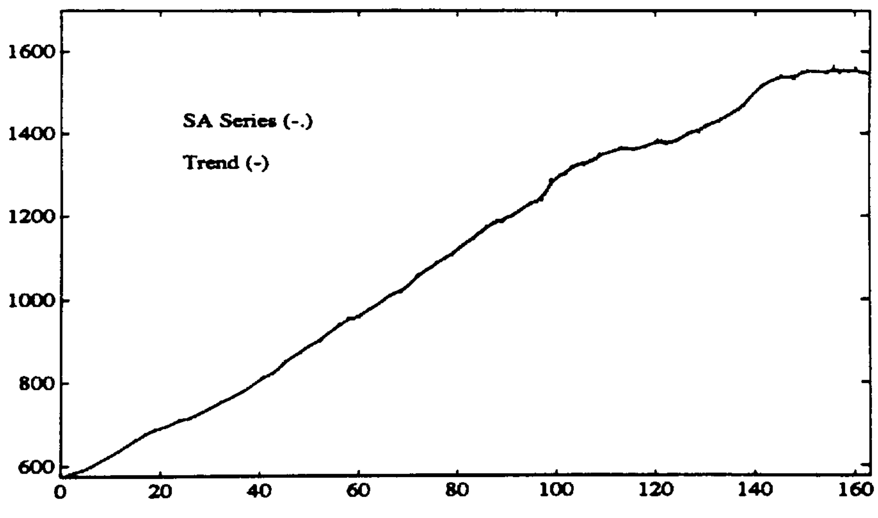
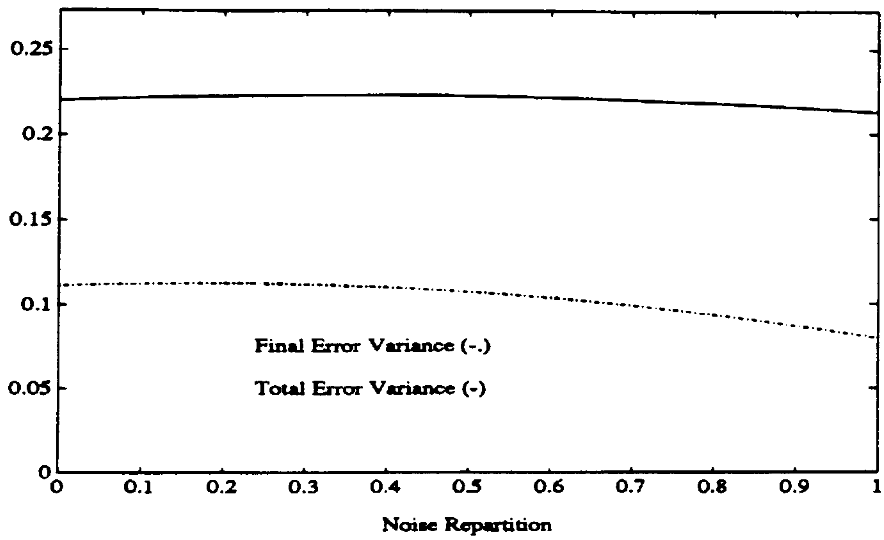


Figure 9.6: M1F: Mean Squared Errors



of $.155V_a$, while the variance of the "inverted" process is $h_0 = 2.00$. Using lemmas 5.1 and 5.3, the errors variances as functions of the noise repartition repartition are then obtained as:

$$V(e_t^\alpha) = .111 + .017\alpha - .048\alpha^2,$$

for the final error variance, while for the concurrent estimator, we have:

$$V(d_t^\alpha) = .220 + .017\alpha - .024\alpha^2.$$

These functions are displayed on figure 9.6, and the estimation errors associated with the five decompositions considered are displayed on table 9.4.

The mean squared error and the lag-0 covariance between historical estimators are minimized at .08 by the specification of a noninvertible trend. The minimax filter yields the maximum error variance of .112. As expected, for concurrent estimates, the lowest error variance (.214) is still obtained on the estimation of a canonical trend. The highest total error variance that yields the minimax filter is reached at .223. If, for the French Ml series, a canonical trend represents the best estimated model, one can easily check that this model specification implies furthermore that the revision error are maximized at .134. With this example, the gain of precision reached by selecting the best estimated model is rather low: the error can be reduced by 28% for historical estimators, and by 4% for concurrent estimators with respect to the error obtained with the minimax filter.

To minimize the mean squared error and the lag-0 covariance between the historical seasonal and the historical nonseasonal estimators, all the noise must be attributed to the seasonal component. The solution of the error minimization problem is thus the maximum noise extraction, and the nonseasonal component is now a canonical trend. The corresponding maximization of the revision error is the price paid for a greater precision of the estimation. The French Ml series yields opposite results to the US exports series.

Table 9.4

M1F: Models for the nonseasonal component:

SA Series/Canonical seasonal.

