Diagnostic checking and Intra-daily effects in Time series models

By
Siem Jan Koopman

London
January 1992

A thesis submitted for a degree of Doctor of Philosophy
The London School of Economics and Political Science

ABSTRACT

A variety of topics on the statistical analysis of time series are addressed in this thesis. The main emphasis is on the state space methodology and, in particular, on structural time series (STS) models. There are now many applications of STS models in the literature and they have proved to be very successful.

The keywords of this thesis vary from - Kalman filter, smoothing and diagnostic checking - to - time-varying cubic splines and intra-daily effects -. Five separate studies are carried out for this research project and they are reflected in the chapters 2 to 6. All studies concern time series models which are placed in the state space form (SSF) so that the Kalman filter (KF) can be applied for estimation. The SSF and the KF play a central role in time series analysis that can be compared with the important role of the regression model and the method of least squares estimation in econometrics. Chapter 2 gives an overview of the latest developments in the state space methodology including diffuse likelihood evaluation, stable calculations, etc.

Smoothing algorithms evaluate the full sample estimates of unobserved components in time series models. New smoothing algorithms are developed for the state and the disturbance vector of the SSF which are computationally efficient and outperform existing methods. Chapter 3 discusses the existing and the new smoothing algorithms with an emphasis on theory, algorithms and practical implications. The new smoothing results pave the way to use auxiliary residuals, that is full sample estimates of the disturbances, for diagnostic checking of unobserved components time series models. Chapter 4 develops test statistics for auxiliary residuals and it presents applications showing how they can be used to detect and distinguish between outliers and structural change.

A cubic spline is a polynomial function of order three which is regularly used for interpolation and curve-fitting. It has also been applied to piecewise regressions, density approximations, etc. Chapter 5 develops the cubic spline further by allowing it to vary over time and by introducing it into time series models. These time-varying cubic splines are an efficient way of handling slowly changing periodic movements in time series.

This method for modelling a changing periodic pattern is applied in a structural time series model used to forecast hourly electricity load demand, with the periodic movements being intra-daily or intra-weekly. The full model contains other components, including a temperature response which is also modelled using cubic splines. A statistical computer package (SHELF) is developed to produce, at any time, hourly load forecasts three days ahead.

ACKNOWLEDGEMENTS

I am very grateful to my parents for their love and invaluable encouragement throughout the years.

I thank Professor Andrew Harvey for his support and supervision. This thesis has never been written without his knowlegde and enthusiasm. Also, I like to thank Professor Piet de Jong for his support and many discussions.

This work has never been accomplished without the love, help and support of many but especially Margreet, Pauline, Karin, Regina, Gerda, Zoila, Arthur, George and Wim. Furthermore, I wish to mention, in alphabetical order, Casey Brace, Dr. Cristiano Fernandes, Professor Jan de Gooijer, Klaus Jackisch, Dr. M. Knott, Rob Luginbuhl, Dr. Esther Ruiz and Dr. Neil Shephard.

For the general support in the Netherlands, I thank the Tinbergen Institute (Dr. Annemarie Rima and Miep Oomes) at the University of Amsterdam. I am indebted to the Electric Power Research Institute (EPRI) in Palo Alto, California, USA, and the Puget Light Power & Sound company, Bellevue, Washington, USA, for their financial support.

Finally, I like to thank my teachers in econometrics at the University of Amsterdam, the Professors J.S. Cramer, J.F. Kiviet and H. Neudecker.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACF	Autocorrelation function
AIC	Akaike information criterion
ARIMA model	Autoregressive-integrated-moving average model
BSM	Basic structural time series model
CCF	Cross-correlation function
EWMA	Exponential weighted moving average
HOT	Householder orthogonal transformation
KF	Kalman filter
LL model	Local level model
LLT model	Local linear trend model
MAPE	Mean absolute percentage error
OLS	Ordinary least squares
RWN model	Random walk plus noise model
SHELF	Structural hourly electricity load forecaster
SSF	State space form
STAMP	Structural time series analyser modeller predictor
STS model	Structural time series model

CHAPTER 1
INTRODUCTION

This thesis deals with a variety of topics concerning the statistical analysis of time series. It is my belief that structural time series models provide the most satisfactory framework for analysing time series since it does not rely on all kinds of subjective judgements to the process of time series modelling. Moreover, the structural approach of time series modelling easily adopts all powerful results related to the state space model and the Kalman filter, it provides a full set of different diagnostic tools to validate the estimated model and it has a close connection with various econometric methodologies of modelling. But most important, the structural time series model is preferable because it provides a direct appeal to the interpretability of unobserved components in time series.

Five separate studies are carried out for this research project and, basically, they all concern the linear state space model and the Kalman filter which, jointly, play a central role in time series analysis such as the regression framework plays a central role in econometrics.

State space form and the Kalman filter

The first part describes briefly but clearly the latest developments in the state space methodology such as diffuse initial conditions, fixed effects in time series models, likelihood evaluation, prediction, diagnostic checking and computational considerations. This part is primarily based on research of Harvey (1989), De Jong (1988a,1991b), Anderson & Moore (1979) and Ansley & Kohn (1985).

Smoothing algorithms for state and disturbance vector

The second part of this thesis is a study on smoothing in time series models and it concentrates on theory, algorithms and

practical implications. Essentially, it proposes to concentrate on the disturbance vector of the SSF. A special smoothing algorithm is developed to calculate the auxiliary residuals, that is the full-sample estimates of the elements of the disturbance vector. A new method to calculate smoothed (full sample) estimates of the state vector is developed which is computationally efficient and outperforms existing methods, see Anderson & Moore (1979), De Jong (1988b,1989) and Kohn & Ansley (1989), for most practical time series models. Furthermore, the new residual smoother leads to an attractive method of estimating covariance parameters of a time series model. This estimation method can be regarded as an essential improvent of the EM algorithm, see Watson & Engle (1983).

Diagnostic checking of innovations and auxiliary residuals

The new smoothing results pave the way to use auxiliary (smoothed) residuals for diagnostic checking of unobserved component time series models. Commonly, diagnostic checking is carried out using the standardised innovations. These residuals can be compared with the recursive residuals in the regression framework, see Brown, Durbin & Evans (1975). Due to the nice properties of the innovations, e.g. they are serially uncorrelated, it is attractive to use them for tests on normality, heteroscedasticity, serial correlation, etc. Auxiliary residuals are obtained from polynomial functions (within the forward time operator) of the innovations and they are not serially uncorrelated even when the parameters of the model are known, see Maravall (1987). Therefore, the diagnostic checks on normality and heteroscedasticity for the auxiliary residuals must be corrected. It seems at first sight that the auxiliary residuals are not very useful for diagnostic checking because they contain no new information and they do not possess special properties. It will be shown that, on the contrary, they are very useful because they present the information, available from innovations, in such a way that several problems in time series modelling, such as detecting outliers and structural change, are tackled straightforwardly. This research project has examined how the auxiliary residuals are related to each other and to the

innovations and how they they can be used to construct test-statistics. A number of examples are given as illustrations.

Time-varying cubic splines

The fourth part of the thesis deals mainly with the structural time series model applied to periodic time series where the observations are affected by intra-daily or intra-weekly effects. The basic structural time series model consists of the unobserved components trend, seasonal and irregular. Many seasonal time series can be effectively described by the basic structural time series model, see Harvey (1989). For example, the periodicity of monthly observations is modelled by time-varying dummy parameters or, alternatively, by time-varying trigonometric terms. For observations at more frequent intervals, e.g. daily or hourly observations, the dummy or trigonometric seasonal component may not be very parsimonious for modelling the periodic effect.

This study explores the idea of time-varying cubic splines to handle frequent periodic movements in time series. The spline function is well-known as a technique to describe a complicated and, perhaps, unknown non-linear function. A particular spline function, the cubic spline, can be easily established in the regression framework, see Poirier (1973). The periodic cubic spline is defined by imposing special begin- and end-conditions on the cubic spline. In the context of time series, it is restrictive to consider the periodic cubic spline as fixed over time. Treating the periodic cubic spline as time-varying and incorporating it in a time series model causes some theoretical problems but they can be solved straightforwardly. The required calculations for imposing a time-varing periodic cubic spline in a time series model are implemented in an efficient algorithm that avoids standard matrix inversions.

Forecasting hourly electricity load demand

A structural time series model with time-varying splines is applied to the problem of forecasting total hourly electricity load demand at an electricity power station in the north-west of the

United States of America. A computer model, which will be referred to as SHELF, is developed to produce, at any time, three days ahead hourly load forecasts. In this illustration, explanatory wheather variables are included in the time series model to improve the forecasts. Two problems arise, (i) the response of weather on the load demand is non-linear and (ii) the response of weather is different at specific hours during the week. These problems are handled by using fixed cubic spline functions.

Preliminaries and notation

A time series is denoted by y_t where t=1,...,n such that the length of the time series is n. The full set of observations is denoted by Y, the sub-set of observations $y_1,...,y_t$ is denoted by Y_t and $Y \equiv Y_n$. The logarithm of the joint density function under normality for the set of observations Y is denoted by l[y] and the, so-called, likelihood criterion is denoted by L[y] so that $l[y] = -\frac{1}{2} L[y] +$ constants. The likelihood criterion conditional on a set of random variables δ is denoted by $L[y|\delta]$ and the concentrated likelihood criterion with respect to δ is written as $L_\delta[y]$. The minimum mean square linear estimator of any unknown x is denoted by E[x] with its mean square error matrix Mse[x]=Cov(x-E[x]). When the estimator is constructed using the data-set Y, the notation becomes E[x|y] and Mse[x|y], respectively. Under normality assumptions with regards to y, the estimator E[x|y] is no longer confined to the class of linear estimators. Thus, under normality, E[x|y] can be referred to as the minimum mean square estimator and, indeed, E[x|y] can be regarded as the conditional expectation of x using y.

A (nx1) column vector v is a stack of n elements and is denoted by $v=(e_1;...;e_n)$. A (1xn) row vector w is written as $w=(e_1,...,e_n)$. The transpose of any matrix M is denoted by M'. Let the (nx2) matrix M consist of, respectively, 2 (nx1) column vectors u and v then M=(u,v). Let the (2xn) matrix N consist of, respectively, 2 (1xn) row vectors w and x then N=(w;x). This notation implies that (A',B')'=(A;B) where A and B are matrices with the same number of columns. Furthermore, the trace of a matrix A is tr A and the determinant of a matrix A is |A|. The notation for a

zero value is 0 but it can be a scalar, vector or matrix.

Finally, a section of a chapter is only referred to by its number when it can be found in the same chapter otherwise a section is denoted by two numbers from which the first number refers to the chapter, for example section 2.3 (chapter 2 section 3). The sub-sections are not numbered. The equations are indicated by two numbers from which the first one is the section number. When an equation is referred from another chapter, it is indicated by three numbers from which the first number is the chapter number and the second number is the section number. The sub-equations are numbered with use of an additional small letter. The tables and figures are all numbered by two numbers from which the first one refers to the chapter. The appendices are indicated by the number of the chapter and a capital letter. A note is indicated by a number surrounded by squared brackets and it can be found at the end of a chapter.

CHAPTER 2

THE STATE SPACE FORM AND THE KALMAN FILTER : theory, algorithms and applications

0. ABSTRACT

This chapter considers the linear state space model and the Kalman filter which, jointly, play an important role in time series analysis. The latest developments in the state space methodology are discussed. These include diffuse initial conditions, fixed effects in time series models, log-likelihood evaluation, prediction, diagnostic checking and computational considerations. This chapter is primarily based on research of Harvey (1989), De Jong (1988a,1991b), Anderson & Moore (1979) and Ansley & Kohn (1985).

Keywords : Diagnostic checking; Diffuse; Kalman filter; Likelihood evaluation; Prediction; Square root filter; State space; Structural time series models.

1. INTRODUCTION

The state space form (SSF) is a natural set up for linear time series models. Although originating in the engineering literature, several authors, e.g. Harvey (1981), have emphasized that the SSF has a potential to be important in the study of time series in statistics and econometrics. A main attraction is its generality: all linear time series models can be placed in the state space form. For example, the linear regression model is naturally embedded within the SSF but also any autoregressive moving average (ARMA) model can be put into this framework. The basic tool for statistical analysis with the SSF is the Kalman filter (KF) which

was developed by Kalman (1960) and Kalman & Bucy (1961). The Kalman recursions give one-step ahead predictions allowing parameter estimation via the prediction error decomposition. Moreover, the Kalman filter is of importance for analytical and practical matters related to estimating initial and fixed effects, smoothing, prediction and other topics in time series analysis.

The literature contains many examples nowadays where the state space methodology is applied successfully. Some good examples in practical time series analysis are Burridge & Wallis (1985), Burmeister et.al. (1986), Engle & Watson (1981) and Harvey & Phillips (1979). Also, the new developments in structural time series models and its applications contribute to the state space methodology, see Harvey (1989) and Harvey & Shephard (1992). Recently, the state space form is modified to deal with initial and fixed effects in time series models explicitly, see De Jong (1991b). This modification paves the way to estimate these effects directly and to compute the likelihood function exactly using the Kalman filter. The expression for the likelihood function depends on the assumptions regarding the initial effects which may be fixed, random or diffuse.

This chapter concerns the state space form, the Kalman filter, likelihood evaluation and some other related issues in time series analysis. The organisation of this chapter is as follows. Section 2 presents the state space form which is adjusted to allow for fixed effects. Several practical time series models are discussed in section 3. The derivation of the Kalman filter is given in section 4 and it is modified in section 5 in order to estimate fixed and initial effects and to evaluate the likelihood under several assumptions.

Several diagostics are available to validate how well the time series model describes the data generation process. Commonly, diagnostic checking is carried out using the one-step ahead prediction errors which are also known as the innovations. These residuals are obtained from the Kalman filter. The methodology of structural time series modelling, where these diagnostics play an important role, is illustrated by using a real time series in section 6. To ensure numerical stability, the Kalman recursions can

be formulated in a square root form. This can be compared with numerical stable approaches to least-squares computations in regression analysis. Section 7 explains a numerical orthogonalization technique (the Householder transformation) and shows how the square root form can be implemented. Finally, a general Pascal computer program is given in section 8 for some required computations discussed in this chapter. Section 9 concludes this chapter.

2. STATE SPACE FORM

The state space form (SSF) consists of the following two equations

$$y_t = Z_t\alpha_t + X_t\beta + G_t\epsilon_t \qquad (2.1a)$$
$$\alpha_{t+1} = T_t\alpha_t + W_t\beta + H_t\epsilon_t \qquad (2.1b)$$

where y_t is a vector of time series for t=1,...,n. The vector α_1 and the vector β are given by

$$\alpha_1 = W_0\beta + H_0\epsilon_0 \qquad \beta = b + B\delta \qquad (2.2)$$

The disturbance vector ϵ_t and the parameter vector δ are both random with distributions

$$\epsilon_t \sim (0 , \sigma^2 I) \qquad \delta \sim (\mu , \sigma^2\Lambda) \qquad (2.3)$$

where δ and ϵ_t are uncorrelated for t=0,...,n. The matrices Z_t, X_t, G_t, T_t, W_t, H_t, for t=1,...,n, and (b,B), W_0 and H_0 are called system matrices and are supposed to be fixed and known. Equation (2.1a) is referred to as the measurement equation and equation (2.1b) is called the transition equation. The vector α_t is the state vector and is not fully defined without a proper definition for the initial state vector α_1 as given in (2.2). To separate known and unknown elements of β but also to incorporate linear restrictions, the vector β is parameterized by (2.2).

The state space form (2.1)-(2.3) is proposed by De Jong (1991b) and it is analytically attractive because it accomodates random, fixed and diffuse effects in time series models explicitly rather than only time-varying effects. Moreover, this SSF handles correlation between the disturbances in the measurement and transition equations. The following remarks clarify the SSF in more detail.

[1] The system matrices of (2.1) are time-varying. A time series model is time invariant if the system matrices Z_t, T_t, G_t and H_t do not change over time for t=1,...,n. The initial system matrices W_0 and H_0 are allowed to differ from the time invariant system matrices W_t and H_t.

[2] It is not restrictive that the parameter vector β appears in the measurement and transition equations. The matrices X_t and W_t may be interpreted as selection matrices. Note that for many time series models the restriction $X_tW_t'=0$ holds.

[3] The matrices G_t and H_t transform the sequence of disturbance vectors into a multivariate random process with zero mean and a specific time-varying covariance structure. For the special case that $H_tG_t'=0$ for t=1,...,n, the measurement equation and the transition equation are said to be uncorrelated.

[4] Under the assumption that the disturbance vector is normally distributed, the related estimators are minimum mean square estimators (mmse). If the normality assumption does not apply, the inference results are still valid but the estimators are optimal within the class of linear estimators, i.e. they are minimum mean square linear estimators (mmsle); see Duncan & Horne (1972) and Anderson & Moore (1979).

[5] It is well known that any linear time series process can be placed into the Markovian representation as given by the transition equation of (2.1b), see Harvey (1981).

[6] The dimension of the state vector $\#[\alpha_t]$ may also vary over time. This is particularly useful for interventions in time series models where specific effects are only valid for some sub-set of observations.

[7] The parameter vector β is a linear combination of the random vector δ. When the covariance matrix Λ is set to zero, the parameter

vector β is fixed but unknown. The initial state vector can still be, partially, random by an appropriate choice of H_0. The SSF may handle partially diffuse initial conditions as discussed by Ansley & Kohn (1985), De Jong (1988a,1991b) and Marshall (1992). The random vector δ is said to be diffuse when the inverse of Λ converges to zero in the Euclidian norm, denoted by $\Lambda \to \infty$. The diffuse assumption reflects parameter uncertainty and it needs to be applied when a time series model is nonstationary. The specification of the initial state vector, such that it depends on a random, fixed or diffuse parameter vector δ, implies a particular method of initializing the Kalman filter as discussed in section 2.5.

[8] The SSF can be regarded as a multivariate regression model with β as the parameter vector. This is not obvious at first sight although it can be shown easily. Let y be the stack of the observation vectors y_t, $y=(y_1;\ldots;y_n)$, and define the matrix T as a lower block triangular matrix with $\frac{1}{2}(n+1)(n+2)$ non-zero blocks and identity matrices on the main block diagonal. The (t,j) block of matrix T is given by the matrix product

$$T_{t,j} = T_{t-1}T_{t-2}\cdots T_j \tag{2.4}$$

for $t=2,\ldots,n+1$ and $j=1,\ldots,t-1$. It follows from (2.1) and (2.2) that

$$y = Z\alpha + X\beta + G\epsilon \tag{2.5a}$$
$$\alpha = T(W\beta + H\epsilon) \tag{2.5b}$$

where $\alpha=(\alpha_1;\ldots;\alpha_{n+1})$, $\epsilon=(\epsilon_0;\epsilon_1;\ldots;\epsilon_n)$, $Z=[\mathrm{diag}(Z_1,\ldots,Z_n),0]$, $X=(X_1;\ldots;X_n)$, $G=[0,\mathrm{diag}(G_1,\ldots,G_n)]$, $W=(W_0;\ldots;W_n)$ and $H=\mathrm{diag}(H_0,\ldots,H_n)$. Equation (2.5b) shows that α is linear in β and, therefore, y also is linear in β. This becomes apparent by substituting (2.5b) into (2.5a) leading to

$$y = A\beta + a \tag{2.6}$$

where $A=ZTW+X$ and $a=(ZTH+G)\epsilon$ such that $\mathrm{Cov}(a)=\sigma^2(ZTH+G)(ZTH+G)'$. The matrices (ZTW+X) and (ZTH+G) are supposed to be known.

3. TIME SERIES MODELS

The state space form (2.1)-(2.3) includes various types of time series models. The following types of models will be discussed in this section, the regression model, the autoregressive moving average (ARMA) model and the structural time series model.

Regression models

Many books are available on the introduction of regression analysis. The strength of the regression model is its simplicity, its optimal properties as formulated in the Gauss-Markov theorem and its effectiveness in empirical analysis. The ordinary linear regression model is given by

$$y_t = x_t'\beta + \epsilon_t \qquad (3.1)$$

where ϵ_t is NID$(0,\sigma^2)$ and β is the vector of fixed parameters. It is seen immediately that (3.1) is a special case of the SSF where $X_t=x_t'$ and $G_t=1$. All other system matrices are equal to zero including the covariance matrix Λ. Any linear restriction can be incorporated by $\beta=b+B\delta$. The dynamic regression model is established in this framework when lagged and differenced explanatory variables are included in X_t.

A regression model with stochastic parameters is given by (3.1) as well but now $\beta=\delta$ and its covariance matrix Λ is assumed to be non-singular. A regression model with time-varying parameters is formulated in terms of the state vector

$$y_t = x_t'\alpha_t + u_t \qquad \text{for } t=1,\ldots,n \qquad (3.2a)$$

$$\alpha_t = \alpha_{t-1} + v_t \qquad \text{for } t=2,\ldots,n \qquad (3.2b)$$

where u_t is NID$(0,\sigma^2)$ and v_t is NID$(0,\sigma^2 V)$. Also the regression model with time-varying parameters can be put into the SSF by $Z_t=x_t'$, $G_t\epsilon_t=u_t$, $G_tG_t'=1$, $W_0=I$, $H_{t-1}\epsilon_{t-1}=v_t$, $H_tH_t'=V$ and $T_t=I$ for $t=1,\ldots,n$. All other system matrices are zero except matrix (b,B). Thus the first parameter vector α_1 is set equal to $\beta=b+B\delta$ where δ can be regarded as

fixed, random or diffuse.

Autoregressive moving average (ARMA) model

A time series y_t is often effectively described by a stochastic process known as the autoregressive integrated moving average model of a specific order (p,d,q), denoted by ARIMA(p,d,q), which is given by

$$z_t = \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + u_t - \theta_1 u_{t-1} - \ldots - \theta_q u_{t-q} \qquad (3.3)$$

where u_t is NID$(0,\sigma^2)$ and y_t is differenced d times, $z_t=(1-L)^d y_t$ where y_t are the observations. Note that L is the lag operator such that $Ly_t=y_{t-1}$. For an extensive discussion on the ARIMA model, the reader is referred to the standard text of Box & Jenkins (1976). Also, many other books on basic time series analysis are available that give an introduction in ARIMA modelling. The ARIMA(p,d,q) model is represented by a time-invariant state space model with

$$Z_t = (1,o') \qquad (3.4a)$$
$$T_t = (\phi_1,\ldots,\phi_r;I,o)' \qquad (3.4b)$$
$$R_t = (1;\theta_1;\ldots;\theta_{r-1}) \qquad (3.4c)$$

and with r=max(p,q+1), I is the identity matrix, o is a column vector of zeroes and #[o]=r-1. The system matrices X_t and W_t are zero. If X_t is non-zero, the SSF is regarded as a regression model with ARMA residuals, see Harvey & Phillips (1979). The initial condition for an ARMA model in the context of the SSF is extensively discussed in the literature, see for example Gardner, Harvey & Phillips (1980). The initial condition for an ARMA model with some non-stationary roots is discussed by De Jong & Chu-Chun-Lin (1991). The state space specification for univariate ARMA models is easily generalised for multivariate ARMA models.

Structural time series models

The structural time series (STS) model is an unobserved components time series model where a time series is decomposed by interpretable components such as trend, season, cycle and irregular.

The STS model can be regarded as a time-varying regression model where the explanatory variables are functions of time. The structural time series model is easily represented in a state space form.

A structural time series model can be generally expressed by

$$y_t = \mu_t + \gamma_t + u_t \tag{3.5}$$

where μ_t is the level or trend component, γ_t is the seasonal component and the irregular u_t is NID$(0,\sigma^2)$. The components trend and seasonal can be modelled in different ways. For example, a deterministic trend is given by $\mu_t=a+bt$ where a is the level constant and b is the trend slope. It is unfortunate that this simple trend component is rarely appropriate for real time series. Especially for economic time series where the components trend and seasonal do change over time. Therefore, it is unreasonable that, in the context of forecasting, all observations receive the same weight at any point in time. This can be solved by allowing the regression coefficients to evolve over time. This implies the discounting of past observations. An appropriate way of modelling a stochastic level is by a random walk process

$$\mu_t = \mu_{t-1} + \eta_t \tag{3.6}$$

where η_t is NID$(0,q_\eta\sigma^2)$. The so-called local level (LL) model (3.5) and (3.6), with $\gamma_t=0$, is easily embedded within the state space model by putting $Z_t=T_t=1$, $G_t=(1,0)$, $H_t=(0,\sqrt{q_\eta})$ and all other system matrices equal to zero. Usually, the initial trend is regarded as diffuse such that $(b,B)=(0,1)$, $W_0=1$ and $H_0=0$. Note that sometimes the LL model is referred to as the random walk plus noise (RWN) model. The ARIMA representation, or reduced form, of the LL model is of order (0,1,1) with MA parameter θ_1 equals to $\frac{1}{2}(2+q_\eta-\sqrt{[4q_\eta+q_\eta^2]})$. The forecast function of the LL model is an exponential weighted moving average (EWMA) scheme with smoothing parameter $(1-\theta_1)$.

The local linear trend (LLT) model is given by (3.5) with $\gamma_t=0$ and the trend component μ_t is modelled as

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \qquad (3.7a)$$

$$\beta_t = \beta_{t-1} + \zeta_t \qquad (3.7b)$$

where ζ_t is NID$(0,\sigma_\zeta^2)$. This model (3.5) and (3.7) is put into a SSF with $Z_t=(1,0)$, $T_t=(1,1;0,1)$, $W_0=B=I$, $G_t=(1,0,0)$, $H_t=[0,\text{Diag}(\sqrt{q_\eta},\sqrt{q_\zeta})]$, the other system elements equal to zero and δ regarded as diffuse. The reduced form of a LLT model is an ARIMA model of order (0,2,2) and the forecasting function has a Holt-Winters forecasting scheme, see Harvey (1989).

The basic structural time series model (BSM) is given by (3.5) and (3.7) where γ_t is modelled as a stochastic seasonal process. There are several methods for modelling γ_t but the most straightforward BSM is based on dummy seasonals. The seasonal dummy regression model consists of a parameter vector with the elements γ_j for j=1,...,s where s is the seasonal length. The implementation of dummy seasonals in a regression model requires the restriction $\Sigma_{j=1}^{s} \gamma_j = 0$ or, equivalently, $\gamma_s = -\Sigma_{j=1}^{s-1} \gamma_j$. The BSM allows the seasonal dummies to evolve over time such that the seasonal effect at time t is $\gamma_t=\gamma_j$ where period j is prevailing at time t and

$$\Sigma_{j=0}^{s-1} \gamma_{t-j} = \omega_t \qquad (3.8)$$

where ω_t is NID$(0,q_\omega\sigma^2)$. Thus the seasonal dummies do not exactly sum up to zero but the expectation of the sum is zero such that the dummies have the flexibility to change over time. This basic structural time series model can also be put into the SSF. For example, consider a quarterly BSM with dummy seasonals (s=4), the system matrices of the time-invariant SSF are given by

$$Z_t = (1,0,1,0,0) \qquad G_t = (1,0,0,0)$$

$$T_t = \text{Diag}(T^{[T]},T^{[S]}) \qquad H_t = (0,H;0)$$

$$T^{[T]} = (1,1;0,1) \qquad H = \text{Diag}(\sqrt{q_\eta},\sqrt{q_\zeta},\sqrt{q_\omega})$$

$$T^{[S]} = (-1,-1,-1;0,1,0;0,0,1) \qquad W_0 = B = I$$

and all other sytem elements are equal to zero, see Harvey (1989) for more details. A monthly BSM with dummy seasonals has an ARMA representation (reduced form) which is very close to the 'airline'

model of Box & Jenkins (1976).

Other approaches also exist to deal with stochastic seasonal components in unobserved components time series models. For example, Harvey (1989) discusses an alternative BSM with time-varying trigonometric terms and Harrison & Stevens (1976) propose a vector of random walks where each element characterizes a particular season. In chapter 5, both approaches are discussed in the context of modelling seasonality by periodic cubic splines.

For a thorough discussion on stochastic properties, applications and many other issues concerning structural time series models, the reader is referred to the elaborate work of Harvey (1989).

4. THE KALMAN FILTER

The Kalman filter is primarily a recursive set of analytical expressions for the minimum mean square linear estimator and the mean square error matrix of the observation vector and state vector using Y_{t-1} and parameter vector δ. Because the likelihood can be expressed in terms of one-step ahead prediction errors, the Kalman filter is an important tool for evaluating the log-likelihood function. This section will give the details of this efficient method of calculating the likelihood. In the following it is assumed that all system matrices are known and that the vector δ is random.

Prediction error decomposition

The log-likelihood function of y conditional on the random parameter vector δ is defined by $l[y|\delta]=-\frac{1}{2}(\#[y]\ \log 2\pi + L[y|\delta])$ where the likelihood criterion $L[y|\delta]$ is given by

$$L[y|\delta] = \log|\Sigma_y| + (y-\mu_y)'\Sigma_y^{-1}(y-\mu_y) \qquad (4.1)$$

where $\mu_y=E[y|\delta]=(ZTW+X)\beta$ and $\Sigma_y=Mse[y|\delta]=Cov(y-\mu_y)=\sigma^2(ZTH+G)(ZTH+G)'$, see equation (2.6). Expression (4.1) is computationally not attractive. An efficient approach to evaluate the likelihood is derived in the two results [2.1] and [2.2]. The first result derives an alternative expression for (4.1) using the prediction error

decomposition. This likelihood formulation can be easily computed via the Kalman filter. The second result contains the derivation of the Kalman filter.

Result [2.1] Denote the minimum mean square linear estimator of α_t, using Y_{t-1} and δ, and its mean square error matrix as, respectively,

$$a_{t|t-1} = E[\alpha_t|Y_{t-1},\delta] \qquad \sigma^2 P_{t|t-1} = Mse[\alpha_t|Y_{t-1},\delta] \tag{4.2}$$

then the likelihood criterion $L[y|\delta]$ of (4.1) can be rewritten as

$$\#[y] \log \sigma^2 + \Sigma_{t=1}^{n}[\log |F_t| + \sigma^{-2} v_t' F_t^{-1} v_t] \tag{4.3}$$

where the prediction error or innovation v_t and its covariance matrix $\sigma^2 F_t$ are defined as

$$v_t = y_t - E[y_t|Y_{t-1},\delta] = y_t - Z_t a_{t|t-1} - X_t\beta \tag{4.4a}$$
$$\sigma^2 F_t = Mse[y_t|Y_{t-1},\delta] = Cov(v_t) = \sigma^2 (Z_t P_{t|t-1} Z_t' + G_t G_t') \tag{4.4b}$$

Equation (4.3) follows from the prediction error decomposition, see Harvey (1981), that allows rewriting the likelihood criterion as a sum of likelihood functions which are conditional on the set of past observations Y_{t-1} and the random vector δ,

$$L[y|\delta] = L[y_1|\delta] + \Sigma_{t=2}^{n} L[y_t|Y_{t-1},\delta] \tag{4.5}$$

The equation for v_t follows from $E[y_t|Y_{t-1},\delta] = Z_t a_{t|t-1} + X_t\beta$ as implied by the measurement equation of the SSF. The equation for $\sigma^2 F_t$ is obtained by introducing the state prediction error vector

$$x_t = \alpha_t - E[\alpha_t|Y_{t-1},\delta] = \alpha_t - a_{t|t-1} \tag{4.6}$$

with $Cov(x_t) = \sigma^2 P_{t|t-1}$. By substituting the measurement equation of (2.1a) into (4.4a), leads to $v_t = Z_t x_t + G_t \epsilon_t$ from which it follows directly that its covariance matrix $\sigma^2 F_t$ is given as asserted. [ ]

Kalman filter

Equation (4.3) provides a viable method for calculating the likelihood criterion. The innovations and their variances are provided by the Kalman filter which is mainly a recursive set of equations to evaluate the estimators $a_{t|t-1}$ and $P_{t|t-1}$. The following result derives the Kalman filter.

Results [2.2] The Kalman filter is given by

$$v_t = y_t - Z_t a_{t|t-1} - X_t\beta \qquad F_t = Z_t P_{t|t-1} Z_t' + G_t G_t' \qquad (4.7a)$$

$$K_t = (T_t P_{t|t-1} Z_t' + H_t G_t') F_t^{-1} \qquad (4.7b)$$

$$a_{t+1|t} = T_t a_{t|t-1} + W_t\beta + K_t v_t \qquad (4.7c)$$

$$P_{t+1|t} = T_t P_{t|t-1} L_t' + H_t M_t' \qquad (4.7d)$$

where $L_t=T_t-K_tZ_t$ and $M_t=H_t-K_tG_t$. The equations (4.7a,b) can be regarded as definitions. The proof of the two latter Kalman recursions is given below.

Because the set of innovations $I_t=\{v_1,\ldots,v_t\}$ is a linear combination of the set of observations $Y_t=\{y_1,\ldots,y_t\}$ and the parameter vector δ, the minimum mean square linear estimator of the state vector can be redefined as

$$a_{t+1|t} = E[\alpha_{t+1}|Y_t,\delta] = E[\alpha_{t+1}|I_t]. \qquad (4.8)$$

The set of innovations I_t are independent. Therefore, as follows from appendix 2A (result 2), the state estimator can be decomposed into

$$a_{t+1|t} = E[\alpha_{t+1}|I_{t-1}] + E[\alpha_{t+1}|v_t] - E[\alpha_{t+1}] \qquad (4.9)$$

where

$$E[\alpha_{t+1}|I_{t-1}] = E[T_t\alpha_t+W_t\beta+H_t\epsilon_t|I_{t-1}] = T_t a_{t|t-1} + W_t\beta \qquad (4.10)$$

An explicit expression for the term $E[\alpha_{t+1}|v_t]$ is obtained using the minimum mean square linear estimation results of appendix 2A. It follows that

$$E[\alpha_{t+1}|v_t] = E[\alpha_{t+1}] + Mse[\alpha_{t+1},v_t]Mse[v_t]^{-1}(v_t-E[v_t])$$
$$= E[\alpha_{t+1}] + Mse[\alpha_{t+1},v_t]\sigma^{-2}F_t^{-1}v_t \qquad (4.11)$$

where

$$Mse[\alpha_{t+1},v_t] = Cov(T_t\{\alpha_t-E[\alpha_t]\} + H_t\epsilon_t \ , \ Z_tx_t + G_t\epsilon_t)$$
$$= \sigma^2 (T_tP_{t|t-1}Z_t' + H_tG_t') \qquad (4.12)$$

By substituting the equations (4.10) and (4.11) into (4.9) and by defining the Kalman gain matrix as $K_t = (T_tP_{t|t-1}Z_t' + H_tG_t')F_t^{-1}$, the recursion for $a_{t+1|t}$ follows immediately. The estimator of the initial state vector is $a_{1|0} = E[\alpha_1|\delta] = W_0(b+B\delta)$.

The recursion for $P_{t+1|t}$ is derived by starting to note that

$$\sigma^2 P_{t+1|t} = Mse[\alpha_{t+1}|Y_t,\delta] = Cov(x_{t+1})$$
$$= Cov(\alpha_{t+1}-E[\alpha_{t+1}],x_{t+1}) \qquad (4.13)$$

where the prediction error $x_{t+1} = \alpha_{t+1} - a_{t+1|t}$ can be evaluated recursively by

$$x_{t+1} = L_tx_t + M_t\epsilon_t \qquad (4.14)$$

This follows from the definitions of α_{t+1} and $a_{t+1|t}$ leading to $x_{t+1} = \alpha_{t+1} - a_{t+1|t} = T_tx_t + H_t\epsilon_t - K_tv_t$ where $v_t = Z_tx_t + G_t\epsilon_t$. From this result, the Kalman equation for $P_{t+1|t}$ follows immediately because

$$\sigma^2 P_{t+1|t} = Cov(T_t\{\alpha_t-E[\alpha_t]\} + H_t\epsilon_t \ , \ L_tx_t + M_t\epsilon_t) \qquad (4.15)$$

which give the recursion for $P_{t+1|t}$ as asserted. The initial mean square error matrix $Mse[\alpha_1|\delta]$ is given by $\sigma^2 P_{1|0} = \sigma^2 H_0H_0'$ [ ]

The update equations for $a_{t+1|t}$ and $P_{t+1|t}$ will be referred to as the Kalman recursions. The derivation of result [2.2] follows mainly the proof of Anderson & Moore (1979) using linear estimation theory. Simpler alternative derivations are at present in the literature. Duncan & Horn (1972) derive the KF by using generalized least squares techniques, see also Harvey (1981). Another simple proof can

be given when Gaussianity is assumed. The proof is based on applying the rules of conditional expectations, see Harvey (1989) and appendix 2B.

The second part of the likelihood criterion (4.3) can be rewritten by $f_n + \sigma^{-2}q_n$ where both scalars are evaluated recursively by

$$f_t = f_{t-1} + \log |F_t| \qquad q_t = q_{t-1} + v_t' F_t^{-1} v_t \tag{4.16}$$

and $f_0=0$ and $q_0=0$. Both recursions are assumed to be included in the Kalman filter.

Steady state

A time invariant SSF is given by

$$y_t = Z\alpha_t + X_t\beta + G\epsilon_t \qquad \alpha_{t+1} = T\alpha_t + W_t\beta + H\epsilon_t \tag{4.17}$$

together with (2.2) and (2.3). The Kalman filter applied to the time invariant SSF is still time varying but can converge into a time invariant filter, that is when

$$P_{t+1|t} = P \tag{4.18}$$

In this occasion, the KF is said to be in a steady-state such that the KF quantities F_t, K_t and $P_{t+1|t}$ are constant over time. In fact, the steady state mean square error matrix P is the solution of the Riccati equation which is given by

$$P - TPT' + (TPZ'+HG')(ZPZ'+GG')^{-1}(TPZ'+HG')' - HH' = 0 \tag{4.19}$$

This Ricatti equation follows from the recursion of $P_{t+1|t}$ which can be rewritten as

$$P_{t+1|t} = T_t P_{t|t-1} T_t' - K_t F_t K_t' + H_t H_t' \tag{4.20}$$

Equation (4.20) is obtained by considering (4.7d) and the definitions of K_t, L_t and M_t. It is not straightforward to get a

solution for (4.19) and, if a solution is available, it is not clear whether P is unique or whether P is a positive semi-definite matrix, see the discussion in Anderson & Moore (1979) and Harvey (1989, section 3.3.3).

The innovation and the estimator of the state are linear in δ

The equations for F_t, K_t and $P_{t|t-1}$ are data-independent. These quantities only depend on the system matrices whereas the vectors v_t and $a_{t+1|t}$ depend on the set of observations Y_t, the parameter vector δ and the system matrices. In view of the linear SSF and the linear KF, the vectors v_t and $a_{t+1|t}$ appear to be linear functions of the set of observations Y_t and of the parameter vector $\beta=b+B\delta$. Because no actual values are given for β, the vectors v_t and $a_{t|t-1}$ cannot be evaluated. This is in contrast with the Kalman quantities F_t, K_t and $P_{t|t-1}$ which are not dependent on β. The following result will show that v_t and $a_{t|t-1}$ are linear functions of parameter vector δ.

Result [2.3] The innovation v_t and state $a_{t|t-1}$ can be expressed explicitly as linear functions of δ

$$v_t = v_t^o + v_t^+\delta = V_t(1;\delta) \qquad (4.21a)$$

$$a_{t|t-1} = a_t^o + a_t^+\delta = A_t(1;\delta) \qquad (4.21b)$$

where

$$v_t^o = y_t - Z_t a_t^o - X_t b \qquad a_{t+1}^o = T_t a_t^o + W_t b + K_t v_t^o$$

$$v_t^+ = - Z_t a_t^+ - X_t B \qquad a_{t+1}^+ = T_t a_t^+ + W_t B + K_t v_t^+$$

$$V_t = (v_t^o, v_t^+) \qquad A_t = (a_t^o, a_t^+)$$

and $A_1 = W_0(b,B)$. Equation (4.21) follows directly from the Kalman equations v_t and $a_{t|t-1}$ by replacing β with its definition $b+B\delta$ and rearranging the KF equations. [ ]

The Kalman filter in stacks and the Cholesky decomposition

It was recognised in the literature, see Harvey (1981), that the Kalman filter performs a Cholesky decomposition on the set of observations such that they are transformed into an orthogonal data-set. The following result gives the details.

Result [2.4] The stack of the innovations $v=(v_1;\ldots;v_n)$ is a Cholesky transformation on $(y-\mu_y)$, where $\mu_y = E[y|\delta] = (ZTW+X)\beta$, as given by

$$v = C[y - \mu_y] \qquad Cov(v) = \sigma^2 F = \sigma^2 C\Sigma_y C' \tag{4.22}$$

where C is the Cholesky (lower block triangular) matrix.

The update equation for $a_{t+1|t}$ can be rewritten as

$$a_{t+1|t} = L_t a_{t|t-1} + N_t\beta + K_t y_t \tag{4.23}$$

where $L_t=T_t-K_tZ_t$ and $N_t=W_t-K_tX_t$. Equation (4.23) is obtained by substituting the equation for v_t into $a_{t+1|t}$. In order to derive expressions for the stack of the vectors v_t and $a_{t|t-1}$, define $a=(a_{1|0};\ldots;a_{n+1|n})$, $K=(0;\mathrm{diag}(K_1,\ldots,K_n)$, $N=(W_0;N_1;\ldots;N_n)$ and note that $N=W-KX$ where X and W are defined in section 2 (8-th remark). Furthermore, define matrix L as a lower block triangular matrix with $\frac{1}{2}(n+1)(n+2)$ non-zero blocks and with identity matrices on the main block diagonal. The (t,s) block of matrix L is given by the matrix

$$L_{t,s} = L_{t-1}L_{t-2}\ldots L_s \tag{4.24}$$

for $t=2,\ldots,n+1$ and $s=1,\ldots,t-1$. Compare matrix L with matrix T, the latter has the same structure as L but its blocks are replaced by the blocks of (2.4). It follows that

$$v = y - Za - X\beta \qquad a = L(N\beta + Ky) \tag{4.25}$$

By substituting a into v and by recalling $y = (ZTW+X)\beta + (ZTH+G)\epsilon$ from (2.6), we have

$$v = Cy - (ZLN+X)\beta = C(y-X\beta) - ZLW\beta = (CZT-ZL)W\beta + J\epsilon \tag{4.26}$$

where $C=I-ZLK$ and $J=C(ZTH+G)$. Because $E[v]=0$, it must hold that $CZT=ZL$ and $v=J\epsilon=C[y-(ZTW+X)\beta]$. The Cholesky result follows now immediately. Note that covariance matrix σ^2F equals σ^2JJ' where $J=C(ZTH+G)=ZLH+CG=ZLM+G$ and $M=H-KG=(H_0;M_1;\ldots;M_n)$. [ ]

The Cholesky decomposition shows, in an alternative way, that the likelihood criterion of (4.1) can be rewritten as (4.5),

$$\begin{aligned} L[y|\delta] &= \log|\Sigma_y| + (y-\mu_y)'\Sigma_y^{-1}(y-\mu_y) \\ &= \log|C||\Sigma_y||C'| + (y-\mu_y)'C'(C\Sigma_y C')^{-1}C(y-\mu_y) \\ &= \#[y]\log\sigma^2 + \log|F| + \sigma^{-2}v'F^{-1}v \ . \end{aligned} \tag{4.27}$$

Note that $|C|=|C'|=1$ and $|A||B|=|AB|$ when A and B are non-singular square matrices. Finally, it follows from the stack notation as well that the innovation is a linear function of parameter vector δ,

$$v = C[y - (ZTW+X)b] - C(ZTW+X)B\delta = v^o + v^+\delta = V(1;\delta) \tag{4.28}$$

where $v^o=(v_1^o;\ldots;v_n^o)=C[y-(ZTW+X)b]$, $v^+=(v_1^+;\ldots;v_n^+)=-C(ZTW+X)B$ and $V=(v^o,v^+)$.

5. LIKELIHOOD EVALUATION AND ESTIMATION

The likelihood criterion $L[y|\delta]$ is given in (4.3) and it is evaluated by the KF. The derivations are given for the concentrated likelihood criterions $L_\delta[y]$ and $L_{\delta;\sigma}[y]$ for the following three cases regarding $\delta \sim (\mu\ ,\ \sigma^2\Lambda)$. (i) It is straightforward to derive the likelihood for the case that the parameter vector is fixed such that $\Lambda=0$. (ii) When Λ is a non-singular covariance matrix, parameter vector δ is said to be random. (iii) A special case of the random assumption is a diffuse δ, that is, the inverse of Λ converges to zero in the Euclidean norm. The likelihood criterions for these cases are derived below as well. The approach of De Jong (1991b) and Marshall (1992) is adopted in this section but the derivations are more transparent.

The likelihood criterion under fixed conditions

When the parameter vector δ is assumed to be fixed such that $\Lambda=0$, the likelihood criterion is similar to (4.3) or (4.27). By replacing v with $V(1;\delta)$, the likelihood criterion is

$$L[y|\delta] = \#[y] \log \sigma^2 + \log |F| + \sigma^{-2}(1;\delta)'Q(1;\delta) \qquad (5.1)$$

where $Q = (q,-s';-s,S) = V'F^{-1}V$. It follows that the maximum likelihood estimator for δ is equal to the generalised least squares estimator

$$\hat{\delta} = S^{-1}s \qquad Mse[\hat{\delta}] = \sigma^2 S^{-1} \qquad (5.2)$$

and the concentrated likelihood criterion, with respect to δ, $L_\delta[y]$ is given by

$$L_\delta[y] = \#[y] \log \sigma^2 + \log |F| + \sigma^{-2}(q - s'S^{-1}s) \qquad (5.3)$$

The maximum likelihood estimator for σ^2 follows from (5.3) and it is given by

$$\hat{\sigma}^2 = (q - s'S^{-1}s)/\#[y] \qquad (5.4)$$

such that the concentrated likelihood criterion, with respect to $(\delta;\sigma^2)$, $L_{\delta;\sigma}[y]$ is given by

$$L_{\delta;\sigma}[y] = \#[y](1 + \log \hat{\sigma}^2) + \log |F| \qquad (5.5)$$

To evaluate matrix $Q = \Sigma_{t=1}^{n} V_t' F_t^{-1} V_t$, a partially modified Kalman filter is applied as given in the next result.

Result [2.5] The modified Kalman filter is primarily extended to calculate the dummy matrices of (4.21). Thus, the modification only concern the vectors v_t and $a_{t|t-1}$. Their equations are replaced by the matrix equation and matrix recursion, respectively,

$$V_t = (y_t,0) - Z_t A_t - X_t(b,B) \qquad (5.6a)$$
$$A_{t+1} = T_t A_t + W_t(b,B) + K_t V_t \qquad (5.6b)$$

where $A_1 = W_0(b,B)$. Also, the matrix recursion

$$Q_t = Q_{t-1} + V_t' F_t^{-1} V_t \qquad (5.7)$$

is added where $Q_0=0$ and $Q_n=Q$. In fact, the latter recursion replaces the recursion for q_t but the recursion for f_t remains, see (4.16) [ ]

This approach of estimating the parameter vector $(\delta;\sigma^2)$ and evaluating the likelihood criterion, for fixed initial conditions, is a generalisation of the results of Rosenberg (1973). It is based on the fact that the (fixed) initial conditions are linear in the observations and, therefore, a generalised least squares procedure can be applied to obtain maximum likelihood estimates. The required computations are relatively straightforward because of the prediction error decomposition. The modified Kalman filter (5.6) is proposed by De Jong (1991b) in what he calls the diffuse Kalman filter (DKF). This name is slightly confusing because it has, strictly speaking, nothing to do with a diffuse δ.