Model specification: $\Delta^2 n_t = (1 - .1.272B + .310B^2)a_{nt}$

$V_n = .373$

Final Error var.	Revision error var.	Total estimation error var.
.111	.109	.220

Canonical Trend:

Model specification: $\Delta^2 n_t = (1 + .057B - .943B^2)a_{nt}$

$V_n = .042$

Final Error var.	Revision error var.	Total estimation error var.
.080	.134	.214

Restricting the order of the trend MA polynomial:

Model specification: $\Delta^2 n_t = (1 - .944B)a_{nt}$

$V_n = .169$

Final Error var.	Revision error var.	Total estimation error var.
.096	.123	.219

Minimax filter on historical estimator.

Model specification: $\Delta^2 n_t = (1 - 1.229B + .269B^2)a_{nt}$

$V_n = .329$

Final Error var.	Revision error var.	Total estimation error var.
.112	.110	.222

Minimax filter on concurrent estimator.

Model specification: $\Delta^2 n_t = (1 - 1.170B + .214B^2)a_{nt}$

$V_n = .282$

Final Error var.	Revision error var.	Total estimation error var.
.111	.112	.223

9.4 Italian Money Supply (M1)

The Italian M1 series (T=246) was modeled as an ARIMA$(0,1,0)\mathrm{x}(0,1,1)_{12}$, and the estimation results are displayed in table 9.5.

Table 9.5: Model fitted: $\Delta\Delta_{12}x_t = (1-\theta_{12}B^{12})a_t$			
Parameters:	θ_{12}	V_a	Q_{20}
Estimators	.55	.149 10^{-3}	19.02
	(.059)		

Five outliers are found at time t=52, 132, 228, 245, 246, with the corresponding t-values of -4.13, 3.32, 3.30, 3.14, -2.79.

The Italian M1 series, the estimators of the SA series and the Trend series are plotted on figures 9.7 and 9.8.

For s_t^0 denoting a canonical seasonal component and n_t^1 a canonical trend, we have: $\nu_{s0}^0 = .223, \nu_{n0}^1 = .390, \xi_{s0}^0 = .198$, $\xi_{n0}^1 = .652$. In that case, lemmas 6.2 and 6.6 point the decomposition with canonical seasonal as been the most accurately estimated, whether the estimation is concurrent or historical.

Looking at the relationship between mean squared error and noise repartition over the range of all admissible decompositions, we have: $V(e_t^0) = .163$, $V(d_t^0) = .341$, and $h_0 = 2.581$, while the variance of the pure noise part of the series is given by: $V_u = .150$. The error variance functions are then obtained as:

$$V(e_t^\alpha) = .163 + .083\alpha - .058\alpha^2,$$

Figure 9.7: Italian M1 Series (1971-1 1991-6)