The likelihood criterion under random conditions

The likelihood criterion L[y], when vector δ is assumed to be random, can be derived using Bayes' rule, $L[y] = L[y|\delta] + L[\delta] - L[\delta|y]$. The first likelihood criterion is given by

$$\begin{aligned} L[y|\delta] &= \#[y] \log \sigma^2 + \log |F| + \sigma^{-2}(1;\delta)'Q(1;\delta) \\ &= \#[y] \log \sigma^2 + \log |F| + \sigma^{-2}(q - 2\delta's + \delta'S\delta) \end{aligned} \qquad (5.8)$$

where $Q=(q,-s';-s,S)$ is evaluated by the modified KF. The definition of $L[\delta]$ is straightforward and is given by

$$L[\delta] = \#[\delta] \log \sigma^2 + \log |\Lambda| + \sigma^{-2}(\mu'\Lambda^{-1}\mu - 2\delta'\Lambda^{-1}\mu + \delta'\Lambda^{-1}\delta) \qquad (5.9)$$

Although it is based on classical results, see any textbook on multivariate statistical theory, it is not straightforward to derive $L[\delta|y]$. The results of appendix 2A need to be applied in the same way as they are applied to the derivation of the Kalman filter in section 4. It is given that $L[\delta|y]=L[\delta|v^o]$ because, as it is observed earlier, the vector v^o is a linear combination of y. Furthermore,

$$\begin{aligned} &E[\delta] = \mu \qquad Cov[\delta] = \sigma^2\Lambda \qquad Cov(\delta,v^o) = -\sigma^2\Lambda v^{+\prime} \\ &E[v^o] = -v^+\mu \qquad Cov[v^o] = \sigma^2(v^+\Lambda v^{+\prime} + F) \end{aligned} \qquad (5.10)$$

and, by applying the result of minimum mean square linear estimation, it follows that

$$E[\delta|v^o] = \mu - \Lambda v^{+\prime}(v^+\Lambda v^{+\prime} + F)^{-1}(v^o + v^+\mu) \tag{5.11a}$$
$$Mse[\delta|v^o] = \sigma^2(\Lambda - \Lambda v^{+\prime}(v^+\Lambda v^{+\prime} + F)^{-1}v^+\Lambda) \tag{5.11b}$$

Applying the standard inversion lemma on $(v^+\Lambda v^{+\prime}+F)$ leads to

$$(v^+\Lambda v^{+\prime} + F)^{-1} = F^{-1} - F^{-1}v^+(\Lambda^{-1} + S)^{-1}v^{+\prime}F^{-1} \tag{5.12}$$

such that (5.11) can be rewritten as

$$\begin{aligned} E[\delta|v^o] &= \mu + \Lambda(I - S(\Lambda^{-1} + S)^{-1})(s - S\mu) \\ &= I\mu + (\Lambda^{-1} + S)^{-1}(s - S\mu) \\ &= (\Lambda^{-1} + S)^{-1}(s + \Lambda^{-1}\mu) \end{aligned} \tag{5.13a}$$

$$\begin{aligned} Mse[\delta|v^o] &= \sigma^2(\Lambda - \Lambda(I - S(\Lambda^{-1} + S)^{-1})S\Lambda) \\ &= \sigma^2(I - (\Lambda^{-1} + S)^{-1}S)\Lambda \\ &= \sigma^2(\Lambda^{-1} + S)^{-1} \end{aligned} \tag{5.13b}$$

The derivation of (5.13) becomes apparent by replacing all identity matrices with the matrix term $(\Lambda^{-1}+S)(\Lambda^{-1}+S)^{-1}$ or its reverse. The likelihood criterion $L[\delta|v^o]=L[\delta|y]$ follows immediately from (5.13) and is given by

$$L[\delta|y] = \#[\delta]\ \log\sigma^2 - \log|\Lambda^{-1} + S| + \sigma^{-2}(\delta-\bar{\mu})'\bar{\Sigma}^{-1}(\delta-\bar{\mu}) \tag{5.14}$$

$\bar{\mu}=E[\delta|y]$ and $\bar{\Sigma}=Mse[\delta|y]$, see (5.13).

The likelihood criterion for y where δ is assumed to be random follows from the equations (5.8), (5.9) and (5.14) and is given by

$$\begin{aligned} L[y] &= \#[y]\ \log\sigma^2 + \log|F| + \log|\Lambda| + \log|\Lambda^{-1}+S| \\ &+ \sigma^{-2}(q + \mu'\Lambda^{-1}\mu - (s + \Lambda^{-1}\mu)'(\Lambda^{-1} + S)^{-1}(s + \Lambda^{-1}\mu)) \end{aligned} \tag{5.15}$$

The likelihood criterion (5.15) can be evaluated using the modified Kalman filter (5.6) provided that actual values are given for μ and

Λ. However, in the case that μ and Λ are known, it is much simpler to get the likelihood from the normal KF which is initialized with $a_{1|0} = W_0(b+B\mu)$ and $P_{1|0} = W_0B\Lambda B'W_0' + H_0H_0'$ and where β is replaced by $b+B\mu$.

The concentrated likelihood criterion, with respect to μ, is given by

$$L_\mu[y] = \#[y]\log \sigma^2 + \log|F| + \log|I + \Lambda S| + \sigma^{-2}(q - s'S^{-1}s) \qquad (5.16)$$

Note that the maximum likelihood estimator for μ is given by $S^{-1}s$. It follows from (5.16) directly that the estimator for Λ is a zero matrix such that the concentrated likelihood criterion $L_{\mu,\Lambda}[y]$ under random conditions is the same as the concentrated likelihood criterion $L_\delta[y]$ under fixed conditions. Finally, the estimator for σ^2 remains equal to $(q - s'S^{-1}s)/\#[y]$.

The likelihood criterion under diffuse conditions

The random vector $\delta \sim (\mu, \sigma^2\Lambda)$, is said to be diffuse if the inverse of Λ converges to zero in the Euclidian norm which implies that $a'\Lambda^{-1}b \to 0$ where a and b can be any non-zero vector. The initial condition must be treated as diffuse for non-stationary time series models but also in cases where time series models have been applied since time immemorial, see De Jong & Chu-Chun-Lin (1991) for a thorough discussion. In general, the diffuse condition reflects the uncertainty regarding the initializiation of the time series model. It can be regarded as a standard assumption for nonstationary time series models.

The diffuse likelihood criterion is defined as the limit of (5.15) when $\Lambda \to \infty$ and where the term $\log|\Lambda|$ is dropped, i.e.,

$$L[y] = \#[y]\log \sigma^2 + \log|F| + \log|S| + (q - s'S^{-1}s)/\sigma^2 \qquad (5.17)$$

This pseudo-likelihood criterion differs from the one for the fixed case with only the term $\log|S|$. At first sight, it seems arbitrarily to drop the term $\log|\Lambda|$ to get the diffuse likelihood. Nevertheless, (5.17) is a likelihood criterion for My where (i) the transformation

matrix M has rank #[y]-#[δ], (ii) Cov(My,δ)=0 and (iii) log|Cov(My)| = (#[y]log σ^2+log|F|+log|S|), see De Jong (1991b). In fact, the likelihood criterion (5.17) is very similar to the marginal likelihood, see for example McCullagh & Nelder (1989). The marginal likelihood is based on a linear transformation as well, say Ny, but only the restrictions (i) and (ii) hold such that the data are invariant to δ. Since log|Cov(Ny)|=log|$\sigma^2 N\Sigma_y N'$|=#[y]log σ^2+log|F| +log|NN'|, it is obvious that the difference between the marginal and the diffuse likelihood is equal to the term log|S|-log|NN'|.

The estimator for δ under diffuse conditions is the limit of (5.13) when $\Lambda\rightarrow\infty$, i.e.,

$$\hat{\delta} = E[\delta|y] = S^{-1}s \qquad Mse[\delta|y] = \sigma^2 S^{-1} \tag{5.18}$$

These inference results do not differ from the earlier inference results for the fixed case. The estimator for σ^2 is unchanged because its estimator for the random case is invariant to Λ.

It can be concluded that the inference results concerning δ do not differ when δ is assumed to be fixed or diffuse, see table 2.1. However, the associated likelihood criterions do differ.

Initialization of the Kalman filter

Consider the Kalman filter as derived in section 4. The parameter vector δ is assumed to be unknown, thus the KF cannot be initialized properly. This initialization problem is widely discussed in the time series literature. The 'main stream' solution is to start the KF with a very large covariance matrix given by $P_{1|0}=\kappa I$ where κ is a large constant, see Harvey & Phillips (1979). This 'big-κ' method reflects the uncertainty with regards to the initial state. It is a numerical solution that may cause computational inaccuracies and, more importantly, it is conceptuallly not appealing. Another proposed solution is to apply the information filter that provides recursive equations for the inverse of $P_{t|t-1}$, see Anderson & Moore (1979). Note that the 'big-κ' method is implied by $P_{1|0}^{-1}=0$. Ansley & Kohn (1985) point out that the information filter is numerically inefficient and cannot be used in

specific cases like ARMA models. It is also possible to use an initial stretch of the data-set to estimate the initial state vector by regression techniques, see Harvey & Pierse (1984). The drawback of their approach is that the initial data set cannot include missing observations and that it is tailored to specific cases. A similar approach is put forward by Bell & Hillmer (1991).

This section has shown how the likelihood function can be evaluated exactly by applying the modified Kalman filter which does not have an initialization problem ($A_1=W_0(b,B)$ is known). Ansley & Kohn (1985) propose another modification of the Kalman filter which also deals with diffuse initial conditions but they use other concepts. Although judgements of this alternative method are rather subjective, their approach is less appealing. De Jong (1991b) clearly points out why. The approach of Ansley & Kohn (1985), as also presented in many of their subsequent papers, is analytically not attractive because their derivations are cumbersome to follow. The approach of De Jong (1991b), adopted in this chapter, is a trivial modification of the KF and the derivation is transparent. The main result is based on Rosenberg (1973). Finally, the approach of Ansley & Kohn (1985) is less general.

The extra computational costs of the modified KF compared to the usual KF is that (i) the vector recursions for v_t and $a_{t+1|t}$ are extended with a number of $\#[\delta]$ equivalent vector recursions and (ii) the scalar recursion for q_t is replaced by the matrix recursion for Q_t, see (5.6). The dimension of this matrix recursion is $(1+\#[\delta])$ x $(1+\#[\delta])$. Note that $\#[\delta]$ is in many cases equal to $\#[\alpha_t]$. For example, the basic structural time series model with a quarterly dummy seasonal component is put in a SSF with dimensions $\#[\alpha_t]=\#[\delta]=5$ and $\#[\epsilon_t]=4$.

In general, it is not required to apply the modified KF for the full data-set. After an initial stretch of the data, the matrix S_t becomes non-singular such that an estimator for δ exists. Note that the length of the initial data set, say d, can never be smaller than $\#[\delta]$. The initial estimator of δ is used to collapse from the modified KF to the usual KF. The collapsed KF is applied to $t=d+1,\ldots,n$ and it starts off with the state $A_{d+1}(1;S_d^{-1}s_d)$ and with its mean square error matrix $P_{d+1|d} + a^+_{d+1}S_d^{-1}a^{+\prime}_{d+1}$. Furthermore, the matrix

recursion for Q_t (5.7) reduces to the scalar recursion for q_t initialised with $q_d - s_d' S_d^{-1} s_d$. The full details are given in De Jong (1991a) and Chu-Chun-Lin (1991). This is the preferred method to initialize the usual KF. However, this method implies that a full-sample estimator for δ cannot be evaluated unless δ is fully included in the state vector of the collapsed KF. In many cases a full-sample estimator of the initial condition is not of particular interest but a full-sample estimator of the 'regression' effects is of interest.

To illustrate matters, redefine the specification of the initial condition and the parameter vector associated with the SSF as

$$\alpha_1 = b_0 + B_0\delta + H_0\varepsilon_0 \qquad \beta = b_1 + B_1\delta$$

such that B_0 'selects' the elements of δ for the initial conditions and B_1 corresponds to the 'regression' effects of a time series model. Assume that the part of δ that is concerned with the initial state vector, i.e. corresponding to B_0, will be collapsed. With regards to the part of δ that is associated with the 'regression' effects, two strategies can be adopted. First, (CKF - I) the appropriate part of δ can be included in the state vector of the usual KF after an initial estimator is available. Second, (CKF - II) the modified KF remains for the part of δ associated with B_1. The second option is computationally superior to the usual KF with δ (partially) in the state vector, this is illustrated for a local linear trend model with two explanatory variables. In table 2.3, the set of boxes show the dimensions of the Kalman filter quantities (each block represents one element). For example, in the case of a modified Kalman filter applied to the LLT model with two explanatory variables, the innovation vector v_t^+ is of dimension 4x1. Table 2.2 gives the number of additions and multiplications for each step of the Kalman filter. It is clear that the strategy CKF - II (a partially collapsed KF) is more efficient than the CKF - I strategy. For a complete discussion on collapsing the Kalman filter, see Chu-Chun-Lin (1991).

6. PRACTICAL TIME SERIES MODELLING

This section discusses a selection of topics concerning practical time series modelling using the state space formulation. The first step is to place the time series model into the SSF. Generally, some elements of the system matrices are unknown. They are commonly referred to as hyperparameters. There are several methods of estimating these unknowns as discussed below.

Diagnostic tools for checking the adequacy of an estimated time series model to describe the data are widely available and they can be classified into graphical procedures and into test-statistics. These diagnostics are applied to the estimated standardised innovations which are derived, together with the estimated state vectors, in this section. An overview of several diagnostic tools is given in this section as well. Finally, an illustrative example is given.

Hyperparameter estimation

The system matrices of the SSF (2.1)-(2.3) are assumed to be known. As it is shown in section 3, this does not hold for most time series models because specific elements of the system matrices are related to unknown parameters of the time series model. For example, the SSF for an ARMA model places the AR parameters in the transition matrix T_t and it places the MA parameters in the matrix H_t. The ARMA parameters are not known apriori. For time-varying and structural time series models, (diagonal) elements of the matrix H_tH_t' are not known.

Any unknown element in a system matrix will be referred to as a hyperparameter. The estimation of hyperparameters takes place by minimizing a certain criterion, for example the 'least squares' or the 'maximum likelihood' criterion, which can be constructed from the innovations, see section 2.5. The decision of which criterion has to be minimized is an important one. In the case of structural time series models, it is appropriate to use the diffuse likelihood criterion as derived in section 2.5.

The appropriate criterion is computed by the Kalman filter and its optimal value is found via some non-linear optimization technique, for example, the method of scoring. A more simple 'grid-search' procedure is mostly not a favourable strategy to find the optimal value because the number of function calls, i.e. the Kalman filter, can be huge. Scoring methods are better because they provide a more precise search direction by taking account of the exact or approximated first derivative of the criterion with respect to the hyperparameters. In appendix 2C the analytic first derivative functions are given for the likelihood criterion where it is assumed that the hyperparameters are only in the time-invariant matrix H_t. This method of estimating the hyperparameters is referred to as estimation in the time domain (TD).

Another method of hyperparameter estimation is based on approximating the estimation criterion via the frequency domain (FD) where the innovations are obtained from the Fourier transform, see Harvey (1981). A very different approach of estimating the hyper-parameters is the EM (expectation-maximization) algorithm which is developed in the context of time series models by Watson & Engle (1983) and Shumway & Stoffer (1982). It is reported by Harvey & Peters (1990) that the EM algorithm performs poorly in estimating the hyperparameters of a structural time series model. A modification of the EM algorithm is given in the next chapter.

Prediction

In the state space approach of time series modelling, the following estimators are of particular interest,

$$\hat{y}_t = E[y_t|Y_{t-1}] \quad \hat{v}_t = y_t - \hat{y}_t \quad \hat{\alpha}_t = E[\alpha_t|Y_{t-1}] \tag{6.1}$$

The estimators of δ and their mean square error matrices, under fixed, random and diffuse conditions, are developed in section 2.5 and are given in table 2.1.

The estimators for y_t and α_t are given by

$$\begin{aligned}\hat{y}_t &= E[y_t|Y_{t-1}] = E\{\ E[y_t|Y_{t-1},\delta]\ |\ Y_{t-1}\ \} \\ &= y_t - E[V_t(1;\delta)\ |\ Y_{t-1}] = y_t - V_t(1;\hat{\delta}_{t-1})\end{aligned} \tag{6.2a}$$

$$\hat{\alpha}_t = E[\alpha_t|Y_{t-1}] = E\{ E[\alpha_t|Y_{t-1},\delta] \mid Y_{t-1} \} =$$
$$= E[A_t(1;\delta) \mid Y_{t-1}] = A_t(1;\hat{\delta}_{t-1}) \qquad (6.2b)$$

with their mean square error matrices

$$Mse[y_t|Y_{t-1}] = Cov(y_t - E[y_t|Y_{t-1},\delta] + E[y_t|Y_{t-1},\delta] - E[y_t|Y_{t-1}])$$
$$= Cov(v_t) + Cov[V_t(1;\delta) - V_t(1;\hat{\delta}_{t-1})]$$
$$= \sigma^2 F_t + v_t^+ Mse[\hat{\delta}_{t-1}]v_t^{+\prime} \qquad (6.3a)$$

$$Mse[\alpha_t|Y_{t-1}] = Cov(\alpha_t - a_{t|t-1} + a_{t|t-1} - \hat{\alpha}_t)$$
$$= Cov(x_t) + Cov[A_t(1;\delta) - A_t(1;\hat{\delta}_{t-1})]$$
$$= \sigma^2 P_{t|t-1} + a_t^+ Mse[\hat{\delta}_{t-1}]a_t^{+\prime} \qquad (6.3b)$$

The prediction error $\hat{v}_t = y_t - E[y_t|Y_{t-1}]$ is given by $V_t(1;\hat{\delta}_{t-1})$ with its covariance matrix $\sigma^2 F_t + v_t^+ Mse[\hat{\delta}_{t-1}]v_t^{+\prime}$. Finally, the estimator for σ^2 using Y_t is given by

$$\hat{\sigma}_t^2 = (q_t - s_t' S_t^{-1} s_t)/(t\#[y_t]) \qquad (6.4)$$

and this result is invariant to any assumptions regarding δ.

All estimation quantities are evaluated by the modified KF. The estimator of δ and its mean square error matrix can be evaluated recursively by

$$\hat{v}_t = V_t(1;\hat{\delta}_{t-1}) \qquad (6.4a)$$
$$V_t^* = V_t Diag(0; S_{t-1}^{-1}) \qquad \hat{F}_t = F_t + V_t^* V_t' \qquad (6.4b)$$
$$\hat{\delta}_t = \hat{\delta}_{t-1} - V_t^{*\prime} \hat{F}_t^{-1} \hat{v}_t \qquad S_t^{-1} = S_{t-1}^{-1} - V_t^{*\prime} \hat{F}_t V_t^* \qquad (6.4c)$$

The equations (6.4) are obtained by applying an inversion lemma to the matrix $S_t = S_{t-1} + v_t^{+\prime} F_t^{-1} v_t^+$. The recursion (6.4) is started off at t=d+1 where d is the first integer for which S_d is non-singular. The estimators cannot be computed by (6.4) for t=1,...,d, unless a generalized inverse of S_t is used. Section 8 discusses more details and implement all required calculations in a Pascal computer program. An alternative recursion for (6.4) is embedded within a square root version of the modified KF, see De Jong (1991a).

When a full collapse has taken place at some point in time, the usual Kalman filter quantities $a_{t|t-1}$ and v_t are already the estimators based on Y_{t-1} and the equations (6.4) are not (and cannot be) applied.

Diagnostic checking

Once the estimators are evaluated, plotting the estimate of α_t or some linear combination, e.g. $Z_t\hat{\alpha}_t$, and plotting the estimate of y_t against time provides some indication of the adequacy of the model. More informative diagnostics are based on the standardised prediction errors $\hat{F}_t^{-\frac{1}{2}}\hat{v}_t$. These residuals are approximately serially independent in a correctly specified model and it is regularly assumed that they are normal distributed with a zero mean and a constant variance. The purpose of diagnostic checking is to validate these propositions. The available tools for this important phase in time series modelling can be classified in diagnostic tests and diagnostic plots. In the following, an overview of some important diagnostics will be given. For simplicity, it is assumed that the observation y_t is a scalar such that $\hat{v}_t$ and $\hat{F}_t$ also are scalars.

DIAGNOSTIC PLOTS

Standardised residual plot: The sequence of residuals

$$u_t = \hat{F}_t^{-1}\hat{v}_t/r \qquad t=d+1,\ldots,n$$

where $r^2 = (n-d-1)^{-1}\Sigma_{t=d+1}^{n}(\hat{F}_t^{-1}\hat{v}_t - m)^2$ and $m = (n-d-1)^{-1}\Sigma_{t=d+1}^{n}\hat{F}_t^{-1}\hat{v}_t$, can be plotted against time and must be checked for any irregularities.

Variance plot : To check if the variance is constant over time, the quantity $\hat{\sigma}_t^2$ of (6.4) might be examined.

Cumulative sum of residuals : Brown et.al. (1975) propose to check the cumulative sum of residuals to check for irregularities. The so-called CUSUM plot consist of the sequence

$$\text{cusum}_t = \text{cusum}_{t-1} + u_t = \Sigma_{j=d+1}^{t}u_j$$

for $t=d+1,\ldots,n$ and with $\text{cusum}_d=0$. If cusum_t does exceed specific

bands, which are associated with some siginificance level, it might indicate that some form of a structural change has taken place. The CUSUM is certainly not a formal test procedure.

Cumulative sum of squared residuals : Brown et.al. (1975) also propose to check the cumulative sum of squared residuals to check if the variance is constant. The so-called CUSUM-SQR plot is based on

$$\text{cusumsqr}_t = (\Sigma_{j=d+1}^{t} u_j^2)/(\Sigma_{j=d+1}^{n} u_j^2)$$

The CUSUM-SQR can be regarded as an additional diagnostic for structural change but is more appropriate as a check for heteroscedasticity.

Correlogram : A graph of the correlations is useful to check whether the residuals are serially correlated. The scaled residual sample autocovariance a_j is given by

$$a_j = \Sigma_{t=d+j+1}^{n} (u_t - m)(u_{t-j} - m)$$

for j=1,...,P and P is a specific constant. The residual sample autocorrelation of lag j is given by a_j/a_0. A plot of a_j/a_0 is called the correlogram. The significance of serial correlation at a specific lag depends on n, see Harvey (1981).

DIAGNOSTIC STATISTICS

Mean and variance : The sample mean m and the sample variance r^2 of the set of innovations $\hat{F}_t^{-\frac{1}{2}}\hat{v}_t$, for t=1,...,n, can be regarded as descriptive statistics.

Box-Ljung portmanteau statistic : This test statistic is mostly used in the context of ARIMA models and is based on a sum of squared autocorrelations, specifically,

$$Q = n^*(n^*+2)\Sigma_{j=1}^{p}(n^*-j)^{-1}a_j^2$$

where $n^*=n-d$. The Q statistic is asymptotically χ_p^2, provided that all system matrices of the SSF are known, see Ljung & Box (1978). If not, the statistic can be used as a diagnostic.

Measures of skewness and kurtosis : The third and fourth moments of u_t are given by, respectively,

$$s = r^{-3}(n-d-1)^{-1}\Sigma_{t=d+1}^{n}(\hat{F}_t^{-1}\hat{v}_t - m)^3$$
$$k = r^{-4}(n-d-1)^{-1}\Sigma_{t=d+1}^{n}(\hat{F}_t^{-1}\hat{v}_t - m)^4$$

which are asymptotically normal distributed

$$s \sim (0 \ , \ 6/\{n-d-1\})$$
$$k \sim (3 \ , \ 24/\{n-d-1\})$$

when the time series model is Gaussian and correctly specified.
Normality test : Bowman and Shenton (1975) propose a test for non-normality based on the statistic

$$N = (n-d-1)(s^2/6 + \{k-3\}^2/24)$$

and is, under the null hypothesis, χ_2^2 distributed in large samples.
Heteroskedasticity : A simple and intuitive diagnostic for heteroskedasticity is

$$H_h = SS(n-h+1,n)/SS(d+1,d+1+h)$$

where $SS(a,b) = \Sigma_{t=a}^{b}u_t^2$ and h is the nearest integer to (n-d)/3. Under the null hypothesis, the statistic H_h has a F(h,h) distribution.

Goodness of fit

A discussion of several goodness of fit measures is given by Harvey (1989). The basic measure is the prediction error variance $\hat{\sigma}^2=(q-s'S^{-1}s)/\#[y]$. The standard relative measure of fit is R^2 which compares $(q-s'S^{-1}s)$ with the sum of squared deviations from the mean, see Maddala (1988). In the context of time series it is better to base the comparison on the first differences of the data (R_D^2) or on the seasonal adjusted data (R_S^2), see Harvey (1989, section 5.5.5). These adjusted measures have a direct connection with Theil's U-statistic. Finally, two rival models may be compared on the basis of information criteria such as AIC, see Harvey (1981).

Illustration : the COAL series

This example concerns the quarterly UK coal consumption from 1960 until 1986 (source of data : UK Department of Energy). The time series is analysed earlier in Harvey (1989) where also the actual values of the COAL series are listed (p. 512). The log of the COAL series is plotted in figure 2.1. The salient features of this time series are a general downwards movement between 1965 and 1976, a clear seasonal pattern and a number of shocks of which the one in 1984 is most clear. These features can be explained as follows. The downwards trend is due to the introduction of gas from the North Sea in the late sixties. The seasonal effect is explained from the fact that the energy demand rises during the winter months for heating. The extreme values are generally caused by various industrial disputes but some are also caused by extreme cold periods. However, the sharp fall in 1984 is due to the prolonged miners' strike during that year.

It is the basic structural time series model that decomposes the COAL series into the salient components as discussed above. Let y_t denote the log of the COAL series and define the BSM model as

$$y_t = \mu_t + \gamma_t + u_t \qquad (6.5)$$

where μ_t is the trend, γ_t is the seasonal component and u_t the irregular which takes account of the random shocks in the series. The observations after 1983 are not considered such that the estimates are not influenced by the miners' strike in 1984.

In order to show that the data cannot be described satisfactorily by deterministic components, a regression model is fitted with a constant and a time-trend ($\mu_t=\alpha+\beta t$) and dummy seasonals. The regression results are given in table 2.4 and the plots of the series with the fitted trend line and the standardised OLS residuals are given in figure 2.2 and 2.3, respectively. Some diagnostics applied to the OLS residuals show that the regression model with a constant trend does not fit the data very well. A prominent indicator is the Box-Ljung test statistic for serial correlation Q(10)=74.49 but also the Durbin-Watson test is far off

its base-value 2.0. The global trend may be appropriate for the middle of the series but it is surely not appropriate for the end of the series, say, after 1975. This will cause serious problems in post-sample forecasting. To what extent the local trend for after 1975 differs with the global trend becomes clear from figure 2.2. The estimated slope of the global trend is -.0173 and the one for the local trend (from 1976 onwards) is -.0053. This notion of global and local properties of time series has led to the development of structural time series models where components such as trend and seasonal are allowed to change over time.

The components of a basic structural time series model follow various stochastic processes as discussed in section 3. The seasonals are based on dummies such that the sum of the four seasonal parameters has expectation zero. It is shown before that the BSM can be placed into a SSF from which three elements of its system matrices are unknown, i.e. the signal-to-noise ratios. These hyperparameters are estimated in the frequency domain by a scoring method using the computer program STAMP[1], see table 2.5. Other estimation methods gave similar estimates of the hyperparameters.

The initial conditions of a BSM are assumed to be diffuse because the model can be applied since time immemorial, see De Jong & Chu-Chun-Lin (1991). The modified KF provides estimates of the initial conditions and, from the mean square error matrix S^{-1}, their standard errors and their t-statistics can be calculated, see table 2.6.

The graph of the standardised innovations $\hat{F}_t^{-\frac{1}{2}}\hat{v}_t$ in figure 2.4 clearly shows that some large shocks disturb the time series. Although the values of the residuals are within the 95% confidence boundaries, i.e. approximately between the values -2 and 2, the set of diagnostics will point out whether the standardised innovations can be regarded as non-informative noise. In fact, the diagnostics reported in table 2.5 suggest that the BSM fits the COAL series not unsatisfactorily. Compare, for example, the Box-Ljung test-statistic with the one for the regression model. The additional diagnostic plots in figure 2.5 confirm this conclusion as well. Finally, the model can be improved by treating some observations, which cause the huge shocks, as missing.

7. NUMERICAL CONSIDERATIONS

Kalman filter computations sometimes lead to numerical instabilities caused by missing observations, badly defined initial conditions and rounding-off errors. A particular problem is the Kalman recursion for $P_{t|t-1}$ that does not rule out a negative definite matrix. To escape from these problems, the square root form of the KF (Sqrt KF) is applied. The Sqrt KF is based on othogonalization transformations such as the Householder, Givens or QR decompositions. A good reference for these techniques and associated matrix computations is Golub & van Loan (1989).

This section includes an exposition of the Householder transformation and gives details of the implementation of the Sqrt KF. The modified Sqrt KF requires an extra adjustment concerning the recursion for Q_t.

Householder transformation

Result [2.7] The Householder orthogonal transformation (HOT) finds, for any non-zero row vector x, a matrix P such that xP is a row vector of zeroes except its first position equals $-\pm\sqrt{(xx')}$. More formally, given a non-zero row vector x, the HOT defines

$$p = x \pm \sqrt{(xx')}e_1 \qquad P = I - 2p'p/pp' \tag{7.4a}$$

such that

$$xP = -\pm \sqrt{(xx')}e_1 \ . \tag{7.4b}$$

Note that the e_1 vector can be replaced by any zero vector with on a particular position a unit value, e.g. e_j. Matrix P will be referred to as the HOT matrix.

To derive the HOT method, let p be any row vector and define square matrix $P = I - 2p'p/pp' = I - 2p'q$ where row vector q is given by $q=p/pp'$. Note that $pq'=1$. It follows, sequentually, that (i) $P = P'$, (ii) $pP = -p$ and $Pp' = -p'$, (iii) $PP = P-2Pp'q = P+2p'q$

$= I$ and (iv) $P = P^{-1}$ and $P'P = I$. Consider a non zero row vector x and post multiply it by matrix P such that $xP=x-2x(p'p/pp')=x-2p(xp'/pp')$. Choose $p=x+\alpha e_1$ where α is an unknown scalar value and e_1 is a row vector of zeroes except its first position is unity. Substitution of this particular p vector into xP leads to

$$xP =[1 - 2(xp'/pp')]x - 2\alpha(xp'/pp')e_1 \qquad (7.3a)$$

where

$$xp' = xx' + \alpha x_1 \qquad (7.3b)$$

$$pp' = xp' + \alpha p_1 = xx' + \alpha x_1 + \alpha x_1 + \alpha^2 = xx' + 2\alpha x_1 + \alpha^2 \qquad (7.3c)$$

and z_1 denotes the first element of a vector z. The HOT aims to find a value for α such that $xp'/pp' = \frac{1}{2}$ and $xP = -\alpha e_1$, see (7.3a). The equations (7.3b) and (7.3c) show that the equation $xp'/pp' = \frac{1}{2}$ is equivalent to $2xx'+2\alpha x_1 = xx'+2\alpha x_1+\alpha^2$ and leads to the solution $\alpha=\pm\sqrt{(xx')}$. [ ]

The row vector x can also be regarded as a row of any matrix X. A sequence of vector HOT operations is able to transform X into a spatial matrix Y denoted by Y←HOT(X). This matrix HOT operation involves a sequence of matrix multiplications, e.g. XQ where $Q=P_1P_2...P_k$ and P_j is a particular HOT matrix based on an exclusive row of X. Note $Q'Q=I$. When a particular part of a row vector must be transformed into zeroes, the HOT matrix is partioned, accordingly. For example, let x=(a,b,c) where non-zero row vector b must be transformed into a zero vector (except its first position). For this case, the HOT matrix is defined by $P = diag(I,R,I)$ such that $r = b \pm \sqrt{(bb')}e_j$, $R = I - 2r'r/rr'$ and $xP = (a,-\pm\sqrt{(xx')}e_j,c)$.

Some practical problems arises during implementation, for example, the choice of plus or minus $\sqrt{(xx')}$. Such matters are discussed extensively in Golub & van Loan (1989). In the numerical literature, the HOT procedure is known as a very flexible and stable technique.

The square root Kalman filter

The square root form of the KF is based on the work of Morf & Kailath (1975). An introduction can be found in Anderson & Moore (1979).

Result [2.8] The Sqrt KF mainly consists of the HOT operation

$$E_t \leftarrow HOT(D_t)$$

where matrix $D_t = [(Z_t;T_t)P^{1/2}_{t|t-1}, (G_t;H_t)]$ and matrix E_t is the spatial matrix $[F^{1/2}_t, 0; K_tF^{1/2}_t, P^{1/2}_{t+1|t}, 0]$. This matrix HOT operation appears to be recursive which is implied by the definitions of D_t and E_t.

Consider the mean square error matrix of the joint vector $(y_t;\alpha_t)$ using the data-set Y_{t-1}, $C_t=Mse[y_t;\alpha_t|Y_{t-1}]$, given by

$$C_t = \sigma^2(Z_t;T_t)P_{t|t-1}(Z_t;T_t)' + (G_t;H_t)(G_t;H_t)' \qquad (7.5)$$

The mean square error matrix (7.5) can be rewritten in two different ways, e.g. D_tD_t' and E_tE_t'. The equivalence $C_t=E_tE_t'$ becomes clear from the Kalman filter equations for F_t, K_t and $P_{t+1|t}$, see (4.20) for the latter. This implies that a matrix Q exists such that $D_tQ = E_t$ and $Q'Q=I$. A matrix HOT operation provides such a matrix. In fact, the Sqrt KF involves only the transormation of D_t into a matrix with the same spatial structure of E_t [ ]

The matrices D_t and E_t do have the same number of rows but, in general, matrix D_t does have more non-zero columns than E_t. Therefore, the matrix E_t is expanded with 'dummy' zero columns to let D_t and E_t have the same number of columns, see table 2.9.

The KF quantities can all be formed using matrix E_t. The mean square error matrix $P_{t|t-1}=P^{1/2}_{t|t-1}P^{1/2\prime}_{t|t-1}$ is now, by definition, positive definite. A systematic scheme of the modified Sqrt KF is given in table 2.7. De Jong (1991a) formulates the recursion for Q_t in a square root form as well. The attraction is that the inverse of S_t is obtained without inverting the matrix S_t straightforwardly. Moreover, the paper of De Jong (1991a) presents an eloborate discussion on stable calculations regarding the modified Kalman filter.

The sqrt KF applied to a structural time series model does not involve many transformations with large HOT matrices. This will be illustrated for the the local level (LL) model, the local linear trend (LLT) model and the basis structural time series model (BSM) with quarterly dummy seasonals. Table 2.9 presents the structure of the D_t and the E_t matrix with o denoting a zero value and * denoting a non-zero value. Table 2.8 gives, for some STS models, the dimensions of the required transformations and the number of times it has to be applied for one KF update. It can be concluded from these tables that the dimension of the state vector determines how many times a HOT matrix must be constructed and that the dimensions of the transformations vary but are in general quite small.

8. COMPUTER PROGRAMS

This section presents a computer program which takes care of all computations discussed in this chapter. The computer language is Pascal. The actual code of the program is given at the end of the thesis. The code can be adjusted in a flexible way so that the user can add other computations. It is recommended to keep the main structure of the program.

The program below only deals with univariate time series models such that the observation, the innovation and their covariances are scalars and the Kalman gain is a vector. It is straightforward to generalise the program for multivariate time series models, this version of the program is avaible on request.

The main program

The system matrices are placed in the variables, TSM (time series model), RGM (regression model), CSM (covariance structure), ICM (initial conditions) and CIM (covariances of initial conditions). In table 2.10 it is shown how these matrices are organised in the computer program. The dimensions of the system matrices all depend on the integers DimSt #[α_t], DimBt #[β], DimEp #[ϵ_t] and DimRg #[δ]. The one-dimension array of real values (Data-record) Y does contain the set observations. The set of Data-records

X contain time-varying system elements such as explanatory variables in a regression (X_t).

The system matrices are in general sparse and may contain specific values such as plus or minus unity, sine and cosine quantities and time-varying values. The computer program takes account of all these special features by defining a special record for system elements given by

```
SysFlt  = Record
            Code     : 0..9;
            Positive : Boolean;
            Rl       : Real;
            Pos      : Integer;
          End;
```

where the field code corresponds to

0 = 0.0	5 = Cos(2π/Rl)
1 = 1.0	6 = Tan(2π/Rl)
2 = Rl	7 = Sqr(Rl)
3 = X[Pos]	8 = Sqrt(Rl)
4 = Sin(2π/Rl)	9 = Exp(Rl)

For example, if a system element has the sequence of field values (4,FALSE,4.0,0) then its value is -Sin(π/2). Another example is (3,TRUE,0.0,2) such that the system element value is found in the second Data-record of X. This allows system elements to be time-varying. The system matrices consist of a double array of system element pointer records. If the system element pointer is NIL then its value is zero. The Kalman filter quantities, that is V_t, F_t, K_t, A_t and $P_{t|t-1}$, are arrays of real values.

Specific procedures are required that read in the observations and the system matrices. The computer program, as discussed in this section, does not deal with the input/output procedures but it will indicate where to place it in the main program. Another important part of the program are two procedures that multiply and add system elements.

Two procedures are discussed in detail below. The first procedure concerns the Kalman update for the state quantities given a specified time series model, the new observation and the previous

state quantities. The other procedure computes the likelihood, it calculates estimates of fixed and diffuse effects and it computes one-step ahead predictions of the state and the observation sequences given the time series model and the set of observations. In the next chapter, some other procedures are discussed which can be included in the main program as well.

The KFupdate procedure

This procedure updates the state vector and its mean square error matrix for a given new scalar observation. It computes the innovation, its variance and the Kalman gain vector as well. The heading of the procedure is

```
KFupdate(        Obs          : Float;
                 DimKF        : Integer;
           Var   KFA,KFP      : Matrix );
```

The new observation is placed in Obs and the dimension of the KF operations regarding V_t and A_t is given by DimKF. Note that when DimKF is equal to zero, the modified KF reduces to the usual KF. For a time-invariant time series model, the KF may reach a steady state at, say, time j such that $P_{t+1|t}=P_{t|t-1}$, for t=j+1,...,n. In this case, the calculations for F_t, K_t and $P_{t+1|t}$, for t=j+1,...,n, are not required. The procedure does indicate where to check for a steady state but it is not implemented in the program. The structures of the matrices KFA and KFP differ slightly before and after the KF update procedure, see table 2.11.

The procedure KFupdate consists of the following parts

1. $M_1 = (Z_t;T_t)A_t$ $\qquad M_2 = (Z_t;T_t)P_{t|t-1}$
2. $KFP = (Z_t;T_t)M_2'$
3. $KFA = M_1 + (X_t;W_t)$
   $KFP = KFP + (H_t;G_t)(H_t;G_t)'$
4. $V_t = (y_t,0) - KFA[0,0..DimKF]$ $\qquad F_t = KFP[0,0]$
5. $K_tF_t = KFP[1..DimSt,0]$ $\qquad K_t'=KFP[0,1..DimSt]/F_t$
6. $A_{t+1} = KFA[1..DimSt,0..DimKF] + K_tV_t$
   $P_{t+1|t} = KFP[1..DimSt,1..DimSt] - K_tF_tK_t'$

When the KF enters a steady state, the computations regarding the matrices M_2 and KFP can be dropped.

The Estimation procedure

Primarily, the procedure Estimation evaluates recursively the scalar q_t, the vector s_t and the matrix S_t^{-1} for t=j+1,...,n. From these quantities, the one-step ahead predictions of the observation and the state vector, including their mean square error quantities, can be evaluated. Note that this includes the one-step ahead prediction error and its variance. Furthermore, at the end of the procedure, the likelihood can be calculated and the estimate of δ can be constructed. The heading of the procedure is

```
Estimation( Var Q       : Matrix;
            Var L       : Float;
            Var d       : Integer);

Var             A,P,Si  : Matrix;
                k       : Vector;
                v,f,σ   : Float;
                t       : Integer;
```

The integer value d indicates the time that the matrix S_t becomes non-singular and the integer value t is the time index. The real value L contains the likelihood value. The structure of the composite matrix Q is given by (q,-s';-s,S) for t=1,...,d and (q,-s';Sj,Si) for t=d+1,...,n where $Si=S^{-1}$ and $Sj=S^{-1}s$. The matrices A and P contain all KF quantities and they act as the intermediates for the KFupdate procedure. The vector k places $V_tS_{t-1}^{-1}$ and the real values v, f and σ places $\hat{v}_t$, $\hat{F}_t$ and $\hat{\sigma}_t^2$ respectively, see equation (6.4). The overall structure of the procedure is given by table 2.12. The computer code indicates where to check for a steady state via the sign !.

9. CONCLUDING REMARKS

This chapter is an overview of the latest developments in the state space methodology. The new developments, such as diffuse likelihood evaluation and estimation of fixed effects in time series models, can be straightforwardly applied to structural time series models as also emphasized by Harvey & Shephard (1992). Furthermore, the general results are important for the new smoothers whiche are developed in the next chapter.

NOTES

[1] The computer program STAMP (Structural Time series Analyser, Modeller and Predictor) is a menu-driven software system for IBM-compatible PCs and its main task consists of hyperparameter estimation, diagnostics, forecasting and full sample estimation of unobserved components for univariate structural time series models. (Information : STAMP, Statistics Department, LSE, Houghton Street, London WC2A 2AE, UK)

APPENDIX 2A
Two results on minimum mean square linear estimation

This appendix gives two important standard results of linear estimation. The proofs can be found in many books on multivariate statistical analysis but also in Anderson & Moore (1979).

1. Minimum mean square linear estimation.
Consider two random vectors $x \sim (\mu_x, \Sigma_x)$ and $y \sim (\mu_y, \Sigma_y)$ which have a non-zero covariance matrix Σ_{xy} and where Σ_y is non-singular. The minimum mean square linear estimator for x, using y, and its mean square error (Mse) matrix are given by

$$\hat{x} = E[x|y] = \mu_x + \Sigma_{xy}\Sigma_y^{-1}(y-\mu_y) \qquad Mse(\hat{x}) = \Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}$$

2. Uncorrelated predictions
Suppose $y=(y_1;\ldots;y_n) \sim (\mu_y, \Sigma_y)$ where Σ_y is block diagonal such that the sequence $y_1,\ldots,y_n$ is mutually uncorrelated, then the minimum mean square linear estimator of $x \sim (\mu_x, \Sigma_x)$ using y is given by

$$\hat{x} = E[x|y] = \Sigma_{t=1}^{n} E[x|y_t] - (n-1)\mu_x$$

and the estimator using $(y_1;\ldots;y_t)$ is given by

$$E[x|y_1,\ldots,y_t] = E[x|y_1,\ldots,y_{t-1}] + E[x|y_t] - \mu_x$$

APPENDIX 2B
Derivation of the Kalman filter under Gaussianity

Consider the state space form (2.1)-(2.3)) and assume Gaussianity, i.e. ϵ_t is normally distibuted, such that the theory of conditional distributions can be applied to derive the Kalman filter. This particular derivation requires some classical results

of conditional distributions which are found in many introductory books on statistics, see for example Mood, Graybill and Boes (1974).

The distribution of the state vector α_t conditional on Y_{t-1} and δ is given by

$$\alpha_t | Y_{t-1}, \delta \sim N(a_{t|t-1} , \sigma^2 P_{t|t-1}) .$$

From the SSF it follows that the conditional distribution for y_t is given by

$$y_t | Y_{t-1}, \delta \sim N(Z_t a_{t|t-1} + X_t \beta , \sigma^2 F_t)$$

where $F_t = Z_t P_{t|t-1} Z_t' + G_t G_t'$. Consider the joint distribution of $(y_t ; \alpha_{t+1})$, conditional on Y_{t-1} and δ, given by

$$N\left(\begin{bmatrix} Z_t a_{t|t-1} + X_t \beta \\ T_t a_{t|t-1} + W_t \beta \end{bmatrix} , \sigma^2 \begin{bmatrix} Z_t P_{t|t-1} Z_t' + G_t G_t' & Z_t P_{t|t-1} T_t' + G_t H_t' \\ T_t P_{t|t-1} Z_t' + H_t G_t' & T_t P_{t|t-1} T_t' + H_t H_t' \end{bmatrix} \right)$$

The conditional distribution for $\alpha_{t+1} | Y_{t-1}, y_t, \delta$ is given by $N(a_{t+1|t}, \sigma^2 P_{t+1|t})$ where

$$\begin{aligned} a_{t+1|t} &= T_t a_{t|t-1} + W_t \beta + K_t v_t \\ P_{t+1|t} &= T_t P_{t|t-1} T_t' - K_t F_t K_t' + H_t H_t' \end{aligned}$$

with $v_t = y_t - E[y_t | Y_{t-1}, \delta] = y_t - Z_t a_{t|t-1} - X_t \beta$ and $K_t = [T_t P_{t|t-1} Z_t' + H_t G_t'] F_t^{-1}$. These update equations are obtained by applying the standard rule on conditional distributions. It follows, by definition, that $a_{1|0} = W_0(b + B\delta)$ and its mean square error matrix $\sigma^2 P_{1|0}$ is equal to $\sigma^2 H_0 H_0'$. The recursions and the definitions for v_t, F_t and K_t form the Kalman filter which is derived under Gaussianity, see also Anderson & Moore (1979) and Harvey (1989).