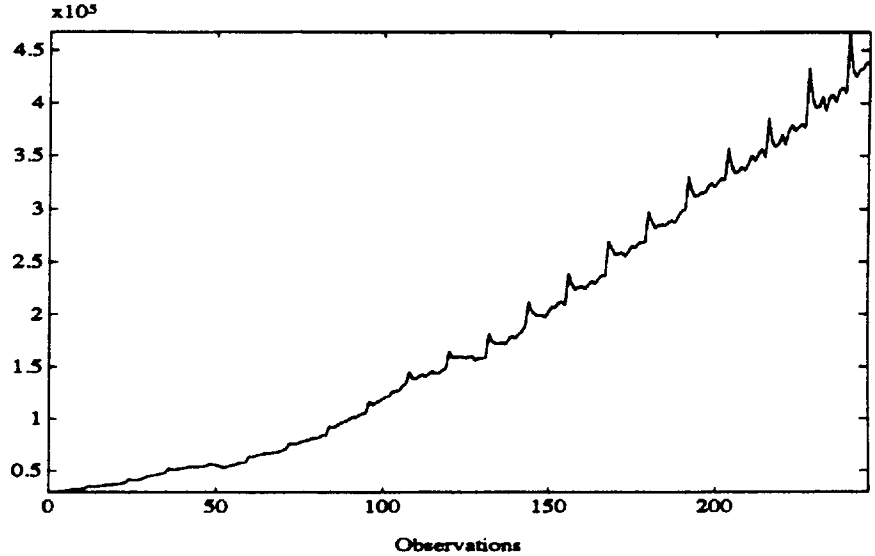


Figure 9.8: M1IT: SA Series / Trend

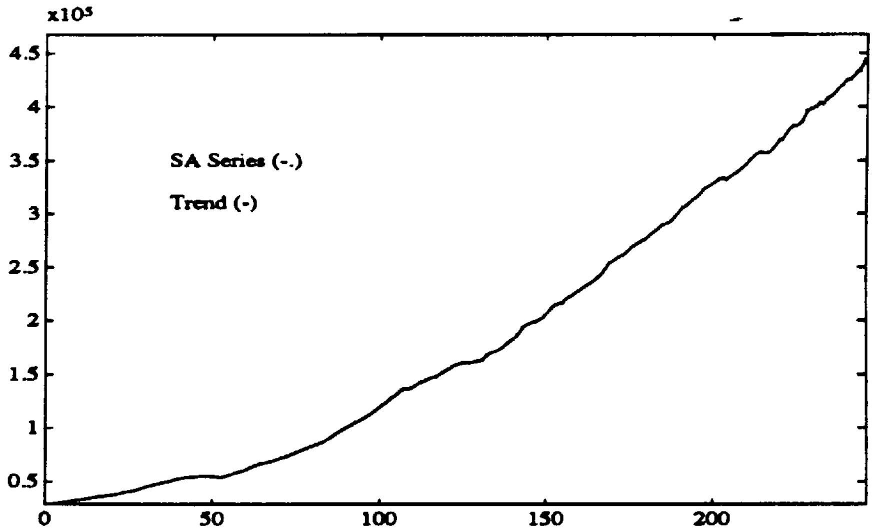
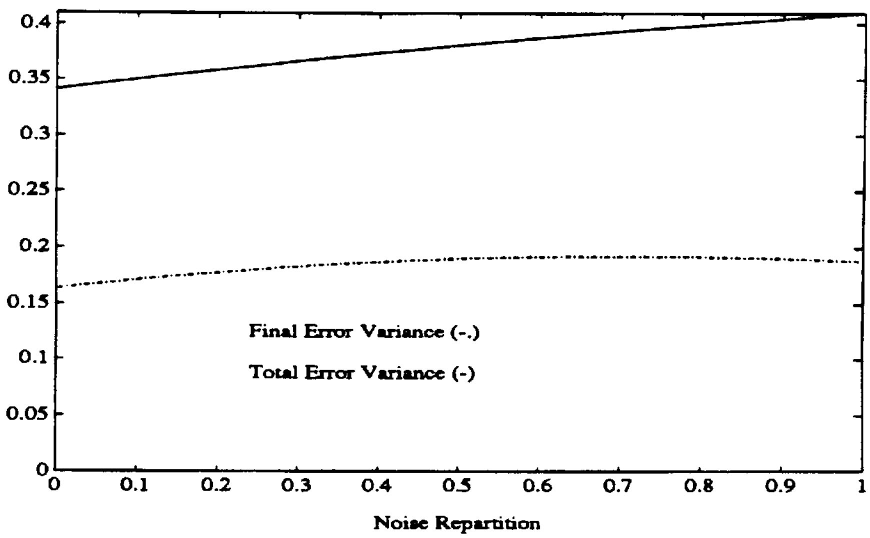


Figure 9.9: M1IT: Mean Squared Errors



for the historical estimator, and for the concurrent estimator:

$$V(d_t^\alpha) = .341 + .091\alpha - .023\alpha^2,$$

These functions are plotted on figure 9.9, and the estimation error associated with the different model specifications considered are presented on table 9.5.

Figure 9.7 confirms that to minimize the error variance and the lag-0 covariance between the historical estimators, all the noise must be attributed to the nonseasonal component. This model specification yields a mean squared error of .162, while a minimax filter yields an error variance of .193. For concurrent estimators, the lowest mean squared error is still reached with a canonical seasonal component at .341. Again, this result could have been anticipated from the discussion of the airline models in section 4.4. The model specification corresponding to the minimax filter for concurrent estimates is equivalent to specifying a noninvertible trend component, and yields an error variance of .409 . Notice that a canonical trend maximizes the revision error at .221. For historical as well as for concurrent estimation, the maximum gain of precision that we can expect from our model selection procedure is in the order of 16%.

Table 9.5

M1IT: Models for the nonseasonal component:

SA Series/Canonical seasonal.

Model specification: $\Delta^2 n_t = (1 - .966B + .032B^2)a_{nt}$

$V_n = .615$

Final Error var.	Revision error var.	Total estimation error var.
.162	.178	.341

Canonical Trend.

Model specification: $\Delta^2 n_t = (1 + .048B - .952B^2)a_{nt}$

$V_n = .149$

Final Error var.	Revision error var.	Total estimation error var.
.188	.221	.409

Restricting the order of the trend MA polynomial:

Model specification: $\Delta^2 n_t = (1 - .951B)a_{nt}$

$V_n = .596$

Final Error var.	Revision error var.	Total estimation error var.
.167	.179	.346

Minimax filter on historical estimator.

Model specification: $\Delta^2 n_t = (1 - .040B - .190B^2)a_{nt}$

$V_n = .521$

Final Error var.	Revision error var.	Total estimation error var.
.193	.202	.395

Minimax filter on concurrent estimator.

Model specification: $\Delta^2 n_t = (1 + .048B - .952B^2)a_{nt}$

$V_n = .149$

Final Error var.	Revision error var.	Total estimation error var.
.188	.221	.409

9.5 Japan Exports Series

For the Japan Export series (1972-1 1992-7), with sample size $T = 247$, a $(2,1,1)(0,1,1)_{12}$ model was fitted. The parameters estimators are presented on the following table :

Table 9.7

Model fitted: $(1 - \phi_1 B - \phi_2 B^2)\Delta\Delta_{12} x_t = (1 - \theta_1 B)(1 - \theta_{12} B^{12}) a_t$

Parameters:	ϕ_1	ϕ_2	θ_1	θ_{12}	V_a	Q_{20}
Estimators	-.778	-.379	-.231	.650	.179 10^{-3}	29.32
	(.162)	(.085)	(.186)	(.081)		

Three outliers are found at t=29, 76, 86, and their t-values are 3.33, -3.03, 2.54. The AR polynomial has conjugate roots at a frequency close to $2\pi/3$, and are thus assigned to the seasonal component.

The Japan Export series and the estimators of the Trend and of the SA series are plotted on figures 9.10 and 9.11.

Denoting s_t^0 and n_t^1 the canonical seasonal and trend components of x_t, and computing the related WK filters, we get: $\nu_{s0}^0 = .439$, $\nu_{n0}^1 = .283$, $\xi_{s0}^0 = .401$, $\xi_{n0}^1 = .434$. Lemma 6.2 and 6.6 allow us to deduce that a canonical trend will yield the most accurate historical estimator, while for concurrent estimation, the most precisely estimated decomposition will be the one with canonical seasonal. This is thus a case of switching solutions.

In order to characterize the error variance for intermediate decompositions, we derive: $V(e_t^0) = .111$, $V(d_t^0) = .202$, and $h_0 = 1.690$. In this model, the maximum variance of the irregular component is : $V_u = .165 V_a$.

Figure 9.10: Japan Exports Series (1972-1 1992-7)

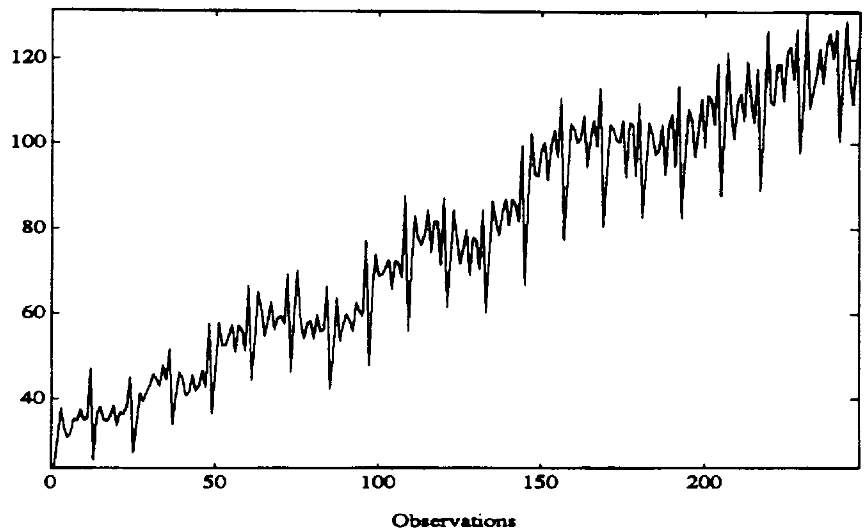


Figure 9.11: XJP: SA Series / Trend

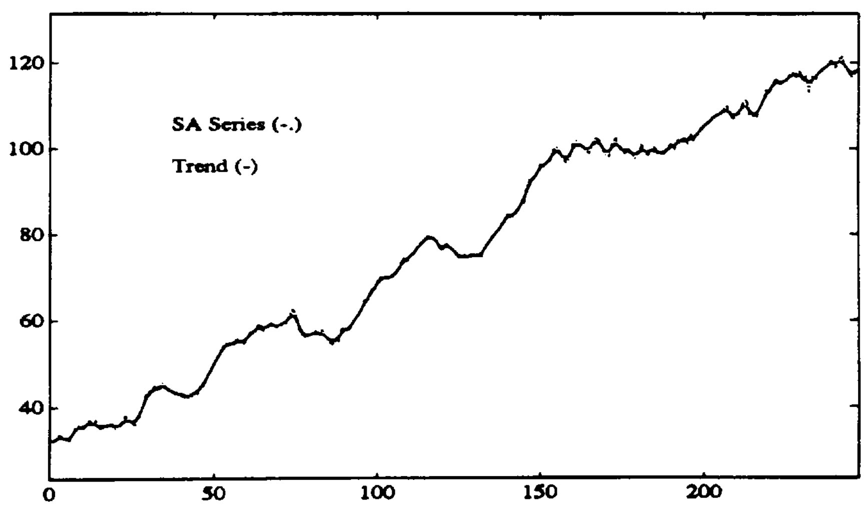
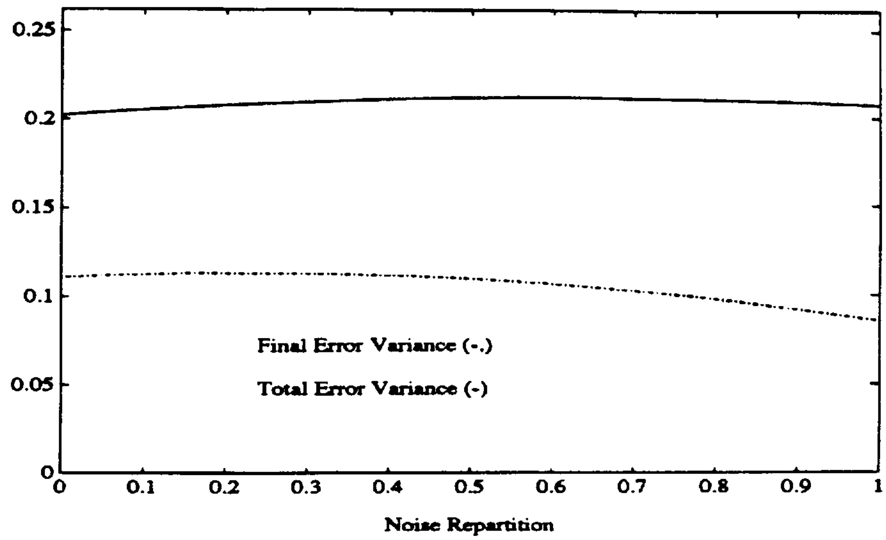