APPENDIX 2C
The Jacobian for a Structural time series model

This appendix derives the first derivative of the likelihood criterion of a structural time series model with respect to a set of unknown parameters. Any structural time series model can be put into the Gaussian time-invariant univariate state space formulation, i.e.

$$y_t = Z\alpha_t + (\sigma,0)\epsilon_t \qquad \alpha_{t+1} = T\alpha_t + H\epsilon_t$$

where ϵ_t is normally distributed with mean zero and covariance matrix I. The observations are supposed to be scalar such that Z is a row vector. The unknown hyperparameters are only found in matrix H which can be regarded as a function of the parameter vector ψ. The likelihood criterion can be expressed as

$$L = \Sigma_{t=1}^{n} L_t \qquad \text{where } L_t = \log F_t + v_t^2/F_t$$

and the standard deviation σ can be concentrated out of the likelihood and estimated directly, see section 2.4. The Kalman filter can be rewritten as the set of equations

$$a_{t|t-1} = Ta_{t-1} \qquad P_{t|t-1} = TP_{t-1}T' + HH'$$
$$v_t = y_t - Za_{t|t-1} \qquad F_t = ZP_{t|t-1}Z' + \sigma^2$$

$$a_t = a_{t|t-1} + P_{t|t-1}Z'F_t^{-1}v_t$$
$$P_t = P_{t|t-1} - P_{t|t-1}Z'F_t^{-1}ZP_{t|t-1}$$

where all quantities are conditional on the initial state vector α_1. To derive the Jacobian vector of L with respect to ψ, denoted by $J(\psi)$, we adopt the matrix differential theorems of Magnus & Neudecker (1988). The differential for L is $dL = \Sigma_{t=1}^{n} dL_t$ where

$$dL_t = (2v_t/F_t)(dv_t) - ([v_t^2-F_t]/F_t^2)(dF_t)$$

and

$$dv_t = -Z\, da_{t|t-1} \qquad da_{t|t-1} = T\, da_{t-1}$$
$$dF_t = Z\, dP_{t|t-1}\, Z' \qquad dP_{t|t-1} = T\, dP_{t-1}\, T' + dH\, H' + H\, dH'$$

In order to evaluate the differentials recursively, we obtain

$$dP_t = dP_{t|t-1} - dP_{t|t-1}\, Z'F_t^{-1}ZP_{t|t-1}$$
$$+ P_{t|t-1}Z'F_t^{-1}\, dF_t\, F_t^{-1}ZP_{t|t-1} - P_{t|t-1}Z'F_t^{-1}Z\, dP_{t|t-1}$$

such that

$$dP_{t+1|t} = L_t\, dP_{t|t-1}\, L_t' + dH\, H' + H\, dH'$$

where $L_t = T - K_tZ$ and $K_t = TP_{t|t-1}Z'F_t^{-1}$. Note that $dF_t^{-1} = -F_t^{-1}\, dF_t\, F_t^{-1}$. By using the previous result, the following results can be derived in the same way

$$da_t = da_{t|t-1} + dP_{t|t-1}\, Z'F_t^{-1}v_t - P_{t|t-1}Z'F_t^{-1}\, dF_t\, F_t^{-1}v_t$$
$$+ P_{t|t-1}Z'F_t^{-1}\, dv_t$$

$$da_{t+1|t} = L_t\, da_{t|t-1} + T\, dP_{t|t-1}\, Z'F_t^{-1}v_t - K_t\, dF_t\, F_t^{-1}v_t$$
$$= L_t\, da_{t|t-1} + L_t\, dP_{t|t-1}\, Z'F_t^{-1}v_t\ .$$

Using the arguments in Magnus & Neudecker (1988), the Jacobian row vector of L with respect to ψ can be written as $J(\psi) = \Sigma_{t=1}^{n} J_t$ where

$$J_t = \partial L_t/\partial\psi' = -[\ 2(v_t/F_t)Z\ ,\ (v_t^2-F_t)(Z \otimes Z)/F_t^2\ ]\ D_t$$

and

$$D_t = \partial(a_{t|t-1}\ ;\ \text{vec}\ P_{t|t-1})\ /\ \partial\psi'$$

which can be evaluated recursively by

$$D_{t+1} = [L_t, (v_t/F_t)(Z \otimes L_t)\ ;\ 0, L_t \otimes L_t]D_t$$
$$+ [0\ ;\ H\otimes I + I\otimes H](\partial\text{vec}\ H/\partial\psi')$$

Note that vec(1,2;3,4;5,6)=(1;3;5;2;4;6) and vec ABC=(C' $\otimes$ A)vec B. For the standard structural time series models the matrix ∂vec $H/\partial\psi'$ is a selection matrix with elements equals to zero or one.

TABLES & FIGURES

Table 2.1

Inference results concerning δ under different assumptions for δ

	$\hat{\delta}_t = E[\delta \mid Y_{t-1}]$	$Mse[\hat{\delta}_t] = Mse[\delta \mid Y_t]$
Fixed	$S_t^{-1}s_t$	$\sigma^2 S_t^{-1}$
Random	$(\Lambda^{-1} + S_t)^{-1}(\Lambda^{-1}\mu + s_t)$	$\sigma^2(\Lambda^{-1} + S_t)^{-1}$
Diffuse	$S_t^{-1}s_t$	$\sigma^2 S_t^{-1}$

Table 2.2

Required number of computations for one Kalman filter step applied to a LLT model with two explanatory variables

	modified KF	CKF - I	CKF - II
Additions	56	54	32
Multiplication	68	50	32

Table 2.3

Dimensions of Kalman filter quantities before and after the collapse for a LLT model with two explanatory variables

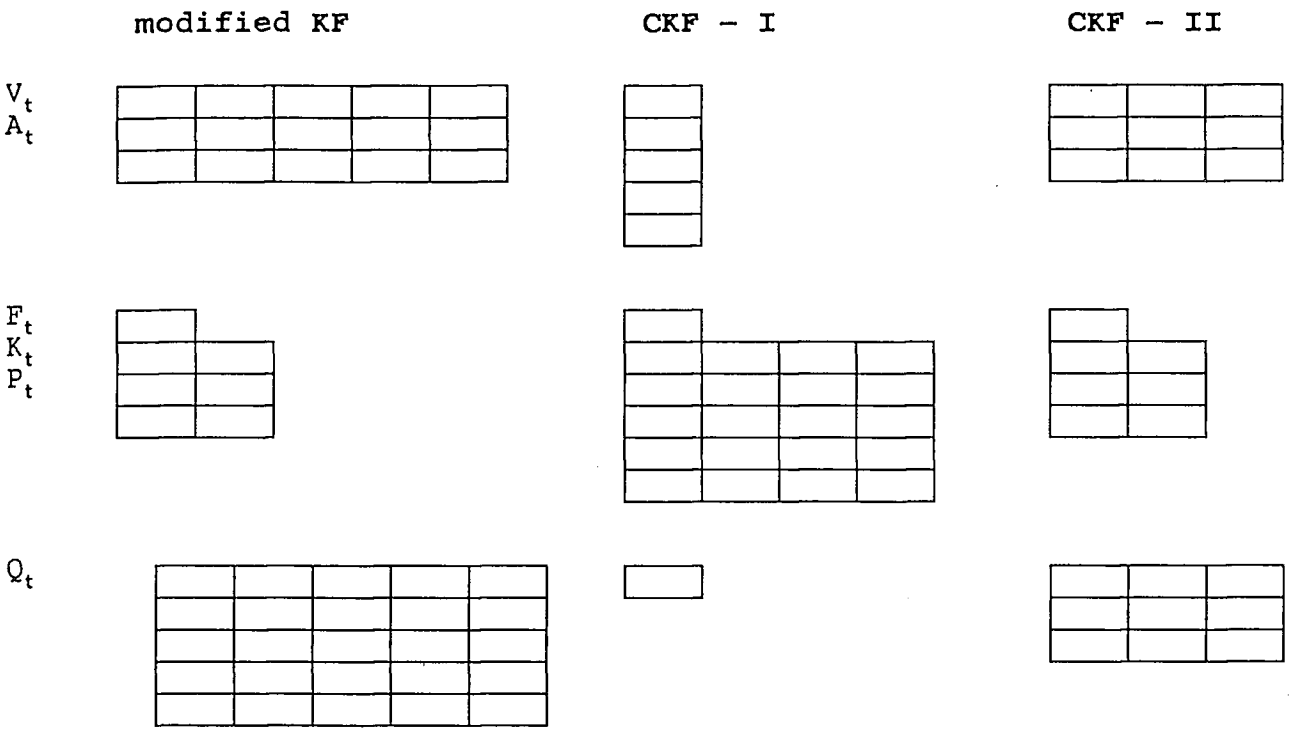


Table 2.4

OLS regression results for COAL

dependent variable : Log of $COAL_t$ t = 1960q1,...,1983q4
number of observations : 96 season : quart. (s=4)
model : deterministic BSM method : OLS

parameter	estimate	stand.err.	t-statistic
level α	5.472	0.031	178.197
slope β	-0.017	0.001	-31.551
seas γ_1	0.262	0.026	9.926
γ_2	-0.106	0.026	-4.014
γ_3	-0.412	0.026	-15.611

R-square	0.937	(adjusted 0.934)
DW-stats	1.183	
BL-stats (10)	74.49	
SumSqrResid	2.024	

Table 2.5

Hyperparameter estimates BSM for COAL

dependent variable : Log of $COAL_t$ t = 1960q1,...,1983q4
number of observations : 96 season : quart. (s=4)
model : BSM with diffuse initial cond. method : FD scoring

hyperparameter	estimate	q ratio
$10^4 \times \sigma_\epsilon^2$	114	
$10^4 \times \sigma_\eta^2$	4.577	0.04015
$10^4 \times \sigma_\zeta^2$	0.125	0.00110
$10^4 \times \sigma_\omega^2$	3.803	0.03336

Table 2.6

Estimates of diffuse initial conditions BSM for COAL

dependent variable : Log of $COAL_t$ t = 1960q1,...,1983q4
number of observations : 96 season : quart. (s=4)
model : BSM with diffuse initial cond. method : modified KF

state element	estimate	stand.err.	t-statistic
level α_1	5.360	0.194	27.643
slope β_1	-0.008	1.046	-0.008
seas γ_1	0.286	0.231	1.239
γ_2	0.196	0.207	0.948
γ_3	-0.360	0.205	-1.760

Table 2.7

Outline of implementation square root KF

Step	Description
[0]	The Sqrt KF is initialized with $A_1 = W_0(b,B)$ and $P^{\frac{1}{2}}_{1\|0} = (H_0, 0)$
[1]	The matrix V_t is calculated and $D_t = [(Z_t;T_t)P^{\frac{1}{2}}_{t\|t-1}, (G_t;H_t)]$
[2]	$E_t \leftarrow HOT(D_t)$
[3]	The matrix A_{t+1} is updated using E_t
[4]	The process [1]-[3] is repeated until end of sample

Table 2.8

Number of Householder transformations for one square root KF-step

Model	Dimension 2	3	4	5	6
Local Level	1	1			
Local Linear Tr		1	2		
Basic Str TSM_4	1	1	3		1

Table 2.9

Zero (o) and non-zero (*) elements for Sqrt KF matrices D and E

Model	Matrix D	Matrix E
Local Level	* * o * * *	* o o * * o
Local Linear Tr	* o * o o * * o * o o * o o *	* o o o o * * * o o * * * o o
Basic Str TSM_4	* o * o o * o o o * * o o o o * o o o * o o o o o * o o o * * * o o o * o o * o o o o o o o o o * o o o o o	* o o o o o o o o * * * * * * o o o * * * * * * o o o * * * * * * o o o * * * * * * o o o * * * * * * o o o

Table 2.10

Organisation of the SSF system matrices in the computer program

Program name	Positions	System matrix
TSM	0,1..DimSt	Z_t
	1..DimSt,1..DimSt	T_t
RGM	0,0	$X_t b$
	0,1..DimRg	$X_t B$
	1..DimSt,0	$W_t b$
	1..DimSt,1..DimRg	$W_t B$
CSM	0,0	$G_t G_t'$
	0,1..DimSt	$G_t H_t'$
	1..DimSt,0	$H_t G_t'$
	1..DimSt,1..DimSt	$H_t H_t'$
ICM	1..DimSt,0	$W_0 b$
	1..DimSt,1..DimRg	$W_0 B$
CIM	1..DimSt,1..DimSt	$H_0 H_0'$

Table 2.11

Organisation of the modified KF matrices in the computer program : before and after KF update

Name	Position	Before KF	After KF
KFA	0,0..DimKF	-	V_t
	1..DimSt,1..DimKF	A_t	A_{t+1}
KFP	0,0	-	F_t
	0,1..DimSt	-	K_t'
	1..DimSt,0	-	K_tF_t
	1..DimSt,1..DimSt	$P_{t\|t-1}$	$P_{t+1\|t}$

Table 2.12

Structure of procedure Estimation

Step	Control	Operations		
0.	Initialize	$Q = 0$	$L = 0$	$t = 0$
		$A = W_0(b,B)$	$P = H_0H_0'$	$d = 0$
1.	KF step	$t = t+1$		
		KFupdate(y[t], #[δ], A, P)		
		$L = L + \log F_t$		
2.	If d=0 then	$Q = Q + V_t'F_t^{-1}V_t$		
		If not $\|Q\| = 0$ then :	$Sinv = S^{-1}$	
			$Sinvs = -S^{-1}s$	
			$d = t$	
	else	$(q;-s) = (q;-s) + V_t'F_t^{-1}v_t^o$		
		$(0;k) = (0;Sinv)V_t'$		
		$v = V_t(1;Sinvs)$	$f = F_t + V_t(0;k)$	
		$Sinvs = Sinvs - k(v/f)$	$Sinv = Sinv - kk'/f$	
		$\sigma = (q - s'Sinvs)/t$		
3.	If t=n then goto 4			
	else goto 1			
4.		$q = (q - s'Sinvs)$		
		$L = L - \log \|Sinv\| + n + \log 2\pi q$		

Figure 2.1 UK Coal consumption between 60Q1 and 85Q4 (in logs)

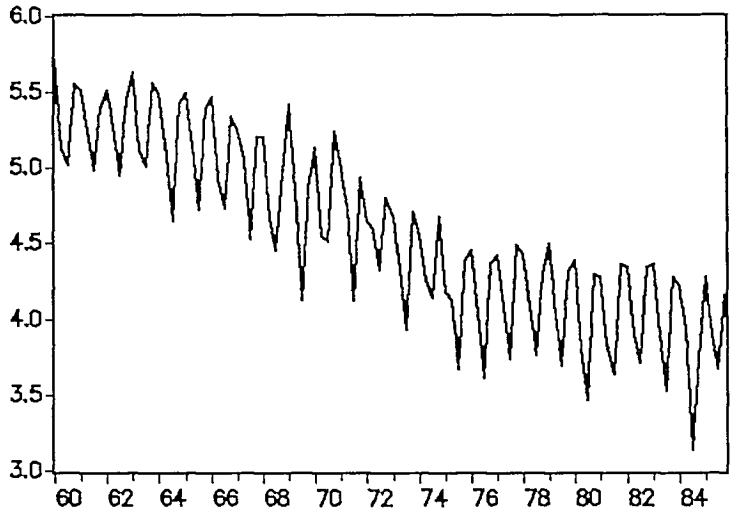


Figure 2.2 UK Coal consumption with global & local fixed trends

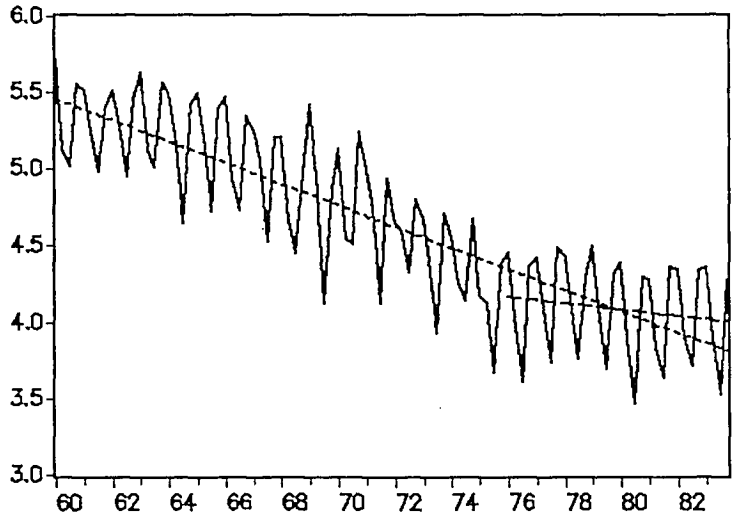


Figure 2.3 Standardized OLS residuals for COAL

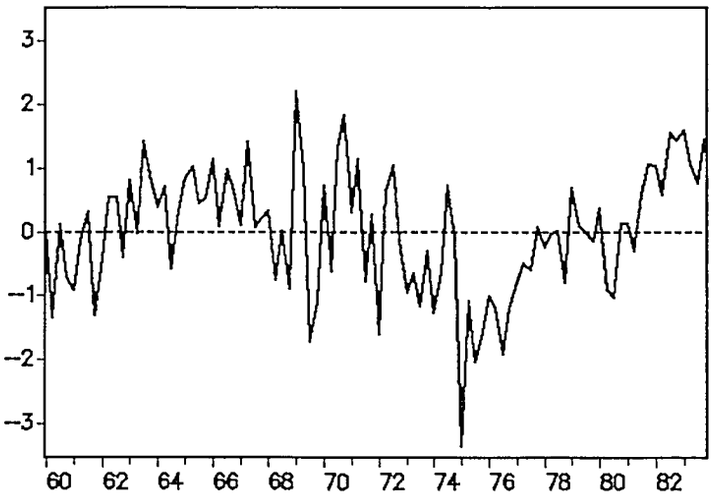


Figure 2.4 Standardized innovations for COAL

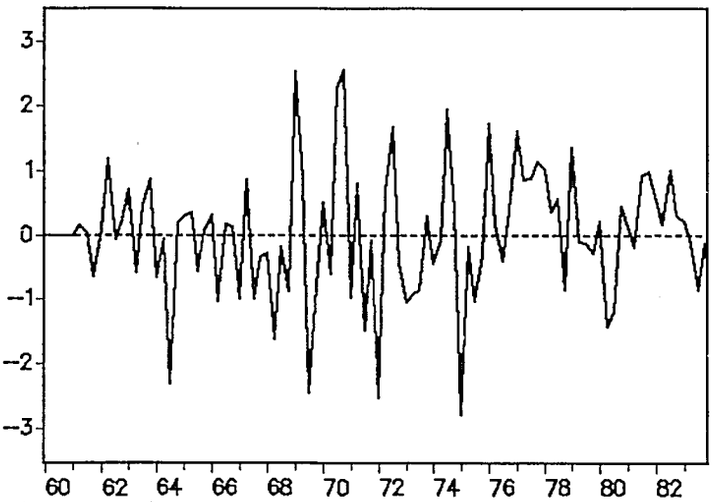


Figure 2.5a Diagnostic plot for COAL : Sigma Square

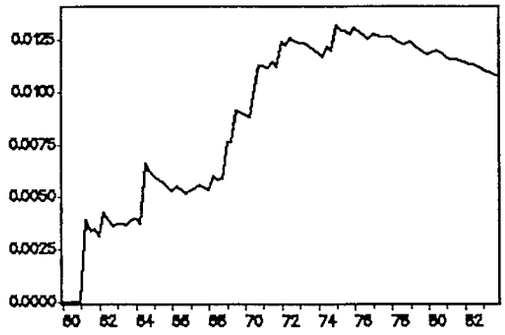


Figure 2.5b Diagnostic plot for COAL : Cusum

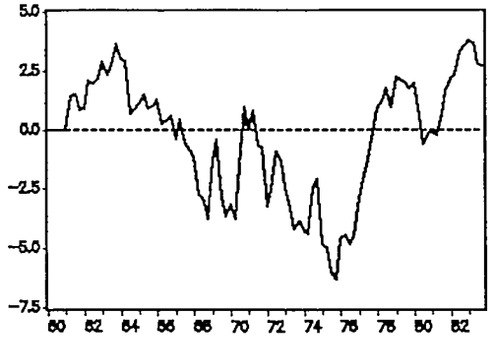
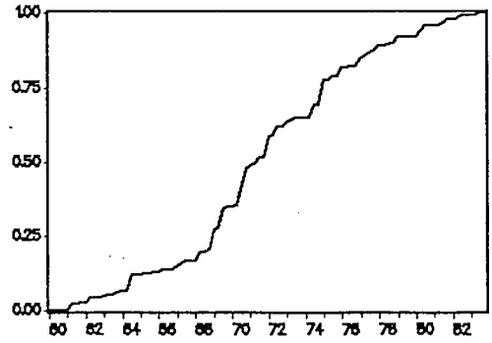


Figure 2.5c Diagnostic plot for COAL : Cusum Square



CHAPTER 3
On Smoothing in Time Series Models : theory, algorithms and applications

0. ABSTRACT

The main objective of this chapter is to develop a new and efficient method for evaluating the smoothed estimator of the state vector and the disturbance vector. For most linear time series models, the new algorithms lead to considerable computational savings as compared with existing methods. The smoothing results have several implications, for example, they can be applied to obtain the theoretical auto- and cross-correlation functions of the estimated residuals of a time series model. Also, the new smoothing results are applied to improve the EM algorithm for estimating parameters in the covariance matrices of a time series model.

Keywords : Auxiliary residuals; Diffuse; EM algorithm; Kalman filter; Signal extraction; Smoothing; State space; Structural time series model.

1. INTRODUCTION

The (modified) Kalman filter yields estimators of the state vector using the set of past observations Y_{t-1}, see the previous chapter. The objective of a smoother is the same but it takes account of all observations in the sample, i.e. Y_n. The full-sample estimator of the state vector is often referred to as the smoothed state vector. Smoothers are important in different areas of time series modelling such as signal extraction, maximum likelihood estimation via the EM algorithm, cross-validation and missing value estimation by interpolation.

An extensive overview of classical smoothing algorithms is

given by Anderson & Moore (1979, Chapter 7). They give three different smoothing algorithms for different purposes. The derivations of these smoothers are messy and they are based on an extension of the state vector with the quantities of interest. For econometric and statistical applications, but also in other areas, the fixed interval smoother is the most important smoothing algorithm. This off-line smoother requires a huge amount of computer storage space and it is computational expensive. In section 2, the classical methods of smoothing will be discussed in detail.

Recently, improved smoothing algorithms have been developed by De Jong (1988b,1989) and Kohn & Ansley (1989) to be referred to as DJKA. They propose a simple backwards dummy vector recursion as the 'heart' of the smoother. Their approach is computationally more efficient compared to the classical smoothing algorithms but no savings on storage space are obtained. Section 3 will show that the DJKA smoother follows immediately from the classical methods of smoothing. Also, the formal derivation of the improved smoothers, as given by DJKA, will be reproduced and discussed.

Section 4 develops a fixed interval smoother which evaluates the smoothed estimators of the disturbance vector and its mean square error matrix. The disturbance vector appears in the measurement equation and in the transition equation of the SSF, see section 2.2. The derivation of the disturbance smoother is direct and straightforward and it is based on minimum mean square linear estimation. The dummy recursion of DJKA does appear as well but now the result is transparent and an interpretation is available for the dummy vector.

Smoothing algorithms can be modified in a similar way as the KF is modified such that they can deal with fixed and diffuse initial conditions. The modifications are trivial and straightforward. The smoothed estimators can be evaluated exactly by using the modified smoothing algorithms. These matters and computational matters of smoothing are discussed at the end of the sections 3 and 4.

One important implication of the disturbance smoother is a more efficient new fixed interval smoothing algorithm for the state vector. In comparing DJKA's method, it saves on computational time

and storage space for important practical time series models which have a SSF with sparse system matrices, see section 5. Another attractive implication of the new disturbance smoother is a considerable improvement of the EM algorithm for estimating the parameters of a time series model by maximum likelihood. The new EM algorithm is discussed in section 6.

The new smoothing results are, in particular, useful for practical time series modelling. For example, it is straightforward to construct the theoretical auto- and cross correlation functions of the residuals of a time series model. Using the same illustration as in chapter 2, section 7 will give the details of how the smoothing methods can be applied in empirical time series analysis. Section 8 gives computer programs for the new smoothing algorithms. Some conclusions and a discussion follow in section 9.

2. CLASSICAL METHODS OF SMOOTHING

Anderson & Moore (1979) discusses three different types of smoothing algorithms that adjust the Kalman filter predictors of the state vector by taking account of an augmented set of observations. The fixed point smoother is concerned with the estimator of the state vector at a specific time-point j using the set of observations Y_j plus the observations which are gradually coming available after time j. Thus it evaluates $E[\alpha_j|Y_t,\delta]$ for $t=j,\ldots,n$ where $0<j<n$ is a fixed integer. This smoother is an on-line filter such that it is applied in parallel with the Kalman filter. The fixed lag smoother is also an on-line smoother but it gives estimators of the set of lagged state vectors, $E[\alpha_{t-1};\ldots;\alpha_{t-p}|Y_t,\delta]$ for some fixed p. The estimators of the full set of state vectors $(\alpha_1;\ldots;\alpha_n)$, using the full set of observations Y_n, are obtained from the fixed interval smoother. This smoother for $E[\alpha_t|Y_n,\delta]$, $t=1,\ldots,n$, is a backwards off-line filter which can be applied after the forwards Kalman filter.

This section discusses only the fixed point and the fixed interval smoothers because they are most important in empirical applications concerning social and economic time series. The

derivations of these classic smoothers are given as presented in Anderson & Moore (1979). Furthermore, it will be shown how the smoothers can be modified as the KF is modified in order to deal with fixed and diffuse conditions regarding δ.

Fixed point smoother

An algorithm that evaluates the estimator of the state vector at a specific time j using the observations upto time $t \geq j$, i.e. $E[\alpha_j | Y_t, \delta]$ for $t=j+1,\ldots,n$, is called a fixed point smoother. It is obtained by including the state vector α_j into the state space formulation, for $t=j,\ldots,n$, and applying the Kalman filter straightforwardly. The augmented SSF is given by

$$y_t = (Z_t, 0)(\alpha_t; \alpha_j) + X_t\beta + G_t\epsilon_t \qquad (2.1a)$$

$$(\alpha_{t+1}; \alpha_j) = (T_t, 0; 0, I)(\alpha_t; \alpha_j) + (W_t; 0)\beta + (H_t; 0)\epsilon_t \qquad (2.1b)$$

for $t=j,\ldots,n$. The usual KF for the augmented SSF (2.1) gives the required estimator $(a_{t+1|t}; a_{j|t})$ for $t=j,\ldots,n$. By decomposing the augmented KF analytically for the two parts of the state vector, leads to (i) the usual KF for the non-augmented SSF and (ii) a new set of recursions to be used in parallel with the usual KF, see Anderson & Moore (1979) and Harvey (1989). The second set of recursions is called the fixed point smoother and it is given by

$$K^*_{j,t} = P_{j,t|t-1}Z_t' F_t^{-1} \qquad P_{j,t+1|t} = P_{j,t|t-1}L_t' \qquad (2.2a)$$

$$a_{j|t} = a_{j|t-1} + K^*_{j,t}v_t \qquad P_{j|t} = P_{j|t-1} - P_{j,t|t-1}Z_t'K^*_{j,t}' \qquad (2.2b)$$

for $t=j,\ldots,n$ and where $L_t=T_t-K_tZ_t$. The two mean square error matrices are defined by $\sigma^2 P_{j|t-1}=\text{Mse}[\alpha_j | Y_{t-1}, \delta]$ and $\sigma^2 P_{j,t|t-1}=\text{Mse}[\alpha_j, \alpha_t | Y_{t-1}, \delta]$ and it is defined that $P_{j,j|j-1} \equiv P_{j|j-1}$. The smoother is started off at $t=j$ and is applied (on-line) in parallel with the KF. The equation for $P_{j|t}$ can be dropped when the smoothed mean square error matrix of α_j is not of interest.

The fixed point smoothing algorithm shows that the smoothed estimator of the state vector α_j is a linear combination of future innovations where the weights are constructed from mean square error quantities. Harvey (1989) shows for the local level (LL) model that,

when the KF has reached a steady state, the weights decline exponentially over time. This is also true for other time series models. The larger the values for the diagonal elements of H_t, the more rapidly the decline takes place via the matrix L_t.

Fixed interval smoother

This smoother is concerned with the full sample estimator of the state vector $E[\alpha_j|Y_n,\delta]=a_{j|n}$ and its mean square error matrix $Mse[\alpha_j|Y_n,\delta]=\sigma^2 P_{j|n}$ for j=1,...,n. The recursive smoothing algorithm is given by

$$a_{j|n} = a_{j|j} + P_{j|j}T_j'P_{j+1|j}^{-1}\ (a_{j+1|n} - a_{j+1|j}) \qquad (2.3a)$$

$$P_{j|n} = P_{j|j} + P_{j|j}T_j'P_{j+1|j}^{-1}\ (P_{j+1|n} - P_{j+1|j})\ P_{j+1|j}^{-1}T_jP_{j|j} \qquad (2.3b)$$

where $a_{j|j} = a_{j|j-1} + P_{j|j-1}Z_j'F_j^{-1}v_j$ and $P_{j|j} = P_{j|j-1} - P_{j|j-1}Z_j'F_j^{-1}Z_jP_{j|j-1}$.

The fixed interval smoother (2.3) is a backwards off-line filter that starts at j=n and it terminates at j=1. Before this smoother can be applied, the KF quantities $a_{j+1|j}$, $P_{j+1|j}$, K_j, v_j and F_j must be available for j=1,...,n. Thus to obtain the full sample estimators of the state vectors, a forwards pass (Kalman filter) and a backwards pass (smoother) are required.

The derivation uses the earlier results of the fixed point smoother. Considering (2.2) gives

$$P_{j,t|t-1} = P_{j|j-1}(L_{t-1}L_{t-2}\cdots L_{j+1}L_j)' = P_{j|j-1}L_{t,j}' \qquad (2.4a)$$

$$K^*_{j,t} = P_{j|j-1}(Z_tL_{t,j})'F_t^{-1} \qquad (2.4b)$$

for j=t-1,...,1 where matrix $L_{t,j}=L_{t-1}L_{t-2}\cdots L_{j+1}L_j$, see equation (2.4.17). The matrix $L_{t,j}$ is the (t,j) block of the lower block triangular matrix L of section 2.4 and it can be evaluated recursively. The exact definition of the important matrix $L_{t,j}$ is given by

$$L_{t,j} \begin{cases} = 0 & t = 1,\ldots,j-1 \\ = I & t = j \\ = L_{t,j+1}L_j & t = j+1,\ldots,n \end{cases}$$

By repeatingly applying the former equation of (2.2b) and using the definition of (2.4b) leads to

$$a_{j|n} = a_{j|j-1} + \Sigma_{t=j}^{n} K_{j,t}^{*} v_t = a_{j|j-1} + P_{j|j-1} \Sigma_{t=j}^{n} (Z_t L_{t,j})' F_t^{-1} v_t \qquad (2.5)$$

and, similarly, $a_{j+1|n} = a_{j+1|j} + P_{j+1|j} \Sigma_{t=j+1}^{n} (Z_t L_{t,j+1})' F_t^{-1} v_t$ such that

$$\Sigma_{t=j+1}^{n} (Z_t L_{t,j+1})' F_t^{-1} v_t = P_{j+1|j}^{-1} (a_{j+1|n} - a_{j+1|j}) \ . \qquad (2.6)$$

From equation (2.5), the result (2.3a) is derived as

$$\begin{aligned} a_{j|n} &= a_{j|j} + P_{j|j-1} \Sigma_{t=j+1}^{n} (Z_t L_{t,j})' F_t^{-1} v_t \\ &= a_{j|j} + P_{j|j-1} L_j' \Sigma_{t=j+1}^{n} (Z_t L_{t,j+1})' F_t^{-1} v_t \\ &= a_{j|j} + P_{j|j-1} L_j' P_{j+1|j}^{-1} (a_{j+1|n} - a_{j+1|j}) \end{aligned} \qquad (2.7)$$

The definitions for $a_{j|j}$ and $P_{j|j}$ are obtained directly from (2.2) for t=j. It can be easily shown that $P_{j|j-1} L_j' = P_{j|j} T_j'$ such that (2.3a) follows as asserted. The recursion for $P_{j|n}$ can be derived by using the same steps of the derivation for $a_{j|n}$. An alternative proof of the fixed interval smoother (2.2) is given by Ansley & Kohn (1982) which is based on geometrical arguments.

Harvey (1989) shows for the local level (LL) model that the smoother recursion (2.3a) reduces to an exponential weighted moving average (EWMA) scheme for j=t,...,n where t is the time-point for which the KF has reached a steady state. The smoothing parameter is based on $P_{j|j} T_j' P_{j+1|j}^{-1}$.

The modified smoother

In a similar way as the Kalman filter is modified to deal with the unknown parameter vector δ, the smoothing algorithms can be adjusted as well. These modified smoothers are required to compute the smooth predictions $E[\alpha_j | Y_t]$ and their mean square error matrices $Mse[\alpha_j | Y_t]$ for t=j,...,n and 1<j<n (fixed point) or for j=1,...,n and t=n (fixed interval). As it is observed in section 2.6, the estimation results of δ do not differ whether δ is fixed or diffuse, see table 2.1.

The estimator $a_{j|t}$ is defined as $E[\alpha_j|Y_t,\delta]$. It follows directly from the results of chapter 2 that $a_{j|t}$, as defined in (2.5a) with n=t, can be rewritten as

$$a_{j|t} = a_{j|j-1} + \Sigma_{i=j}^{t} K_{j,i}^{*} v_i = A_{j|t}(1;\delta) \tag{2.8a}$$

where

$$A_{j|t} = (a_{j|t}^{o}, a_{j|t}^{+}) = A_j + \Sigma_{i=j}^{t} K_{j,i}^{*} v_i \tag{2.8b}$$

for t=j,...,n. The modified fixed point smoother replaces the recursion of $a_{j|t}$ in (2.2b) by

$$A_{j|t} = A_{j|t-1} + K_{j,t}^{*} v_t \tag{2.9}$$

for t=j,...,n. Note that $A_{j|j-1}=A_j$. The implication for this modified fixed point smoother is that an additional number of #[δ] vectors have to be calculated.

The modified fixed interval smoother follows directly from the previous results and it is given by

$$A_{j|n} = A_{j|j} + P_{j|j} T_j' P_{j+1|j}^{-1} (A_{j+1|n} - A_{j+1}) \tag{2.10}$$

such that $a_{j|n}=A_{j|n}(1;\delta)$. Indeed, equation (2.3b) remains the same. Again, the modification consists of calculating an additional number of #[δ] vectors, compare (2.3a).

The predictor $\hat{a}_{j|t}=E[\alpha_j|Y_t]$ and its mean square error matrix $\sigma^2\hat{P}_{j|t}=Mse[\alpha_j|Y_t]$ for t=j,...,n and 1<j<n (fixed point) or for j=1,...,n and t=n (fixed interval), are obtained by

$$\hat{a}_{j|t} = E[\alpha_j|Y_t] = E[A_j(1;\delta)|Y_t] = A_j(1;\hat{\delta}_t) \tag{2.11a}$$

$$\begin{aligned}\sigma^2\hat{P}_{j|t} &= Mse[\alpha_j|Y_t] = Cov[\alpha_j - \hat{a}_{j|t}] \\ &= Cov[\alpha_j - a_{j|t} + a_{j|t} - \hat{a}_{j|t}] \\ &= \sigma^2 P_{j,t} + Cov[A_{j|t}(0;\delta-\hat{\delta}_t)] = \sigma^2 P_{j|t} + a_{j|t}^{+} Mse[\hat{\delta}_t] a_{j|t}^{+\prime}\end{aligned} \tag{2.11b}$$

where the estimators regarding a fixed, random and diffuse δ are given in table 2.1, see section 2.5. These smoothed estimation

results for the classical smoothing methods are not reported earlier but they follow directly from earlier results.

In section 2.5 the initialization of the Kalman filter and the approach of collapsing the modified KF are discussed. When a collapse takes place at time r, such that matrix S_r is non-singular, the modified fixed point smoother collapses in a similar fashion at time t=r. It reduces to (2.11) for t=r and, consequently, the smoother (2.2) is applied for t=r+1,...,n. In case of the fixed interval smoother, the equation of $a_{j|n}$, see (2.3), does apply for j=n,...,1 but for j=r,...,1, the terms v_j and $a_{j|j-1}$ must be replaced by $V_j(1;\hat{\delta}_r)$ and $A_j(1;\hat{\delta}_r)$, respectively.

Computational matters

The fixed point smoother (2.2) is an on-line filter that must be applied in parallel with the Kalman filter. Even its modified form is a relatively efficient recursion and no storage is required. On the other hand, the computational and storage costs for the (modified) fixed interval smoother are severe. The main burden of the computational effort consists of the calculation of the inverse of $P_{t+1|t}$ for t=1,...,n. Since these matrices are not sparse, the number of calculations for inverting is huge. Also, it is required that the modified Kalman filter quantities A_t and P_t are stored for t=1,...,n. To avoid the re-calculation of V_t, F_t and K_t during the backwards smoothing algorithm, these quantities may be stored as well. In the case of univariate time series models, the storage requirements for the modified fixed interval smoother, where V_t, F_t and K_t are to be stored as well, consist mainly of a number of $(\#[\alpha_t]+1)$ columns of dimension $(\#[\alpha_t]+\#[\delta]+1)$ for every observation (t=1,...,n).

In the next chapters, fixed interval smoothers for the state and disturbance vectors are developed which are much more efficient because they avoid computing the inverse of the matrix P_t for t=1,...,n. Therefore, it is not useful to compare the computational requirements of these new efficient methods with the classical fixed interval smoothing algorithm as discussed in this section.

3. THE STATE SMOOTHER

As was observed in the previous section, the classical fixed interval smoother is computationally demanding. De Jong (1988b,1989) and Kohn & Ansley (1989), to be referred as DJKA, propose a more efficient fixed interval smoother. The DJKA's smoother will be referred to as the state smoother. The algorithm is based on two dummy recursions and it does not require any matrix inversion.

The aim of the state smoother is to evaluate the estimator $a_{t|n}=E[\alpha_t|Y_n,\delta]$ and its mean square error matrix $\sigma^2 P_{t|n}=Mse[\alpha_t|Y_n,\delta]$, for $t=1,\ldots,n$. It is given by the equations

$$r_{t-1} = Z_t'F_t^{-1}v_t + L_t'r_t \qquad N_{t-1} = Z_t'F_t^{-1}Z_t + L_t'N_tL_t \tag{3.1a}$$

$$a_{t|n} = a_{t|t-1} + P_{t|t-1}r_{t-1} \qquad P_{t|n} = P_{t|t-1} - P_{t|t-1}N_{t-1}P_{t|t-1} \tag{3.1b}$$

The smoother (3.1) is an off-line backwards filter which starts off at $t=n$ where vector $r_n=0$ and matrix $N_n=0$. The Kalman quantities $a_{t|t-1}$, $P_{t|t-1}$, K_t, v_t and F_t must be stored but the latter three quantities can be re-calculated as well during the backwards smoother. Indeed, the predictors in (3.1) cannot be evaluated unless the parameter vector δ is known.

The derivation of the state smoother (3.1) can be reproduced by starting from the earlier results of section 2 or by using the results of minimum mean square linear estimation directly. Both approaches are presented below.

A simple derivation of the state smoother

The derivation of the state smoother of DJKA can be given very simply by using the results of the previous section. Consider (2.5) and define

$$r_{t-1} = \Sigma_{j=t}^{n}(Z_jL_{j,t})'F_j^{-1}v_j \tag{3.2}$$

such that the equation for $a_{t|n}$ follows immediately. The weighted sum of future innovations r_t can be evaluated recursively, because

$$r_t = \Sigma_{j=t+1}^{n} (Z_j L_{j,t+1})' F_j^{-1} v_j$$

and

$$\begin{aligned} r_{t-1} &= Z_t' F_t^{-1} v_t + \Sigma_{j=t+1}^{n} (Z_j L_{j,t})' F_j^{-1} v_j \\ &= Z_t' F_t^{-1} v_t + L_t' \Sigma_{j=t+1}^{n} (Z_j L_{j,t+1})' F_j^{-1} v_j \end{aligned} \qquad (3.3)$$

such that the recursion for r_{t-1} in (3.1) follows directly. Three miscellaneous results are of importance, i.e.

$$r_n = 0 \qquad (3.4a)$$

$$\mathrm{Cov}(r_t, v_t) = 0 \qquad (3.4b)$$

$$r_{t-1} = P_{t|t-1}^{-1} (a_{t|n} - a_{t|t-1}) \qquad (3.4c)$$

Result (3.4c) follows from equation (2.5c) but also from the derived equation for $a_{t|n}$.

The mean square error matrix $\sigma^2 P_{t|n} = \mathrm{Mse}[\alpha_t | Y_n, \delta]$ is also evaluated by using a dummy recursion. Define $\sigma^2 N_t = \mathrm{Cov}(r_t)$ and from (3.4c) it follows that

$$\begin{aligned} N_{t-1} &= P_{t|t-1}^{-1} \mathrm{Cov}([\alpha_t - a_{t|t-1}] - [\alpha_t - a_{t|n}]) P_{t|t-1}^{-1} \\ &= P_{t|t-1}^{-1} (P_{t|t-1} - P_{t|n}) P_{t|t-1}^{-1} \end{aligned}$$

such that the equation for $P_{t|n}$ in (3.1b) and the initial condition $N_n = 0$ follows immediately. The recursion for N_{t-1} in (3.1a) emerges from the definition of r_{t-1} in (3.1a) and from result (3.4b) that leads to

$$\sigma^2 N_{t-1} = \mathrm{Cov}(r_{t-1}) = Z_t' F_t^{-1} \mathrm{Cov}(v_t) F_t^{-1} Z_t + L_t' \mathrm{Cov}(r_t) L_t \ . \qquad (3.6)$$

This completes the proof of the state smoother. In view of the arguments used in this derivation of the state smoother, which are very close to the arguments used by Anderson & Moore (1979), it is remarkable that this efficient state smoother is not reported earlier than the late eighties by DJKA. Even more surprisingly, this (simple) state smoother is rarely used in applied research.

General derivation of state smoother

The DJKA state smoother can also be derived by using the results of minimum mean square linear estimation. This 'stand-alone' proof is split up in four results from which the final two results give the actual proof as adopted from De Jong (1989). Before these results are given, some definitions of chapter 2 are restated here

$$x_t = \alpha_t - a_{t|t-1} \qquad L_t = T_t - K_t Z_t \qquad M_t = H_t - K_t G_t$$
$$x_{t+1} = L_t x_t + M_t \epsilon_t \qquad Cov(x_{t+1}) = \sigma^2 P_{t+1|t} = \sigma^2 (L_t P_{t|t-1} T_t' + M_t H_t')$$
$$v_t = Z_t x_t + G_t \epsilon_t \qquad Cov(v_t) = \sigma^2 F_t = \sigma^2 (Z_t P_{t|t-1} Z_t' + G_t G_t')$$

and the composite matrix $L_{t,j}$ is defined in the previous section.

Result [3.1] More general expressions for the prediction errors are given by

$$x_t = L_{t,j} x_j + \Sigma_{i=j}^{t-1} L_{t,i+1} M_i \epsilon_i$$
$$v_t = Z_t L_{t,j} x_j + Z_t \Sigma_{i=j}^{t-1} L_{t,i+1} M_i \epsilon_i + G_t \epsilon_t$$

for j=1,...,t-1. The equation for x_t follows from repeated substitution of its recursion and the equation for v_t is obtained by substituting x_t into $v_t = Z_t x_t + G_t \epsilon_t$.

Result [3.2] The covariance matrix between the two different prediction errors $Cov(x_t, v_j)$ is equal to $Cov(\alpha_t, v_j)$ and is given by

$$Cov(\alpha_t, v_j) \begin{cases} = 0 & \text{for } j = 1,\ldots,t-1 \\ = \sigma^2 P_{t|t-1} Z_t' & \text{for } j = t \\ = \sigma^2 P_{t|t-1} (Z_j L_{j,t})' & \text{for } j = t+1,\ldots,n \end{cases}$$

These covariances follow from the definition of v_t in result [3.1] and

$$Cov(\alpha_t, v_j) = Cov(\alpha_t, x_t)[Z_j L_{j,t}]' \qquad \text{for } j = t,\ldots,n$$

which lead immediately to the asserted result. Note that $L_{t,t}=I$.

Result [3.3] The state smoother is given by

$$a_{t|n} = a_{t|t-1} + P_{t|t-1}r_{t-1} \qquad P_{t|n} = P_{t|t-1} - P_{t|t-1}N_{t-1}P_{t|t-1}$$

where

$$r_{t-1} = \Sigma_{j=t}^{n}(Z_jL_{j,t})'F_j^{-1}v_j \qquad N_{t-1} = \Sigma_{j=t}^{n}(Z_jL_{j,t})'F_j^{-1}(Z_jL_{j,t})$$

and $r_n=0$ and $N_n=0$. The smoothed estimator of the state vector $a_{t|n}$ and its mean square error matrix $\sigma^2 P_{t|n}$ are derived by

$$E[\alpha_t|Y,\delta] = E[\alpha_t|Y_{t-1},v_t,\ldots,v_n] = a_{t|t-1} + \Sigma_{j=t}^{n}E[\alpha_t|v_j]$$
$$= a_{t|t-1} + \Sigma_{j=t}^{n}\mathrm{Cov}(\alpha_t,v_j)F_j^{-1}v_j = a_{t|t-1} + \Sigma_{j=t}^{n}P_{t|t-1}(Z_jL_{j,t})'F_j^{-1}v_j$$

$$\mathrm{Mse}[\alpha_t|Y,\delta] = \mathrm{Cov}(\alpha_t,\{\alpha_t-a_{t|n}\}) = \mathrm{Cov}(\alpha_t,x_t) - \mathrm{Cov}(\alpha_t,r_{t-1})P_{t|t-1}$$
$$= P_{t|t-1} - \Sigma_{j=t}^{n}\mathrm{Cov}(\alpha_t,v_j)F_j^{-1}(Z_jL_{j,t})\ P_{t|t-1}$$
$$= P_{t|t-1} - P_{t|t-1}\ \Sigma_{j=t}^{n}(Z_jL_{j,t})'F_j^{-1}(Z_jL_{j,t})\ P_{t|t-1}$$

The state smoother follows immediately. The derivation of $a_{t|n}$ is based on the uncorrelated prediction results of appendix 2A.

Result [3.4] The dummy vector r_t and dummy matrix N_t are evaluated recursively as given by (3.1a). The recursions follow from the definitions of r_t and N_t, see result [3.3], as it is shown for r_t in (3.3). It also follows directly that $\mathrm{Cov}(r_t)=\sigma^2 N_t$.

Modified state smoothing

The modified form of the state smoother computes the full sample predictions of the state vector $\hat{a}_{t|n}=E[\alpha_t|Y_n]$ and its mean square error matrix $\sigma^2\hat{P}_{t|n}=\mathrm{Mse}[\alpha_t|Y_n]$. The derivation is similar to that given in the previous section. The smoothed state is rewritten as

$$a_{t|n} = (A_t + P_tR_{t-1})(1;\delta)$$

where

$$R_{t-1} = (r_{t-1}^{o},r_{t-1}^{+}) = \Sigma_{j=t}^{n}(Z_jL_{j,t})'F_j^{-1}V_j$$

which follows straightforwardly. The matrix R_t can be evaluated recursively in a similar way by

$$R_{t-1} = Z_t' F_t^{-1} V_t + L_t' R_t \qquad (3.8)$$

and it follows that $r_t = R_t(1;\delta)$ and $R_n = 0$. The matrices V_t and A_t are obtained from the modified Kalman filter.

The smoothed predictions are given by

$$\begin{aligned} \hat{a}_{t|n} &= E[\alpha_t | Y_n] = E[(A_t + P_t R_{t-1})(1;\delta) | Y_n] \\ &= (A_t + P_t R_{t-1})(1;\hat{\delta}_n) \end{aligned} \qquad (3.9a)$$

$$\begin{aligned} Mse[\hat{a}_{t|n}] &= Mse[\alpha_t | Y_n] = Cov[\alpha_t - a_{t|n} + a_{t|n} - \hat{\alpha}_t] \\ &= \sigma^2 P_{t|n} + Cov[(A_t + P_t R_{t-1})(0;\delta - \hat{\delta})] \\ &= \sigma^2 (P_t - P_t N_{t-1} P_t) + (a_t^+ + P_t r_{t-1}^+) Mse[\hat{\delta}_n] (a_t^+ + P_t r_{t-1}^+)' \end{aligned}$$

Note that $\hat{\delta}_n = E[\delta | Y_n] = S^{-1}s$ and $Mse[\hat{\delta}_n] = Mse[\delta | Y_n] = \sigma^2 S^{-1}$ when δ is supposed to be fixed or diffuse, see section 2.6.