Figure 9.12: XJP: Mean Squared Errors



The historical error variance is then related to the noise repartition α through:

$$V(e_t^{\alpha}) = .111 + .020\alpha - .046\alpha^2,$$

and for the concurrent estimators:

$$V(d_t^{\alpha}) = .202 + .032\alpha - .027\alpha^2,$$

These functions are plotted in figure 9.12, and the errors sizes for five different decompositions are given on table 9.8.

Figure 9.12 ilustrates that, for historical estimators, the error variance and lag-0 cross-covariance are minimized at .085 with a noninvertible trend component. At the oppsite, a minimax filter yields a maximum error variance of .112. From specifying a canonical trend in order to minimize the historical error, it follows that the revision error is maximized at .122. This implies the switching of solutions, in the sense that now the lowest concurrent estimation error variance (.207) is obtained on a trend concentrating all the noise of the model (the seasonally adjusted series). Conversely, the highest error in concurrent estimators is reached at .212 with the minimax filter. Specifying a canonical trend instead of using the minimax filter induces thus a gain of precision of 25% on the historical estimators, while, on the concurrent estimator, the gain of precision may only be of 4%.

Here, the model specification minimizing both the error and the covariances between the historical estimators corresponds to a canonical trend, which however maximizes the revision error. The magnitude of the revision error in this case implies that the error in concurrent estimators is minimized on the other canonical decomposition.

Table 9.8 XJP: Models for the nonseasonal component:		
SA Series/Canonical seasonal.		
Model specification: $\Delta^2 n_t = (1 - 1.238B + .263B^2)a_{nt}$		
$V_n = .421$		
Final Error var.	Revision error var.	Total estimation error var.
.111	.091	.202
Canonical Trend.		
Model specification: $\Delta^2 n_t = (1 + .035B - .965B^2)a_{nt}$		
$V_n = .056$		
Final Error var.	Revision error var.	Total estimation error var.
.085	.122	.207
Restricting the order of the trend MA polynomial.		
Model specification: $\Delta^2 n_t = (1 - .965B^2)a_{nt}$		
$V_n = .223$		
Final Error var.	Revision error var.	Total estimation error var.
.103	.108	.211
Minimax filter on historical estimator.		
Model specification: $\Delta^2 n_t = (1 - 1.179B + .207B^2)a_{nt}$		
$V_n = .362$		
Final Error var.	Revision error var.	Total estimation error var.
.112	.095	.208
Minimax filter on concurrent estimator.		
Model specification: $\Delta^2 n_t = (1 - 1.015B + .049B^2)a_{nt}$		
$V_n = .247$		
Final Error var.	Revision error var.	Total estimation error var.
.106	.106	.212

Chapter 10

Conclusion

For the general signal/nonsignal decompositions, this dissertation analyses the incidence on the estimation errors of the specification of the unobserved component models. The analysis involves the class of ARIMA models, but in fact the results are valid for any linear stochastic processes. The choice of the ARIMA models as a tool for the discussion has been motivated in Chapter 2 through a brief review of the steps leading to the development of model-based approaches considering stochastic components.

Whatever is the approach undertaken, the unobserved component analysis cannot avoid the identification problem. We have seen in chapter 3 that the underidenfication of the components turns out to be simply a problem of allocating a certain amount of white noise between the components. Every noise repartition yields a particular decomposition. When reviewving the identification criteria in use in the statistical literature, we have seen that one of them, the canonical decomposition, assigns all the noise to one component and lets the other one noninvertible. In the two-components decompositions, there are thus two possible canonical decompositions. Since any intermediate decomposition shares the noise between the components, the range of the admissible decompositions may be seen as lying between these two.

Next, we have presented in Chapter 4 the estimation procedure in unobserved components models. We have seen that estimation error depends on the time at which the estimation of the signal is computed. We obtained in particular an error in the historical or final estimator, an error in the concurrent estimator, and more generally the error in any preliminary estimator. Some important properties of the final estimator have been derived, and we established the result that estimators cross-covariances and final error covariances are identical. Also, it appeared that the estimation errors and the stochastic properties of the estimators depend on the selected models for the components.

In Chapter 5, the analytical expression giving the estimation error variance for any admissible decomposition and for any type of error has then been derived. The coefficients involved in these expressions can be straightforwardly obtained from the overall model. It has been shown that for historical, concurrent or preliminary estimators, the error variance is always minimized with a canonical decomposition. Since these decompositions are not unique, simple rules are derived to indicate which component must be held canonical in order to obtain the best estimated model. These rules involve the B^0 coefficients of the WK filters designed to estimate the components in their canonical form. They simply consists in assigning all the noise of the model to the component with greatest B^0 coefficient, the other one being specified noninvertible. This procedure also yields final estimators with a minimum lag-0 cross-covariance. Moreover, we have shown that canonical decompositions still display these optimal properties of maximizing the precision of the estimators when rates of growth are considered.

We believe that this work has practical relevance. While unobserved components models are not identified, much of the statistical literature on unobserved components skips the identification stage by postulating "reasonable" models for the components. However, any particular choice of a specification is arbitrary and difficult to motivate, and how the identifying assumptions affect the components properties is not well-known. We thus thought it worthwhile to investigate how the estimators were affected by such *a priori* choices.

This research may be seen in the line of Bell and Hillmer's (1984) discussion about seasonally adjusting time series. In attempting to evaluate different seasonal adjustment methods, they reached the conclusion that it was impossible to compare different approaches, saying that "different methods produce different adjustments because they make different assumptions about the components and hence estimate different things". As a consequence, they recommended that the debate should center on the assumptions made about the components, arguing that "efforts would be better spent evaluating the assumptions underlying adjustment methods, rather than trying to evaluate methods by looking at adjusted data". Then, they tried to use the revision error in the concurrent estimator to evaluate the different assumptions. As we have seen in section 6.2, for the models that they considered, no particular approach gives the best results. Everything depends on the stochastic properties of the series under analysis, so that their next proposition for evaluating a method for seasonal adjustment was that it "...should be consistent with an adequate model for the observed data". They clearly supported model-based procedures, but the problem remained to supply a motive for a particular decomposition, that is to evaluate all the different assumptions that are possible to make on the components. Our dissertation brings a simple answer to this problem: since the decompositions differ only by a white noise allocation, select the decomposition which can be best estimated. Through Bell and Hillmer discussion, it appeared very important to explicitly relate the possible assumptions to a simple criterion. The choice of the error variance in the estimator as a criterion for evaluating different assumptions seems perfectly justified and it also greatly simplifies the debate because the error variance always can be expressed as a second order polynomial with coefficients that can be easily derived from the observed series model. Such a choice has the advantage of being able to take into account the objectives of the analysis: consider the final estimation error if you are interested in historical estimators, or focus on the error in concurrent estimators if you are interested in concurrent analysis, and similarly with preliminary estimators and forecasts.

References

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

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

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

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

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

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

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

Conference on Applied Time Series Analysis of Economic Data.

Beveridge, W.H. (1921), "Wheather and Harvest Cycles", *Economic Journal*, 31, 429-452.

Beveridge, W.H. (1922), "Wheat Prices and Rainfall in Western Europe", *Journal of the Royal Statistical Society*, 85, 412-459.

Beveridge, S. and Nelson, C.R. (1981), "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'", *Journal of Monetary Economics*, 7, 151-174.

Box, G.E.P and Jenkins, G.M (1970): *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

Box, G.E.P, Hillmer S.C., and Tiao G.C. (1978), "Analysis and Modeling of Seasonal Time Series", in *Seasonal Analysis of Time Series*, ed. A. Zellner, Washington, D.C: U.S. Department of Commerce, Bureau of the Census, 309-334.

Box, G.E.P., Pierce, D.A., and Newbold, P., (1987), "Estimating Trend and Growth Rates in Seasonal Time Series", *Journal of the American Statistical Association*, 82, 276-282.

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

Burridge, P. and Wallis, K.F. (1984), "Unobserved Components Models for Seasonal Adjustment Filters", *Journal of Business and Economics Statistics*, 2, 350-359.

Burridge, P. and Wallis, K.F., (1985), "Calculating the Variance of Seasonally Adjusted Series", *Journal of the American Statistical Association*, 80, 391, 541-552.

Buys Ballot, C, H. D. (1847), *Les Changements Périodiques de Température*, Utrecht: Kemink et Fils.

Cleveland, W.P. (1972), "The Inverse Autocorrelations of a Time Series and their Applications", *Technometrics*, 14, 277-298.

Cleveland, W.P. and Tiao, G.C. (1976) "Decomposition of Seasonal Time Series: a Model for the X-11 Program", *Journal of the American Statistical Association*, 71, 581-587.

Cochrane, J.H. (1988), "How Big is the Random Walk in GNP", *Journal of Political Economy*, 96, 893-920.

Dagum, E.B. (1980), "The X11 ARIMA Seasonal Adjustment Model", Statistics Canada, Catalogue 12-564E.

Durbin, J. (1960), "The Fitting of Time-Series Models", *Review of the International Statistical Institute*, 28, 233-243.