Computational matters

The state smoother of DJKA is a fixed interval smoother for the state vector but it is much more efficient than the classical method as described in section 2. The clearest improvement is that matrix inversions are avoided. The dummy recursions for r_t and N_t are computationally efficient. For many practical time series models, the system matrices of the SSF, such as Z_t and T_t, are sparse, see section 2.3. Computations concerning sparse matrices can be done very efficient as it is shown by the computer programs of section 2.8. Note that matrix L_t is the composite matrix $T_t - K_t Z_t$ such that the product $L_t'X$ partially involves sparse matrix computations, see also the next section.

The computational consequences of the state smoother are given in table 3.1 for state space models with different state dimensions. The computational costs and the required storage space are considered for every observation y_t which is supposed to be a scalar. Thus the number of additions and multiplications are counted for the recursions R_t and N_t and for the equations of $A_{t|n} = A_t + P_{t|t-1} R_{t-1}$ and $P_{t|n}$. The required storage capacity is measured by counting the number of elements to be stored. Two strategies regarding storing the Kalman filter quantities can be followed : (i) storing A_t and P_t

from the modified KF and re-calculating V_t, F_t and K_t; (ii) storing A_t, P_t, V_t, F_t and K_t. The latter option is considered in table 3.1. The sparsity of the system matrices are not exploited. Table 3.1 reports the computational costs of the modified state smoother as well. In this case, the number of additions and multiplications are multiplied by (1+#[δ]) with regard to the equations r_{t-1} and $a_{t|n}$. Also, the number of elements to be stored are multiplied by (1+#[δ]) with regard to v_t and $a_{t|t-1}$.

4. THE DISTURBANCE SMOOTHER

The full-sample estimator of the state vector is of interest in time series applications. In most cases where the smoothed state vectors are required, the classical fixed-interval smoothing algorithm is used to compute them. A much more efficient state smoother will be developed in section 5. The applications are often related to signal extraction and to the decomposition of time series into orthogonal components. For these cases the smoothed state vector is the important feature. However, the full-sample estimator of the disturbance vector is rarely considered. The smoothed estimator of the measurement disturbance vector $G_t\epsilon_t$ plays a similar role as the best linear unbiased estimated (BLUE) disturbance in regression analysis. The smoothed transition residuals, i.e. the full-sample estimators of $H_t\epsilon_t$ for t=1,...,n, have never been used in empirical situations. Chapter 4 will show that these transition residuals have a potential to play an important role in time series analysis. In the following, the smoothed estimator of the disturbance vector ϵ_t and its mean square error matrix are derived.

This section is organized as follows. Firstly, a simple derivation is given for only the smoothed estimators for the disturbances $G_t\epsilon_t$ and $H_t\epsilon_t$ where it is assumed that $H_tG_t'=0$. This derivation will show that the dummy vector r_t is in fact the scaled smoothed estimator of the transition residual vector $H_t\epsilon_t$. Secondly, a general derivation, where the assumption $H_tG_t'=0$ is dropped, is given of the smoothing algorithm for the disturbance vector ϵ_t and its (lagged) mean square error matrix. The results are based on

minimum mean square linear estimation arguments. Furthermore, as in the previous sections, a discussion takes place on the modified smoother and some computational matters.

The smoothed disturbance vector

In many practical time series it holds that $H_tG_t'=0$ such that both equations are uncorrelated. The results below assume this restriction.

Consider the recursion for r_t that can be split up into two parts

$$e_t = F_t^{-1}v_t - K_t'r_t \qquad r_{t-1} = Z_t'e_t + T_t'r_t \tag{4.1}$$

which follows from the definition $L_t=T_t-K_tZ_t$. It will be shown that the quantities e_t and r_t are interpreted as the scaled smoothed disturbances of the measurement and transition equation, respectively, under the assumption that $H_tG_t'=0$ which implies that $K_t=T_tP_{t|t-1}Z_t'F_t^{-1}$ and $H_tM_t'=H_tH_t'$. The smoothed measurement disturbance is given by

$$\begin{aligned} &E[G_t\epsilon_t|Y_n,\delta] = y_t - Z_ta_{t|n} - X_t\beta \\ &= v_t - Z_tP_{t|t-1}r_{t-1} = v_t - (F_t - G_tG_t')e_t - F_tK_t'r_t \\ &= F_te_t - (F_t - G_tG_t')e_t = G_tG_t'e_t \end{aligned} \tag{4.2a}$$

and the smoothed transition residual is given by

$$\begin{aligned} &E[H_t\epsilon_t|Y_n,\delta] = a_{t+1|n} - (T_ta_{t|n} + W_t\beta) \\ &= K_tv_t + P_{t+1|t}r_t - T_tP_{t|t-1}r_{t-1} \\ &= K_tv_t + P_{t+1|t}r_t - T_tP_{t|t-1}(Z_t'F_t^{-1}v_t + L_t'r_t) = H_tH_t'r_t \end{aligned} \tag{4.2b}$$

The derivations (4.2) uses the definition of $a_{t|n}$ and the definitions of (4.1). Until now the vector r_t was regarded as a dummy vector but the above result shows that it is in fact the scaled smoothed transition residual. Note that in the previous section the vector r_t is defined as a weighted sum of future innovations. Furthermore, it is shown that e_t is the scaled smoothed measurement residual. Note that the signal of the time series is given by $Z_t\alpha_t+X_t\beta=y_t-G_t\epsilon_t$ and its

estimator is given by

$$\hat{y}_t = y_t - E[G_t\epsilon_t|Y_n,\delta] = y_t - G_tG_t'e_t \qquad (4.3)$$

It is clear that this is a much more efficient method of evaluating the signal than computing the signal by $\hat{y}_t = Z_ta_{t|n} + X_t\beta$ because the calculation of the smoothed state vector is required for the latter case. Note again that all the results above hold under the assumption $H_tG_t'=0$ for $t=1,\ldots,n$.

A general proof of the disturbance smoother

The aim of the disturbance smoother is to evaluate the estimator $u_t=E[\epsilon_t|Y_n,\delta]$ and its (lagged) mean square error matrices for every time series model in state space form without the restriction $H_tG_t'=0$. The disturbance smoother equations are given by

$$u_t = G_t'F_t^{-1}v_t + M_t'r_t \qquad C_t = G_t'F_t^{-1}G_t + M_t'N_tM_t \qquad (4.4a)$$

$$C_t^* = G_t'F_t^{-1}Z_t + M_t'N_tL_t \qquad (4.4b)$$

$$r_{t-1} = Z_t'F_t^{-1}v_t + L_t'r_t \qquad N_{t-1} = Z_t'F_t^{-1}Z_t + L_t'N_tL_t \qquad (4.4c)$$

where the definitions for r_t and R_t are given by result [3.3]. The mean square error matrices $Mse[\epsilon_t|Y,\delta]$ and $Mse[\epsilon_t,\epsilon_j|Y,\delta]$ are given by

$$Mse[u_t] = \sigma^2(I - C_t) \qquad Mse[u_t,u_j] = -\sigma^2 C_t^*L_{t,j+1}M_j \qquad (4.5)$$

respectively, for $t=j-1,\ldots,1$. The required Kalman filter quantities are reduced to only v_t, F_t and K_t. Therefore, the storage space is limited considerably as compared to the state smoother. The computations can be done efficiently by re-organizing the equations appropriately, see below.

The derivation of the smoother (4.4) is based on minimum mean square linear estimation results. The recursive evaluation of r_t and N_t is derived in result [3.4]. The following two results, with some references to earlier results, forms the derivation of (4.4).

Result [4.1] The covariance matrices between the disturbance vector ϵ_t and the sequence of innovations are given by

$$\begin{aligned} & = 0 && \text{for } j = 1,\ldots,t-1 \\ \mathrm{Cov}(\epsilon_t, v_j) & = \sigma^2 G_t' && \text{for } j = t \\ & = \sigma^2 (Z_j L_{j,t+1} M_t)' && \text{for } j = t+1,\ldots,n \\ \mathrm{Cov}(\epsilon_0, v_j) & = \sigma^2 (Z_j L_{j,1} H_0)' && \text{for } j = 1,\ldots,n \end{aligned}$$

These covariances follow almost immediately by writing the innovation as in result [3.3] with s=1, i.e.

$$v_j = Z_j L_{j,1} H_0 \epsilon_0 + Z_j \Sigma_{t=1}^{j-1} L_{j,t+1} M_t \epsilon_t + G_j \epsilon_j$$

and note that $x_1 = H_0 \epsilon_0$. [ ]

Result [4.2] The smoothed estimator of the disturbance vector ϵ_t is given by

$$\begin{aligned} & u_t = E[\epsilon_t | Y, \delta] = E[\epsilon_t | v_1, \ldots, v_n] = \Sigma_{j=1}^{n} E[\epsilon_t | v_j] \\ & = \Sigma_{j=1}^{n} \mathrm{Cov}(\epsilon_t, v_j) F_j^{-1} v_j = G_t' F_t^{-1} v_t + \Sigma_{j=t+1}^{n} (Z_j L_{j,t+1} M_t)' F_j^{-1} v_j \\ & = G_t' F_t^{-1} v_t + M_t' \Sigma_{j=t+1}^{n} (Z_j L_{j,t+1})' F_j^{-1} v_j \end{aligned}$$

and leads to the former equation of (4.4a) using the definition of r_t. In a similar way it can be shown that

$$u_0 = E[\epsilon_0 | Y, \delta] = H_0' r_0$$

The mean square error matrix $\mathrm{Mse}[\epsilon_t, | Y, \delta]$ is given by

$$\begin{aligned} & \mathrm{Mse}[u_t] = \mathrm{Cov}(\epsilon_t, \epsilon_t - u_t) = \mathrm{Cov}(\epsilon_t, \epsilon_t - G_t' F_t^{-1} v_t) - \mathrm{Cov}(\epsilon_t, r_t) M_t = \\ & \sigma^2 (I - G_t' F_t^{-1} G_t) - \{\Sigma_{j=t+1}^{n} \mathrm{Cov}(\epsilon_t, v_j) F_j^{-1} (Z_j L_{j,t+1})\} M_t \\ & \sigma^2 (I - G_t' F_t^{-1} G_t) - M_t' \{\Sigma_{j=t+1}^{n} (Z_j, L_{j,t+1})' F_j^{-1} (Z_j L_{j,t+1})\} M_t \end{aligned}$$

and is evaluated by using the latter equation of (4.4a). The mean square error matrix $\mathrm{Mse}[\epsilon_t, \epsilon_s | Y, \delta]$, for s=1,...,t-1, is given by

$$\text{Mse}[u_t,u_s] = \text{Cov}(\epsilon_t,\epsilon_s-u_s) = \text{Cov}(\epsilon_t,\epsilon_s-G_s'F_s^{-1}v_s) - \text{Cov}(\epsilon_t,r_s)M_s$$
$$= -\text{Cov}(\epsilon_t,r_s)M_s = -\{\Sigma_{j=s+1}^{n}\text{Cov}(\epsilon_t,v_j)F_j^{-1}(Z_jL_{j,s+1})\}M_s$$

(here note that $\text{Cov}(\epsilon_t,v_j)=0$ for $j=1,\ldots,t-1$ and $s<t$)

$$= -\text{Cov}(\epsilon_t,v_t)F_t^{-1}(Z_tL_{t,s+1})M_s - \Sigma_{j=t+1}^{n}\text{Cov}(\epsilon_t,v_j)F_j^{-1}(Z_jL_{j,s+1})M_s$$
$$= -\sigma^2\{G_t'F_t^{-1}(Z_tL_{t,s+1}) + M_t'\Sigma_{j=t+1}^{n}(Z_jL_{j,t+1})'F_j^{-1}(Z_jL_{j,s+1})\}M_s$$
$$= -\sigma^2\{G_t'F_t^{-1}Z_t + M_t'[\Sigma_{j=t+1}^{n}(Z_jL_{j,t+1})'F_j^{-1}(Z_jL_{j,t+1})]L_t\}L_{t,s+1}M_s$$

and is evaluated using the equation for C_t^* of (4.4b). This completes the derivation of the disturbance smoother. Finally note that

$$\text{Mse}[u_0] = \sigma^2(I - H_0'N_0H_0) \qquad \text{Mse}[u_t,u_0] = -\sigma^2 C_t^*L_{t,1}H_0$$

[ ]

A simple alternative derivation of the disturbance smoother can be given when Gaussianity is assumed, see appendix 3A. It is noted that the disturbance smoother (4.4) is a backwards recursive algorithm that starts off with $r_n=0$ and $N_n=0$ and terminates at $t=1$. The various calculations can be done very efficiently. This is shown by the next result.

Result [4.3] To re-organise the equations of (4.4), define

$$e_t = F_t^{-1}v_t - K_t'r_t \qquad D_t = F_t^{-1} + K_t'N_tK_t \ . \tag{4.6}$$

and note that $\text{Cov}(e_t)=\sigma^2 D_t$. The backwards residual smoother becomes

$$u_t = J_t'b_t \qquad C_t = J_t'B_tJ_t \tag{4.7a}$$
$$C_t^* = J_t'B_tI_t \tag{4.7b}$$
$$r_{t-1} = I_t'b_t \qquad N_{t-1} = I_t'B_tI_t \tag{4.7c}$$

where

$$J_t = (G_t;H_t) \qquad I_t = (Z_t;T_t)$$
$$b_t = (e_t;r_t) = (-K_t';I)r_t + (I;0)F_t^{-1}v_t \tag{4.8a}$$
$$\sigma^2 B_t = \text{Cov}(b_t) = \text{Cov}(e_t;r_t) = \sigma^2(D_t,-K_t'N_t;-N_tK_t,N_t) \tag{4.8b}$$

It follows that $\text{Cov}(u_t,r_{t-1})=\sigma^2 C_t^*$. The equations follow directly from

the earlier results by some minor manipulation. [ ]

The efficient disturbance smoother (4.6)-(4.7) can be regarded as efficient because the main part is concerned with the system matrices of the state space formulation which are in general sparse, see the discussion on computational matters below.

Under the assumption that $H_t G_t'=0$, the same result for smoothed residuals does appear as in (4.2), but without this assumption, the stack of the measurement and transition smoothed residual becomes

$$J_t u_t = J_t J_t' (e_t; r_t) \tag{4.9}$$

where $J_t=(G_t;H_t)$. The mean square error matrix of the stack follows immediately from (4.5), for example, $Mse(J_t u_t)$ is given by

$$\sigma^2 [J_t J_t' - J_t J_t' (D_t, -K_t' N_t; -N_t K_t, N_t) J_t J_t']$$

The estimator of the signal of the time series model is given by

$$E[y_t | Y, \delta] = y_t - G_t u_t = y_t - G_t G_t' e_t - G_t H_t' r_t \tag{4.10}$$

such that (4.10) is similar to (4.3) when $H_t G_t'=0$.

Specific results of the disturbance smoother (4.6)-(4.7) are earlier recognised by DJKA. Their state smoother and related results include the recursions for r_t and N_t and they do recognise that vector e_t is the scaled measurement smoothed residual with covariance matrix $\sigma^2 D_t$, under the assumption that $H_t G_t'=0$. DJKA derive the equations for e_t and D_t by using the approach as adopted in (4.2a). However, it has never been recognised that r_t is the scaled transition smoothed residual, under the assumption $H_t G_t'=0$. These smoothed residuals have proved to be important in empirical time series modelling, for example, in detecting irregularities in time series such as outliers and structural changes, see Chapter 4. Also, the (lagged) mean square error matrices of the residual vector have not been derived earlier. Other implications of the disturbance smoother results are discussed in the next sections.

Modified disturbance smoothing

The modified disturbance smoother computes the predictions $\hat{u}_t=E[\epsilon_t|Y_n]$ and the associated mean square error matrices. Therefore, the disturbance smoother must be adjusted in a similar way as the modified Kalman filter. In the previous section it is shown how the recursion for r_t can be modified, to deal with the parameter vector δ, by replacing it with R_t such that $r_t=R_t(1;\delta)$. The next equations are required in addition,

$$E_t = F_t^{-1}V_t - K_t'R_t \qquad E_t = (e_t^o, e_t^+)$$
$$U_t = J_t'(E_t;R_t) \qquad U_t = (u_t^o, u_t^+)$$
$$R_{t-1} = I_t'(E_t;R_t)$$

such that $e_t=E_t(1;\delta)$ and $u_t=U_t(1;\delta)$. The modifications follow directly from the equations (4.7) and (4.8) and with replacing v_t by $V_t(1;\delta)$. The equations for C_t and C_t^* remain the same.

The smoothed prediction quantities are given by

$$\hat{u}_t = U_t(1;S^{-1}s) \tag{4.12a}$$
$$Mse[\hat{u}_t] = \sigma^2(I - C_t + u_t^+S^{-1}u_t^{+\prime}) \tag{4.12b}$$
$$Mse[\hat{u}_t,\hat{u}_j] = \sigma^2(-C_t^*L_{t,j+1}M_j + u_t^+S^{-1}u_j^{+\prime}) \tag{4.12c}$$

for t=1,...,n and j=1,...,t-1. These results are valid when δ is supposed to be fixed or diffuse. If δ is supposed to be random, S^{-1} must be replaced by $(\Lambda^{-1}+S)^{-1}$ and s must be replaced by $(\Lambda^{-1}\mu+s)$, see table 2.1. The derivation is given by

$$\hat{u}_t = E[\epsilon_t|Y] = E\{E[\epsilon_t|Y,\delta]|Y\} = E[U_t(1;\delta)|Y] = U_t(1;E[\delta|Y])$$

$$Mse[\hat{u}_t,\hat{u}_j] = Mse[\epsilon_t,\epsilon_j|Y] = Cov(\epsilon_t-u_t+u_t-\hat{u}_t,\epsilon_j-u_j+u_j-\hat{u}_j)$$
$$= Mse[u_t,u_j] + Cov[U_t(0;\delta-\hat{\delta}),U_j(0;\delta-\hat{\delta})]$$
$$= Mse[u_t,u_j] + u_t^+Mse[\delta|Y]u_j^{+\prime}$$

for j=1,...,t. The smoothed predictions are derived as asserted.

computational matters

The computational costs for the disturbance smoother is much less compared with the state smoother because the former does not require the storage and the calculations concerning the Kalman quantities $a_{t|t-1}=A_t(1;\delta)$ and $P_{t|t-1}$. The disturbance smoother, as represented by equations (4.7) and (4.8), has the nice property that most computations only involve system matrices which are usually sparse. Moreover, they also depend on the dimension of the disturbance vector which is in many cases less than the dimension of the state vector.

The computational consequences of the disturbance smoother are presented in table 3.2 in a similar fashion as in table 3.1. Again, the number of additions and multiplications are counted for the various smoothing equations. Also, the required storage space, concerning the Kalman quantities V_t, F_t and K_t, is measured by counting the number of elements to be stored. The sparsity of the system matrices is not taken into account. The additional computational costs for some equations, when the modified smoother is applied, is proportional to the dimension of $(1;\delta)$.

5. A NEW EFFICIENT METHOD FOR STATE SMOOTHING

It is discussed earlier that the state smoother of section 3 is important for various features in time series analysis. For example, the estimated unobserved components of a structural time series model are extracted by a state smoothing algorithm. This section presents a new efficient method for state smoothing which is based on the results of the previous section. The new smoother, to be referred as the K smoother, is, compared with the state smoother of DJKA (see section 3), computationally very efficient for many practical time series models. The gains only correspond to the smoothed state vector, the new results do not consider the mean square error matrix of the smoothed state.

This section discusses also a very different smoothing approach based on a chalenging article of Whittle (1991). This approach suggests a smoothing algorithm that requires no storage of

any Kalman filter quantity at all. It is disappointing that a straightforward implementation of this method is not numerically stable and cannot be used in practice. This section shows that an adjustment of Whittle's smoother leads to the same method of state smoothing as discussed in this section.

Computational matters are also considered in this section and it will be shown that the new method of state smoothing is much more efficient and, therefore, faster than the state smoother of section 3 for various time series models and, in particular, for structural time series models.

Efficient state smoothing

The new state smoother is based on the idea that the transition equation provides an approach of state smoothing as well because

$$a_{t+1|n} = E[\alpha_{t+1}|Y,\delta] = T_t E[\alpha_t|Y,\delta] + W_t\beta + H_t E[\epsilon_t|Y,\delta] \qquad (5.1a)$$

$$= T_t a_{t|n} + W_t\beta + H_t u_t \qquad (5.1b)$$

where $a_{1|n} = E[\alpha_1|Y,\delta] = W_0\beta + H_0E[\epsilon_0|Y,\delta] = W_0\beta + H_0u_0$. The sequence of smoothed residuals is obtained from the efficient disturbance smoother of the previous section. Thus, the usual KF and the usual disturbance smoother are required including the storage of v_t, F_t and K_t. It will be argued below that no additional storage space is required.

In practical applications, since δ is generally unknown, the modified form of the state smoother (5.1) must be applied. The smoothed prediction of the state vector is defined as $\hat{a}_{t|n}=E[\alpha_t|Y]$ and it follows that

$$\hat{a}_{t+1|n} = T_t\hat{a}_{t|n} + W_t\hat{\beta} + H_t\hat{u}_t \qquad (5.2)$$

where $\hat{\beta}=b+B\hat{\delta}$ and $\hat{a}_{1|n}=W_0\hat{\beta}+H_0\hat{u}_0$ for $t=0,\ldots,n$. The forwards recursion (5.2) is in fact the new method of smoothing. Its derivation is very simple and it follows immediately. The 'input' of this recursion consists of $\hat{\delta}$ and $\hat{u}_t$ for $t=0,\ldots,n$. As it is observed in the previous chapter, the estimator $\hat{\delta}$ is a result of the modified KF. The

smoothed predictions $\hat{u}_t$ are obtained from the backwards residual smoother as given by section 4. Thus the proposed K smoother is a three-step method :

1. The modified Kalman filter (a forwards pass)
2. The disturbance smoother (a backwards pass)
3. The state recursion (5.1) (a forwards pass)

Note the following three remarks. (i) The modified KF must be applied primarily because the full-sample estimator of δ ($\hat{\delta}=S^{-1}s$) is required. (ii) The disturbance smoother can keep its usual (non-modified) form, as presented by (4.7) and (4.8), except that v_t must be replaced by $V_t(1;S^{-1}s)$ where V_t is obtained from the modified KF. This specific disturbance smoother give the sequences $\hat{e}_t$, $\hat{r}_t$ and $\hat{u}_t=G_t'\hat{e}_t+H_t'\hat{r}_t$ for $t=0,\ldots,n$. The modified smoother quantities E_t and R_t, which form together $U_t=G_t'E_t+H_t'R_t$, do not have to be computed because the mean square error matrix $Mse[\hat{u}_t]$ is not required. (iii) Finally, the initialization for (5.2) is computed by using the estimator $\hat{\delta}$.

Whittle's approach of smoothing

Whittle (1991) uses path integrals as a flexible framework for estimation and prediction in time series analysis. The attraction of his approach is that the optimization criterions (represented by an integral) can be maximized freely without constraints. This direct approach of estimation is a very non-classical approach in statistical inference but it has a potential to become very important in future, see Whittle (1991, including the remarks of the participants in the discussion). One application of this approach (although it has nothing to do with path integrals) shows that considering the likelihood function directly leads to attractive results with regards to smoothing. However, the smoothing results are limited to the smoothed state and disturbance vector, expressions for mean square error matrices are not obtained. The result of Whittle (1991, section 11) for a local level (or random walk plus noise) model is generalised below for any time series model in state space form.

It is assumed, for simplicity, that $H_tG_t'=0$, $(X_t;W_t)=0$ and $\sigma^2=1$ for $t=1,\ldots,n$. Furthermore, it is assumed that the initial condition is known and fixed. These restrictions can be relaxed straightforwardly. The logarithm of the joint density function can be given directly in terms of the state space model, i.e.

$$\Sigma_{t=1}^{n}\{\ \log|G_tG_t'| + (y_t-Z_t\alpha_t)'(G_tG_t')^{-1}(y_t-Z_t\alpha_t) + \log|H_tH_t'| + (\alpha_{t+1}-T_t\alpha_t)'(H_tH_t')^{-1}(\alpha_{t+1}-T_t\alpha_t)\ \}$$

Under appropriate regularity conditions (Gaussianity, etc), the maximum of this log-density with regards to the sequence of state vectors α_t, for $t=1,\ldots,n$, gives the minimum mean square (full sample) linear estimator of α_t. By taking the first derivative with respect to α_t and restricting it to zero, the 'first-order' condition is obtained and given by

$$(H_{t-1}H_{t-1}')^{-}(\alpha_t-T_{t-1}\alpha_{t-1}) = Z_t'(G_tG_t')^{-}(y_t-Z_t\alpha_t) + T_t'(H_tH_t')^{-}(\alpha_{t+1}-T_t\alpha_t)$$

for $t=1,\ldots,n$. The solution of this system of restrictions is denoted by $\hat{\alpha}_t$. Now, define $\hat{e}_t=(G_tG_t')^{-}(y_t-Z_t\hat{\alpha}_t)$ and $\hat{r}_t=(H_tH_t')^{-}(\hat{\alpha}_{t+1}-T_t\hat{\alpha}_t)$ such that the sequence of first order conditions can be represented by

$$\hat{r}_{t-1} = Z_t'\hat{e}_t + T_t'\hat{r}_t$$

for $t=1,\ldots,n$. This result is equivalent to the recursion (4.7c). The vectors $\hat{e}_t$ and $\hat{r}_t$ are directly defined as the scaled smoothed measurement and transition residual, respectively. Note that it is assumed that $H_tG_t'=0$. The definitions for $\hat{e}_t$ and $\hat{r}_t$ leads to a method to evaluate these quantities plus the smoothed prediction of the state. Thus from the definition for $\hat{r}_t$ it follows that

$$\hat{\alpha}_t = T_t^{-1}(\hat{\alpha}_{t+1} - H_tH_t'\hat{r}_t)$$

such that $\hat{e}_t$ can be calculated by its definition and, consequently, $\hat{r}_{t-1}$ is computed using its recursion. This leads to a backwards recursive algorithm ($t=n,\ldots,1$) that is able to evaluate the

smoothed state vector without using any KF quantity except the state estimate $\hat{a}_{n+1|n}$. This one-step ahead prediction of the state is required to initialize the recursion by computing $\hat{\alpha}_n$. The inverse of T_t does exists in many applications but, in general, this may cause problems and T_t^{-1} must be replaced by a generalized inverse. Finally, note again that $\hat{r}_n=0$.

This method of smoothing is, at first sight, a revolutionary result that allows efficient smoothing without storage. It is even more an attractive result because the derivation has used very simple arguments. This contributes to Whittle's approach of estimation and prediction in time series analysis by having a direct appeal to the optimization criterion. It is unfortunate that a straightforward implementation of Whittle's smoother is numerically unstable. The reason is that the same filter can be derived in a reverse time order and, therefore, the roots of the polynomial cannot be solved uniquely. This problem will be illustrated for the local level (LL) model in chapter 4. The unstability can be avoided by restricting the smoother at its begin- and end-points as Whittle puts forward in the discussion with Harvey, see Whittle (1991). Some details has been given by Whittle in a personal correspondence but the required adjustments leads to the Kalman filter approach as given in section 4.

A straightforward suggestion to make Whittle's smoother stable will end in the efficient state smoother as described earlier in this section. The quantity $\hat{e}_t$ must not be computed using its definition but must be replaced by $\hat{e}_t=F_t^{-1}\hat{v}_t-K_t'\hat{r}_t$ as given in section 3 and 4. Furthermore, the backwards recursion for $\hat{\alpha}_t$ can be easily reversed to a forwards recursion such that the inverse of T_t is not required. This solution requires the storage space as discussed earlier.

Computational matters

Before the computational costs of the state smoothers are compared, more details are given of the DJKA state smoother and the K state smoother (5.1). Both approaches can be carried out in its usual form or in its modified form.

DJKA's smoother applies the Kalman filter and, alongside, stores the state vector $a_{t|t-1}$ and its mean square error matrix $P_{t|t-1}$. In addition, v_t, F_t and K_t may be stored to avoid re-calculation. The KF is followed by the backwards recursion for the state vector as given by

$$u_t = F_t^{-1}v_t - K_t'r_t \qquad r_{t-1} = Z_t'u_t + T_t'r_t$$
$$a_{t|n} = a_{t|t-1} + P_{t|t-1}r_{t-1}$$

where all the KF quantities are available or have to be re-calculated. The modified form of the DJKA's smoother requires the modified KF (with storage) and the backwards smoother as given by

$$\hat{u}_t = F_t^{-1}V_t(1;\hat{\delta}) - K_t'\hat{r}_t \qquad \hat{r}_{t-1} = Z_t'\hat{u}_t + T_t'\hat{r}_t$$
$$\hat{a}_{t|n} = A_t(1;\hat{\delta}) + P_t\hat{r}_{t-1}$$

such that the matrices V_t and A_t must be stored in stead of the vectors v_t and $a_{t|t-1}$, respectively.

The proposed smoothing method of this section also requires the Kalman filter but only the quantities K_t, v_t and F_t have to be stored. In general, the dimensions of these quantities are much less compared to $a_{t|t-1}$ and $P_{t|t-1}$. The second step is the backwards recursion for u_t and r_t as given above. The vector r_t can be stored by using the space of K_t. Thus, the KF quantity K_t is replaced by r_t during the backwards recursion and no extra storage space is required. The dimensions for K_t and r_t are the same for univariate time series models but, in general, the required storage space for K_t is always larger than for r_t. If necessary, the storage for v_t can be replaced by e_t. These quantities always have the same dimension. Finally, the forwards recursion (5.2) is applied. Note that the procedure for the modified form of the K state smoother is similar but it requires the storage of V_t in stead of v_t during the forwards KF.

Table 3.3 considers four different models : the local level model, the local linear trend model, the basic structural time series model and a time-varying regression model. Section 2.3 discusses these models and gives their state space formulation. In

table 3.3.a the additions and multiplications are counted for one recursion of r_t. The same is done for the DJKA smoother and the recursion (5.1). The table 3.3.b does report the same but for the modified form. The number of elements to be saved for the recursion r_t (in its usual and modofied) form is reported in table 3.3.c. Also, the extra required saving space for the DJKA smoother and the recursion (5.1) is given. For the DJKA smoother it is supposed that the quantities K_t, v_t and Ft are stored as well such that they do not have to be re-calculated. It is concluded from table 3.3 that the K smoother outperforms the DJKA smoother for many time series models.

6. A NEW EM ALGORITHM FOR ESTIMATING COVARIANCES OF THE SSF

The estimation of the hyper parameters of the state space formulation can be done by a scoring method where the KF is used to calculate the estimation criterion, see section 2.6. An alternative method is proposed by Watson & Engle (1983) and Shumway & Stoffer (1982), to be referred to as WESS. Their method is based on the EM algorithm and it is explored for use in the context of time series models. The way in which the EM algorithm can be applied to hyper-parameter estimation for structural time series models is described by Harvey & Peters (1990). They report that the EM algorithm tended to be very time consuming due to slow convergence. Although it has to be said that the EM algorithm performs well in the context of multivariate data analysis, in the case of estimation of a time series model, it is advocated to adopt a mixture of EM (to get close to the maximum of the likelihood criterion) and a scoring method (to get quickly to the optimum), see Watson & Engle (1983).

In this section it is shown that the computational effort for the EM algorithm can be reduced considerably when the hyper parameters are only found in the covariance matrix $(G_t;H_t)$. This new result is based on the argument that the state vector is a cumulative sum of the initial state plus the disturbances. The derivation of the improved EM algorithm is based on the smoothing results of section 3.3. It is assumed that the covariance matrix is time invariant such that $(G_t;H_t)=(G;H)$. For simplicity reasons,

additional assumptions are $HR'=0$, $(X_t;W_t)=0$ for $t=1,\ldots,n$ and $B=0$ such that the usual KF can be applied which is initialized by the big k method, see section 2.5. The latter three assumptions can be relaxed without violating the arguments used below.

The EM algorithm for time series models

The EM algorithm is derived by setting up the likelihood directly in terms of the state space formulation in a similar way as Whittle (1991) does to derive his estimation methodology, see section 3.5. The expectation step (E-step) is to evaluate the likelihood conditional on the set of observations, i.e. $E[L[y]|Y]$ what reduces, for time series models, to smoothing. The maximization step (M-step) obtains the estimates of the hyper parameters by taking first and (approximated) second derivatives from the 'expected' likelihood in order to get the normal equations.

Alternatively, by using more intuitive arguments, the EM algorithm can be derived as follows. Given the identities

$$\sigma^2 GG' = \mathrm{Cov}(y_t - Z_t\alpha_t) \qquad \sigma^2 HH' = \mathrm{Cov}(\alpha_{t+1} - T_t\alpha_t)$$

and given the set of observations $Y=\{y_1,\ldots,y_n\}$, the estimators for GG' and HH' are derived by, respectively,

$$\begin{aligned} GG' &= n^{-1}\Sigma_{t=1}^{n}\ \sigma^{-2}\mathrm{Cov}[\ (y_t-Z_ta_{t|n}) - Z_t(\alpha_t-a_{t|n})\ ] \\ &= n^{-1}\Sigma_{t=1}^{n}\ [\ \sigma^{-2}(y_t-Z_ta_{t|n})(y_t-Z_ta_{t|n})' + Z_tP_{t|n}Z_t'\ ] \end{aligned}$$

$$\begin{aligned} HH' &= n^{-1}\Sigma_{t=1}^{n}\ \sigma^{-2}\mathrm{Cov}[\ (a_{t+1|n}-T_ta_{t|n}) + (\alpha_{t+1}-a_{t+1|n}) - T_t(\alpha_t-a_{t|n})\ ] \\ &= n^{-1}\Sigma_{t=1}^{n}\ [\ \sigma^{-2}(a_{t+1|n}-T_ta_{t|n})(a_{t+1|n}-T_ta_{t|n})' \\ &\quad + P_{t+1|n} - T_tP_{t,t+1|n} - P_{t+1,t|n}T_t' + T_tP_{t|n}T_t'\ ] \end{aligned}$$

Note that these estimators can be regarded as maximum likelihood estimators, see WESS.

The computational consequences of the EM algorithm are severe. The main computational burden is that a fixed interval smoothing algorithm must be applied to obtain the smoothed state vector $a_{t|n}$ and its mean square error matrix $P_{t|n}$ for $t=1,\ldots,n$. Of course, the state smoother of DJKA, see section 3.3, is preferred

rather than the classical smoothing method. However, even the DJKA state smoother requires a huge amount of storage and many computational operations, see table 3.1. Another implication of the EM algorithm is that the cross mean square error matrix $P_{t+1,t|n}$ is required. Watson & Engle (1983) propose to augment the state vector for the Kalman filter and the state smoother by $(\alpha_{t+1};\alpha_t)$ such that the cross mean square error matrix $P_{t+1,t|n}$ is obtained automatically. The number of extra computations and the increase in storage costs are huge. A special smoother for $P_{t+1,t|n}$ is developed by Shumway & Stoffer (1982). An improved (more efficient) algorithm for $P_{t+1,t|n}$ is proposed by De Jong & MacKinnon (1988). However, all suggestions imply a severe increase in computational costs.

An improved EM algorithm

The new EM algorithm is more efficient mainly because it avoids the state smoother plus the algorithm for $P_{t+1,t|n}$ and, therefore, the huge computational and storage requirements. Instead, the improved EM algorithm applies the disturbance smoother of section 3.3 which is, for many time series models, more efficient than the state smoother. Without giving facts, it is clear that the approach below reduces the computational costs for the EM algorithm considerably. Note that the new results only apply to estimating hyperparameters in the covariance matrices of the SSF.

The arguments for the derivation of the new EM algorithm are similar as before. Given the identities

$$\sigma^2 GG' = \mathrm{Cov}(G\epsilon_t) \qquad \sigma^2 HH' = \mathrm{Cov}(H\epsilon_t)$$

and given the set of observation Y, it follows that

$$\begin{aligned} GG' &= n^{-1}\Sigma_{t=1}^{n}\ \sigma^{-2}\mathrm{Cov}[\ Gu_t - G(\epsilon_t - u_t)\ ] \\ &= n^{-1}\Sigma_{t=1}^{n}\ \sigma^{-2}\mathrm{Cov}[\ GG'e_t - G(\epsilon_t - u_t)\ ] \\ &= GG' + GG'\{\ n^{-1}\Sigma_{t=1}^{n}\ [\ \sigma^{-2}e_te_t' - D_t\ ]\ \}GG' \end{aligned}$$

$$\begin{aligned} HH' &= n^{-1}\Sigma_{t=1}^{n}\ \sigma^{-2}\mathrm{Cov}[\ Hu_t - H(\epsilon_t - u_t)\ ] \\ &= n^{-1}\Sigma_{t=1}^{n}\ \sigma^{-2}\mathrm{Cov}[\ HH'r_t - H(\epsilon_t - u_t)\ ] \\ &= HH' + HH'\{\ n^{-1}\Sigma_{t=1}^{n}\ [\ \sigma^{-2}r_tr_t' - N_t\ ]\ \}HH' \end{aligned}$$

Note that from section 3.3 it follows that $GC_tG'=GG'D_tGG'$ and $HC_tH'=HH'N_tHH'$ under the assumption that $HG'=0$. These estimators are appealing and appear to be recursive in a natural way. The maximum likelihood estimators are found if the summation term within the curly brackets is close to a zero matrix. This discrepancy function is expected to close to zero for theoretical reasons. The outline of the EM algorithm is given in table 3.4.

The computational attraction of the new EM algorithm is due to the efficiency of the disturbance smoother but also to the fact that only the disturbance smoother is required. So, a matrix like $P_{t,t+1|n}$ is not required. This emphasizes the importance of the derived results in section 3.

7. PRACTICAL SMOOTHING

In section 2.6 a structural time series model is fitted to the COAL series. In this section, the same time series is considered again but now it is shown what specific role the smoothing results may play in applied time series analysis.

Signal extraction

The signal of y_t is defined as $Z_t\alpha_t + X_t\beta$ and its estimator is evaluated by the disturbance smoother that also gives (lagged) mean square error quantities, e.g.

$$\hat{y}_t = y_t - G_t\hat{u}_t \qquad \text{Mse}(\hat{y}_t) = G_t\text{Cov}(\hat{u}_t)G_t'$$
$$\text{Mse}(\hat{y}_t,\hat{y}_j) = G_t\text{Cov}(\hat{u}_t,\hat{u}_j)G_j'$$

The plot of the signal together with the actual series is informative and it may help the time series analysist to check if the model describes the time series satisfactory. The signal of the COAL series is presented in figure 3.1.

Full-sample estimates of the unobservables

Full sample estimates of the unobserved components trend, seasonal and irregular are obtained from a state smoother such as

the efficient K smoother, see section 5. In case of unobserved components models, the full-sample estimates of the unobservables are an important diagnostic in the process of model selection. When the estimated unobservable components appear to be different as one might expect from theoretical arguments, the particular model specification must be reconsidered.

In the case of the COAL series, the extracted components of the BSM reflect the features of the time series as discussed in section 2.7. The estimated trend component in figure 3.2 moves clearly downwards between 1975 and 1986. As expected, the seasonal pattern in figure 3.3 does not change dramatically over the years. The shock movements do appear much clearer in the estimated irregular than in the standardised innovations, compare the figures 2.4 and 3.4. When the confidence intervals for the estimated irregular are plotted as well, possible interventions for outliers can be selected easily. A further discussion on outliers takes place in the next chapter.

Theoretical auto- and cross correlations of the residuals

The class of structural time series models, discussed in section 2.3, all assume that $G_t G_t'=1$, $H_t G_t'=0$ and $H_t H_t'$ is a diagonal matrix which is not neccessarily of full rank such that zeroes may be found on the diagonal. In other words, the unobserved components (trend, seasonal, cycle, irregular) are assummed to be orthogonal and, therefore, the associated disturbances are uncorrelated. It is a misunderstanding that the set of estimated disturbances, to be referred to as the auxiliary residuals, are expected to be uncorrelated as well. It can be shown by using classical results of signal extraction, see for example Maravall (1987) and explored in the context of structural time series models in the next chapter (section 4.2), that in a double infinite sample, the auto- and cross-correlation values for auxiliary residuals can be quite high. This is also true for finite samples as will be shown below.

The theoretical auto- and cross-correlation functions for the auxiliary residuals of any structural time series model can be calculated using the recursion for the covariance matrix $\sigma^2 C_t$ and the (t-j) lagged covariance matrix $\sigma^2 C_t^* L_{t,j+1} M_j$ of equation (4.4). These

quantities are time-varying but for time-invariant time series models they are approximately constant in the middle of the series. This is shown for the variance of the measurement residual of a local level (LL) model with different signal-to-noise ratios q in figure 3.5. Define $c_{0,ik}$ as the (i,k) element of the covariance matrix $\sigma^2 C_t$ where t is close to $\frac{1}{2}n$. Similarly, define $c_{t-j,ik}$ as the (i,k) element of the lagged covariance matrix $\sigma^2 C_t^* L_{t,j+1} M_j$ where t is close to $\frac{1}{2}n$, j=t-1,...,t-P and P can be any moderate integer value between 0 and $\frac{1}{2}n$. For example, it follows directly that the auto-correlation value at lag m for the i-th auxiliary residual is given by $c_{m,ii}/c_{0,ii}$.

Studies are at present in the literature where the existence of auto- and cross-correlation in auxiliary residuals are a source of concern for the validity of the decomposition of time series into orthogonal components, see Garcia-Ferrer & del Hoyo (1991). Although they are aware that auto- and cross-correlations may exist in the set of auxiliary residuals, they 'expect' it to be rather low in a well-specified model. Table 3.6 shows that, for different signal-to-noise ratio values, the theoretical auto- and cross correlation values of the auxiliary residuals can be relatively high. The auto- and cross correlogram of the auxiliary residuals may be compared with their theoretical counterparts.

Harvey, in his comment on Garcia-Ferrer & del Hoyo (1991), makes it perfectly clear that 'large' auto- and cross-correlograms for the auxiliary residuals do not tell us anything about the validity of the assumptions underlying a structural time series model. Misspecification is related to the ARIMA representation (reduced form). A structural time series model is valid by construction when it is given that the implied reduced form is consistent with the underlying data generation process. Of course, other decompositions can be constructed as well such as, for example, the canonical decomposition, see Pierce (1979) and Hillmer & Tiao (1982). The question of which decomposition is appropriate can only be solved by an appeal to theoretical arguments concerning the type of properties one wishes unobserved components to possess, see Harvey (1989, Chapter 6).

It is concluded that the auxiliary residuals are serially correlated even for a correctly specified model with known

parameters, see table 3.6. One may be alerted to misspecification by the fact that the correlograms of the auxiliary residuals are very different to their implied auto-correlations. It remains to be seen if this diagnostic check is worthwhile for detecting model misspecification as Maravall (1987) has suggested. In the case of the COAL series, table 3.5 reports theoretical auto-correlations of the set of 4 auxiliary residuals for some lags together with the associated correlogram values. They all match relatively close.

The next chapter will develop a new methodology of tracing irregularities in time series (outliers and structural change) mainly by checking the plots of the auxiliary residuals and applying some simple test diagnostics to these residuals. The required tools for this approach are all developed in this chapter. This emphasizes again the importance of the new smoothing results for empirical time series analysis.

8. COMPUTER PROGRAMS

This section presents four additional procedures which concern the computations discussed in this chapter. All procedures can be placed in the main program of section 2.8. The actual code of the program is given the end of this thesis. The procedures can be adjusted in a flexible way. Again, it is recommended to keep the main structure of the program. Note that the main program only deals with univariate time series models such that the observation, the innovation and their covariances are scalars and the Kalman gain is a vector.

SmoothUpdate

The SmoothUpdate procedure updates the matrices R_t and N_t and calculates the vector E_t, the scalar D_t and the vector $-N_tK_t$, given the Kalman quantities V_t, F_t and K_t. The appropriate equations can be found in section 4, equations (4.6)-(4.8). From these matrices the smoothing computations, as discussed in this section, can be done straightforwardly. The heading of the procedure SmoothUpdate is

```
SmoothUpdate(     DimKF              : Integer;
                Var DSE,DSD,DSU,DSC  : Matrix);
```

The dimension of the parameter vector δ for the modified KF update is given by DimKF. Note that if DimKF is equal to zero, the modified disturbance smoother reduces to the usual smoother. The structure of the matrices DSE and DSD are different at start and at finish of the procedure, see table 3.7. The matrices DSU and DSC contain the information for the smoothed disturbance vector $u_t=U_t(1;\delta)$ and its covariance matrix C_t, respectively. All computations concerning covariances can be dropped when the KF has been in a steady state.

The structure of the SmoothUpdate procedure consists of the following five steps

1. $DSE[0,0..DimKF] = V_tF_t^{-1} - K_t'R_t$
   $DSD[0,0] = F_t^{-1} + K_t'N_tK_t$
   $DSD[0,1..DimSt] = -K_t'N_t$ $\qquad$ $DSD[1..DimSt,0] = -N_tK_t$
2. $DSU = (G_t;H_t)(G_t;H_t)'DSE$ $\qquad$ $M_2 = (G_t;H_t)(G_t;H_t)'DSD$
3. $DSC = (G_t;H_t)(G_t;H_t)'M_2'$
4. $M_1 = (Z_t;T_t)'DSE$ $\qquad$ $M_2 = (Z_t;T_t)'DSD$
5. $DSE = (E_t;M_1)$ $\qquad$ $DSD = (Z_t;T_t)'M_2'$

AuxResiduals

This procedure gives a sequence of auxiliary residuals, associated with the measurement equation or with a specific element of the state vector, and the sequence of mean square errors for $t=1,\ldots,n$. The heading is

```
AuxResiduals(     Nr     : Integer;
                  Q      : Matrix;
                Var r,vr : Data);
```

where Nr indicates which residual is requested. When Nr equals zero, the measurement residual is requested. The Data-types r and vr consist of the auxiliary residuals and their mean square errors, respectively. Before this routine can be called, the procedure Estimation must be applied to store the Kalman filter quantities V_t, F_t and K_t. Also the matrix Q of the procedure Estimation must be

available. In fact, this procedure is the implementation of equation (4.12a) and the implementation of the diagonal elements of the mean square error matrix (4.12b). The structure of this procedure is straightforward and speaks for itself.

AuxCrossMse

This procedure gives the cross mean square error values between two requested auxiliary residuals for t=1,...,n. The heading is

```
AuxCrossMse(     Nr1,Nr2      : Integer;
                 Q            : Matrix;
            Var cc            : Data);
```

The structure is very similar to the previous procedure and it is in fact the implementation of the non-diagonal elements of the mean square matrix of equation (4.12b). The requirements of the procedure AuxResiduals must be valid for this procedure as well.