Engle, R.F. (1978), "Estimating Structural Models of Seasonality", ed. A. Zellner, in *Seasonal Analysis of Economic Time Series*, Washington, D.C.: Dept. of Commerce, Bureau of the Census, 281-297.

Findley D.F. (1985), "An analysis and generalisation of M. Watson's minimax procedure for determining the component estimates of a seasonal time series", Washington, D.C.: Statistical Research Division, Bureau of the Census.

Friedman, M. (1957), *A Theory of the Consumption Function*, New York: National Bureau of Economic Research.

Garcia-Ferrer A., Del Hoyo J. (1992): "On Trend Extraction Models : Interpretation, Empirical Evidence and Forecasting Performances", *Journal of Forecasting*, 11, 645-665.

Gersch, W. and Kitagawa, G. (1983), "The Prediction of Time Series with Trends and Seasonalities", *Journal of Business and Economics Statistics*, 1, 253-264.

Gilbart, J.W. (1852), "On the Laws of the Currency in Ireland, as Exemplified in the Changes That Have Taken Place in the Amount of Bank Notes in Circulation in Ireland, Since the Passing of the Acts of 1845", *Journal of the Statistical Society of London*, 15, 307-325.

Gourieroux, C. and Monfort, A. (1990), *Series Temporelles et Modèles dynamiques*, Paris: Economica.

Granger, C.W.J. (1978), "Seasonality: Causation, Interpretation and Implications", in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington D.C.: U.S. Dept. of Commerce, Bureau of the Census, 33-46.

Harrisson, P.J. and Steven, C.F. (1976), "Bayesian Forecasting", *Journal of the Royal Statistical Society*, Ser. B, 38, 205-247.

Harvey, A.C. (1985), "Trends and Cycles in Macroeconomic Time Series", *Journal of Business and Economic Statistics*, 3, 3, 216-227.

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

Harvey, A.C. and Koopman, S.J. (1992), "Diagnostic checking of Unobserved Components Time Series Models", *Journal of Business and Economic Statistics*, 10, 377-390.

Harvey, A.C., and Todd, P.H.J. (1983), "Forecasting Economic Time Series with Structural and Box-Jenkins Models : a Case Study", *Journal of Business and Economic Statistics*, 1, 299-307.

Haussman, J.A., and Watson, M.W. (1985), "Errors-in-Variables and Seasonal Adjustment Procedures", *Journal of the American Statistical Association*, 80, 531-540.

Hibbert Committee (1988), "Report of the Working Party on Seasonal Adjustment of the Monetary Aggregates", *CSO* 8.

Hillmer, S.C. (1985), "Measures of Variability for Model-Based Seasonal Adjustment Procedures", *Journal of Business and Economic Statistics*, 3,1,60-68.

Hillmer S.C. , and Tiao G.C. (1982) "An ARIMA-Model-Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, 77, 63-70.

Hodrick, R. and Prescott, E. (1980), "Post-War U.S. Business Cycles: An Empirical Investigation", Carnegie Mellon University Manuscript.

Hotta, L.K. (1989), "Identification of Unobserved Components Models", *Journal of Time Series Analysis*, 10, 259-270.

Jenkins, G.M. (1979), "Practical Experience with Modeling and Forecasting Time Series", ed. Anderson, O.D., *Forecasting*, Amsterdam: North-Holland.

Joy, A. and Thomas, W. (1928), "The Use of Moving Averages in the Measurement of Seasonal Fluctuations", *Journal of the American Statistical Association*, 22, 241-252.

Kendall, M. (1976), *Time Series*, London: Griffin and Co.

King, W.W.I. (1924), "An Improved Method for Measuring the Seasonal Factor", *Journal of the American Statistical Association*, 19, 301-313.

Kohn, R. and Ansley, C.F (1986), "Estimation, Prediction and Interpolation for ARIMA Models with Missing Data", *Journal of the American Statistical Association*, 81, 751-761.

Kohn, R. and Ansley, C.F. (1987), "Signal Extraction for Finite Nonstationary Time Series", *Biometrika* , 74,411-421.

Kuznets, S. (1933), "Seasonal Variations in Industry and Trade", New York: National Bureau of Economic Research.

Lucas, R.E. (1976), "Understanding Business Cycles", in *Stabilization of the Domestic and International Economy*, ed. Brunner, K. and Meltzer, A., Amsterdam: North-Holland.

Macaulay, F.R. (1931), *The smoothing of Time Series*, New York: National Bureau of Economic Research.

Mann, H.B. and Wald, A. (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, 11, 173-220.

Maravall, A. (1985), "On Structural Time series Models and the Characterization of Components", *Journal of Business and Economic Statistics*, 3, 4, 350-355.

Maravall, A. (1986), "Revisions in ARIMA Signal Extraction", *Journal of the American Statistical Association*, 81, 736-740.

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

Maravall, A. (1988b), "The Use of ARIMA Models in Unobserved Components Estimation", in *Dynamic Econometric Modeling*, ed. Barnett, W., Berndt, E. and White, H., Cambridge: Cambridge University Press.

Maravall, A. (1989), "On the Dynamic Structure of a Seasonal Component", *Journal of Economic Dynamics and Control*, 12, 589-593.

Maravall, A. (1993a), "Stochastic Linear Trends: Models and Estimators", *Journal of Econometrics*, 54, 1-33.

Maravall, A. (1993b), "Unobserved Components in Economic Time Series", forthcoming in *Handbook of Applied of Econometrics*, ed. M.H. Pesaran, T. Schmidt, M. Wickens, Oxford Basil Blackwell.

Maravall, A. (1994), "Use and Misuse of Unobserved Components in Economic Forecasting", *Journal of Forecasting*, 13, 157-177.

Maravall, A. and Gómez, V. (1992), "Signal Extraction in Economic Time Series: Program SEATS", EUI Working Paper ECO No. 92/65, Department of Economics, European University Institute.

Maravall, A. and Mathis, A. (1994), "Encompassing Univariate Models in Multivariate Time Series": A Case Study", *Journal of Econometrics*, 61, 197-223.

Maravall, A. and Pierce, D.A. (1983), "Preliminary-Data Error and Monetary Aggregate Targeting", *Journal of Business and Economics Statistics*, 1,3,179-186.

Maravall, A. and Pierce, D.A. (1986), "The Transmission of Data Noise into Policy Noise in U.S. Monetary Control", *Econometrica*, 54, 4, 961-979.

Maravall, A. and Pierce, D.A. (1987), "A prototypical Seasonal Adjustment Model", *Journal of Time Series Analysis*, 8, 177-193.

Nelson, C.R. and Plosser, G.J. (1982), "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10, 139-162.

Nerlove, M. (1964), "Spectral Analysis of Seasonal Adjustment Procedures", *Econometrica*, 32, 241-286.

Nerlove, M., Grether, D.M, and Carvalho, J.L. (1979), *Analysis of Economic Time Series: A Synthesis"*, New York: Academic Press.

Ng, C.N. and Young, P.C. (1990), "Recursive Estimation and Forecasting of Nonstationary Time Series", *Journal of Forecasting*, 9, 173-204.

Osborn, D. (1993), "Moving Average Detrending of Integrated Processes", forthcoming in *Journal of Econometrics.*

Pagan A. (1975), "A Note on the Extraction of Components from Time Series", *Econometrica*, 43, 1, 163-168.

Persons, W. M. (1919), "Indices of Business Conditions", *Review of Economics and Statistics*, 1, 5-107.

Persons, W.M. (1923), "Correlation of Time Series", *Journal of the American Statistical Association*, 18, 713-726.

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

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

Pierce, D.A. (1980), "Data Revisions in Moving Average Seasonal Adjustment procedures", *Journal of Econometrics*, 14, 1, 95-114.

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.

Phillips, P.C.B. and Perron, P. (1988), "Testing for Unit Roots in Time Series Regression", *Biometrika*, 75, 335-346.

Robb, R.A. (1929), "The Variate Difference Method of Seasonal Variation", *Journal of the American Statistical Association*, 24, 250-257.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica*, 39, 577-592.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program", Technical Paper 15, Washington, D.C.: Bureau of the Census.

Slutsky, E. (1937), "The Summation of Random Causes as a Source of Cyclical Processes", *Econometrica*, 5, 105-146.

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

Stock, J.H. and Watson, M.W. (1989), "New Indexes of Leading and Coincident Economic Indicators", National Bureau of Economic Research, *Macroeconomic Annual*, 351-394.

Stock, J.H. and Watson, M.W. (1991), "A Probability Model for Predicting Recessions with Leading Indicators", in *Leading Economic Indicators: New Approaches and Forecasting Records*, ed. Lahiri, K., and Moore, G.H., Cambridge: Cambridge University Press.

Stock, J.H., and Watson, M.W. (1993), " A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Performances", in *New Business Cycle Indicators and Forecasting*, ed. Stock, J.H. and Watson, M.W., Chicago: Chicago University Press.

Walker, A.M. (1962), "Large Sample Estimation of Parameters for Autoregressive Processes with Moving Average Residuals", *Biometrika*, 49,117-131.

Watson, M.W. (1986), "Univariate Detrending Methods with Stochastic Trends", *Journal of the Monetary Economics*, 18,49-75.

Watson, M. W. (1987), "Uncertaincy in model-based-seasonal adjustment procedures and construction of minimax filters", *Journal of the American Statistical Association*, 82, 398, 395-408.

Weber, A.A. (1992), "The Role of Policymakers' Reputation in the EMS Disinflations. An Empirical Evaluation", *European Economic Review*, 36, 1473-1492.

Whittle, P. (1952), "Tests of Fits in Time Series", *Biometrika*, 39, 309-318.

Whittle P. (1963), *Prediction and Regulation using Least-Square Methods*, London: English Universities Press.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Uppsala: Almqvist and Wicksell,

Yule, G.U. (1927), "On a Method of Investigating Periodicities in Disturbed Series, With Special Reference to Wolfer's Sunspot Numbers", *Philosophical Transactions of the Royal Society*, Ser. A, 226, 267-298.