AuxCrossCorrFnc

This procedure runs the smoother up to the middle of the series and calculates from there the cross correlation function between two selected auxiliary residuals for a given maximum lag. The heading is

```
AuxCrossCorrFnc(     Nr1,Nr2    : Integer;
                     Q          : Matrix;
                     MaxLag     : Integer;
                Var ccf         : Vector);
```

This procedure can be regarded as a way of obtaining useful information from equation (4.12c). The requirements are the same as for the previous two procedures.

9. CONCLUDING REMARKS

This chapter has developed a disturbance smoother for the SSF which produces the full sample estimates of the disturbance vector (auxiliary residuals) and its mean square error matrix. The auxiliary residuals can be very important in practical time series modelling, see the next chapter. Other implications of the new smoother is a very efficient state smoother for the SSF which produces the full sample estimates of the state vector. There are clear computational improvements for most practical time series models compared with existing smoothers. Also, the new results lead to a new EM algorithm for estimating hyperparameters in the covariance structure of the SSF.

APPENDIX 3A
Smoothing under Gaussianity

A straightforward approach to derive the disturbance smoother under Gaussianity is to use the 'stack-argument'. This implies that the SSF is formulated by the multivariate regression model (2.2.6) where it is assumed that

$$\epsilon \sim N(0 , \sigma^2 I)$$

The smoothed disturbance vector is defined as $u = E[\epsilon|Y,\delta]$. It holds as well that $u = E[\epsilon|v]$ because the innovations v are a full-rank linear combination of the set of observations Y and the parameter vector δ. In order to obtain the conditional distribution of ϵ conditional on v, the joint distribution of $(\epsilon;v)$ is required. The classical multivariate results provide the appropriate formulas to get the conditional distribution of $\epsilon|v$ from the joint distribution $(\epsilon;v)$, see any introductory book on multivariate statistical theory or Harvey (p.165, 1989).

It is shown in result 2.4 that the stack of the innovations $v=(v_1;\ldots;v_n)$ can be expressed as a linear combination of the disturbance vector

$$v = J\epsilon$$

where matrix $J = ZLM + G$ is lower block triangular with on the main diagonal blocks the sequence G_t. Note that $Cov(v)=\sigma^2F=\sigma^2JJ'$ is a block diagonal matrix. The definitions of the matrices Z, L, M and G can be found in sections 2.2 and 2.4. The joint distribution of the stacked vector $(\epsilon;v)$ is given by

$$(\epsilon;v) \sim N[(0;0) , \sigma^2(I,J';J,F)]$$

It follows that

$$\epsilon|v \sim N[\ U'F^{-1}v\ ,\ \sigma^2U'F^{-1}U\ ]$$
$$\sim N[\ u\ ,\ \sigma^2C\ ]$$

Taking into account the special structure of matrix J, it follows that

$$u = G'F^{-1}v + M'r \qquad\qquad r = Z'L'F^{-1}v$$

Note that the matrix L is lower block triangular with on the main diagonal identity matrices. The matrices G, M and Z are block diagonal. By exploiting the structure of these matrices, the equations for u and r lead to the following recursion which is recogised earlier

$$u_t = G_t'F_t^{-1}v_t + M_t'r_t \qquad r_{t-1} = Z_t'F_t^{-1}v_t + L_t'r_t$$

The recursion for the covariance matrix of u_t can be obtained in a similar way and are given by equation (3.4.4). All smoothing results of section 3.3 can be obtained by using this approach, see Koopman (1991). The state smoother can be derived under Gaussianity using the 'stack-argument' as well, see De Jong (1988).

TABLES & FIGURES

Table 3.1

Computational consequences of state smoother (#[y_t]=1 #[α_t]=k #[δ]=d)

CALCULATIONS

Dim State	Number Additions			Nr Multiplications		
Equation	k	5	10	k	5	10
$L_t=T_t - K_tZ_t$	k^2	25	100	k^2	25	100
$r_{t-1}=Z_t'F_t^{-1}v_t + L_t'r_t$ (modified)	k^2 $(d+1)k^2$	25	100	k^2+k+1 $(d+1)(k^2+k+1)$	31	111
$N_{t-1}=Z_t'F_t^{-1}Z_t - L_t'N_tL_t$	$2k^3-k^2$	225	1900	$2k^3+k^2+k$	280	2110
$a_{t\|n}=a_{t\|t-1} + P_{t\|t-1}r_{t-1}$ (modified)	k^2 $(d+1)k^2$	25	100	k^2 $(d+1)k^2$	25	100
$P_{t\|n}=P_{t\|t-1} - P_{t\|t-1}N_{t-1}P_{t\|t-1}$	$2k^3-k^2$	225	1900	$2k^3$	250	2000

STORAGE

Dim State	k	5	10
KF quantities			
v_t (modified)	1 d+1	1	1
F_t K_t	k+1	6	11
$a_{t\|t-1}$ (modified)	k $(d+1)k$	5	10
$P_{t\|t-1}$	k^2	25	100

Table 3.2

Computational consequences of disturbance smoother
(#[y_t]=1 #[α_t]=k #[ϵ_t]=m #[δ]=d)

CALCULATIONS

Dimensions Equation	Number Additions (k,m)	 (7,4)	Nr Multiplications (k,m)	 (7,4)
$b_t = (e_t; r_t)$ (modified)	k (d+1)k	7	k+1 (d+1)(k+1)	8
$B_t = (D_t, -K_t'N_t; -N_tK_t, N_t)$	k^2	49	k^2+k+1	57
$u_t = J_t'b_t$ (modified)	km (d+1)mk	28	(k+1)m (d+1)(k+1)m	32
$C_t = J_t'B_tJ_t$	(m+k+1)mk	336	(m+k+1)m(k+1)	384
$C_t^* = J_t'B_tI_t$	(2k+1)mk	420	(2k+1)m(k+1)	480
$r_{t-1} = I_t'b_t$ (modified)	k^2 (d+1)k^2	49	(k+1)k (d+1)(k+1)k	56
$N_{t-1} = I_t'B_tI_t$	$2k^3+k^2$	735	$2k^3+3k^2+k$	840

STORAGE

Dimensions KF quantities	(k,m)	(7,4)
v_t (modified)	1 (d+1)	1
F_t K_t	k+1	8

Table 3.3

Comparing computational costs of DJKA and K state smoother

COMPUTATIONS

	recursion r_t		DJKA smoother		K smoother	
	add	mult	add	mult	add	mult
(a) Usual state smoother						
LLM	2	2	1	1	1	1
LLT	4	3	4	4	3	2
BSM_4	10	6	25	25	6	3
BSM_{12}	26	14	169	169	14	3
TVR_6	12	13	36	36	6	6
(b) Modified state smoother						
LLM	3	3	2	2	1	1
LLT	6	5	8	8	3	2
BSM_4	15	11	50	50	6	3
BSM_{12}	39	27	338	338	14	3
TVR_6	18	19	72	72	6	6

(c) STORAGE REQUIREMENTS

	recursion r_t		DJKA smoother		K smoother	
	usual	modif	usual	modif	usual	modif
LLM	3	4	2	3	0	0
LLT	4	6	6	10	0	0
BSM_4	7	12	30	55	0	0
BSM_{12}	15	28	182	351	0	0
TVR_6	8	14	42	78	0	0

Table 3.4

Outline of the new EM algorithm

Step	Description
0.	Initial values are set for the hyper parameters
1.	The KF is applied and the quantities v_t, F_t and K_t are stored
2.	The variance σ^2 is etimated by $\hat{\sigma}^2 = n^{-1}\Sigma_{t=1}^{n} v_t' F_t^{-1} v_t$
3.	The disturbance smoother is applied and the cumulative sums $J_1^G = \Sigma_{t=1}^{n} e_t e_t'$ $J_2^G = \Sigma_{t=1}^{n} D_t$ $J_1^H = \Sigma_{t=1}^{n} r_t r_t'$ $J_2^H = \Sigma_{t=1}^{n} N_t$ can be calculated recursively.
4.	The new estimates for GG' and HH' are, respectively, $GG' + GG'\{(\hat{\sigma}^{-2}J_1^G - J_2^G)/n\}GG'$ $HH' + HH'\{(\hat{\sigma}^{-2}J_1^H - J_2^H)/n\}HH'$
5.	The iterative process from 1 to 4 is terminated when the matrices within the curly brackets of step 4 is close enough to zero

Table 3.5

Theoretical and sample autocorrelations of COAL auxiliary residuals

Lag	$\hat{\epsilon}_t$ Theory	$\hat{\epsilon}_t$ Sample	$\hat{n}_t$ Theory	$\hat{n}_t$ Sample	$\hat{\zeta}_t$ Theory	$\hat{\zeta}_t$ Sample	$\hat{\omega}_t$ Theory	$\hat{\omega}_t$ Sample
1	-0.062	-0.111	0.735	0.710	0.984	0.975	-0.415	-0.406
2	-0.076	-0.160	0.503	0.427	0.945	0.913	-0.172	-0.208
3	-0.042	-0.003	0.311	0.236	0.890	0.830	-0.290	-0.153
4	-0.167	-0.263	0.141	0.048	0.824	0.735	0.744	0.572
5	-0.014	-0.073	0.060	0.012	0.755	0.637	-0.288	-0.294

Table 3.6

Theoretical auto- and cross-correlations for the auxiliary residuals of a Local Level model with different signal to noise ratios (q) +

q	theta	Lags 0	1	2	3	4	5	6
$\rho_{\hat{\epsilon}}$								
0.001	0.969	1.00	-0.02	-0.02	-0.01	-0.01	-0.01	-0.01
0.010	0.905	1.00	-0.05	-0.04	-0.04	-0.04	-0.03	-0.03
0.100	0.730	1.00	-0.14	-0.10	-0.07	-0.05	-0.04	-0.03
1.000	0.382	1.00	-0.31	-0.12	-0.05	-0.02	-0.01	-0.00
10.00	0.084	1.00	-0.46	-0.04	-0.00	-0.00	-0.00	-0.00
100.0	0.010	1.00	-0.50	-0.00	-0.00	-0.00	-0.00	-0.00
$\rho_{\hat{\eta}}$								
0.001	0.969	1.00	0.97	0.94	0.91	0.88	0.85	0.83
0.010	0.905	1.00	0.90	0.82	0.74	0.67	0.61	0.55
0.100	0.730	1.00	0.73	0.53	0.39	0.28	0.21	0.15
1.000	0.382	1.00	0.38	0.15	0.06	0.02	0.01	0.00
10.00	0.084	1.00	0.08	0.01	0.00	0.00	0.00	0.00
100.0	0.010	1.00	0.01	0.00	0.00	0.00	0.00	0.00
$\rho_{\hat{\epsilon}\hat{\eta}}$								
0.001	0.969	0.12	0.12	0.12	0.11	0.11	0.11	0.10
0.010	0.905	0.22	0.20	0.18	0.16	0.15	0.13	0.12
0.100	0.730	0.37	0.27	0.20	0.14	0.10	0.08	0.06
1.000	0.382	0.56	0.21	0.08	0.03	0.01	0.00	0.00
10.00	0.084	0.68	0.06	0.00	0.00	0.00	0.00	0.00
100.0	0.010	0.70	0.01	0.00	0.00	0.00	0.00	0.00

+ calculations are based on sample size n=100.

Table 3.7

Organisation of the modified disturbance smoother matrices in the computer program : before and after the smooth update

Name	Position	Before smooth	After smooth
DSE	0,0..DimKF	V_t	E_t
	1..DimSt,1..DimKF	R_t	R_{t-1}
DSD	0,0	F_t	D_t
	0,1..DimSt	K_t'	$-K_t'N_t$
	1..DimSt,0	-	$-N_tK_t$
	1..DimSt,1..DimSt	P_t	N_{t-1}
DSU	1..DimSt,1..DimKF	-	U_t
DSC	1..DimSt,1..DimSt	-	C_t

Figure 3.1 Signal for COAL

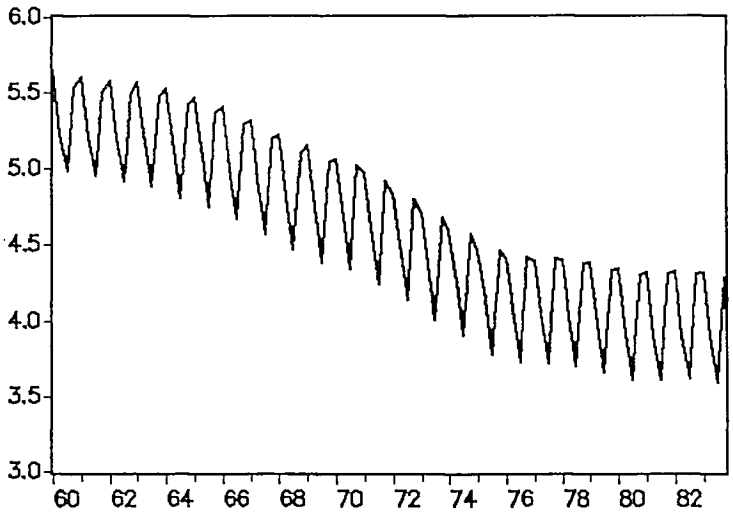


Figure 3.2 Trend for COAL

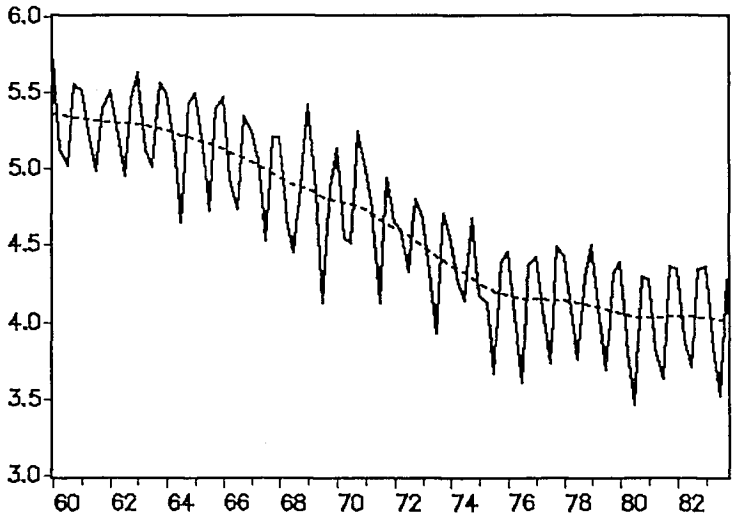


Figure 3.3 Seasonal component for COAL

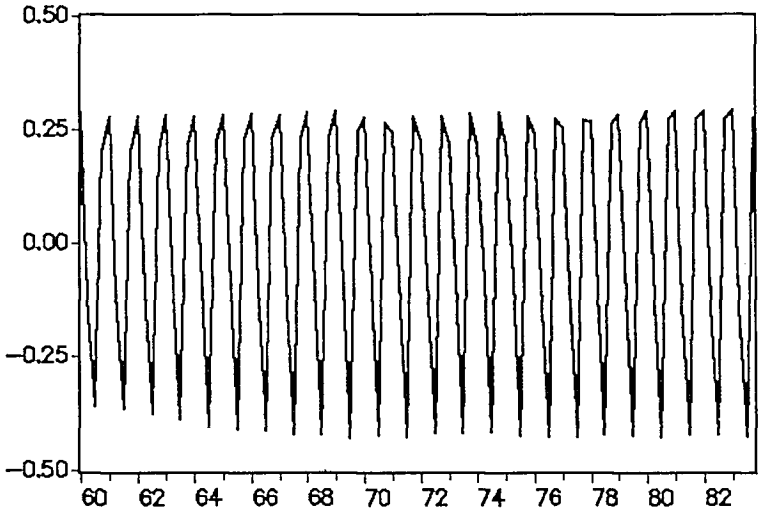


Fig 3.4 Standardized Irregular for COAL

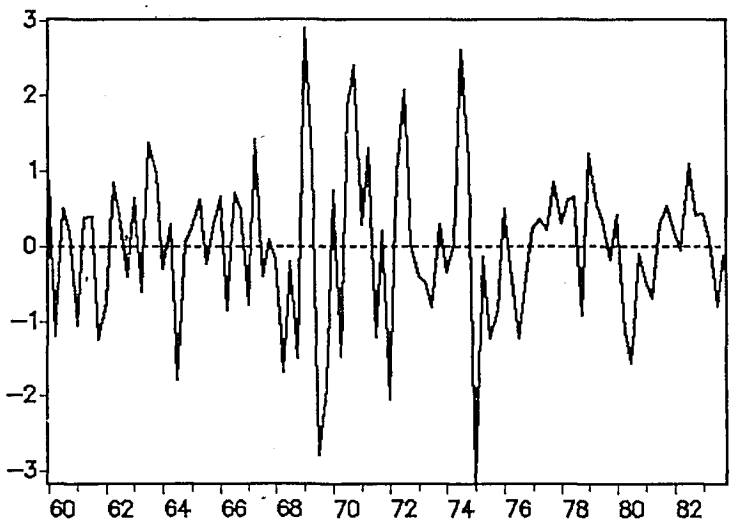


Fig 3.5a Mean square error of irregular in a LL model (q=1.0)

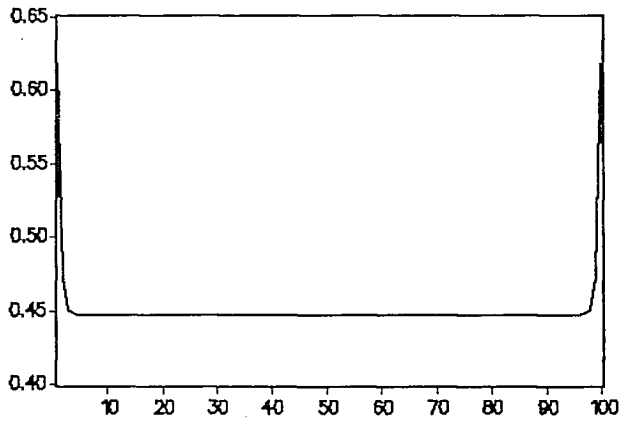


Fig 3.5b Mean square error of irregular in a LL model (q=0.01)

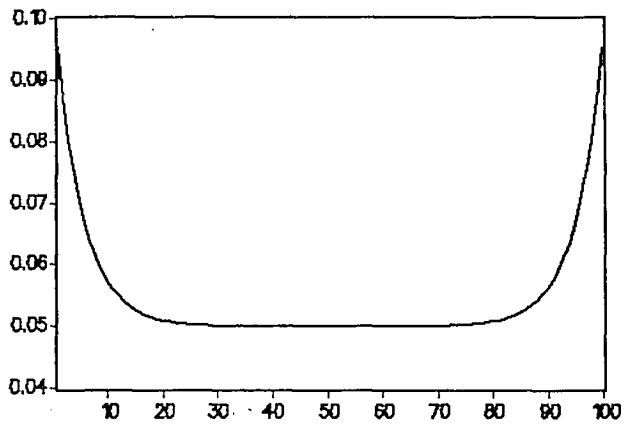
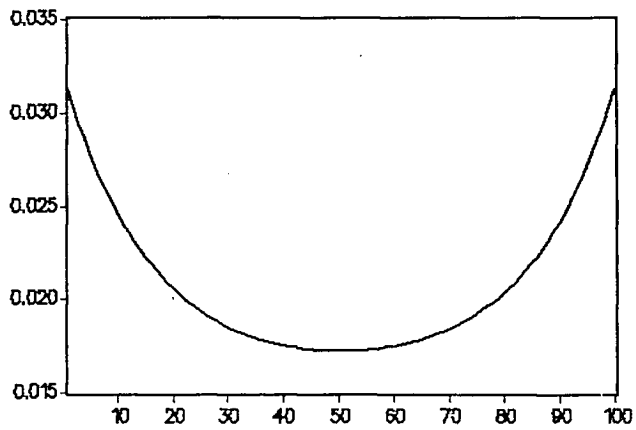


Fig 3.5c Mean square error of irregular in a LL model (q=0.001)



CHAPTER 4
Diagnostic checking of unobserved components time series models

0. ABSTRACT

Diagnostic checking of the specification of time series models is normally carried out using the innovations, that is the one-step ahead prediction errors. In an unobserved components model, other sets of residuals are available. These auxiliary residuals are estimators of the disturbances associated with the unobserved components. They can often yield information which is less apparent from the innovations, but suffer from the disadvantage that they are serially correlated, even in a correctly specified model with known parameters. This chapter shows how the properties of the auxiliary residuals may be obtained, how they are related to each other and the innovations, and how they can be used to construct test statistics and diagnostics. Applications are presented showing how residuals can be used to detect and distinguish between outliers and structural change.

Keywords : Misspecification; Outliers; Signal extraction; Smoothing; Structural change; Structural time series model.

1. INTRODUCTION

Diagnostic checking of the specification of a time series model is normally carried out using the one-step ahead prediction errors, see section 2.6. In an unobserved components model, other residuals are available. These auxiliary residuals are full-sample estimators of the disturbances associated with the observation y_t and with the unobserved components. In the state space framework, these disturbances are given by $(G_t;H_t)\epsilon_t$ and expressions for the auxiliary residuals $(G_t;H_t)u_t$ can be obtained from section 3.4. As is shown in chapter 3, the auxiliary residuals are functions of innovations, but

they present the information in a different way. This can lead to the discovery of features of the fitted model which are not apparent from the innovations themselves. Unfortunately, the auxiliary residuals suffer from the disadvantage that they are serially correlated, even in a correctly specified model with known parameters. This follows from section 3.4 since the lagged mean square error matrix $\mathrm{Mse}[u_t,u_j]$, see equation (3.4.5), is not equal to zero.

The aim of this chapter is to show how the properties of auxiliary residuals may be obtained, how they are related to each other and to the innovations, and how they can be used to construct test-statistics and diagnostics. The methods extend straightforwardly to models containing observed explanatory variables.

Section 2 derives the properties of the auxiliary residuals using the classical approach based on a doubly infinite sample. This follows Maravall (1987), except that in his paper attention is restricted to the irregular component in the decomposition of an autoregressive integrated moving average (ARIMA) model. Although the results are general for any time series model in SSF, the emphasis is on structural time series models. It is our conviction that these models provide the most satisfactory framework for exploring many important issues in time series analysis, in particular, issues concerning outliers and structural change.

It is clear from the results of section 3.4, which can be applied to any time series model placed in the SSF, that various relationships exist between the auxiliary residuals themselves and the innovations in finite samples. Section 3 studies these finite sample relationships in more detail for structural time series models by using the direct approach of Whittle (1991) as presented in section 3.5. It is shown that the finite sample relationships are similar to those observed for infinite samples. Note that the (theoretical) auto- and cross-correlations of the auxiliary residuals for any time series model can be obtained from the disturbance smoother as described in section 3.7.

The interpretation of the auxiliary residuals means that they are potentially useful, not only for detecting outliers and

structural changes, but also for distinguishing between them. The work of Kohn & Ansley (1989), which is concerned only with the measurement residuals to detect outliers, is extended. Section 4 discusses how the Bowman-Shenton test can be modified to take account of the serial correlation in the auxiliary residuals. The behaviour of these adjusted tests in small samples is investigated by some Monte-Carlo experiments. Section 5 gives several illustrations for some data sets.

2. PROPERTIES OF RESIDUALS IN INFINITE SAMPLES

Classical results in signal extraction can be used to derive the properties of various auxiliary residuals in a doubly infinite sample. Let the observed univariate time series, y_t, be the sum of m+1 mutually uncorrelated ARIMA processes, μ_{it}, that is

$$y_t = \Sigma_{i=0}^{m} \mu_{it} = \Sigma_{i=0}^{m} \alpha_i(L)\xi_{it} \qquad (2.1)$$

where $\alpha_i(L)$ is an infinite polynomial in the lag operator, and the disturbance ξ_{it} is a mutually and serially uncorrelated random variable, with zero mean and constant variance, σ_i^2, for i=0,...,m. The polynomial $\alpha_i(L)$ may be restricted to a rational function of lag polynomials as given by

$$\alpha_i(L) = \theta_i(L)/\phi_i(L) \qquad (2.2)$$

where $\theta_i(L)$ and $\phi_i(L)$ are finite lag polynomials and they are referred to as moving-average and autoregressive polynomials, respectively. The autoregressive polynomial is allowed to contain unit roots. The reduced form, i.e. the overall model consistent with (2.1), is given by

$$y_t = \alpha(L)\xi_t \qquad (2.3)$$

where ξ_t is white noise with constant variance σ^2. The infinite lag polynomial $\alpha(L)$ may be defined as a rational function of two finite

lag polynomials, e.g. $\alpha(L)=\theta(L)/\phi(L)$. Classical results on signal extraction as developed by Bell (1984) include the minimum mean square linear estimator (mmsle) of μ_{it} which is given by

$$\begin{aligned}\hat{\mu}_{it} &= \{\sigma_i^2 |\alpha_i(L)|^2\} / \{\Sigma_{j=0}^{m} \sigma_j^2 |\alpha_j(L)|^2\} y_t \\ &= \{\sigma_i^2 |\alpha_i(L)|^2\} / \{\sigma^2 |\alpha(L)|^2\} y_t \\ &= \{\sigma_i^2 \alpha_i(L)\alpha_i(F)\} / \{\sigma^2 \alpha(L)\alpha(F)\} y_t \end{aligned} \quad (2.4)$$

where $F=L^{-1}$ is the lead operator. Since the mmsle of ξ_{it} can be expressed by $\hat{\xi}_{it} = \alpha_i^{-1}(L)\hat{\mu}_{it}$, it follows from (2.3) and (2.4) that

$$\begin{aligned}\hat{\xi}_{it} &= \{\sigma_i^2 \alpha_i(F)\} / \{\sigma^2 \alpha(L)\alpha(F)\} y_t \\ &= \{\sigma_i^2 \alpha_i(F)\} / \{\sigma^2 \alpha(F)\} \xi_t \end{aligned} \quad (2.5)$$

Under the assumption that the infinite polynomials are rational functions, the last expression may be written as

$$\hat{\xi}_{it} = \{\sigma_i^2 \phi(F)\theta_i(F)\} / \{\sigma^2 \theta(F)\phi_i(F)\} \xi_t \quad (2.6)$$

Unit roots in $\phi_i(F)$ will cancel with unit roots in $\phi(F)$ and so, if time is reversed, $\hat{\xi}_{it}$ is seen to be an ARMA process, driven by the innovations ξ_t. The process is stationary but due to the possibility of unit roots in $\theta(F)$, not necessarily strictly invertible.

The autocovariance function (ACF) of $\hat{\xi}_{it}$ may be evaluated from a knowledge of the ARMA process implied by (2.6). In the case of an unobserved components time series model as (2.1), the ACF follows straightforwardly, see Maravall (1987). In the context of structural time series models, a reduced form, or ARMA representation, exists but it is only for the local level model that an exact relationship exists between the hyperparameter (signal-to-noise ratio) and the reduced form parameter, see Harvey (1989). The classical signal extraction results are now applied to some of the principal structural time series models.

Local level

The local level (LL) model (2.3.5)-(2.3.6), where $\gamma_t=0$, is discussed in section 2.3 and is given by

$$y_t = \mu_t + \epsilon_t \qquad \mu_t = \mu_{t-1} + \eta t$$

where ϵ_t and η_t are mutually uncorrelated white noise processes with variances σ_ϵ^2 and σ_η^2, respectively. The signal-to-noise ratio is defined as $q_\eta = \sigma_\eta^2/\sigma_\epsilon^2$. The reduced form of the LL model is based on $\Delta y_t = \eta_t + \epsilon_t - \epsilon_{t-1}$ and can be expressed as the ARIMA(0,1,1) model

$$\Delta y_t = (1 - \theta L)\xi_t \qquad 0 \le \theta \le 1 \tag{2.7}$$

The autocorrelations of both specifications are given by, respectively,

$$\rho(1) = -1 / (2 + q_\eta) \qquad 0 \le q_\eta < \infty \tag{2.8a}$$
$$= -\theta / (1 + \theta^2) \tag{2.8b}$$

and $\rho(\tau)=0$ for $\tau=2,\ldots,\infty$. This leads to the relationship $\theta = \frac{1}{2}\{2 + q_\eta - \sqrt{(4q_\eta + q_\eta^2)}\}$. Thus the LL model can be written as

$$y_t = \eta_t/\Delta + \epsilon_t \tag{2.9}$$

or, in terms of (2.1), m=1, $\alpha_0(L)=1$ and $\alpha_1(L)=1/(1-L)$. Note that the reduced form implies $\alpha(L)=(1-\theta L)/(1-L)$. Applying equation (2.6) gives

$$\hat{\epsilon}_t = \{(1-F)/(1-\theta F)\}(\sigma_\epsilon^2/\sigma^2)\xi_t \tag{2.10a}$$

and

$$\hat{\eta}_t = \{1/(1-\theta F)\}(\sigma_\eta^2/\sigma^2)\xi_t \tag{2.10b}$$

Thus both $\hat{\epsilon}_t$ and $\hat{\eta}_t$ depend on future innovations and, if time is reversed, the sequence $\hat{\eta}_t$ follows an AR(1) process while $\hat{\epsilon}_t$ follows a strictly noninvertible ARMA(1,1) process. Their theoretical autocorrelations are given by

$$\rho_{\hat{\epsilon}}(\pm 1) = -\tfrac{1}{2}(1 - \theta) \qquad \rho_{\hat{\epsilon}}(\pm\tau) = \theta\rho_{\hat{\epsilon}}(\tau \mp 1) \tag{2.11a}$$
$$\rho_{\hat{\eta}}(\pm 1) = \theta \qquad \rho_{\hat{\eta}}(\pm\tau) = \theta^\tau \tag{2.11b}$$

for $\tau=2,3,\ldots,\infty$. Note that the effect of serial correlation is to make the variance of $\hat{\epsilon}_t$ less than that of ϵ_t, see Maravall (1987).

The theoretical cross-correlations are obtained by comparing both equations of (2.10) such that

$$\hat{\eta}_t = \hat{\eta}_{t+1} + q_\eta \hat{\epsilon}_t \qquad 0 \leq q_\eta < \infty \tag{2.12}$$

The cross-covariance is $\mathrm{Cov}(\hat{\epsilon}_t, \hat{\eta}_{t-\tau}) = \mathrm{Cov}\{(\hat{\eta}_t - \hat{\eta}_{t+1})\hat{\eta}_{t-\tau}/q_\eta\}$, for $\tau=0, \pm1, \pm2, \ldots$, and so

$$\rho_{\hat{\epsilon}\hat{\eta}}(\tau) = \theta^{\tau}\sqrt{\{\tfrac{1}{2}(1-\theta)\}} \qquad 0 \leq \theta \leq 1 \tag{2.13}$$

while

$$\rho_{\hat{\epsilon}\hat{\eta}}(-\tau) = -\rho_{\hat{\epsilon}\hat{\eta}}(\tau-1) \qquad \tau = 1,2,\ldots \tag{2.14}$$

As σ_ϵ^2 becomes smaller, θ tends towards zero and $\rho_{\hat{\epsilon}\hat{\eta}}(0)$ tends toward 0.707. Thus, although ϵ_t and η_t are assumed to be uncorrelated, their estimators may be quite highly correlated.

Local linear trend

The local linear trend (LLT) model is discussed in section 2.3 and is given by

$$y_t = \mu_t + \epsilon_t \qquad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \qquad \beta_t = \beta_{t-1} + \varsigma_t \tag{2.15}$$

where η_t and ς_t are mutually uncorrelated white noise processes with variances $\sigma_\eta^2 = q_\eta \sigma_\epsilon^2$ and $\sigma_\varsigma^2 = q_\varsigma \sigma_\epsilon^2$, respectively. The reduced form is an ARIMA(0,2,2) model such that $\alpha(L) = (1-\theta_1 L - \theta_2 L)/(1-L)^2$ in (2.3). A relationship between the q-values and the MA parameters can be found in the same way as it is found for the LL model but, additionally, some discrepancy function has to be minimized because the relationship is non-linear. In the formulation of (2.1), the LLT model is given by

$$y_t = \eta_t/(1-L) + \varsigma_{t-1}/(1-L)^2 + \epsilon_t \tag{2.16}$$

such that $m=3$, $\alpha_0(L)=1$, $\alpha_1(L)=1/(1-L)$ and $\alpha_2(L)=L/(1-L)^2$. Applying (2.6) leads to

$$\hat{\epsilon}_t = \{(1-F)^2/\theta(F)\}(\sigma_\epsilon^2/\sigma^2)\xi_t \qquad (2.17a)$$

$$\hat{\eta}_t = \{(1-F)/\theta(F)\}(\sigma_\eta^2/\sigma^2)\xi_t \qquad (2.17b)$$

$$\hat{\varsigma}_t = \{F/\theta(F)\}(\sigma_\varsigma^2/\sigma^2)\xi_t \qquad (2.17c)$$

where $\theta(F)=1-\theta_1 F-\theta_2 F^2$. The residuals $\hat{\epsilon}_t$, $\hat{\eta}_t$ and $\hat{\varsigma}_t$ follow ARMA(2,2), ARMA(2,1) and AR(2) processes, respectively, with $\hat{\epsilon}_t$ and $\hat{\eta}_t$ being strictly non-invertible. The three processes are stationary provided that $\sigma_\varsigma^2>0$, see Harvey (1989). The autocorrelation functions for the residuals can be obtained using (2.17) only if the appropriate values for the $\theta(L)$ polynomial are determined.

The relationships between the residuals of the LLT model are given by

$$\hat{\eta}_t = \hat{\eta}_{t+1} + q_\eta \hat{\epsilon}_t \qquad 0 \le q_\eta < \infty \qquad (2.18a)$$

$$\hat{\varsigma}_t = \hat{\varsigma}_{t+1} + (\sigma_\varsigma^2/\sigma_\eta^2)\hat{\eta}_{t+1}$$
$$= 2\hat{\varsigma}_{t+1} - \hat{\varsigma}_{t+2} + q_\varsigma \hat{\epsilon}_{t+1} \qquad 0 \le q_\varsigma < \infty \qquad (2.18b)$$

where (2.18a) is similar to (2.12) for the LL model. From these interrelationships, the cross-correlations of the disturbances can be determined.

In typical applications, the variance σ_ς^2 is relatively small. As a result the MA polynomial $\theta(L)$ for the LLT model will have one, and possibly two, of its roots close to unity. The sequence $\hat{\varsigma}_t$ will therefore tend to exhibit very strong positive positive serial correlation. This effect is counteracted in the other auxiliary residuals by the presence of unit roots in the moving average.

Basic structural time series model

The three methods of modelling a seasonal component, γ_t, are described in Harvey (1989, Ch.2), see also section 2.3. All specifications can be expressed in the form

$$\Sigma_{j=0}^{s-1} \gamma_{t-j} = \theta_\omega(L)\omega_t \qquad (2.19)$$

where ω_t denotes a white noise disturbance with variance $\sigma_\omega^2=q_\omega\sigma_\varepsilon^2$, s is the number of seasons and $\theta_\omega(L)$ is a polynomial of finite order at most s-2. Adding the seasonal γ_t into the LLT model yields the basic structural time series model (BSM).

The BSM with dummy seasonals, as discussed in section 2.3, is specified as (2.19) with $\theta_\omega(L)$ equals unity. The BSM may be written as the set of orthogonal unobserved components (2.1) with m=3 and

$$\alpha_0(L) = 1 \qquad \alpha_1(L) = 1/(1-L) \tag{2.20a}$$
$$\alpha_2(L) = L/(1-L)^2 \qquad \alpha_3(L) = \theta_\omega(L)/S(L) \tag{2.20b}$$

where $S(L)=1+L+\ldots+L^{s-1}$. The reduced form of the BSM is defined as (2.3) where

$$\alpha(L) = \theta(L)/\{(1-L)(1-L)^s\} \tag{2.21}$$

and $\theta(L)$ is of order s+1. A relationship between the q-values of the BSM with dummy seasonals and the θ-values of the lag-polynomial $\theta(L)$ cannot be made unless s-2 θ-values are restricted to some known values. Applying (2.6) yields

$$\hat{\varepsilon}_t = \{(1-F)(1-F^s)/\theta(F)\}(\sigma_\varepsilon^2/\sigma^2)\xi_t \tag{2.22a}$$
$$\hat{\eta}_t = \{(1-F^s)/\theta(F)\}(\sigma_\eta^2/\sigma^2)\xi_t \tag{2.22b}$$
$$\hat{\zeta}_t = \{S(F)F/\theta(F)\}(\sigma_\zeta^2/\sigma^2)\xi_t \tag{2.22c}$$
$$\hat{\omega}_t = \{(1-F)^2\theta_\omega(F))/\theta(F)\}(\sigma_\omega^2/\sigma^2)\xi_t \tag{2.22d}$$

The residuals $\hat{\varepsilon}_t$, $\hat{\eta}_t$ and $\hat{\zeta}_t$ bear exactly the same relationship to each other as in the LLT model, see (2.18), but in addition, it holds that

$$S(F)\hat{\omega}_t = q_\omega\theta_\omega(F)\hat{\varepsilon}_t \qquad 0 \le q_\omega < \infty \tag{2.23}$$

and $q_\omega=\sigma_\omega^2/\sigma_\varepsilon^2$.

Analytical expressions for the theoretical auto- and cross-correlation functions of the auxiliary residuals can be obtained from the implied (time-reverse) ARMA processes (2.22) although it is not easy to obtain them for the BSM case unless $\theta(L)$ is known.

However, numerical values for the theoretical acf's and ccf's can be computed for specific parameter values using the disturbance smoother as described in section 3.4. For a quarterly BSM with dummy seasonals, $q_\eta=1$, $q_\zeta=0.1$ and $q_\omega=0.1$, the first ten auto-correlations are as shown in table 4.1. The acf's of the irregular and the level residuals are not too dissimilar to what one might expect in a LL model with $q_\eta=1$ although, if anything, the serial correlation in the level is somewhat reduced by the presence of other components. The high positive serial correlation in the slope residual, to which attention was drawn at the end of the discussion on the LLT model, is clearly apparent. The seasonal residual shows a strong pattern of serial correlation, the most prominent feature of which is the high values at the seasonal lags four and eight. Table 4.1 reports the cross-correlations as well. The relatively pronounced patterns for $\rho_{\hat{\epsilon}\hat{\eta}}$ and $\rho_{\hat{\eta}\hat{\zeta}}$ suggested by the analysis for the LLT model are still apparent, while the relationships involving $\hat{\omega}_t$ show seasonal effects.

3. PROPERTIES OF RESIDUALS IN FINITE SAMPLES

The derived relationships between auxiliary residuals of structural time series models, such as (2.12), are valid for doubly infinite samples. However, exact relationships can be derived for finite samples using the techniques as explored in the sections 3.4, 3.5 and 3.7. The next sub-section below shows how this can be done for the local level (LL) model using the idea of Whittle (1991), see section 3.5. Furthermore, it will be shown, for the local linear trend (LLT) model, that the relationships also follow from the disturbance smoother.

Relationships between auxiliary residuals of LL model

Consider the LL model, $y_t=\mu_t+\epsilon_t$ where $\mu_t=\mu_{t-1}+\eta_t$ for $t=1,\ldots,n$, and suppose that the disturbances ϵ_t and η_t are normally distributed. The initial state is supposed to be normal as well with mean μ and a finite variance $\sigma^2 p$ and it is independent of the disturbances. The logarithm of the joint-density of the observations $y=(y_1;\ldots;y_n)$ and the states $(\mu_0;\mu_1;\ldots;\mu_n)$ is, neglecting constants,

$$J = (-2\sigma^2)^{-1}\{ \Sigma_{t=1}^{n}(y_t-\mu_t)^2 + p^{-1}(\mu_0-\mu)^2 + q^{-1}\Sigma_{t=1}^{n}(\mu_t-\mu_{t-1})^2 \} \quad (3.1)$$

where q is the signal-to-noise ratio. Partially differentiating J with respect to the state μ_t provides a means of evaluating the smoothed (full-sample) estimators $\hat{\mu}_t$, which are, under certain regularity conditions, the expected values of the state conditional on the observations, see section 3.5. Thus by setting the first derivative with respect to μ_t equal to zero, the following backwards recursion appears

$$\hat{\mu}_{t-1} = 2\hat{\mu}_t - \mu_{t+1} - q(y_t - \hat{\mu}_t) \quad (3.2)$$

for t=n-1,...,1. The initialization is obtained from $\partial J/\partial \mu_n$ such that

$$\hat{\mu}_{n-1} = \hat{\mu}_n - q(y_n - \hat{\mu}_n) \quad (3.3)$$

and so (3.2) can be started at t=n by setting $\hat{\mu}_{n+1}$ equal to $\hat{\mu}_n$. Letting $p\to\infty$ gives the end-condition for a diffuse prior, namely

$$\hat{\mu}_2 = \hat{\mu}_1 - q(y_1 - \hat{\mu}_1) \quad (3.4)$$

Although (3.2) looks, at first sight, to be an extremely attractive way of computing the smoothed estimators of the states μ_t, it is argued in section 3.5 that it is numerically unstable. This becomes even clearer from the equations (3.3) and (3.4) which are in time reverse order (take for (3.3) n=3 or n=1). Therefore, the computation of the states $\hat{\mu}_1$ and $\hat{\mu}_2$ using recursion (3.2), initialized by (3.3), is almost certainly violating (3.4). Nevertheless, (3.2) is useful for the theoretical insights it provides. The residuals are defined by

$$\hat{\epsilon}_t = y_t - \hat{\mu}_t \qquad \hat{\eta}_t = \hat{\mu}_t - \hat{\mu}_{t-1} \quad (3.5)$$

and (3.2) can be re-written as

$$\hat{\eta}_t = \hat{\eta}_{t+1} + q\hat{\epsilon}_t \quad (3.6)$$

for t=n,...,2 and with starting value $\hat{\eta}_{n+1}=0$. Thus $\hat{\eta}_t$ is a backwards cumulative sum (cusum) of the smoothed residuals $\hat{\epsilon}_t$ as given by

$$\hat{\eta}_t = q \sum_{j=t}^{n} \hat{\epsilon}_j \tag{3.7}$$

for t=2,...,n. The end condition (3.4) tells us that $\hat{\eta}_2 = -q\hat{\epsilon}_1$ and so on setting t=2 in (3.7) it follows that

$$\sum_{t=1}^{n} \hat{\epsilon}_t = 0 \tag{3.8}$$

It will be recalled that the ordinary least squares regression residuals have the 'sum-up-to-zero' property when a constant term is included.

Relationships between auxiliary residuals of LLT model

In general, the calculation of the auxiliary residuals is carried out by placing the model in SSF and applying the Kalman filter and the disturbance smoother. The algorithm described in section 3.4 enables the computations to be carried out relatively quickly in a numerically stable manner. Structural time series models generally contain nonstationary components and these are handled by means of a diffuse initial condition. The smoother can be modified such that the calculations associated with the initial conditions are carried out exactly, see section 3.4.

For example, consider the LLT model,

$$y_t=\mu_t+\epsilon_t \qquad \mu_t=\mu_{t-1}+\beta_{t-1}+\eta_t \qquad \beta_t=\beta_{t-1}+\varsigma_t$$

for t=1,...,n and where the disturbances ϵ_t, η_t and ς_t are mutually uncorrelated noise terms, see section 2.3. Note that, for the LLT model in SSF, the state vector is given by $(\mu_t;\beta_t)$ and the disturbance vector is given by $(\epsilon_t;\eta_{t+1};\varsigma_{t+1})$. The disturbance smoother for the LLT model is

$$u_t = v_t/f_t - K_t r_t \tag{3.9a}$$
$$r_{t-1} = (1;0)u_t + (1,0;1,1)r_t \tag{3.9b}$$

and

$$\hat{\epsilon}_t = u_t \tag{3.10a}$$

$$\hat{\eta}_t = (q_\eta, 0) r_{t-1} \qquad \hat{\varsigma}_t = (0, q_\varsigma) r_{t-1} \tag{3.10b}$$

The initialization is different from that for the LL model using Whittle's approach. It is assumed that $\hat{\eta}_{n+1}=\hat{\varsigma}_{n+1}=0$ ($r_n=0$) such that (3.9) and (3.10) are valid for t=n,...1. The relationships between the residuals are found by combining (3.9b) and (3.10) and by some minor manipulations,

$$\hat{\eta}_t = \hat{\eta}_{t+1} + q_\eta \hat{\epsilon}_t \tag{3.11a}$$

$$\hat{\varsigma}_t = \hat{\varsigma}_{t+1} + (q_\varsigma/q_\eta)\hat{\eta}_t \tag{3.11b}$$

$$\hat{\varsigma}_{t-1} = 2\hat{\varsigma}_t - \hat{\varsigma}_{t+1} + q_\varsigma \hat{\epsilon}_{t-1} \tag{3.11c}$$

This set of relationships for the auxiliary residuals is equivalent to the infinite results of (2.18). Thus the recursion for $\hat{\eta}_t$ is equivalent as for the LL model and is a backwards cusum of the $\hat{\epsilon}_t$'s. Note that, for the exact smoother, it holds that $\Sigma_{t=1}^n \hat{\epsilon}_t = \hat{\eta}_1 = (q_\eta, 0) r_0$ which is typically close to zero but not equal to zero. The backwards cusum (3.11b) implies that $\hat{\varsigma}_n=0$ and

$$\hat{\varsigma}_t = (q_\varsigma/q_\eta) \Sigma_{j=t+1}^n \hat{\eta}_j \tag{3.12}$$

for t=n-1,...,1. Finally, from (3.11c) or (3.12) we have

$$\hat{\varsigma}_t = q_\varsigma \Sigma_{j=t+1}^n \Sigma_{i=j}^n \hat{\epsilon}_i \tag{3.13}$$

for t=n-1,...,1.

Theoretical auto- and cross-correlation functions

In section 2 it is shown how the auto- and cross-correlation structure of the auxiliary residuals can be obtained directly from the large sample theory when the parameters of the reduced form are known. The disturbance smoother provides recursions to evaluate the theoretical variances, covariances, auto-covariances and cross-covariances of the smoothed (auxiliary) residuals for finite samples. These quantities are time-varying. For time-invariant time

series models, the various variances and covariances of the auxiliary residuals near the middle of the series are constant, see the discussion in section 3.7. However, the variance and covariance quantities of the disturbance smoother are different at the beginning and at the end of the finite sample, see figure 3.5.

Table 4.2 reports for different sample sizes the auto- and cross-correlations of auxiliary residuals from a LL model with different signal-to-noise ratios. These correlations are obtained from the variance quantities of the disturbance smoother near the middle of the sample. It shows that for even small sample sizes, the disturbance smoother (at the middle of the sample) gives the same auto- and cross-correlations as for very large (infinite) samples. Note that if q is relatively large, the memory of the moving average of the reduced form is very small and therefore only a few observations are needed to get the exact theoretical correlations.

Note that the covariance matrix differs from the mean square error matrix of the auxiliary residual vector. It is well known that $\text{Mse}(\hat{x})=\text{Cov}(x-\hat{x})$. For example, the Kalman filter evaluates the matrix $\sigma^2 F_t$ which is the mean square error matrix of the estimator of the observation y_t, that is $\sigma^2 F_t=\text{Mse}(\hat{y}_t)=\text{Cov}(y_t-\hat{y}_t)=\text{Cov}(v_t)$. Thus, $\sigma^2 F_t$ can also be regarded as the covariance matrix of the innovation. This equivalence does not exist for the auxiliary residuals. The mean square error matrix of the estimator of the disturbance vector is given by $\text{Mse}(u_t)=\text{Cov}(\varepsilon_t-u_t)=\text{Cov}(\varepsilon_t,\varepsilon_t-u_t)=\sigma^2 I-\text{Cov}(u_t)=\sigma^2 (I-C_t)$ where ε_t is the disturbance vector of the SSF. The mean square error matrix of u_t increases at both ends of the finite sample such that the covariance matrix $\sigma^2 C_t$ decreases at both ends of the sample. The covariance and mean square error quantities are of interest for different purposes. For example, the auxiliary residuals are standardised by using the mean square error matrix but their auto- and cross-correlation structure is obtained from the covariance quantities.

4. DIAGNOSTICS

The issues of outliers, structural changes and other irregularities in time series are discussed extensively in the literature and they are still a matter of debate. Several detection procedures are developed in the context of ARMA models, for example, Chang, Tiao & Chen (1988) present an approach based on likelihood ratio criteria that is able to detect two different types of outliers. An additive outlier (AO) only influences a specific observation while an innovational outlier (IO) affects a sequence of future observations. The innovational outlier does not indicate a structural change (or level change). The former affects future observations through a (stationary) dynamic structure and the latter influences the future observations permanently, see Tsay (1988). A methodology of identifying (additive and innovative) outliers by an iterative procedure is explored and illustrated by Tsay (1986). Another example of a detection procedure, developed within the ARIMA framework, is the 'leave-k-out' methodology as discussed by Bruce & Martin (1989). This approach is primarily based on model identification using the correlogram which can be seriously distorted when outliers are present. Therefore, this methodology can be regarded as cumbersome.

Within the context of a structural time series model, an outlier arises at time t if the value taken by y_t is not consistent with what might reasonable be expected given the model specification and the way in which this fits the other observations. The best indicator of an outlier should be $\hat{\epsilon}_t$; compare Kohn & Ansley (1989). Note that an outlier at time t will not affect the innovations before time t. Therefore it makes sense that $\hat{\epsilon}_t$ depends only on the innovations which are affected by the outlier.

The simplest kind of structural change is a permanent shift in the level of a series which is of a greater magnitude than might reasonably be expected given the model specification and the other observations. Within the context of the local level model, such a shift might be best detected by an outlying value of $\hat{\eta}_t$. Again only the innovations at time t and beyond are affected by such a shift and $\hat{\eta}_t$ combines these innovations in the most appropriate way.

A sudden change in the slope is likely to be more difficult to detect than a shift in the level. As already noted, the $\hat{\varsigma}_t$'s will typically be very strongly correlated so a break will spread its effect over several $\hat{\varsigma}_t$'s. Furthermore the high serial correlation means that the variances of the normality and kurtosis statistics discussed below will need to be increased considerably giving the tests rather low power. Sudden shifts in the seasonal pattern are also likely to be difficult to detect, primarily because it may not be possible to associate a shift with a single prominent disturbance in the seasonal stochastic equation. Coupled with this is the strong serial correlation which the corresponding auxiliary residuals will tend to display. A change in the seasonal pattern is therefore more likely to be signalled by a group of large residuals. Therefore, it may also be useful to examine the estimated seasonal factors directly. More research on this topic is under progress.

The proposed detection procedure is to plot the auxiliary residuals after they have been standardized. In a Gaussian model, indications of outliers and/or structural change arise for values greater than two in absolute value. The standardized innovations may also indicate outliers and structural change, but will not normally give a clear indication as to the source of the problem.

In the following, the concentration is on a procedure to detect unusually large residuals. Mostly these approaches are based on a test for excess kurtosis and, if the test is combined with a test for skewness, can be extended to the Bowman-Shenton test for normality. In order for such tests to be asymptotically valid, it is necessary to adjust them for serial correlation.

Test based on skewness and kurtosis

Let z_t be a stationary Gaussian time series with autocorrelations, ρ_τ, $\tau=0,1,2,\ldots$, and variance σ_z^2. Lomnicki (1961) explores asymptotically results for higer order moments when serial correlation is present. Consider the estimated moments about the sample mean

$$m_\alpha = n^{-1} \Sigma_{t=1}^{n} (z_t - \bar{z})^\alpha \qquad \alpha = 2,3,4$$

and define

$$\kappa(\alpha) = \Sigma_{\tau=-\infty}^{\infty} \rho_\tau^\alpha \qquad \alpha = 2,3,4$$

Then if μ_α denotes the theoretical α-th moment,

$$\sqrt{n}\ (m_\alpha - \mu_\alpha) \xrightarrow{L} N\ (\ 0\ ,\ \alpha!\ \kappa(\alpha)\sigma_z^{2\alpha}\ )$$

This result enables asymptotically valid test statistics based on higher order moments to be constructed as follows.

[a] excess kurtosis test

The measure of kurtosis is $k = m_4 / m_2^2$, compare its definition in section 2.6. Since m_2 is a consistent estimator of σ_z^2, it follows that the excess kurtosis test statistic

$$K = (k - 3) / \sqrt{[24\kappa(4)/n]}$$

is asymptotically N(0,1) under the null hypothesis. An outlier test is carried out as a one-sided test on the upper tail.

[b] normality test

The measure of skewness is $s = m_3 / m_2^{3/2}$. Combining this with the measure of kurtosis gives the Bowman-Shenton normality test which when corrected for serial correlation takes the form

$$N = [n\ s^2\ /\ 6\ \kappa(3)\ ] + [n\ (k-3)^2\ /\ 24\ \kappa(4)\ ]$$

Under the null hypothesis N is asymptotically χ_2^2 distributed.

The normality and excess kurtosis tests may be applied to innovations, see section 2.6, and auxiliary residuals. In contrast to serial correlation tests, no amendments are needed to allow for the estimation of unknown parameters; compare section 3.7. The serial correlation correction terms, i.e. the $\kappa(\alpha)$'s terms for α=2,3,4 , needed for the auxiliary residuals can be computed using the general disturbance smoother of section 3.4. The results in

section 4.2 are useful in that they enable one to get some idea of the likely size of $\kappa(\alpha)$ where

$$\kappa(\alpha) = \Sigma_{\tau=-\infty}^{\infty} \rho_\tau^\alpha = 1 + 2 \Sigma_{\tau=1}^{\infty} \rho_\tau^\alpha$$

In the case of the LL model, for $\hat{\epsilon}_t$, it holds that

$$\begin{aligned}\kappa(\alpha) &= 1 + 2\{ \rho^\alpha + (\theta\rho)^\alpha + (\theta^2\theta)^\alpha + (\theta^3\rho)^\alpha \ldots \} \\ &= 1 + 2 \rho^\alpha / (1 - \theta^\alpha) \\ &= 1 + 2^{1-\alpha} \{-(1 - \theta)\}^\alpha / (1 - \theta^\alpha)\end{aligned}$$

where $\rho=\rho_{\hat{\epsilon}}(1)$, see (2.11a). This is greater than or equal to unity for $\alpha=4$, but less than or equal to unity for $\alpha=3$. When θ is unity, $\kappa(\alpha)$ takes the values -0.75 and 1.125 for $\alpha=3$ and 4, respectively. On the other hand, for $\hat{\eta}_t$, where $\rho=\rho_{\hat{\eta}}(1)$, see (2.11b),

$$\begin{aligned}\kappa(\alpha) &= 1 + 2 \rho^\alpha / (1 - \theta^\alpha) \\ &= (1 + \theta^\alpha) / (1 - \theta^\alpha)\end{aligned}$$

This is unity for a random walk, that is $\theta=0$, and goes monotonically towards infinity as q_η tends towards zero, that is θ tends to unity.

The kurtosis test statistic for $\hat{\epsilon}_t$ always becomes smaller after being corrected for serial correlation because its $\kappa(4)=1.125$. This is also true for the normality statistic when applied to the level residual $\hat{\eta}_t$ because its $\kappa(\alpha)>1$ for $\alpha=3,4$ and $0<q_\eta<\infty$. The normality test statistic for the irregular may, however, increase because its $\kappa(3)=-0.75$. For the irregular the correction factors are relatively small. The high correction factors for the level residual when θ is close to unity may appear to make the detection of structural change difficult. However, if level shifts are introduced into an otherwise well behaved series, the effect is likely to be an increase in the estimate of the relative variance of η_t, and hence a corresponding decreases in θ and in $\kappa(\alpha)$.

For more complex models, the correction factors can be computed numerically using the disturbance smoother of section 3.4. Table 4.3 shows the $\kappa(\alpha)$ values for the four sets of auxiliary residuals from a specific quarterly BSM. The correction factors are

calculated using the first 20 autocorrelations.

Some Monte Carlo experiments

A series of simulation experiments was applied to examine the performance of the test statistics discussed in the previous section. The experiments were conducted on the local level (LL) model using a sample size of n=150 and different values of the signal to noise ratio, q_η. The white noise disturbances ϵ_t and η_t were generated using the Box-Muller algorithm in Knuth (1981). The results presented in table 4.4 are based on one thousand replications and show the estimated probabilities of rejection for tests at a nominal 5% level of significance.

Table 4.4 gives the estimated size of the tests. It is known that, for independent observations, the size of the Bowman-Shenton test can be some way from the nominal size for small samples and Granger & Newbold (1977, p.314-315) cite evidence which suggests that serial correlation may make matters even worse. However, their remarks are concerned with a test statistic in which the correction factors are based on the correlogram, whereas in our case the correction factor is based on the estimator of a single parameter, θ. The figures in table 4.4 indicate that the estimated type I errors are not too far from the nominal values for both the innovations and the auxiliary residuals.

Table 4.4 shows the estimated powers of the tests when an outlier was inserted three quarters of the way along the series. The magnitude of the outlier was five times σ_ϵ^2. As can be seen, the powers of the tests based on the irregular residual are higher than those based on the innovation. As hoped, the power of the tests based on the level residual are much lower. The kurtosis test is slightly more powerful than the normality test.

A shift in the level, up by five times σ_ϵ^2, was introduced three quarters of the way along the series to generate the results in table 4.4. The tests based on the level residual are now more powerful.

Overall the results are very encouraging. They suggest that the tests have acceptable sizes for moderate samples even when serial correlation corrections have to be made. Furthermore, the

tests based on auxiliary residuals are reasonably effective in detecting, and distinguishing between outliers and structural change.

5. APPLICATIONS

The way in which outliers and structural changes may be detected is illustrated below. In all cases parameter estimation was carried by non-linear optimization of the likelihood, using the method of scoring in the frequancy domain, see section 2.6.

United States exports to Latin America

The monthly series of US exports to Latin America (LAXL), taking logs, contains a number of outliers which are easily detected by examining the irregual component, $\hat{\epsilon}_t$, from a basic structural model; see the comments by Harvey on Bruce & Martin (1990). In fact, the principal outliers, which turn out to be due to dock strikes, are easily seen in a plot of the series and also appear quite clearly in the innovations. If the data are aggregated to the quarterly level (the graph is given in figure 4.1 and estimation results are reported in table 4.5), the outliers are less apparent in the innovations, though they still emerge clearly in the irregular component, see figure 4.2. The kurtosis statistic for the innovations is K = 2.18 and the normality statistic is N = 5.14. The normality statistic is therefore not statistically significant at the 5% level, while the kurtosis is significant on a one-sided test at the 5% level, but not at the 1% level. For the irregular, on the other hand, the raw K and N statistics are 7.32 and 80.08, respectively. After correction for serial correlation these become K = 7.54 and N = 86.41, both of which are significant.

Since σ_η^2 is estimated to be zero, all the movements in the trend stem from the slope disturbance. The (corrected) K and N statistics for the associated auxiliary residuals are only 0.31 and 0.22, respectively. The auxiliary residual diagnostics therefore point clearly to the presence of outliers.

Car drivers killed and seriously injured in Great Britain

Monthly observations of car drivers killed and seriously injured in Great Britain (CDKSI) were used by Harvey & Durbin (1986) in their study of the effects of the seat belt legislation which took effect from the beginning of February 1983. The seat belt law led to a drop in the level of the series. It will be shown below how this structural change would be detected by the auxiliary residuals.

In order to avoid the large fluctuations associated with the oil crisis of 1974, a BSM was estimated using data from July 1975 to the end of the series December 1984, see figure 4.3. The slope and seasonal variances were both estimated to zero, and so the fitted model is basically a local level model with deterministic slope and seasonals, see table 4.6. The theory at the end of sub-section 4.1 therefore applies directly, with $q = 0.118$ and $\eta = 0.710$. The correction factors for the irregular are $\kappa(3) = 0.99$ and $\kappa(4) = 1.00$ while for the level they are $\kappa(3) = 2.12$ and $\kappa(4) = 1.69$.

The kurtosis and normality statistics are shown in table 4.6, with the uncorrelated figures in parantheses. The innovation statistics clearly indicate excess kurtosis, and the auxiliary residual diagnostics point to this as emanating from a change in the level., with the K and N statistics both being statistically significant at the one percent level. The plot of the innovations in figure 4.4 shows large values in December 1981 and Februari 1983 at -3.28 and -3.97, respectively. In the irregular residuals, shown in figure 4.4 as well, both these months are -2.84 but such a value is not excessively large compared with those for some of the other months. In the level residuals, on the other hand, Februari 1983 is -4.46 while December 1981 is only -1.76.

The residuals therefore point clearly to a structural break at the beginning of 1983. The role of December 1981 is less clear. It could be treated as an outlier. In fact, Harvey & Durbin (1986) noted that December 1981 was a very cold month. However, even when the model is re-estimated with an intervention variable for the seat belt law, it does not give rise to a particularly large irregular residual, though, curiously enough, the corresponding innovation is still quite high.

A final point with respect to this example concerns checks for serial correlation as given in section 3.7. For the innovations the Box-Ljung statistic based on the first ten sample autocorrelations is Q(10) = 8.58. Thus no serial correlation is indicated. The correlograms and theoretical autocorrelation functions for both the irregular and the level residuals are quite similar and hence give no further hint of model misspecification. Nor do the sample and theoretical cross-correlations. Of course, evidence of dynamic misspecification can be masked by outliers and structural breaks, but in this instance there was still no evidence of serial correlations after the inclusion of interventions.

Number of Marriages in the UK

West & Harrison (1989) analyse an interesting set of observations: the number of marriages in the UK between 1965 and 1970 for each quarter, see figure 4.5. Many marriages take place in the third quarter of each year because many couples prefer to marry on a day when there is a reasonable chance of good weather. The relatively large number of marriages in the first quarter of the years 1965, 1966, 1967 and 1968 is caused by the UK income tax law system during these years. The UK tax law provides higher tax free allowances for married couples from the current tax year onwards (running from April to March). After 1968 the married person's allowance for the current year was abolished, the claims could be made after the current tax year.

The 24 observations can be modelled with a quarterly BSM with the estimated hyperparameters as reported in table 4.7. The standardized innovation at the first quarter of 1969 is very large for this model. The source of the problem can be found by examining the auxiliary residuals. The plots in figure 4.6 and the diagnostics reported in table 4.7 clearly points to the seasonal component.

A change in the seasonal pattern is modelled by a regression intervention in the transition equation of the SSF. The seasonal interventions are only made for the seasons 1, 2 and 4. It is estimated that the number of marriages in the first quarter has reduced by 43,550. All diagnostics of the BSM with interventions are satisfactory and $\hat{\sigma}^2$ is equal to 16.69.

Consumption of Spirits in the UK

The per capita consumption of spirits in the UK for 1870 to 1938 is a classical data-set which is analysed for the first time by Prest (1949). It has become widely known since it was used as one of the testbeds for the d-statistic in Durbin & Watson (1951). The time series was analysed later, for different purposes and using different approaches, by Fuller (1976), Tsay (1986) and Kohn & Ansley (1989).

The time series, as presented in figure 4.7, can be explained, at least partly, by relative price of spirits and income per capita, as presented in figure 4.7 as well. However, a regression formulated in this way shows significant serial correlation even if a time trend is included, see Durbin & Watson (1951). The regression model with a stochastic trend component, i.e.

$$y_t = \mu_t + X_t\delta + \epsilon_t \qquad (5.1)$$

where μ_t is a local linear trend, as in (2.3.7), and δ is a fixed vector of parameters and ϵ_t is NID$(0,\sigma^2)$, provides a good fit in many respects. It is more parsimonious than the regression model with a quadratic time trend and a first-order autoregressive disturbance reported in Fuller (1976, p.426). The stochastic trend in (5.1) can be interpreted as reflecting changes in tastes. Finally, this regression model can be placed in the SSF (2.1)-(2.3), straightforwardly.

The estimates reported in table 4.8 are for the period 1870 to 1930. As can be seen, the slope is stochastic and so there is a set of three auxiliary residuals, see figure 4.8. The associated test statistics are in table 4.8 as well. Kohn & Ansley (1989) estimate the model without a slope component, so μ_t is just a random walk. Indeed, estimating such a model might not be unreasonable for preliminary data analysis if we wish to focus attention on structural changes which affect the level. However, in this particular case, the kurtosis statistics in table 4.8 are high for both the irregular and level residuals and the presence of the slope makes very little difference.

The plots shown in figure 4.8 indicate a shift in the level

in 1909, with a number of candidates for outliers during World War I. We fitted a level intervention first. The 1918 outlier than stood most clearly in the irregular. On estimating with a 1918 intervention, 1915 stood out most clearly. This led to a model with a 1909 level intervention together with outlier interventions at 1915 and 1918. All the diagnostics in this model are satisfactory. Table 4.9 shows the estimated coefficients of the explanatory variables and compares them with the coefficients obtained from the model without interventions. There is a clear improvement in goodness of fit and this is reflected in the t-statistics shown in parentheses. The innovation diagnostic in the intervention model are entirely satisfactory. It is particularly interesting to note the reduction in the value of the Box-Ljung test-statistic based on the first ten residual autocorrelations, Q(10). In the model without any interventions there were high autocorrelations at lags 8,9 and 10 which had no obvious explanation.

Reference back to Prest (1949), who originally assembled the SPIRITS data set, reveals that the figures for 1915 to 1919 were estimates based on consumptions in the British Army. Thus they may be considerably less reliable than the other observations and taking them all out by intervention variables may not be unreasonable. The results are shown in the last column of table 4.10. The changes in the coefficients of income and price are due to the influence of the observations corresponding to the additional interventions rather than the fact that they may be outliers; see Kohn & Ansley (1989).

On the basis of the level residual, see figure 4.8, there is a case for a structural change in 1919. However, the general unreliability of the observations in 1915 to 1919 makes it difficult to estimate such a change with any degree of confidence. None of the other results change significantly when the 1919 outlier intervention is replaced by a level shift intervention.

The explanation for the level change in 1909 may be found in the social, economic and political situation during this year in the UK. In this remarkable year, the Chancellor of the Exchequer Lloyd George, later PM, presented his budget in which the Government announces their plans of rising income taxes for the well-off and increasing taxes on land-ownership, on luxury goods and on goods

like spirits and tobacco. Some influence of these tax reforms can be found in the downwards shock of the relative price of spirits after 1909. But the overall shock effect, which affected the demand for complementary goods as well, cannot be modelled differently than by an intervention.

The fall in the level of 1909 is highly significant in both of our intervention models, and indicates a permanent reduction, other things being equal, of around 9 %. It is this feature which is detected by our techniques and which is the prime source of the difference between our model and that of Kohn & Ansley (1989). They identify 1909 as a possible outlier. Their preferred model has outlier interventions for the years 1915 to 1919 and 1909. Fitting this model, including variations such as the inclusion of a stochastic slope and using time domain in stead of frequency domain estimation, resulted in a poorer fit than our model and somewhat different coefficients for the explanatory variables.

Tsay (1986) uses the ARIMA approach of detecting outliers, as earlier mentioned in section 4, and applies these techniques to the SPIRITS data set as well. His conclusions are similar to ours. In his terminology, innovative outliers are recognised in 1909, 1910, 1911 and 1912, and additive outliers are found for 1915 and 1918. Note that innovative outliers do not influence residuals permanently, the influence for future residuals damps out exponentially over time. The final model of Tsay (1988) is the same regression model as proposed by Fuller (1976) but with ARMA(1,1) residuals, an innovative outlier at 1909 and the two earlier mentioned additive outliers. The two parameters concerned with the innovative outlier do relate to each other in such a way that the memory of the innovative outlier is long and, therefore, the effect of this innovative outlier approximates the effect of a level intervention as employed to our model. It must be stressed that, although the methodology used by Tsay (1988) leads to the same conclusions, our detection method based on the auxiliary residuals is straightforward and much easier to use. It is also more robust since it does not rely on a model identification procedure based on the correlogram which can be seriously distorted when outliers are present in the time series.

7. CONCLUDING REMARKS

The auxiliary residuals are serially correlated and correlated with each other even when the model is correctly specified. Nevertheless, it seems that they are an useful tool for detecting outliers and shifts in the level in structural time series models. Plots of the auxiliary residuals can be very informative and these can be supplemented with tests for normality and kurtosis corrected to allow for the implied serial correlation. The examples and Monte Carlo experiments illustrate that the techniques work quite well in practice and, finally, they are very simple to apply as well.

TABLES & FIGURES

Table 4.1

Theoretical auto- and cross-correlations for the auxiliary residuals of a quarterly BSM with $q_\eta=1$, $q_\zeta=0.1$ and $q_\omega=0.1$

AUTO-CORRELATIONS

Lag	$\hat{\epsilon}$	$\hat{\eta}$	$\hat{\zeta}$	$\hat{\omega}$
0	1.00	1.00	1.00	1.00
1	-0.29	0.28	0.88	-0.44
2	-0.14	-0.02	0.70	-0.14
3	0.02	-0.12	0.52	-0.24
4	-0.18	-0.24	0.37	0.65
5	0.07	-0.09	0.28	-0.25
6	0.03	-0.05	0.21	-0.14
7	0.04	-0.05	0.15	-0.14
8	-0.11	-0.11	0.10	0.42
9	0.05	-0.02	0.07	-0.14
10	0.03	0.00	0.06	-0.13

CROSS-CORRELATIONS

Lag	$\hat{\epsilon}\hat{\eta}$	$\hat{\epsilon}\hat{\zeta}$	$\hat{\epsilon}\hat{\omega}$	$\hat{\eta}\hat{\zeta}$	$\hat{\eta}\hat{\omega}$	$\hat{\zeta}\hat{\omega}$
0	0.60	0.11	0.06	0.24	-0.07	0.07
1	0.25	-0.01	0.06	0.38	-0.00	0.03
2	0.08	-0.05	-0.32	0.37	0.07	0.03
3	0.10	-0.10	0.35	0.31	-0.31	0.07
4	-0.12	-0.04	-0.02	0.19	0.11	-0.08
5	-0.03	-0.02	-0.02	0.15	0.09	-0.03
6	-0.00	-0.02	-0.22	0.13	0.06	0.01
7	0.05	-0.05	0.22	0.10	-0.20	0.05
8	-0.08	0.01	0.01	0.05	0.06	-0.05
9	-0.02	0.00	-0.03	0.04	0.07	-0.02
10	0.01	-0.00	-0.14	0.04	0.04	0.01

Table 4.2

Theoretical auto- and cross-correlations for the auxiliary residuals of a LL model based on samples of different sizes

sample sizes :		n=6	n=10	n=20	n=50	n=100	n=∞
	ratio q=1.0						
$\rho^{\wedge}_{\epsilon}$	lag						
	1	-.3261	-.3094	-.3090	-.3090	-.3090	-.3090
	2	-.1631	-.1190	-.1180	-.1180	-.1180	-.1180
	3		-.0476	-.0451	-.0451	-.0451	-.0451
	4		-.0238	-.0172	-.0172	-.0172	-.0172
$\rho^{\wedge}_{a}$	lag						
	1	.3734	.3818	.3820	.3820	.3820	.3820
	2	.1245	.1454	.1459	.1459	.1459	.1459
	3		.0545	.0557	.0557	.0557	.0557
	4		.0238	.0172	.0172	.0172	.0172
$\rho^{\wedge\wedge}_{\epsilon a}$	lag						
	0	.5111	.5550	.5559	.5559	.5559	.5559
	1	.1071	.2102	.2123	.2123	.2123	.2123
	2		.0755	.0811	.0811	.0811	.0811
	3		.0163	.0310	.0310	.0310	.0310
	ratio q=0.1						
$\rho^{\wedge}_{\epsilon}$	lag						
	1	-.2109	-.1546	-.1359	-.1351	-.1351	-.1351
	2	-.1917	-.1226	-.0996	-.0986	-.0986	-.0986
	3		-.1030	-.0732	-.0720	-.0720	-.0720
	4		-.0936	-.0542	-.0525	-.0525	-.0525
$\rho^{\wedge}_{a}$	lag						
	1	.4933	.6529	.7264	.7298	.7298	.7298
	2	.2349	.4399	.5285	.5327	.5327	.5327
	3		.2709	.3835	.3888	.3888	.3888
	4		.1290	.2769	.2837	.2837	.2837
$\rho^{\wedge\wedge}_{\epsilon a}$	lag						
	0	.4780	.4042	.3692	.3675	.3675	.3675
	1	.3847	.3050	.2699	.2682	.2682	.2682
	2		.2363	.1975	.1958	.1958	.1985
	3		.1911	.1449	.1429	.1429	.1429
	ratio q=0.01						
$\rho^{\wedge}_{\epsilon}$	lag						
	1	-.1767	-.1088	-.0642	-.0483	-.0476	-.0476
	2	-.1750	-.1057	-.0601	-.0438	-.0430	-.0430
	3		-.1036	-.0566	-.0397	-.0389	-.0389
	4		-.1026	-.0536	-.0360	-.0352	-.0352
$\rho^{\wedge}_{a}$	lag						
	1	.3867	.4962	.6978	.8920	.9048	.9049
	2	.1924	.3679	.6037	.8057	.8187	.8188
	3		.2432	.5157	.7274	.7409	.7409
	4		.1210	.4328	.6564	.6703	.6704
$\rho^{\wedge\wedge}_{\epsilon a}$	lag						
	0	.5159	.4087	.2863	.2215	.2181	.2181
	1	.5045	.3929	.2664	.2007	.1974	.1973
	2		.3811	.2491	.1820	.1786	.1786
	3		.3731	.2343	.1651	.1616	.1616

Table 4.3

Correction factors for a quarterly BSM with $q_\eta=1$, $q_\zeta=0.1$ and $q_\omega=0.1$

Auxiliary residual	$\kappa(3)$	$\kappa(4)$
Irregular	0.93	1.02
Level noise	1.01	1.02
Slope noise	3.53	2.90
Seasonal noise	1.49	1.53

Table 4.4

Estimated rejection probabilities for tests at a nominal 5% level of significance for a local level model with sample size n=150 *

	q = 2.0				q = 0.5			
(a) NO MISSPECIFICATION								
Residual	N-cor	N-unc	K-cor	K-unc	N-cor	N-unc	K-cor	K-unc
Innovation	0.062	0.062	0.077	0.077	0.055	0.055	0.077	0.077
Irregular	0.038	0.036	0.058	0.062	0.039	0.039	0.060	0.062
Level noise	0.034	0.038	0.061	0.062	0.037	0.064	0.053	0.065
(b) SINGLE OUTLIER at t=112								
Residual	N-cor		K-cor		N-cor		K-cor	
Innovation	0.49		0.56		0.87		0.90	
Irregular	0.76		0.79		0.97		0.97	
Level noise	0.25		0.30		0.26		0.31	
(c) STRUCTURAL SHIFT in level at t=112								
Residual	N-cor		K-cor		N-cor		K-cor	
Innovation	0.42		0.45		0.83		0.85	
Irregular	0.15		0.19		0.27		0.34	
Level noise	0.47		0.49		0.94		0.95	

* based on Monte-Carlo experiments with 1000 replications

Table 4.5

Hyperparameter estimates and diagnostics for LAXL

dependent variable : Log of LAXL [+] t = 1966q1,...,1983q4
number of observations : 72 season : quarterly (s=4)
model : BSM with diffuse initial cond. method : FD scoring

HYPERPARAMETER ESTIMATION

hyperparameter	estimate	q ratio
σ_ϵ^2	0.00314	
σ_η^2	0.0	0.0
σ_ζ^2	0.00084	0.2675
σ_ω^2	0.00001	0.0032

DIAGNOSTICS

Auxiliary residual	Kurtosis K	Normality N
Innovation	2.18	5.14
Irregular	7.53 *	86.41 *
Slope residual	0.31	0.22
Seasonal residual	0.11	0.45

[+] US exports to Latin America (see Bruce & Martin, 1989)
* Significant at 1 % level

Table 4.6

Hyperparameter estimates and diagnostics for CDKSI

dependent variable : Log of CDKSI [+] t = 1975m1,...,1984m12
number of observations : 120 season : monthly (s=12)
model : BSM with diffuse initial cond. method : FD scoring

HYPERPARAMETER ESTIMATION

hyperparameter	estimate	q ratio
σ_ϵ^2	0.00425	
σ_η^2	0.000495	0.1165
σ_ζ^2	0.0	0.0
σ_ω^2	0.0	0.0

DIAGNOSTIC STATISTICS

Auxiliary residual	Kurtosis K	Normality N
Innovation	2.51 *	12.61 *
Irregular	0.50	0.86
Level residual	4.80 *	38.04 *

[+] Car drivers killed and seriously injured in Great Britain (see Harvey & Durbin, 1986)
* Significant at 1 % level

Table 4.7

Hyperparameter estimates and diagnostics for MARRIAGE

depend. variable : MARRIAGE [+] t = 1965q1,...,1970q4
number of observations : 24 season : quarterly (s=4)
model : BSM with diffuse initial cond. method : FD scoring

HYPERPARAMETER ESTIMATION

hyperparameter	estimate	q ratio
σ_ϵ^2	0.0	
σ_η^2	1.41630	0.18523
σ_ζ^2	0.0	0.0
σ_ω^2	76.4621	1.0

DIAGNOSTICS

Auxiliary residual	Kurtosis K	Normality N
Innovation	6.48 *	63.12 *
Level residual	0.85	3.08
Seasonal residual	7.72 *	79.87 *

[+] Number of marriages in UK (see West & Harrison, 1989)
* Significant at 1 % level

Table 4.8

Hyperparameter and regression estimates and diagnostics for SPIRITS

dependent variable : Log of SPIRITS [+] t = 1870,...,1930
explanatory vars : Log of INCOME
Log of PRICE
number of observations : 61 season : yearly
model : LLT with diffuse initial cond. method : FD scoring

HYPERPARAMETER ESTIMATION

hyperparameter	estimate	q ratio
σ_ϵ^2	0.000161	
σ_η^2	0.000069	0.4286
σ_ζ^2	0.000037	0.2298

REGRESSION ESTIMATION

fixed effect	estimate	t-test
Income	0.69	5.28
Price	-0.95	-13.6

GOODNESS OF FIT AND DIAGNOSTIC STATISTICS [°]

$\hat{\sigma}$	0.000229
R_D^2	0.71
Box-Ljung Q_{10}	13.06
Normality	5.87
Kurtosis	2.21
Heteroskedasticity	2.47

Auxiliary residual	Kurtosis K	Normality N
Irregular	7.53 *	69.76 *
Level residual	5.19 *	31.65 *
Slope residual	0.32	0.45

[+] Consumption of Spirits in UK (see Durbin & Watson, 1951)
[°] Goodness of fit and diagnostics are discussed in section 2.6
* Significant at 1 % level

Table 4.9

Hyperparameter and regression estimates and diagnostics for SPIRITS

dependent variable : Log of SPIRITS [+] t = 1870,...,1930
explanatory vars : Log of INCOME
Log of PRICE
interventions : Structural shift 1909
Outlier 1915 1918

number of observations : 61 season : yearly
model : LLT with diffuse initial cond. method : FD scoring

HYPERPARAMETER ESTIMATION

hyperparameter	estimate	q ratio
σ_ε^2	0.0	
σ_η^2	0.000117	
σ_ζ^2	0.000014	0.1197

REGRESSION ESTIMATION

fixed effect	estimate	t-test
Income	0.66	7.82
Price	-0.73	-15.2
1909 Structural shift	-0.09	-7.90
1915 Outlier	0.05	5.33
1918 -	-0.06	-7.47

GOODNESS OF FIT AND DIAGNOSTIC STATISTICS [°]

$\hat{\sigma}$	0.000166
R_D^2	0.91
Box-Ljung Q_{10}	5.25
Normality	1.53
Kurtosis	1.23
Heteroskedasticity	0.83

Auxiliary residual	Kurtosis K	Normality N
Irregular	-	-
Level residual	0.93	0.94
Slope residual	0.03	0.01

[+] Consumption of Spirits in UK (see Durbin & Watson, 1951)
[°] Goodness of fit and diagnostics are discussed in section 2.6
[*] Significant at 1 % level

Table 4.10

Hyperparameter and regression estimates and diagnostics for SPIRITS

dependent variable : Log of SPIRITS [+] t = 1870,...,1930
explanatory vars : Log of INCOME
Log of PRICE
interventions : Structural shift 1909
Outlier 1915 1916 1917 1918 1919

number of observations : 61 season : yearly
model : LLT with diffuse initial cond. method : FD scoring

HYPERPARAMETER ESTIMATION

hyperparameter	estimate	q ratio
σ_ε^2	0.0	
σ_η^2	0.000079	
σ_ζ^2	0.000030	0.3797

REGRESSION ESTIMATION

fixed effect	estimate	t-test
Income	0.58	6.45
Price	-0.53	-6.31
1909 Structural shift	-0.09	-8.69
1915 Outlier	0.06	4.34
1916 -	0.004	0.30
1917 -	-0.05	-2.79
1918 -	-0.10	-6.41
1919 -	-0.01	-1.25

GOODNESS OF FIT AND DIAGNOSTIC STATISTICS [°]

$\hat{\sigma}$	0.000144
R_D^2	0.93
Box-Ljung Q_{10}	4.93
Normality	0.73
Kurtosis	0.84
Heteroskedasticity	0.75

Auxiliary residual	Kurtosis K	Normality N
Irregular	-	-
Level residual	0.46	0.24
Slope residual	0.25	0.18

[+] Consumption of Spirits in UK (see Durbin & Watson, 1951)
[°] Goodness of fit and diagnostics are discussed in section 2.6
[*] Significant at 1 % level

Figure 4.1 US Exports to Latin America (in logs)

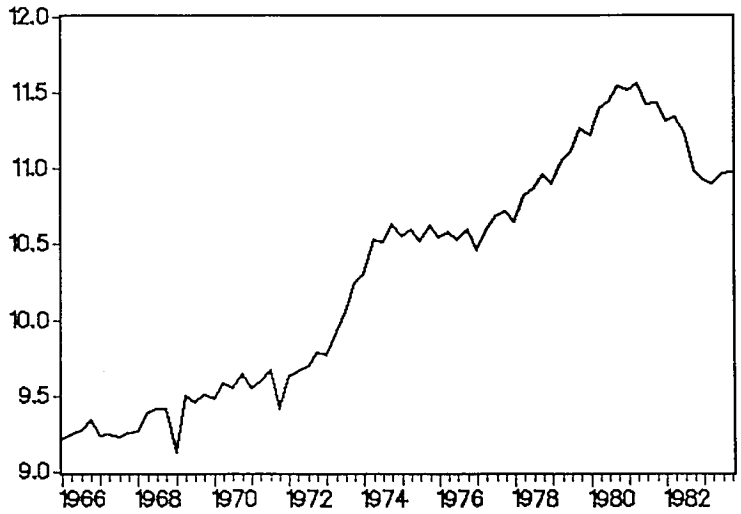


Figure 4.1a Innovations for LAXL

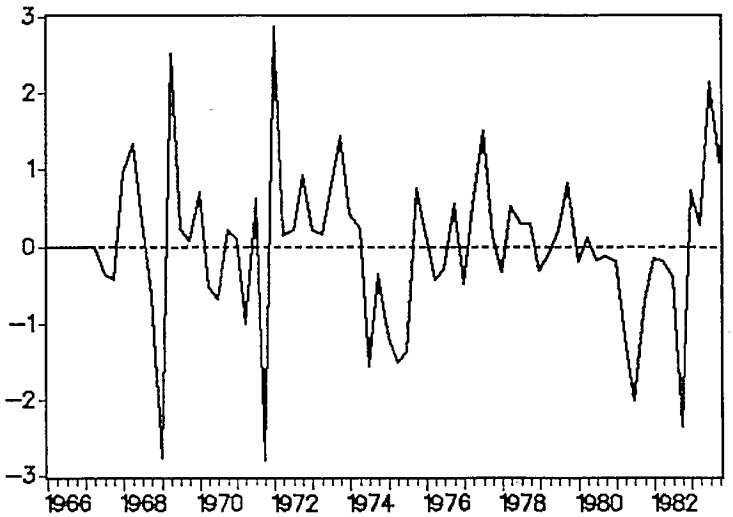


Figure 4.2b Auxiliary residuals for LAXL : Irregular

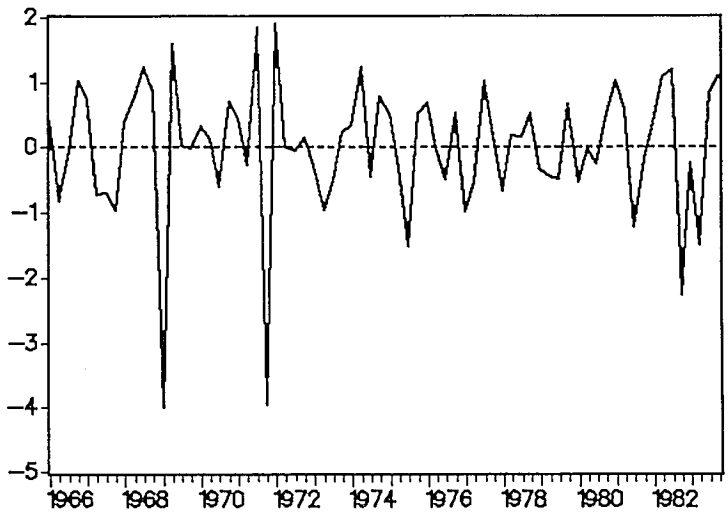


Figure 4.2c Auxiliary residuals for LAXL : Slope residual

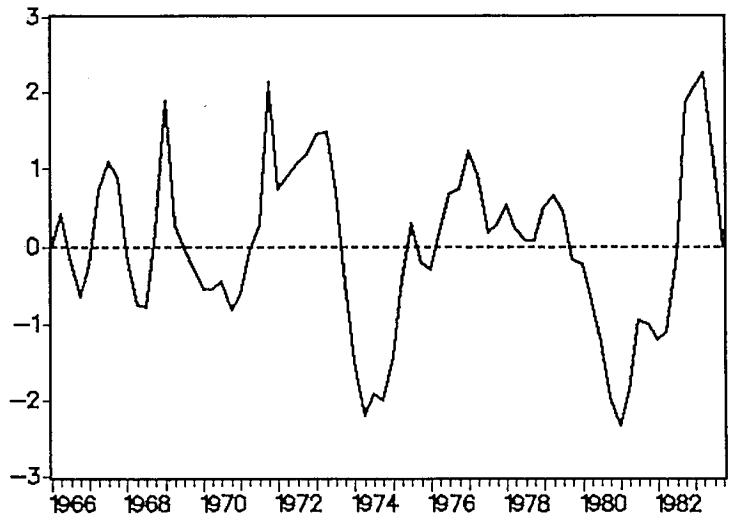


Figure 4.3 Car drivers killed and seriously injured in GB (in logs)

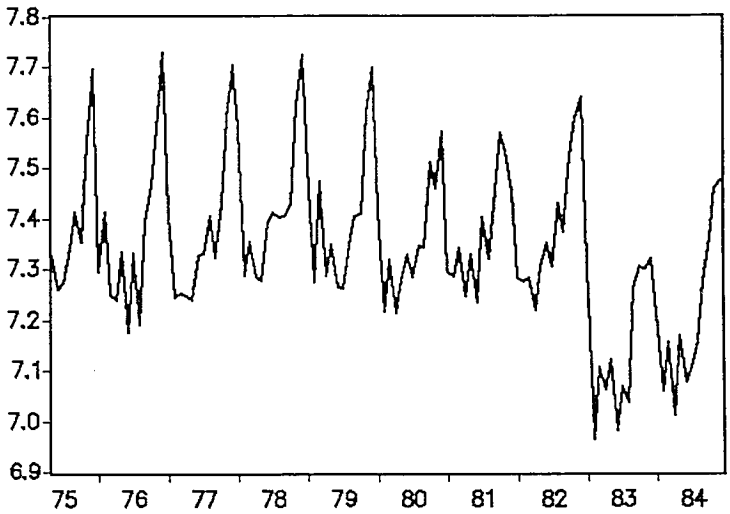


Figure 4.4a Innovations for CDKSI

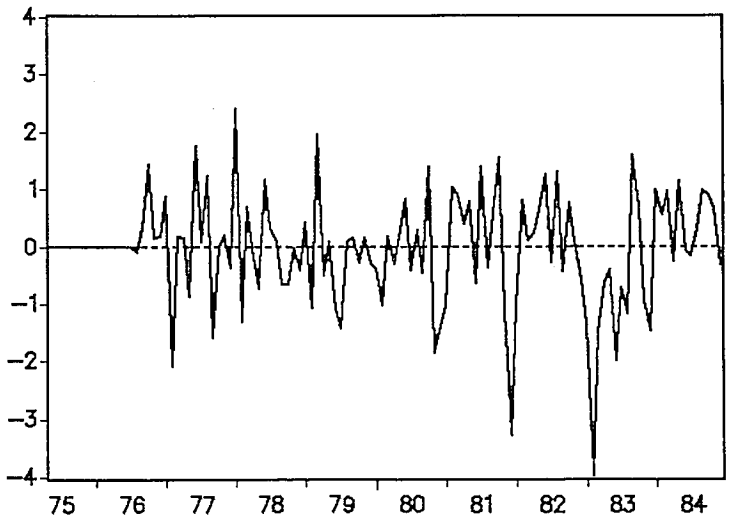


Figure 4.4b Auxiliary residuals for CDKSI : Irregular

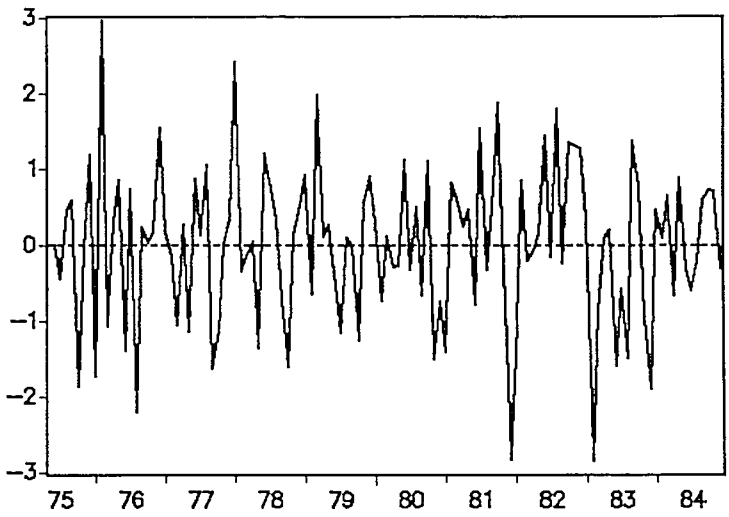


Figure 4.4c Auxiliary residuals for CDKSI : Level residual

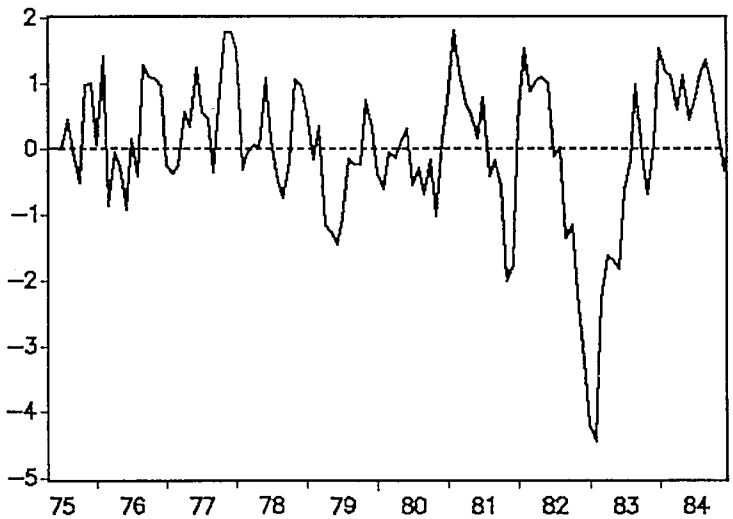


Figure 4.5 Number of marriages in UK

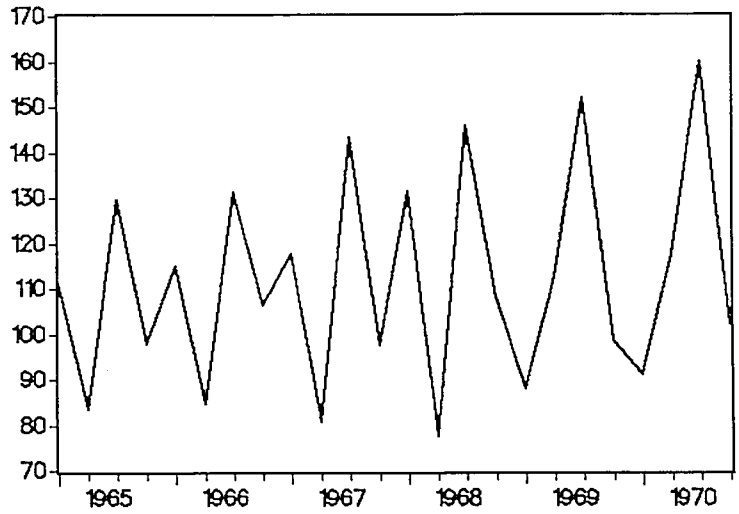


Figure 4.6a Innovations for MARRIAGE

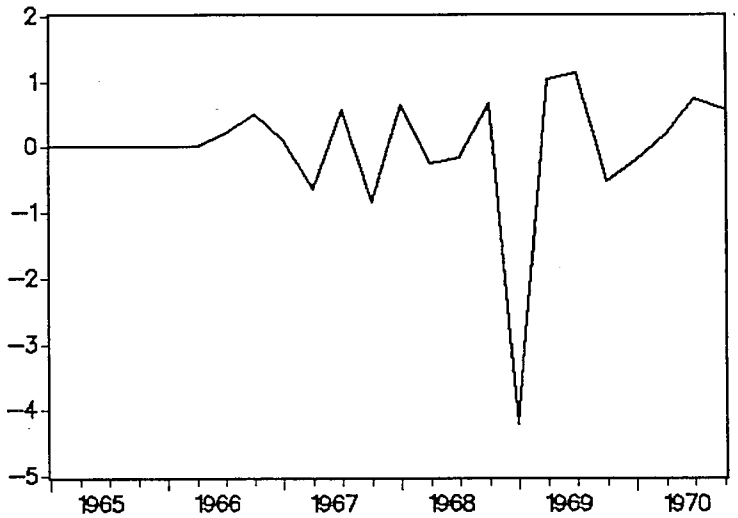


Figure 4.6b Auxiliary residuals for MARRIAGE : Irregular

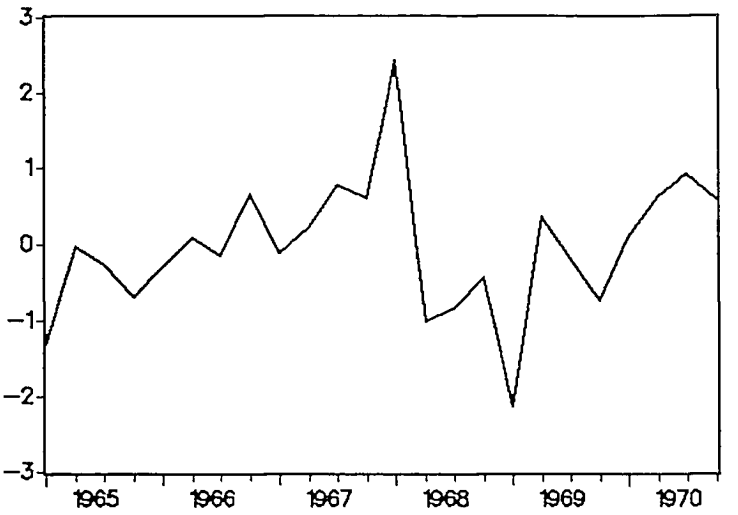


Figure 4.6c Auxiliary residuals for MARRIAGE : Seasonal residual

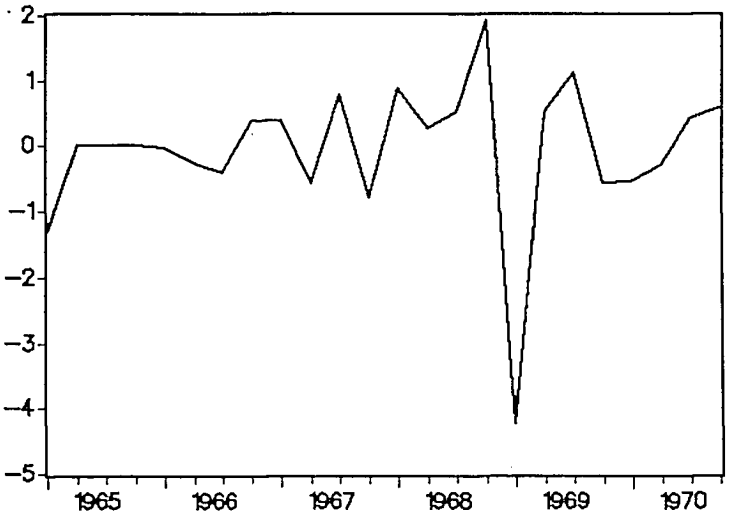


Figure 4.7a Consumption of spirits in UK (in logs)

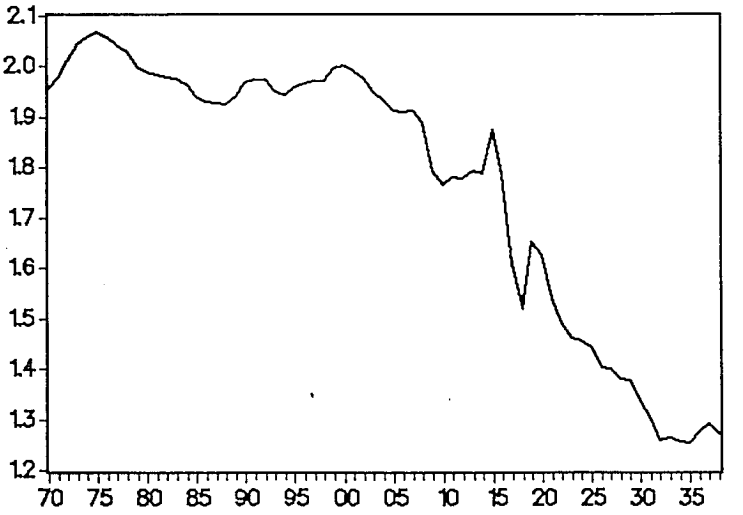


Figure 4.7b Relative price of spirits in UK (in logs)

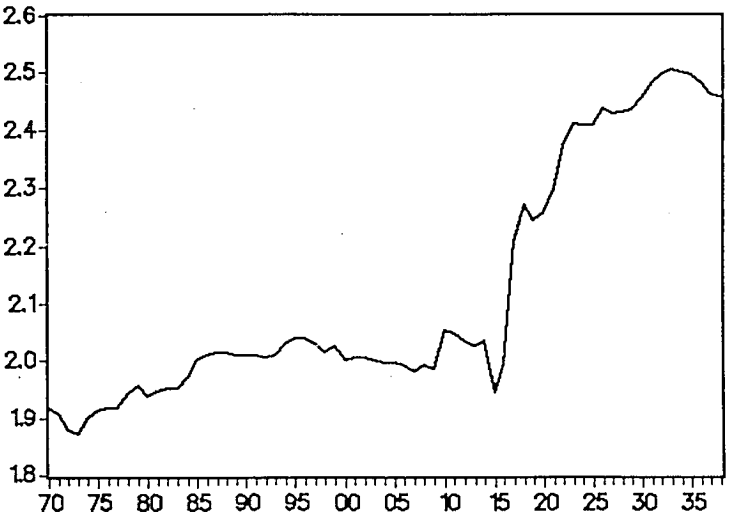


Figure 4.7c Income per capita in UK (in logs)

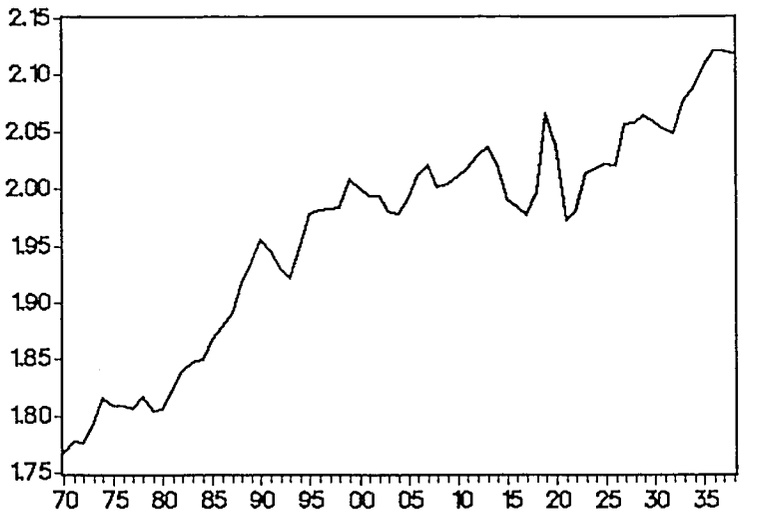


Figure 4.8a Innovations for SPIRITS

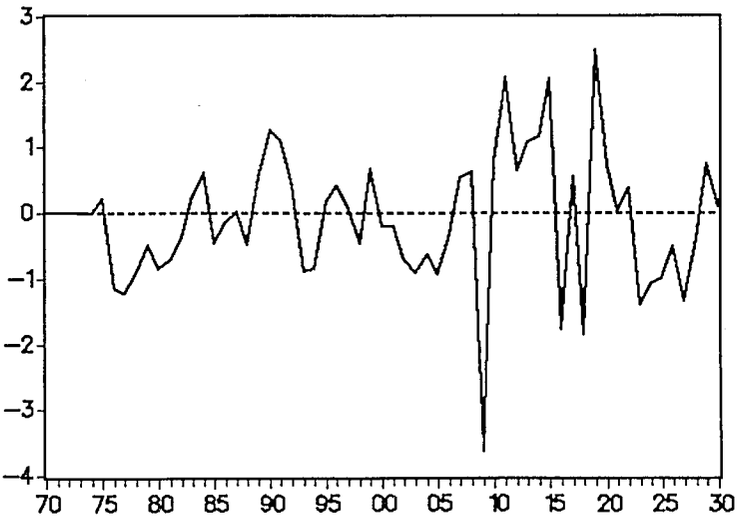


Figure 4.8b Auxiliary residuals for SPIRITS : Irregular

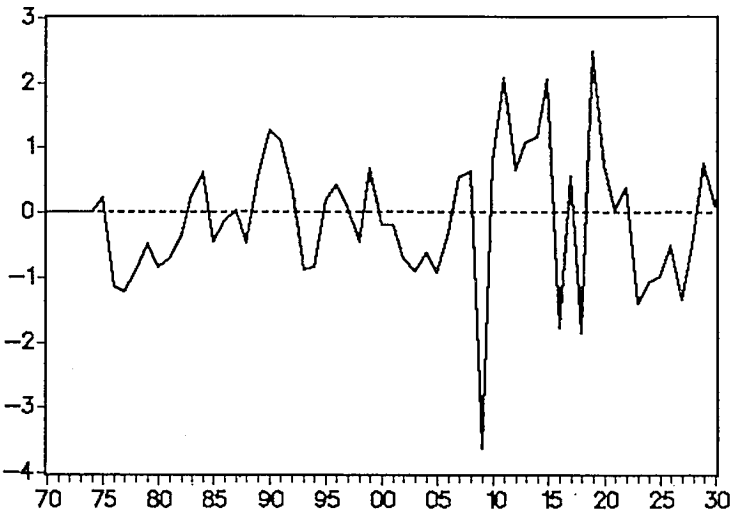
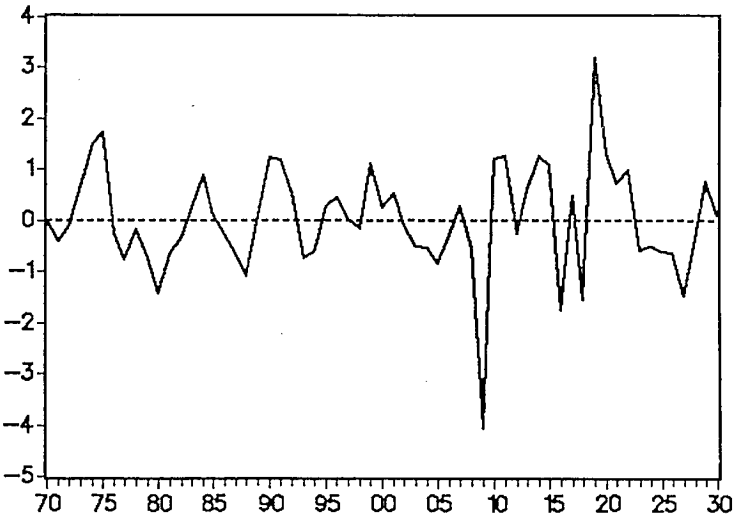


Figure 4.8c Auxiliary residuals for SPIRITS : Level residual



CHAPTER 5
CUBIC SPLINES IN TIME SERIES MODELS

0. ABSTRACT

A non-linear function can be approximated by a cubic spline function which can be regarded as a polynomial function of order three. So the first three derivatives of a cubic spline function exist but only the first two derivatives are continuous. Splines are generally used for interpolation and curve fitting but also for piecewise regressions. This chapter deals with cubic splines mainly in the context of time series models. They are adopted to describe seasonal or cyclical movements in time series but also to approximate non-linear responses of explanatory variables. An important contribution of this chapter is to let cubic spline functions vary over time and to place them into the state space form.

Keywords : Cubic splines; Piecewise regression; Periodic splines; Seasonality; Time-varying.

1. INTRODUCTION

The spline method is a stable and flexible interpolation technique which is especially useful to describe a pattern of data points generated from a particular unknown function. In fact, it draws a smooth line through a set of points in a XY diagram, see figure 5.1. The spline function can be chosen as smooth as one desires. This depends mainly on the number of existing derivatives. For example, if the first p derivatives of a spline function exist, the spline function is of order p. If p is large, the spline function is very smooth. However, a spline function with a relatively small value for p produces already a smooth line. The

linear spline function has p equal to 1, the quadratic spline function has p equal to 2, and the cubic spline function has p equal to 3. The latter is of primary concern in this chapter.

Several applications of splines can be found in the literature of numerical analysis and they are mostly concerned with interpolation and curve-fitting. Other applications are related to particular problems in regression analysis such as structural breaks which can be modelled via a piecewise linear regression model as well, see Johnston (1984). Spline functions allow for a smooth change from one parameter value to another and preserve continuity of derivatives upto a certain order. This chapter will develop more theoretical results concerning cubic splines in the context of time series.

Periodic movements are an important feature in time series analysis. Seasonal effects tend to be the rule rather than the exception with monthly or quarterly time series. When observations are available at more frequent intervals, periodic movements may exist within the week or the day. Commonly, a periodic pattern over s time intervals is modelled with s-1 dummy variables or, alternatively, with s-1 trigonometric terms. Two important issues arise with this approach. Firstly, periodic movements are almost never fixed in real time series, they may evolve over time. Secondly, if s is large, the model with dummy or trigonometric terms become cumbersome and not very parsimonious. The use of a cubic spline is in many of these cases a more appropriate solution for modelling periodic time series. The main purpose of this chapter is to address these problems and to provide a solution by developing the idea of time-varying splines.

The organisation of this chapter is as follows. The next section will derive the cubic spline function as a linear system of equations. The derivation follows mainly the steps which can be found in various books on introductory numerical analysis. An illustration describes the method of interpolation using cubic splines. Statistical applications with splines are widely available in the literature. Section 3 gives a straightforward approach of calculating a polygon through a histogram using cubic splines. Also, cubic splines are incorporated in regression models to approximate

non-linear responses of explanatory variables, illustrated by an example that deals with the non-linear response of the demand of electricity on temperature levels. Moreover, it introduces distributed lags for these regression models and it provides a parsimonious solution. Section 4 incorporates periodic splines into regression models. This spline is restricted to a 'sum-up-to-zero' constraint. An efficient method of computing the periodic spline is proposed as well. The main novelty in this chapter is the development of time-varying splines as can be found in section 5. Some essential computer programs are described in section 6. This chapter is concluded with some general remarks.

2. CUBIC SPLINES

This section presents the derivation of a cubic spline function g(x) that approximates an unknown function f(x) by a set of cubic polynomial functions. Thus the functional form of f(x) is unknown but it is assumed that some particular values of f(x) are known. The set of known pair values $\{x_i^+, y_i^+=f(x_i^+)\}$, for i=0,...,k, will be referred to as the set of knot points associated with the, so-called, mesh $\{x_0^+,\ldots,x_k^+\}$. The (k+1) knot points can be graphically reproduced in a XY diagram, see figure 5.1. The spline approximation method provides the smooth line through the knot points.

The spline function g(x) is based on a set of polynomial functions. The spline function is said to be of order p when the first p derivative functions exist and the first (p-1) derivative functions are continuous. The value of p can be interpreted as an order of smoothness. In many applications it is appropriate to have p equals to 3. This particular spline function is referred to as a cubic spline and can be derived in an analytically attractive way. It is shown below that the cubic spline function is derived as a set of linear equations.

Assume that (i) the mesh is of ascending order such that $x_0^+<x_1^+<\ldots<x_k^+$, (ii) the distance z_i is defined as $z_i=x_i^+-x_{i-1}^+$ for i=1,...,k and (iii) the i-th derivative of g(x) is denoted by $\nabla^i(x)$. In order to let the third derivative function exist, the second

derivative function can be taken as a linear function for every area between the mesh-elements. Note that it is not required that the third derivative is continuous, it only has to exist. An example of a second derivative function $\nabla^2(x)$ for a cubic spline is graphically reproduced in figure 5.2b. The functional form of $\nabla^2(x)$ is locally, for $x^+_{j-1} \leq x \leq x^+_j$, given by

$$\nabla^2_j(x) = \{(x^+_j - x)/z_j\}a_{j-1} + \{(x - x^+_{j-1})/z_j\}a_j \qquad (2.1)$$

for $j=1,\ldots,k$. The values a_0 and $a_j = \nabla^2_j(x^+_j)$ for $j=1,\ldots,k$ are not known. The second derivative $\nabla^2(x)$ of the cubic spline function $g(x)$ is continuous as required. The third derivative function is discontinuous and it is given by

$$\nabla^3_j(x) = (a_j - a_{j-1})/z_j \qquad (2.2)$$

for $x^+_{j-1} \leq x \leq x^+_j$ and $j=1,\ldots,k$. This can be interpreted as the slope of the second derivative function.

In the following, the cubic spline function $g(x)$ is derived in four steps starting off from the definition of the second derivative function. The first step derives the first derivative function and the cubic spline function $g(x)$ using standard rules of integration. It follows that the functions $g(x)$ and $\nabla(x)$ are unknown because they are expressed in terms of the unknown values a_j for $j=0,\ldots,k$. The second step restricts the cubic spline function $g(x)$ to cross the knotpoints as required. The continuity restriction for the first derivative function is enforced in the third step. The first three steps provide a system of $(k-1)$ equations with $(k+1)$ unknown values for a_j, $j=0,\ldots,k$. The last step solves this under-identification by treating a_0 and a_k as known constants. By solving the system of $(k-1)$ equations, the functions $g(x)$, $\nabla(x)$, $\nabla^2(x)$ and $\nabla^3(x)$ can be evaluated and the associated graphs can be drawn, see figures 5.1 and 5.2.

Step (i)

The primitive function of $\nabla^2_j(x)$ and $\nabla_j(x)$ is derived by using standard integration rules. Thus, we have $\nabla_j(x) = \int \nabla^2_j(x)\,dx$ and

$$\nabla_j(x) = -\{\tfrac{1}{2}(x_j^+-x)^2/z_j\}a_{j-1} + \{\tfrac{1}{2}(x-x_{j-1}^+)^2/z_j\}a_j - b_j + c_j \qquad (2.3)$$

where c_j-b_j is a constructed integration constant. Again, the primitive of $\nabla_j(x)$ is $g_j(x)$ and is derived in a similar fashion,

$$g_j(x) = \{(x_j^+-x)^3/6z_j\}a_{j-1} + \{(x-x_{j-1}^+)^3/6z_j\}a_j + (x_j^+-x)b_j + (x-x_{j-1}^+)c_j + d_j \qquad (2.4)$$

where d_j is the integration constant. Note that the cubic spline function $g(x)=g_j(x)$ for $x_{j-1}^+ \le x \le x_j^+$ and $j=1,\ldots,k$.

Step (ii)

It is required that the cubic spline crosses the knotpoints and, therefore, it is enforced that $g_j(x_j^+)=y_j^+$ and $g_j(x_{j-1}^+)=y_{j-1}^+$ for $j=1,\ldots,k$. This implies that the constants b_j, c_j and d_j must be chosen as

$$b_j = y_{j-1}^+/z_j - \{z_j/6\}a_{j-1} \qquad (2.5a)$$
$$c_j = y_j^+/z_j - \{z_j/6\}a_j \qquad (2.5b)$$
$$d_j = 0 \qquad (2.5c)$$

Substituting the constants b_j and c_j into the first derivative function leads to

$$\nabla_j(x) = -\{\tfrac{1}{2}(x_j^+-x)^2/z_j - z_j/6\}a_{j-1} + \{\tfrac{1}{2}(x-x_{j-1}^+)^2/z_j - z_j/6\}a_j + (y_j^+-y_{j-1}^+)/z_j \qquad (2.6)$$

and substituting the constants b_j, c_j and d_j into the spline function $g(x)$ leads to

$$g_j(x) = [(x_j^+-x)\{(x_j^+-x)^2-z_j^2\}/6z_j]a_{j-1} + \{(x_j^+-x)/z_j\}y_{j-1}^+ + [(x-x_{j-1}^+)\{(x-x_{j-1}^+)^2-z_j^2\}/6z_j]a_j + \{(x-x_{j-1}^+)/z_j\}y_j^+ \qquad (2.7)$$

for $x_{j-1}^+ \le x \le x_j^+$ and $j=1,\ldots,k$.

Step (iii)

Finally, it is also required that the first derivative of the

cubic spline function is continous. This is enforced by the restriction $\nabla_j(x_j^+)=\nabla_{j+1}(x_j^+)$ where

$$\begin{aligned}\nabla_j(x_j^+) &= \{2z_j/6\}a_j + \{z_j/6\}a_{j-1} + \{y_j^+-y_{j-1}^+\}/z_j \\ -\nabla_{j+1}(x_j^+) &= \{2z_{j+1}/6\}a_j + \{z_{j+1}/6\}a_{j+1} + \{y_j^+-y_{j+1}^+\}/z_{j+1}\end{aligned} \tag{2.8}$$

for j=1,..,k-1. By some minor manipulation, the set of restrictions leads to the following set of (k-1) equations

$$\begin{aligned}&\{z_j/(z_j+z_{j+1})\}a_{j-1} + 2a_j + \{z_{j+1}/(z_j+z_{j+1})\}a_{j+1} = \\ &6y_{j-1}^+/\{z_j(z_j+z_{j+1})\} - 6y_j^+/z_jz_{j+1} + 6y_{j+1}^+/\{z_{j+1}(z_j+z_{j+1})\}\end{aligned} \tag{2.9}$$

for j=1,...,k-1.

Step (iv)

The previous steps lead to a sequence of k+1 unknown values, i.e. a_j for j=0,...,k, and a system of k-1 restrictions. This system of linear restrictions cannot be solved unless two linear restrictions are added. A number of possibilities are discussed in Poirier (1973,1976). A straightforward and, in many cases, an appropriate solution is to treat the unknowns a_0 and a_k as known constants. A cubic spline function g(x) is referred to as a natural spline function when it is supposed that $a_0=a_k=0$.

The final set of restrictions can be represented in matrix notation by

$$Pa = Qy^+ \tag{2.10}$$

where $a=(a_0;...;a_k)$, $y^+=(y_0^+;...;y_k^+)$ and the matrices P and Q are tridiagonal matrices with the non-zero elements chosen appropriately according to (2.9). The solution for a is given by $P^{-1}Qy^+$ and this enables us to calculate g(x) for any $x_0^+\leq x\leq x_k^+$, see figure 5.1.

Since, the parameter vector a is expressed as a linear combination of y^+, it follows that the solution for g(x) is also linear in y^+. In matrix notation,

$$g(x) = r'y^+ + s'a = w'y^+ \tag{2.11}$$

where row vector w' is equal to $r'+s'P^{-1}Q$ and r and ' are column vectors with the non-zero elements chosen appropriately according to (2.7). If x corresponds to an element of the mesh, $\{x_0^+,\ldots,x_k^+\}$, then all elements of w are zero except the j-th element is equal to unity such that $g(x_j^+)=y_j^+$. In that case, s=0 and r=w.

In order to calculate the elements of the w vector for a particular x-value, the inverse of P is required. This can be done very efficiently because the matrix is tridiagonal, see Press et.al. (1989, p.48). Appendix 5A gives the details to calculate the vector w based on a particular x value and a set of knots.

An illustration

Suppose, weather data for the year 1990 of a small weather station in England is lost. Only the average monthly temperature values for January, May, August and December are found. These values are 30F, 45F, 65F and 35F, respectively. The head of the weather station decides to approximate the missing monthly averages by a natural cubic spline.

The month January is coded with number 1 and the month December is coded with number 12 such that the knot points are (1,30), (5,45), (8,65) and (12,35). The w' vectors for the set of x-values 1,2,...,12 are, respectively,

1	1.0000	0.0000	0.0000	0.0000
2	0.6798	0.4338	-0.1287	0.0150
3	0.3877	0.7941	-0.2059	0.0241
4	0.1517	1.0074	-0.1801	0.0211
5	0.0000	1.0000	0.0000	0.0000
6	-0.0517	0.7386	0.3497	-0.0365
7	-0.0365	0.3497	0.7386	-0.0517
8	0.0000	0.0000	1.0000	0.0000
9	0.0211	-0.1801	1.0074	0.1517
10	0.0241	-0.2059	0.7941	0.3877
11	0.0150	-0.1287	0.4338	0.6798
12	0.0000	0.0000	0.0000	1.0000

and the monthly interpolated temperature values are, respectively, 30.0, 32.1, 34.8, 38.9, 45.0, 53.1, 60.8, 65.0, 63.3, 56.6, 46.7 and 35.0. These values are graphically reproduced in figure 5.3.

3. STATISTICAL APPLICATIONS OF CUBIC SPLINES

Cubic splines can be applied to several problems in statistical theory. In this section, three different applications are discussed. Firstly, the cubic spline is used to obtain a frequency polygon from a frequency histogram. This application provides several ideas which can be explored in a future research project. The second application considers piecewise regressions with cubic splines as proposed by Poirier (1973). This approach is especially useful to approximate non-linear effects in regression models. The last application develops piecewise regressions with cubic splines for regression models including distributed lags. Specific applications of cubic splines in the context of time series models are discussed in the sections 4 and 5.

The frequency polygon as a cubic spline based on a histogram

The frequency distribution for a variable x is given in table 5.1 and can be graphically reproduced as a frequency histogram, see figure 5.4. This graph is discontinuous. When a continuous graphical presentation of the frequency distribution is preferable, the polygon might be considered. This is the line through the midpoints of the classes. Van Casteren (1991) discusses some fundamental objections against this common polygon (cp) : " The cp is not constructed such that the surface under the polygon within each class is proportional to the frequency of that class ... the cp is too flat ... the cp may allocate positive surface to imaginary classes ". He proposes an alternative polygon based on a linear spline that takes account of these incompatibilities.

In the following, a more smooth polygon is proposed based on cubic splines. Spline functions are earlier used as a frequency polygon, for example, Boneva et.al. (1971) present an elaborate discussion on the, so-called, histo-splines. However, the approach below is analytically transparent and the arguments are straightforward to comprehend.

In this application, the mesh-values of the cubic spline can

be chosen freely and, therefore, they are known, but the corresponding y-values of the knot-points are not known. The number of elements in the mesh and their particular values can be chosen independently from the given classes within the total range of x-values. However, to reproduce the frequency distribution as well as possible, in a continuous fashion, the knot positions are placed at the beginning of all the classes and at the end of the last class, in our example {0,10,20,30,50}. The cubic spline is constructed in such a way that the surface under the cubic spline is equal to the surface of the histogram within all classes. The surface under the cubic spline between x_{j-1}^+ and x_j^+ is obtained from the primitive function $G_j(x) = \int g_j(x)dx$ for $x_{j-1}^+ \leq x \leq x_j^+$ where $j=1,\ldots,k$ and k is the number of classes (in our example k=4). Applying standard integration gives

$$\begin{aligned} G_j(x) = {} & [z_j(x_j^+-x)^2/12 - (x_j^+-x)^4/24z_j]a_{j-1} \\ & - [z_j(x-x_{j-1}^+)^2/12 - (x-x_{j-1}^+)^4/24z_j]a_j \\ & - y_{j-1}^+(x_j^+-x)^2/2z_j + y_j^+(x-x_{j-1}^+)^2/2z_j + e_j \end{aligned} \qquad (3.1)$$

where e_j is the integration constant. The surface under the cubic spline between x_{j-1}^+ and x_j^+ is calculated by $G_j(x_j^+) - G_j(x_{j-1}^+)$ which is equivalent to

$$z_j(y_j^+ + y_{j-1}^+)/2 - z_j^3(a_j + a_{j-1})/24 \qquad (3.2)$$

The surface (3.2) is restricted to be the same as the surface f_j of the histogram between x_{j-1}^+ and x_j^+ for $j=1,\ldots,k$. This set of k equations with two (k+1)x1 unknown vectors is identified when it is assumed further that $a_0=0$. The total set of equations can be put into matrix form by

$$Ry^+ - Sa = f$$

where R and S are spatial matrices of simple structure provided by the equations (3.2) and the restriction $a_0=0$, the vector f is the stack of histogram surface values and a zero. In our example,

$$R = 5\times\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix} \quad S = \tfrac{1}{3}5^3\times\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 8 & 8 \end{bmatrix} \quad f = \begin{bmatrix} 0 \\ 200 \\ 400 \\ 300 \\ 200 \end{bmatrix}$$

In the previous section, it is shown that the cubic spline is based on the set of restriction $Pa = Qy^+$ where y^+ is supposed to be known. However, for this application, y^+ is a vector of $(k+1)x1$ unknown values. To let the cubic spline start and finish at zero at either ends of the histogram, the restrictions $y_0^+=y_k^+=0$ are added which imply a slightly different structure for the matrices P and Q. An expression for y in terms of a is easily obtained by $y^+ = Q^{-1}Pa$ which can be substituted into (3.2) such that

$$a = (RQ^{-1}P - S)^{-1}f \tag{3.3a}$$

$$y^+ = Q^{-1}P(RQ^{-1}P - S)^{-1}f \tag{3.3b}$$

The explicit solutions for the vectors a and y in our example are (0.0;-.459;-.207;.134;-.101) and (0.0;36.173;38.271;21.117;0.0), respectively. These vectors ensure that, for any x value between x_0^+ and x_k^+, that is 0 and 50, the cubic spline function can be evaluated. The polygon is graphically reproduced as a smooth line through the histogram by the cubic spline in figure 5.4.

Some specific issues arise when a cubic spline is used as a polygon. Firstly, it is possible that the cubic spline get negative at some intervals. This may happen when a class has a low frequency value and one of its two neighboring classes do have a high frequency. These undesirable negative cubic spline values can be avoided easily by forcing the negative knot points to be zero. A slightly different set of mesh values might be required. Secondly, the ogive can be calculated as well and is in fact the function $G_j(x)$. Thirdly, the sample mean and variance of x based on the calculated cubic spline polygon can be evaluated easily. More research has started, based on Monte Carlo simulations, to find out if this mean and variance are better estimates of the population mean and variance than their sample counterparts based on the histogram. All these issues are addressed in Koopman (1992).

Approximating non-linear effects in regression models

The concept of a piecewise linear regression can be found in several standard econometrics textbooks, for example Johnston (1984). Basically, if the response of an explanatory variable is non-linear, the piecewise solution approximates the response locally by a linear function within a certain range of values corresponding to the explanatory variable. A linear spline is easily implemented by the method of restricted least squares which enforces the response function to be continuous, see Johnston (1984, p.392). The piecewise regression with cubic splines generalises the linear case such that, in addition, the first and second derivative functions are continuous. This option is preferable because it provides a smoother response function. It is shown below, following Poirier (1973), that the concept of a piecewise regression with cubic splines can be parameterized as a multiple linear regression model.

Suppose a set of n pairs of observations $\{x_t,y_t\}$, t=1,...,n, are available and that a non-linear regression model is appropriate of the form

$$y_t = f(x_t) + \epsilon_t \qquad (3.4)$$

where ϵ_t is a mutually uncorrelated disturbance term with zero mean and variance σ^2. The unknown non-linear function f(x) is approximated by the cubic spline function g(x). The spline is based on the mesh $x_0^+<x_1^+<\ldots<x_k^+$. An element of the mesh (x_i^+) does not have to correspond neccesarily to a x-value of the set of pair observations (x_t). Thus the corresponding y-values of the mesh, i.e. $\gamma_0^+,\ldots,\gamma_k^+$, are supposed to be unknown. Following the derivation of a cubic spline in the previous section, the function $f(x_t)$ can be approximated by $g(x_t)$ which can be expressed as

$$g(x_t) = w_t'\gamma^+ \qquad (3.5)$$

where $\gamma^+=(\gamma_0^+;\ldots;\gamma_k^+)$ and the (k+1)x1 vector w_t depends on the mesh and x_t as is shown in the previous section. This leads to the linear multiple regression model

$$y_t = w_t'\gamma^+ + \epsilon_t \qquad (3.6)$$

It is the simplicity and the well-known properties of the linear regression model that makes the cubic spline an attractive approximation to complicated non-linear functions in regression models. Of course, given the assumptions on ϵ_t, the ordinary least squares (OLS) estimate

$$\hat{\gamma}^+ = (\sum_{t=1}^n w_t w_t')^{-1}(\sum_{t=1}^n w_t y_t) \qquad (3.7)$$

can be regarded as the minimum mean square linear estimate (MMSLE) of the cubic spline parameter vector γ^+.

An illustration

Figure 5.5 presents a scatter plot of the demand for electricity in mega-watts against the temperature in Fahrenheit at one o'clock pm for all weekdays in 1990, except holidays (n=251). The data were obtained from the Puget Sound, Power & Light company[1]. The scatterplot clearly shows that the linear regression response line is not appropriate because, obviously, the response of electricity demand on temperature is non-linear. The demand of electricity depends on temperature mainly due to heating but at high temperatures the demand increases due to air-conditioning. This non-linear response is observed earlier in Bunn & Falmer (1985). Engle et.al. (1986) propose to fit a non-parametric spline.

An appropriate treatment of this problem is to consider the non-linear regression model $y_t = f(x_t) + \epsilon_t$ where y_t is the demand for electricity, x_t is the observed temperature and ϵ_t is a normally distributed disturbance term with zero mean and variance σ^2. The function f(x) can be approximated by a cubic spline based on the mesh {10,65,90}. It is shown above that this approach leads to the linear multiple regression model $y_t = w_t'\gamma + \epsilon_t$ where vector γ is fixed but unknown and the vector of natural cubic spline weights w_t corresponds to x_t, see (3.5). The square root of the average squared residual $q=[n^{-1}\Sigma_{t=1}^n(y_t-\hat{y}_t)^2]^{\frac{1}{2}}$ is 125.65. For the ordinary regression model $y_t = \alpha + \beta x_t + \epsilon_t$, this value is q=197.58. An adjusted Akaike information criterion (AIC) can be defined by 2(log q + k/n), where k is the number of parameters. The AIC is usually applied for model selection between non-nested models. For the two estimated

regression models, the AIC values are 9.69 and 10.59, respectively, such that it is concluded that the cubic spline regression model is preferable in the sense of a minimum AIC. The cubic spline function is plotted in figure 5.6.

Approximating dynamic non-linear response functions

The contemporaneous non-linear response model of the previous section can be generalised by including lagged response functions. Consider the model

$$y_t = \sum_{i=0}^{M} f_i(x_{t-i}) + e_t \qquad (3.8)$$

where M is maximum lag window. The (M+1) non-linear response functions can be approximated by cubic splines, such that

$$y_t = \gamma_t + e_t \qquad (3.9)$$

where $\gamma_t=\sum_{i=0}^{M} \gamma_{i,t}$ and $\gamma_{i,t}=g_i(x_{t-i})=w_{i,t}'\gamma_i^+$ for $i=0,\ldots,M$, the vector $w_{i,t}$ is based on a mesh of dimension (k_i+1) and the value x_{t-i}, the parameter vector γ_i^+ has dimension (k_i+1). This multiple linear regression model with distributed lags can be estimated by methods based on ordinary least squares or maximum likelihood. However, severe problems may arise with the estimation of the parameters because of collinearity which leads to imprecise estimates. Moreover, the number of parameters can be quite large with this specification such that a more parsimonious solution may be looked for.

The regression model (3.9) can be simplified when it is assumed that

$$\gamma_i^+ = \lambda_i \gamma^+ \qquad (3.10)$$

and $k_i=k$ for $i=0,\ldots,M$. The scalar λ_i rescales the fixed parameter vector γ^+. This implies that the shape of the non-linear response function is constant for all lags and that only the intensity of the

response is different. For a number of applications, this assumption seems to be valid upto some degree. Note, that this restriction also implies that $w_{i,t}=w_{0,t-i}$, for i=1,...,M, such that the contemporaneous vector of spline weights is redefined by $w_t=w_{0,t}$. The restricted component γ_t of the dynamic regression model (3.9) becomes

$$\gamma_t = \sum_{i=0}^{M} \lambda_i w_{t-i}'\gamma^+ \qquad (3.11)$$

which contains a set of (k+M+2) unknown parmeters. A further reduction in the number of parameters is obtained by requiring that $\sum_{j=0}^{M}\lambda_j = 1$. This restriction is enforced to conform to some normalizing argument. This implies that (k+M+1) number of free parameters remain to be estimated. Two problems arise with estimation. (i) Since no restrictions are introduced in the lag structure, such as polynomial distributed lags proposed by Almon (1965), collinearity is likely to be high between the vectors of weights w_t, $w_{t-1},\ldots,w_{t-M}$. (ii) The model (3.9) and (3.11) cannot be estimated straightforwardly because the two sets of parameters do not relate to each other in a linear fashion. Both problems will be addressed below.

(i) To avoid multicollinearity in the regression model (3.9) and (3.11) and, therefore, to increase numerical stability, it is useful to re-parameterize γ_t by

$$\gamma_t = \lambda^*(w_t'\gamma^+) + \sum_{i=0}^{M-1} \lambda_i^*(\Delta w_{t-i}'\gamma^+) \qquad (3.12)$$

where Δ is the first difference operator, $\lambda^* = \Sigma_{j=0}^{M}\lambda_j = 1$ and $\lambda_i^* = -\Sigma_{j=i+1}^{M}\lambda_j$ for i=0,...,M-1. Thus, γ^+ can be considered as the total (non-linear) multiplier.

(ii) The estimation of the regression model (3.9) and (3.11) can be carried in a similar iterative estimation process as proposed by Cochrane & Orcutt (1949) in the context of regression models with AR disturbances. It involves a sequence of least squares estimations based on two different linear regression models. The first regression specification RM1 is given by

$$y_t = w_t^{*\prime}\gamma^+ + \epsilon_t \qquad (3.13)$$

where $w_t^* = w_t + \sum_{i=0}^{M-1} \lambda_i^* \Delta w_{t-i}$. The second regression specification RM2 is defined in terms of the set of unknown scalars $\{\lambda_0^*;\ldots;\lambda_{M-1}^*\}$, that is

$$y_t - (w_t'\gamma^+) = \sum_{i=0}^{M-1} \lambda_i^* z_{i,t} + \epsilon_t \qquad (3.14)$$

where $z_{i,t} = \Delta w_{t-i}'\gamma^+$ for i=0,...,M-1. A least squares estimation applied to RM1, where $\lambda_i^* = M^{-1}$ for i=0,...,M-1, followed by a least squares estimation applied to RM2, where γ^+ is replaced by its estimate from the previous regression applied to RM1, starts up a process of 'flip-flop' estimations applied to RM1 and RM2 which can be terminated when the estimates of the parameters converge to a set of constant values. This approach of estimating lagged splines, applied to the case of forecasting hourly electricity demand where temperature is the explanatory variable, is discussed in the next chapter.

4. PERIODIC CUBIC SPLINES AND SEASONALITY

Cyclical and seasonal movements do occur regularly in time series and are commonly modelled by a set of dummy variables or trigonometric terms. This approach is used in many cases of applied time series analysis and it has proved to be successful. However, when the periodic pattern repeats itself over a large number of time intervals, the classic approach is far from parsimonious. For example, when weekly observations are available for many years and the data show a periodic movement which repeats itself every year, the classic approach requires 51 dummy or trigonometric terms. It will be shown below that for these cases a seasonal model can be formulated with much less parameters by using periodic cubic splines.

periodic cubic splines

Periodic variations can be described by cubic splines. Because the periodic pattern repeats itself several times, it is only required to describe one period by a spline which, subsequently, can be applied to all periods. The cubic spline becomes a continuous cyclical function when the knot points at the start and at the end of the cubic spline are the same. This is enforced by the assumption that $y_0^+=y_k^+$. Moreover, the first and second derivatives at either side of the cubic spline are restricted to be equal such that the continuity conditions for the derivatives upto order 2 still hold. It will be shown below that these additional restrictions remove the need for the arbitrarily restrictions $a_0=a_k=0$ as applied to natural cubic splines in order to have an identified system of equations.

Thus, in the case of periodic splines, it is required that

$$\nabla_1(x_0^+)=\nabla_k(x_k^+) \qquad \nabla_1^2(x_0^+)=\nabla_k^2(x_k^+) \tag{4.1}$$

These continuity conditions remove the need for any further assumptions to solve the system (2.9) with (k-1) equations and (k+1) unknowns because the latter condition leads to a reduction of one unknown, that is $a_0=a_k$, and the former condition add the k-th equation as given by

$$\{z_k/(z_k+z_1)\}a_{k-1} + 2a_k + \{z_1/(z_k+z_1)\}a_1 = 6y_{k-1}^+/\{z_k(z_k+z_1)\} - 6y_k^+/z_kz_1 + 6y_1^+/\{z_1(z_k+z_1)\} \tag{4.2}$$

Note that it is also assumed that $y_0^+=y_k^+$. This linear system of k equations and k unknowns can also be expressed in the matrix formulation $Pa=Qy^+$ but now the vectors are defined as $a=(a_1;\ldots;a_k)$ and $y^+=(y_1^+;\ldots;y_k^+)$. The periodic spline function can now be evaluated by a linear operation, that is $g(x) = r'y^+ + s'a = w'y^+$ where the vectors r and s are based on (2.7) but taking into account the restrictions $a_0=a_k$ and $y_0^+=y_k^+$. The weight vector w for a periodic cubic spline is still defined as $w'=r'+s'P^{-1}Q$. However, the matrices P and Q have lost their tridiagonal structure because of the equation

(4.2). However, the sparsity of matrix P can still be exploited to calculate its inverse efficiently and to avoid the use of a standard matrix inverse procedure. This is important for the computer algorithm that calculates the weight vector for a periodic cubic spline. Technical details are given in Appendix 5B.

Regression model with seasonal effects

Suppose that a sequence of observations follow a pattern which repeat itself over a stretch of s observations and that the appropriate regression model is the deterministic basic structural time series model as given by

$$y_t = \alpha + \beta t + \gamma_t + \epsilon_t \qquad (4.3)$$

where ϵ_t is a disturbance term with zero mean and variance σ^2. The level constant α and the fixed slope parameter β are unknown. The seasonal effect γ_t is usually modelled by a set of dummy variables, see Johnston (1984). This approach is rather satisfactory provided that s is not large. For example, s being equal to 4 (quarterly) or 12 (monthly) does not give any serious cause for concern. However, when s is equal to 24 (hourly) or 52 (weekly), a more parsimonious solution is needed. In the following, the periodic cubic spline is considered as an alternative to dummy and trigonometric variables.

In order to parameterize the seasonal pattern by a periodic cubic spline, consider the periodic spline of the previous section and take $x_j=j$ for $j=1,\ldots,s$ and $k<s=x_k^+$. The seasonal component is defined by $\gamma_t=\gamma_j$ when a j-th seasonal effect is appropriate and

$$\gamma_j=w_j'\gamma^+ \qquad (4.4)$$

where w_j is the periodic spline weight vector depending on the mesh and the index value j. The kx1 vector $\gamma^+=(\gamma_1^+;\ldots;\gamma_k^+)$ is the spline parameter vector which elements correspond to the mesh of the periodic spline.

To avoid the problem of multi-collinearity with the constant term α, the periodic effects are restricted to sum up to zero. This

problem does exist for a dummy and trigonometric seasonal model as well. For the periodic cubic spline, the restriction leads to

$$\sum_{j=1}^{s}\gamma_j = \sum_{j=1}^{s} w_j'\gamma^+ = w_*'\gamma^+ = 0 \qquad (4.5a)$$

where w_* is the kx1 vector

$$w_* = \sum_{j=1}^{s} w_j \qquad (4.5b)$$

This restriction is equivalent to the equation

$$\gamma_k^+ = -\sum_{j=1}^{s-1}(w_{*j}/w_{*k})\gamma_j^+ = -(w_{**}/w_{*k})'\gamma_*^+ \qquad (4.6)$$

where w_{*i} is the i-th element of the vector w_* and the (k-1)x1 vectors w_{**} and γ_*^+ are defined as $w_{**}=(w_{*1};\ldots;w_{*k-1})$ and $\gamma_*^+=(\gamma_1^+;\ldots;\gamma_{k-1}^+)$, respectively. The previous steps lead to the following specification of the j-th seasonal effect

$$\gamma_j = z_j'\gamma_*^+ \qquad (4.7)$$

where z_j is a (k-1)x1 vector given by $z_j=w_{j*}-(w_{jk}/w_{*k})w_{**}$ and $w_{j*}=(w_{j1};\ldots;w_{j(k-1)})$ such that w_{ji} is the i-th element of w_j and $w_j=(w_{j*};w_{jk})$. The calculation of vector z_j is straightforward when the vector sequence w_j is available for j=1,...,s.

In the special case that k=s, the vector w_j is a vector of zeroes except its j-th element is equal to unity. Thus, the stack of all weight vectors w_j', for j=1,...,s, is equal to the identity matrix, e.g. W=I. It is not surprising that in this case the regression model with periodic splines collapses to the dummy seasonal regression model with $\gamma_j=\gamma_j^+$ for j=1,...,s.

An alternative method to model seasonality in a regression model is to use a set of (s-1) trigonometric terms as mentioned earlier. It is possible to cut down the number of trigonometric terms by dropping the ones, related to some frequency level, which are not significant in the seasonal pattern of the observed data. This not a viable option when the seasonal pattern exhibits sharp peaks as, for example, typically happens with sales of certain

consumer goods which tend to be concentrated in the weeks immediately before Christmas. It might be an entirely satisfactory option for a slowly changing seasonal effect such as average weekly temperature. The latter case is discussed as a part of the next illustration.

Illustration

Consider a time series of weekly averages of daily temperature measurements at noon from January 1985 to August 1990 as given in figure 5.7. Thus, s=52 and the number of observations is n=347. The cyclical behaviour is the prominent feature of the time series. A regression model with a set of 52 seasonal dummies can be considered or, alternatively, the seasonal variation can be modelled by a periodic cubic spline based on the mesh of week numbers {10,22,32,42,52} such that k=5. The seasonal patterns generated by seasonal dummies and by a periodic cubic spline are given in figure 5.8. The periodic cubic spline follows the seasonal dummies closely and smoothly.

A simple F-test shows that the null hypothesis of the 'restricted' model with a periodic cubic spline is not rejected. The test is constructed as follows. Denote q_0^2 as the average of the squared residual for the periodic cubic spline regression model, $q_0^2=4.04$, and denote q_1^2 as the average of the squared residual of the dummy seasonal regression model, $q_1^2=3.88$. The latter model has n-s degrees of freedom and the former model is of the same class of the latter model but with s-k linear restrictions imposed on it. The test criterion

$$F^* = \{ (q_0^2 - q_1^2)/q_1^2 \} * (n-s)/(s-k) ,$$

is, under the null hypothesis, F-distributed with (s-k,n-s)=(47,295) degrees of freedom. The test value F^* is 0.26 which is much less than the critical value of the F-distribution 1.48 with 95% confidence. Note that this F-ratio has a direct relationship with other test criterions such as the AIC, see Maddala (1988, table 12.4, p.431).

The cyclical variation can be modelled by a reduced set of trigonometric terms as well. To compare both specifications, a

regression model is fitted with four trigonometric terms, i.e.

$$y_t = \beta'(1;\cos 2\pi t/s;\sin 2\pi t/s;\cos 4\pi t/s;\sin 4\pi t/s) + \epsilon_t$$

where β is a (5x1) vector of coefficients. The average squared residual of the estimated trigonometric regression model is equal to $q^2=4.043$. This result is very close to the cubic spline model with five knots. It is concluded that the trigonometric and the cubic spline solution are competitive when the seasonal variation is smooth.

5. TIME-VARYING CUBIC SPLINES

The illustration of the previous section shows that in specific cases cyclical or seasonal variations in time series can be modelled successfully by a periodic cubic spline. Typically, in the context of time series models, it is argued before that components such as trend and seasonal may change over time, see the discussion on the COAL series in section 2.7. The illustration of the previous section shows clearly a periodic variation in the time series of average temperatures but it is far from a constant periodic behaviour. Therefore, it is important to let this periodic cubic spline change over time.

Consider the basic structural time series model where, instead of a dummy or trigonometric seasonal, the seasonal component is modelled as a periodic cubic spline as discussed in the previous section. However, now the spline is supposed to evolve over time. The spline vector γ^+ is assumed to follow a vector of random walks, that is

$$\gamma_t^+ = \gamma_{t-1}^+ + \chi_t \qquad (5.1)$$

where χ_t is a disturbance vector with zero mean and covariance matrix $\sigma^2 V$. Note that seasonal effect γ_t is equal to $w_j'\gamma_t^+$ where period j is prevailing at time t. This specification can be compared with a time-varying regression model.

Again, the restriction is imposed that the sum of the

seasonal components sum up to zero. Therefore, it is enforced that $w_*'\gamma_t^+=0$ where w_* is defined in (4.5b). This restriction leads to

$$w_*'\gamma_t^+ = w_*'\gamma_{t-1}^+ + w_*'\chi_t \tag{5.2}$$

such that also the term $w_*'\chi_t$ must be restricted to zero. In the previous section it is shown how the 'sum-to-zero' restriction for γ^+ is implemented by dropping the last element of γ^+ and transforming the weight vector w_i into z_i. Similarly, the i-th seasonal effect at time t is measured by

$$\gamma_{it} = z_t'\gamma_{*t}^+ \tag{5.3}$$

where z_i is specified in the previous section and the (k-1)x1 periodic spline vector γ_{*t}^+ is given by $(\gamma_{1t}^+;\ldots;\gamma_{(k-1)t}^+)$ and γ_{it}^+ is the i-th element of the vector γ_t^+. The time-variation for the periodic spline vector is now reduced to

$$\gamma_{*t}^+ = \gamma_{*(t-1)}^+ + \chi_{*t} \tag{5.4}$$

where χ_{*t} is a disturbance vector with zero mean and covariance matrix $\sigma^2 V_*$.

The 'sum-to-zero' restriction for the disturbance vector χ_t implies that

$$\chi_{kt} = -(w_{**}/w_{*k})'\chi_{*t} \tag{5.5}$$

where χ_{it} is the i-th element of the vector χ_t and $\chi_{*t}=(\chi_{1t};\ldots;\chi_{(k-1)t})$ such that $\chi_t=(\chi_{*t};\chi_{kt})$. The cumulative spline weight vector w_* can be decomposed by $(w_{**};w_{*k})$ where $w_{**}=(w_{*1};\ldots;w_{*k-1})$ and w_{*i} is the i-th element of vector w_*, see equation (2.6). This restriction imposes a special covariance matrix. The matrix V can be partioned into $(V_*,v_k;v_k',v_{kk})$ where $\sigma^2 V_*$ is the covariance matrix of disturbance vector χ_{*t} and

$$v_k = \mathrm{Cov}(\chi_{*t},\chi_{kt})/\sigma^2 = -V_*(w_{**}/w_{*k})$$
$$v_{kk} = \mathrm{Var}(\chi_{kt})/\sigma^2 = (w_{**}/w_{*k})'V_*(w_{**}/w_{*k})$$

since the covariance matrix $\sigma^2 V$ is not necessarily symmetric anymore, the elements of the periodic spline vector γ_t^+ do not possess the same statistical properties. This is undesirable and, therefore, the covariance matrix $\sigma^2 V_*$ must be chosen in such a way that V is a symmetric matrix. This is achieved by assuming

$$V_* = I - w_{**}w_{**}'/w_*'w_* \tag{5.6}$$

because this implies a symmetric matrix V, i.e. $V=I-w_*w_*'/w_*'w_*$. To show this is true, it only has to be shown that $(v_k;v_{kk})$ is equivalent to the last column of V. It follows that

$$\begin{aligned} v_k &= -V_*(w_{**}/w_{*k}) = w_{**}\{(w_{**}'w_{**}/w_*'w_*) - 1\}/w_{*k} = -w_{**}(w_{*k}/w_*'w_*) \\ v_{kk} &= (w_{**}/w_{*k})'V_*(w_{**}/w_{*k}) = w_{**}'w_{**}\{1 - (w_{**}'w_{**}/w_*'w_*)\}/w_{*k}^2 = \\ &= w_{**}'w_{**}/w_*'w_* = 1 - w_{*k}^2/w_*'w_* \end{aligned}$$

Thus, all elements of the time-varying periodic spline vector γ_t^+ have the same properties if $V_*=I-w_{**}w_{**}'/w_*'w_*$.

It is mentioned in the previous section that, for the special case when k=s, the stack of spline weight vectors, W, is an identity matrix. In the case of time-varying periodic cubic splines for the seasonal component, the model reduces to the seasonal model as advocated by Harrison & Stevens (1976). Therefore, the time-varying periodic cubic spline can be regarded as a generalization of the seasonal component of Harrison & Stevens (1976).

The next chapter provides an elaborate illustration of how a time-varying cubic spline can be incorporated in a time series model. The cubic spline can be placed in the SSF (2.1)-(2.2) easily. The γ_t vector is the state vector and the transition matrix T_t is equal to the unity matrix. The covariance structure must be constructed such that $G_tG_t'=1$, $H_tH_t'=I-w_{**}w_{**}'/w_*'w_*$ and $H_tG_t'=0$. Finally, the weight spline vector z_t is placed in Z_t.

6. COMPUTER PROGRAMS

The appendices 5A and 5B give some more details of how to construct the cubic spline weights efficiently. The implementation of all required calculations is provided in a set of Pascal computer procedures which are given in appendix 5C.

The NatPinvQ procedure

Appendix 5A shows that the matrices P and Q are tridiagonal. This procedure calculates the matrix product $P^{-1}Q$ which is needed to get the weight vector for a natural cubic spline $w'=r'+s'P^{-1}Q$. The main part of the routine is provided by Press et.al (1989) and deals with inverting a tridiagonal matrix. The comments in the code of the procedure speak for themselves.

The PerPinvQ procedure

This procedure calculates the matrix product $P^{-1}Q$ which is needed to get the weight vector for a periodic cubic spline $w'=r'+s'P^{-1}Q$. The matrices P and Q are sparse, not tri-diagonal, and the procedure adopts the efficient method of calculating the inverse of P as given in appendix 5B.

The NatSplineWeight procedure

The weight vector w for a natural cubic spline is calculated according to the equation (2.7) and the set of restriction $Pa=Qy^{+}$. The matrix producy $P^{-1}Q$ is obtained from procedure NatPinvQ.

The PerSplineWeight procedure

The weight vector w for a periodic cubic spline is calculated as discussed in section 4. The matrix product $P^{-1}Q$ is obtained from procedure PerPinvQ.

The SeasSpline procedure

This procedure transforms a set of periodic spline weight vectors in order to let the spline sum up to zero, see discussion in section 4. The last elements of all the weight vectors are replaced by unity values.

7. CONCLUDING REMARKS

Cubic splines can be used in several statistical applications. For example, they can be implemented in a regression model for structural change and they can be used to describe non-linear responses and smooth seasonal variations. Especially for the case when the cubic spline is used to model periodic movements in time series, it is suggested that the parameters must be allowed to change over time. Section 5 of this chapter has given the technical details to let cubic splines be time-varying. The necessary computations are implemented in a set of efficient computer procedures.

NOTES

[1] The Puget Sound, Power & Light company, Bellevue, Washington State, USA has provided hourly data from 1 January 1985 to 31 August 1991 of load demand in mega-watts and of temperature in Fahrenheit. Casey Brace has been extremely helpful in supplying the data to us.

APPENDIX 5A

Calculation of weight vector for natural cubic spline

In this appendix the details are given of how to construct the weight vector for a natural cubic spline in order to evaluate the spline function $g(x)=w'\gamma^{+}$ for a specific value x. Suppose a natural cubic spline is based on a mesh of 9 elements such that k=8. The weight vector w is constructed in section 2 as $w'=r'+s'P^{-1}Q$ where the matrices P and Q are, respectively, given by

$$\begin{array}{ccccccccc}
2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
z_1/u_1 & 2 & z_2/u_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & z_2/u_2 & 2 & z_3/u_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & z_3/u_3 & 2 & z_4/u_3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & z_4/u_4 & 2 & z_5/u_4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & z_5/u_5 & 2 & z_6/u_5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & z_6/u_6 & 2 & z_7/u_6 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & z_7/u_7 & 2 & z_8/u_7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2
\end{array}$$

$$\begin{array}{ccccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6/z_1u_1 & -6/z_1z_2 & 6/z_2u_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 6/z_2u_2 & -6/z_2z_3 & 6/z_3u_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 6/z_3u_3 & -6/z_3z_4 & 6/z_4u_3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 6/z_4u_4 & -6/z_4z_5 & 6/z_5u_4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 6/z_5u_5 & -6/z_5z_6 & 6/z_6u_5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 6/z_6u_6 & -6/z_6z_7 & 6/z_7u_6 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 6/z_7u_7 & -6/z_7z_8 & 6/z_8u_7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}$$

and $u_i=z_i+z_{i+1}$ for i=1,...,7. The inverse of a tridiagonal matrix as P can be calculated recursively as explained in Press et.al. (1988). The vectors r and s correspond to a specific value x. Assume that x is a value between x_3^{+} and x_4^{+}. For this case the vectors r and s are, respectively, given by

$$[0;0;k/z_4;l/z_4;0;0;0;0]$$
$$[0;0;(k/6z_4)(k^2-z_4^2);(l/6z_4)(l^2-z_4^2);0;0;0;0]$$

where $k=x_4^{+}-x$ and $l=x-x_3^{+}$. The vector w is now calculated straightforwardly.

APPENDIX 5B
Calculation of weight vector for periodic cubic spline

In this appendix the details are given of how to construct the weight vector for a periodic cubic spline in order to evaluate the spline function $g(x)=w'\gamma^{+}$ for a specific value x. Suppose a periodic cubic spline is based on a mesh of 8 elements such that k=8. The weight vector w is constructed in section 4 as $w'=r'+s'P^{-1}Q$ where the matrices P and Q are, respectively, given by

$$\begin{array}{cccccccc}
2 & z_2/u_1 & 0 & 0 & 0 & 0 & 0 & z_1/u_1 \\
z_2/u_2 & 2 & z_3/u_2 & 0 & 0 & 0 & 0 & 0 \\
0 & z_3/u_3 & 2 & z_4/u_3 & 0 & 0 & 0 & 0 \\
0 & 0 & z_4/u_4 & 2 & z_5/u_4 & 0 & 0 & 0 \\
0 & 0 & 0 & z_5/u_5 & 2 & z_6/u_5 & 0 & 0 \\
0 & 0 & 0 & 0 & z_6/u_6 & 2 & z_7/u_6 & 0 \\
0 & 0 & 0 & 0 & 0 & z_7/u_7 & 2 & z_8/u_7 \\
z_1/u_8 & 0 & 0 & 0 & 0 & 0 & z_8/u_8 & 2
\end{array}$$

$$\begin{array}{cccccccc}
-6/z_1z_2 & 6/z_2u_1 & 0 & 0 & 0 & 0 & 0 & 6/z_1u_1 \\
6/z_2u_2 & -6/z_2z_3 & 6/z_3u_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 6/z_3u_3 & -6/z_3z_4 & 6/z_4u_3 & 0 & 0 & 0 & 0 \\
0 & 0 & 6/z_4u_4 & -6/z_4z_5 & 6/z_5u_4 & 0 & 0 & 0 \\
0 & 0 & 0 & 6/z_5u_5 & -6/z_5z_6 & 6/z_6u_5 & 0 & 0 \\
0 & 0 & 0 & 0 & 6/z_6u_6 & -6/z_6z_7 & 6/z_7u_6 & 0 \\
0 & 0 & 0 & 0 & 0 & 6/z_7u_7 & -6/z_7z_8 & 6/z_8u_7 \\
6/z_1u_8 & 0 & 0 & 0 & 0 & 0 & 6/z_8u_8 & -6/z_1z_8
\end{array}$$

and $u_i=z_i+z_{i+1}$ for i=1,...,7 and $u_8=z_1+z_8$. The matrix P has lost its tridiagonal structure but is still sparse. This can be exploited to preserve an efficient method of calculating its inverse. The matrix P can be expressed as $P = P_* + U'V$ where P_* is the tridiagonal part of matrix P, U=(0,$6/z_1u_1$;$6/z_1u_8$,0), V=(1,0;0,1) and 0 is a row vector of seven zeroes. The matrix inversion lemma

$$P^{-1} = P_*^{-1} - P_*^{-1}U'(VP_*^{-1}U' + I)^{-1}VP_*^{-1}$$

can be applied which can be rewritten as

$$P^{-1} = P_*^{-1} - P_*^{-1}XP_*^{-1}$$

and X is a 8×8 matrix of zeroes except the corner elements are non-zero, that is

$$X_{1,1} = 6x_{2,1}/z_1u_1$$
$$X_{1,8} = 6x_{2,2}/z_1u_1$$
$$X_{8,1} = 6x_{1,1}/z_1u_8$$
$$X_{8,8} = 6x_{1,2}/z_1u_8$$

where $x_{i,j}$ is the (i,j) element of the 2×2 matrix $(VP_*^{-1}U' + I)^{-1}$. The calculation of P^{-1} is now straightforward. This method of inverting P does require the inverse of the tridiagonal matrix P_* and the inverse of a 2×2 matrix. Note that $(a,b;c,d)^{-1}=(d,-b;-c,a)/(ad-bc)$ where a,b,c and d are scalars.

Finally, assume that x is a value between x_3^+ and x_4^+. For this case the vectors r and s are, respectively, given by

$$[0;0;k/z_4;l/z_4;0;0;0;0]$$
$$[0;0;(k/6z_4)(k^2-z_4^2);(l/6z_4)(l^2-z_4^2);0;0;0;0]$$

where $k=x_4^+-x$ and $l=x-x_3^+$. The vector w is now calculated straightforwardly.

TABLES & FIGURES

Table 5.1

Frequency distribution of variable x

Class	Frequency
0 - 10	20
10 - 20	40
20 - 30	30
30 - 50	10
total	100

Figure 5.1 A cubic spline through a set of points

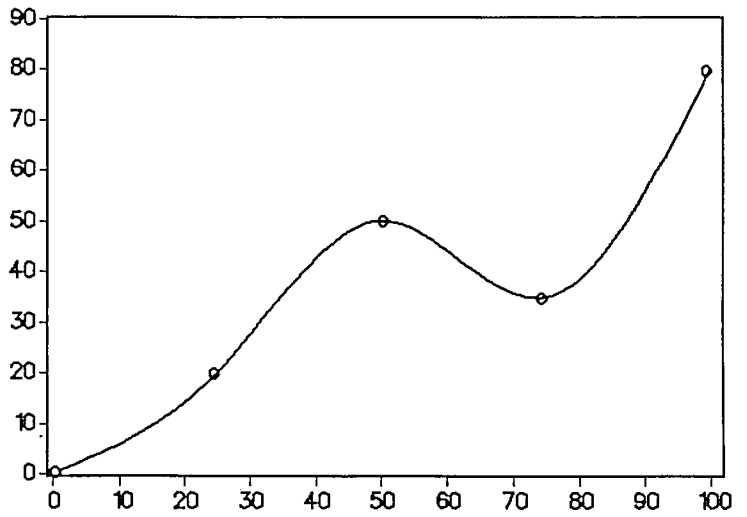


Figure 5.2a First derivative of a cubic spline

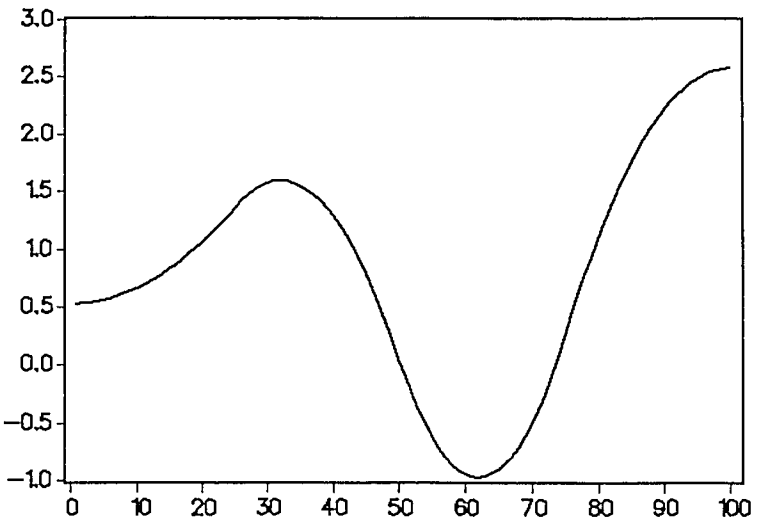


Figure 5.2b Second derivative of a cubic spline

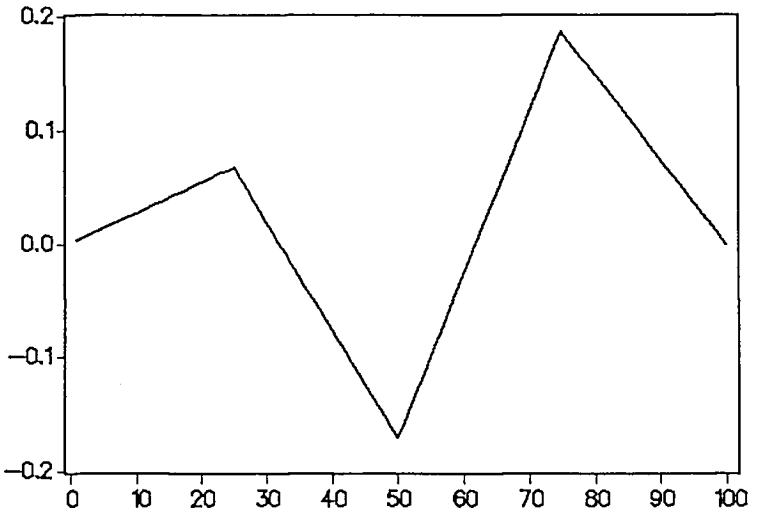


Figure 5.2c Third derivative of a cubic spline

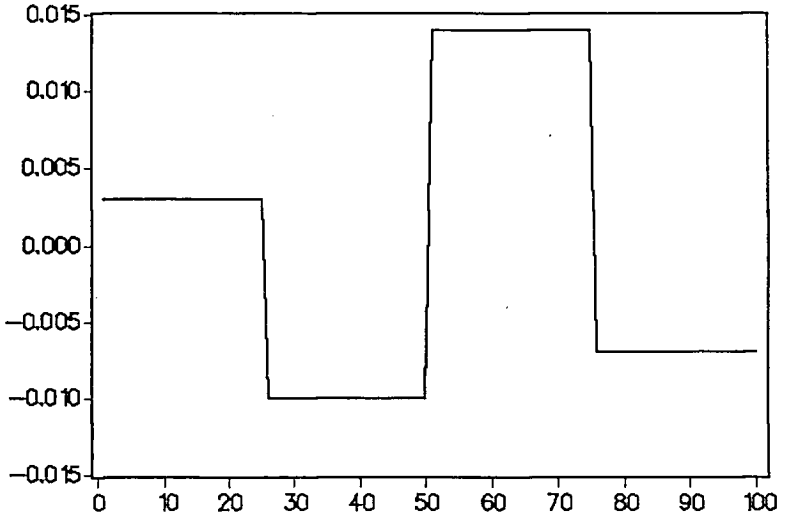


Figure 5.3 Interpolated monthly averages of temperature

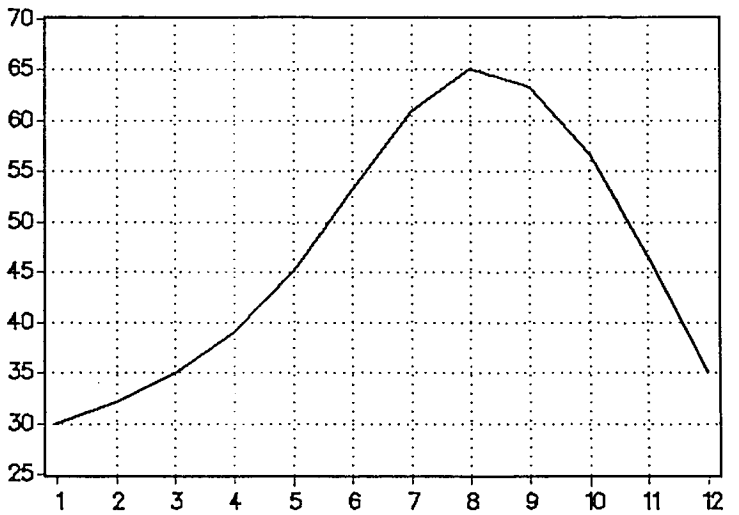


Figure 5.4 Frequency histogram and cubic spline polygon

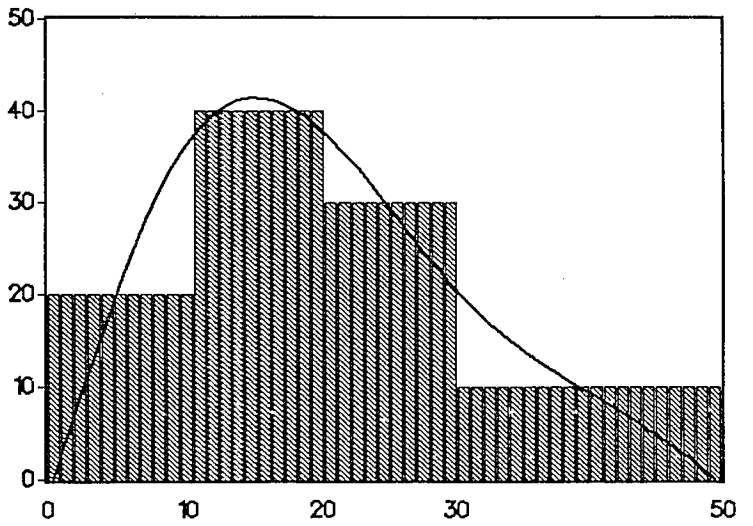


Figure 5.5 Scatterplot for all weekdays in 1990 at 1:00 pm

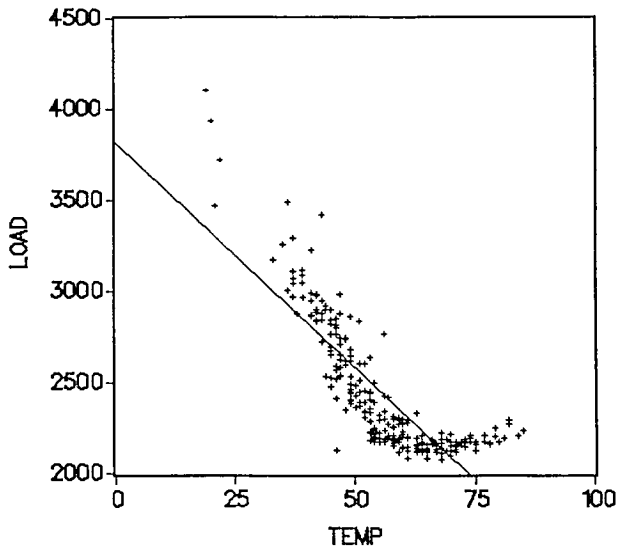


Figure 5.6 Cubic spline for response of load demand to temperatur

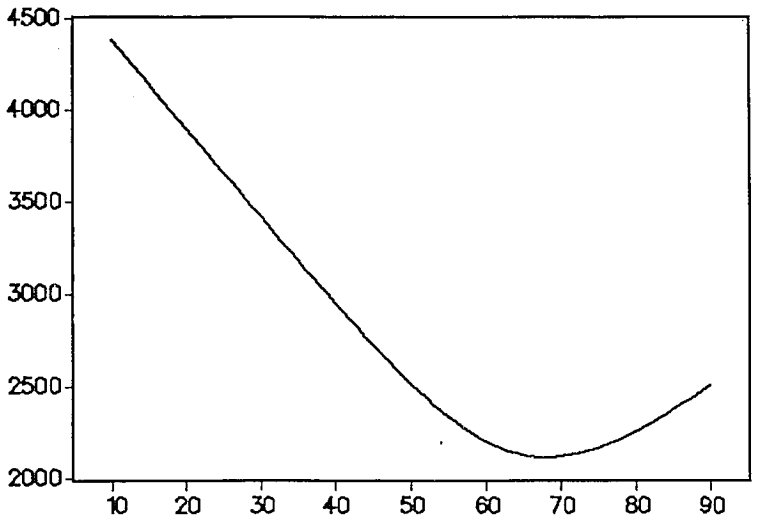


Figure 5.7 Weekly averages of temperature at noon

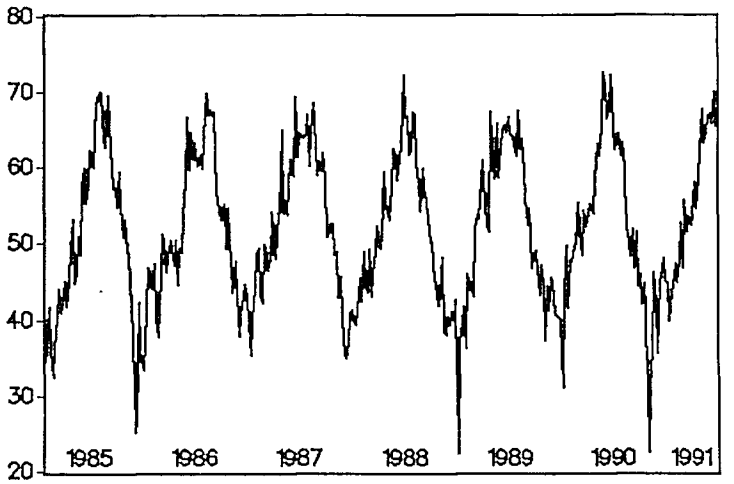
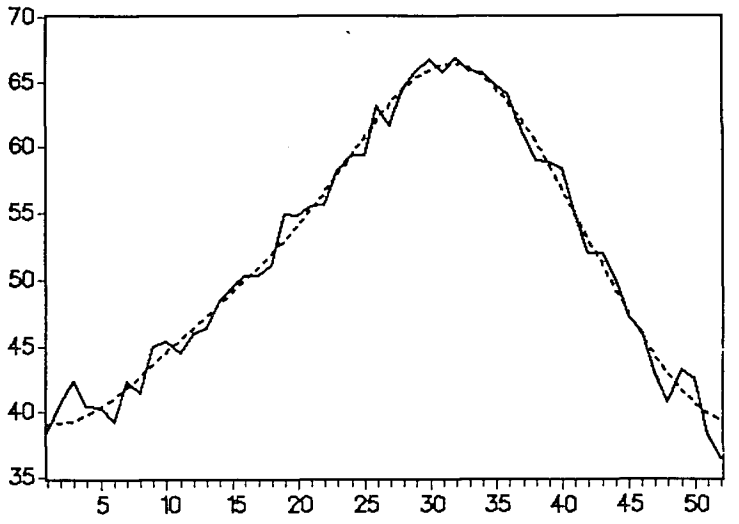


Figure 5.8 Weekly pattern of temp described by dummies & spline



CHAPTER 6
SHORT TERM FORECASTING OF PERIODIC TIME SERIES USING TIME-VARYING SPLINES

0. ABSTRACT

This chapter shows that time-varying splines can effectively be used for modelling a changing periodic pattern. This method is relatively parsimonious compared with other approaches. The time-varying spline is embedded within a structural time series model which is applied to forecast hourly electricity demand, with the periodic movements being intra-daily or intra-weekly. The full model contains other components, including a (non-linear) temperature response which is also modelled using splines.

Keywords : Cubic splines; Forecasting; Load curve; Non-linear regression.

1. INTRODUCTION

The theoretical results developed in the previous chapter will be applied to the problem of modelling intra-daily and/or intra-weekly periodic (seasonal) effects in time series. This approach is illustrated with a time series of hourly demand for electricity. The data are provided by the Puget Sound Power & Light company [1].

Time-varying seasonal effects can be captured by seasonal ARIMA models and for some formulations, such as the 'airline' model, it is possible to extract the seasonal component, see chapter 4 and Hillmer & Tiao (1982). An alternative approach, which lends itself much more readily to the kinds of problems discussed in this chapter, is based on structural time series models, see chapter 4 and Harvey (1989). There are now many applications of the basic structural model in the literature. A case which shows up time-varying seasonality quite clearly is the consumption of electricity

in the UK, reported in Harvey (1989, p95-98), where the change in seasonality in the early 1970's can be explained by the availability of gas from the North Seas and the subsequent substitution of electricity by gas in heating.

Intra-daily effects arise in a variety of applications. In the context of electricity demand the intra-daily pattern is known as the load curve. A parsimonious way of modelling the load curve is highly desirable for hourly observations, and become even more important when observations are made for every half or quarter hour. It is important to allow such splines to evolve over time. The intra-daily pattern may change over a period of several years due to new technology. It certainly changes within the year as can be seen in figure 6.1 where typical load curves are plotted for a selection of months in 1990 [1]. As one might expect there are marked changes over the year with the intra-daily variations in demand being much greater in the winter.

This chapter is organised as follows. Using the technology developed in the previous chapter, the next section shows how to model intra-daily and intra-weekly effects in a time series. Section 3 incorporates this approach, together with other unobserved components, in a structural time series model. Some alternative methods for modelling hourly load demand are addressed in chapter 4. The illustration is given in section 5 and the developed computer program SHELF, that is the Structural Hourly Electricity Load Forecaster, is discussed in section 6.

2. STATISTICAL TREATMENT OF INTRA-DAILY EFFECTS

In many cases the intra-daily effects change over the year with the seasons. Since the seasons are repeated each year, the intra-daily effects behave like a cycle movement in the long run. To model the different patterns throughout the year, ideally, the intra-daily effects require different models for different times of the year. Estimation requires a model which effectively allowed one to average over past years. To transfer smoothly from one type of model to another, periodic splines can be considered. Indeed, such

an approach is cumbersome to implement, it can be far off from a parsimonious solution and it may not be very satisfactory when there are only a few years of data available. However, in some cases it might be a viable option and further empirical research in this direction can be fruitful.

The approach of this research is to allow the intra-daily effect to accomodate seasonal and other changes by a slow movement over time. The time-varying spline technique enables this to be done with a relatively small number of parameters. With hourly data it is important to economise on the number of parameters since, without any restriction, a stochastic intra-weekly effect will contribute 167 elements to the state vector of the state space formulation. Obviously, this problem becomes more acute if observations are available at more frequent intervals.

To model intra-daily or intra-weekly effects parsimoniously, two approaches can be considered which are based on two different entities, i.e. the day or the week. The former approach interprets the intra-weekly effect as a cumulation of seven identical intra-daily effects where some days or some of the same hours at specific days may be different and require a correction. The latter approach takes the week as one entity and restrict some parts of the weeks to be the same. Both approaches will be explored in this section.

Intra-daily effects

The standard intra-daily effect can be appropriately described by a periodic cubic spline, see section 5.3. The main practical problem is to determine the set of mesh-values, e.g. the x-positions of the knots. Generally, sections of the periodic pattern displaying sharp peaks require relatively more knots than do less variable sections. However, it must be stressed that a certain amount of experimentation is needed to determine a good mesh. There is really no systematic way of going about this problem, although the starting point is obviously to use prior knowledge, an examination of unrestricted estimates or a good investigation of the graph to get an idea of the pattern. Next, an illustration discusses our approach in more detail.

Illustration

In the context of electricity demand within a day, the main peaks are around breakfast time and, to a lesser extent, in the early evening. Figure 6.2 shows an example of the intra-daily pattern of the demand for electricity (Friday, 11 March 1988 [1]). The set of mesh values for the periodic cubic spline can be chosen as (4, 6, 7, 8, 9, 15, 16, 17, 18, 19, 20, 23, 24), compare Hendricks et.al. (1979). The corresponding regression model requires thirteen parameters and represents a considerable saving over the 24 needed for an unrestricted intra-daily component (dummies). The regression results are reported in table 6.1. Note that the square root average squared interpolation error is q=5.148 for this example.

As discussed in section 5.4, the intra-daily pattern can also be described by a reduced number of trigonometric terms. The corresponding regression model with a constant and twelve trigonometric terms gives a worse fit in terms of the square root average squared interpolation error because now q=15.675. Figure 6.3 presents the graphs of two series of interpolation errors, one resulting from a periodic cubic spline and one resulting from trigonometric terms.

Unfortunately, the same intra-daily pattern will not normally apply to all days of the week. In particular, Saturdays and Sundays may be different to weekdays, and in the case of the illustration above, there is no doubt that they are different. One way of handling this problem is to set up a time-varying spline to give an intra-daily correction factor to atypical days. Continuity is enforced by setting up splines which are constrained to be zero at the beginning and end of the day. This constraint is reflected in the vector of spline weights. Consider the periodic spline $\gamma_j=w_j'\gamma^+$ for j=1,...,s. To enforce this spline to be zero at time s, it is restricted that $\gamma_k^+=0$ because $w_s=(0;...;0;1)$. This restriction is equivalent to dropping all last elements of w_j, for j=1,...,s, such that the vectors of spline weights are now all of dimension k-1.

Using a correction factor is likely to be particularly appealing when the difference between the intra-daily pattern and the standard intra-daily pattern is relatively smooth. This implies

that fewer knots are needed as compared with a full intra-daily model. The same argument also leads one to consider modelling Sunday by a Saturday correction factor plus a further correction factor, again constrained to be zero at its end points, for the difference between Saturday and Sunday. The use of correction factors means that the intra-daily effects do not sum up to zero over a non-standard day. Instead the sum of the correction factors gives the total amount by which the day in question differs from a non-standard day.

There are two disadvantages to the use of correction factors. The first is that they lead to discontinuity in first and second derivatives of the periodic cubic spline at the point where they join the standard intra-daily spline. Therefore, the transfer from a standard day to an atypical day may be less smooth. The second is that when one non-standard day follows another, as with Sunday and Saturday, the zero end point constraints imply that the intra-daily effect at the point where the correction factors meet should be as for a standard day. This can be unrealistic in some applications. However, the second problem may be solved by having one single correction factor for the whole weekend (which may include Friday afternoon and Monday morning).

Illustration (cont)

Figure 6.4 presents a typical weekly pattern of hourly electricity demand. A periodic spline for a standard day is fitted based on the mesh as given earlier. The interpolation errors are plotted in figure 6.5a. It is obvious that the standard day spline is not appropriate for the Weekend and a correction spline is needed. After some initial trials, it is conceived that a correction factor for Saturday may be based on the mesh (1, 4, 6, 7, 8, 9, 10, 15, 18, 20, 24). The interpolation errors after a standard day correction and a Saturday correction applied to Saturday and Sunday are plotted in figure 6.5b. Finally, an additional correction factor for Sunday may be based on the mesh (1, 6, 9, 10, 11, 14, 17, 20, 24). The final interpolation errors for Sunday are plotted in 6.5c. It is this flexibility of cubic splines which makes them preferable in practice to a set of trigonometric terms.

This approach based on correction splines seems to work satisfactorily, see figure 6.6. The salient feature of the correction factor for Saturday is that the morning peak is less pronounced and occurs an hour or so later. This is exactly what we would expect given our prior knowledge of behaviour on Saturday mornings. The correction spline for Sunday can be interpreted along these lines as well.

Thus, for this illustration concerning electricity demand, it is possible to base an additional Sunday correction factor on only nine knots, while Saturday requires eleven. The two correction factors meet each other at midnight when demand is low. This means that it is not particular important that first and second derivatives are discontinuous and that the knots at midnight are restricted to be the same as for standard days.

Intra-weekly effects

Another solution to the problem of different patterns for different days is to set up a periodic spline for the whole week. The zero sum constraint is now imposed over the whole week and s=168 when hourly observations are available. This approach does not suffer from the disadvantages noted at the end of the previous sub-section, but at first sight it would appear to be very inefficient if the patterns for some of the days are the same. Fortunately, it is possible to take account of this problem. Thus, the vector γ_t^+ may contain sets of parameters corresponding to the knots in each of the seven days. In that case, it is possible to work with a parameter vector of reduced dimension which only contains the elements which are different. Typically, three sets of parameters will be distinguished, e.g. the sets corresponding to weekdays, Saturdays and Sundays. The condensed parameter vector is denoted by γ_t° such that

$$\gamma_t^+ = R\gamma_t^\circ \qquad (2.1)$$

where R can be regarded as a selection matrix of zeroes and ones. Now if z_j is the vector of periodic cubic spline weights, which

enforce the 'sum-to-zero' restriction (see section 5.4), then the effect associated with the j-th hour in the week is

$$\gamma_t = z_j' R \gamma_t^\circ = z_j^{\circ\prime} \gamma_t^\circ \qquad (2.2)$$

where period j is prevailing at time t for j=1,...,s. However, since some of the weight vectors $z_j^\circ = R' z_j$ will be the same, less than the full complement of weight vectors need to be stored. Furthermore, although the vector z_j can become large, it needs to be computed only once and then post-multiplied by R to yield the sequance of vectors z_j° for j=1,...,s.

This solution is very flexible. The net effect of this approach is that days which depend on the same section of γ_t^+ will not have the same first and second derivatives at the knots because the restriction matrix R does only involve vector γ_t^+. This does mean that, although the level of the knots are the same, the patterns for identical days can be (slightly) different since the parameters will partly depend on adjacent days. Suppose Wednesday and Friday depend on the same section of γ_t°, the realized patterns on these days may differ because Friday is influenced by the weekend days. This additional flexibility is almost certainly an advantage.

Illustration (cont)

Again, consider figure 6.4 which shows a typical weekly load pattern. It is obvious that the daily patterns for Tuesday, Wednesday and Thursday are the same and that they are clearly different from the Saturday and Sunday patterns. The Monday morning and Friday afternoon/evening are slightly different from their counterparts at a normal weekday due to the influence of the adjacent weekends. The weekspline vector γ^+ may be based on the mesh

Sunday	(4, 7, 8, 9,10,13,17,18,20,23,24)
Monday	(4, 6, 7, 8, 9,15,16,17,18,19,20,23,24)
Tuesday	(4, 6, 7, 8, 9,15,16,17,18,19,20,23,24)
Wednesday	(4, 6, 7, 8, 9,15,16,17,18,19,20,23,24)
Thursday	(4, 6, 7, 8, 9,15,16,17,18,19,20,23,24)
Friday	(4, 6, 7, 8, 9,15,16,17,18,19,20,23,24)
Saturday	(4, 6, 7, 8, 9,12,16,18,20,23,24)

The restricted spline γ° is based on the spline vector γ^{+} and a set of constraints to ensure that Sunday, Saturday, Monday-morning (am hours) and Friday afternoon (pm hours) are based on different sections of the spline vector than the sections for standard days.

3. SHORT TERM FORECASTING WITH STRUCTURAL TIME SERIES MODELS

This section considers the modelling of observations which are measured within the day, say each hour or each minute. The main purpose of the model is to forecast future realizations two or three days ahead. Our approach is based on structural time series models which are discussed throughout the earlier chapters; for a full discussion, see Harvey (1989). A structural time series model for intra-daily effects and with explanatory variables included can be expressed by

$$y_t = \mu_t + \gamma_t + \delta_t + \epsilon_t \qquad (3.1)$$

where μ_t denotes the underlying level which may include other unobserved components such as slope and cycle, γ_t is the intra-daily effect, δ_t refers to the total of explanatory responses and ϵ_t is the white noise disturbance term. In a Gaussian model, the disturbances driving the various components are assumed to be normally distributed. The complement of components μ_t, γ_t and ϵ_t will be referred to as the univariate part of the model.

The model can be handled by placing it in the state space form. The Kalman filter and associated recursive algorithms provide the basis for updating, prediction and smoothing, see chapters 2 and 3. In addition the Kalman filter is used to construct the likelihood function. Maximization of the likelihood yields estimators of the hyperparameters (variance parameters). Finally, the Kalman filter provides the ideal environment to produce multi-steps ahead forecasts straightforwardly.

Univariate components

The model as it stands can be specified in different ways. The level μ_t picks up both short and long run movements because ϵ_t is white noise. Since the primary concern is short term forecasting, it will be appropriate for most cases to let μ_t follow a random walk rather than the more elaborate local linear trend. However, a more general model is obtained when μ_t represents solely the long term trend and ϵ_t is generalized to become a stationary process, such as a first or second order autoregression, in order to describe the short-term dynamics.

In the latter case, μ_t may represent seasonal and cyclical movements which can be modelled by a full set or a limited number of trigonometric terms. For example, a time-varying trigonometric cycle is modelled by

$$\psi_t = \cos(2\pi/p)\psi_{t-1} + \sin(2\pi/p)\psi^*_{t-1} + \kappa_t \qquad (3.2a)$$

$$\psi^*_t = -\sin(2\pi/p)\psi_{t-1} + \cos(2\pi/p)\psi^*_{t-1} + \kappa^*_t \qquad (3.2b)$$

where p is the length of the cycle, κ_t and κ^*_t are disturbances with mean zero and variance σ^2_κ. Also a stochastic slope term such as β_t can be included in the level. Thus a possible specification for the long term trend may be $\mu_t=\mu^*_t + \psi_t$ where μ^*_t is a local linear trend and ψ_t is the cycle term (3.2). Typically, the long term dynamics will evolve very slowly over time and, therefore, their variance parameters tend to be very close to zero (almost deterministic). Alternatively, the long term trend can be based on a periodic spline or on a full seasonal component.

Explanatory variables

The introduction of explanatory variables into structural time series models is quite straightforward if these variables enter linearly, see Harvey (1989, Chapter 7). Suppose, a set of h explanatory variables are available and their total effect on the dependent variable y_t in the model (3.1) is denoted by δ_t where

$$\delta_t = X_t\delta^+ \tag{3.3}$$

where X_t is the (hx1) row vector of explanatory variables at time t and δ^+ is the regression vector of coefficients. In the context of time series, explanatory variables may respond differently to the dependent variable y_t at various stages of the seasonal cycle, say winter and summer. It is difficult to extract the response pattern because the number of available observations are in general not large. For time series models which concern short term dynamics, the different responses at various sections in the periodic variation may be more pronounced and, therefore, they may be easier to identify. The following method is developed to extract intra-daily response patterns of explanatory variables. The results below can be easily generalised for an intra-weekly response pattern.

If the j-th time of the day, j=1,...,s, is prevailing and only one explanatory variable is considered (h=1), then

$$\delta_t = X_t\delta_j^+ \tag{3.4a}$$

where the scalar δ_j^+ may be parameterized as

$$\delta^+_j = (\delta^*_j + 1)\delta^x \tag{3.4b}$$

such that

$$\Sigma^s_{j=1}\delta^*_j = 0 \qquad \text{and} \qquad \Sigma^s_{j=1}\delta^+_j = s\delta^x \tag{3.5}$$

Thus if $\delta^*_j = 0$ for j=1,...,s, the response of the explanatory variable is the same throughout the day. In this specification, the unknown scalar δ^x can be regarded as the scaling parameter and the set of weights δ^*_j determines the response of the explanatory variable within the day. The shape of the response can be approximated by a periodic cubic spline which is specified as (5.4.7) such that it sums up to zero. The spline is assumed to be based on a mesh with k elements. Thus $\delta^*_j = z_j'\gamma^x$ and the response effect becomes

$$\delta_t = X_t\delta^x\ (1 + z_j'\gamma^x) \tag{3.6}$$

where the j-th time of the day prevails. The set of parameters for the regression response δ_t has reduced from s to k elements. The final result is that the response of an explanatory variable is specified as usual but rescaled by a set of weights which sum up to one within the day. The weights follow a pattern described by a periodic cubic spline. This approach is easily generalised for h explanatory variables when it is assumed that their responses to y_t show the same intra-daily pattern. In this case, the specification (3.6) remains but X_t and δ^x are a (hx1) row and column vector, respectively. This specification contains (h+k-1) parameters. Since any non-linear response function can be approximated by a cubic spline which can be specified as a multiple regression, compare $X_t\delta^x$, it is straightforward to generalize the solution above of modelling different responses within the day for non-linear explanatory variables.

The two parameters δ^x and γ^x do not relate to each other in a linear fashion. This causes problems in linear estimation. In referring back to the discussion of section 5.3, where distributed lags are included in a non-linear regression model, a solution arises because estimation may proceed along these lines. The two different response specifications are given by

$$\delta_t = X_t^*\delta^x \qquad \text{where } X_t^* = X_t(1 + z_j'\gamma^x) \qquad (3.7a)$$

$$\delta_t = x_t + z_j^{*\prime}\gamma^x \qquad \text{where } x_t = X_t\delta^x \text{ and } z_j^* = x_t z_j \qquad (3.7b)$$

and the iterative estimation process can be initialized with setting γ^x equal to a zero vector.

There is no difficulty, in principle, in letting the parameter vector δ^x change over time according to a multivariate random walk. The time-varying parameters become part of the state vector and the covariance structure (hyperparameters) governing the rate at which they can change is estimated numerically along with the other hyperparameters. Finally, without the restrictions implied by (3.4) it may be considered to use a bivariate spline, see Poirier (1976), to allow the shape of the response to change within the day. This would require many more parameters and could become quite complex and too elaborate.

Atypical days

In many social economic time series with weekly, daily, hourly or more frequent observations, the influences of atypical days such as Christmas, New Year, Easter and Bank holidays are pronounced. Some public holidays can be handled by treating them as though they were Sundays. However, this is unlikely to be satisfactory for specific holidays such as Christmas and New Year. The best thing to do in such cases is probably to have a special intra-daily correction factor which is separate from the rest of the model and is just estimated from past observations, in terms of deviations from the underlying level, on that particular holiday. Forecasts of future values can then be made by adding this correction factor to the predicted underlying level. The observed holiday values are probably best treated as missing observations, since to include them could easily introduce distortions into future estimates. This may also apply to some days after a public holiday.

4. VARIOUS APPROACHES TO MODELLING HOURLY ELECTRICITY LOAD DEMAND

Intra-daily or intra-weekly effects arise in a variety of applications. In the context of electricity demand the intra-daily pattern is known as the load-curve. A parsimonious way of modelling the load curve is highly desirable for hourly observations, and becomes even more important when observations are made every half or quarter hour. Our preferred solution is based on the short-term forecasting model developed in the previous section. It is shown in the next section that this approach provides valuable interpretable statistical information and has a very good short-term forecasting performance.

The collection of papers in Bunn & Falmer (1985) gives some indication of the type of models which have been employed in short-term forecasting of energy up to now. The main approaches seems to be based on ARIMA models, regression, exponential smoothing or some mixtures of these.

For daily observations the ARIMA 'airline' model, based on the $\Delta\Delta_7$ operator, is sometimes used. For hourly observation, the 'airline' model becomes

$$\Delta\Delta_{168} y_t = (1+\theta L)(1+\Theta L^{168})\xi_t \qquad (4.1)$$

which might provide an approximation to the structural time series model for hourly observations, see section 2.3. However, it is unlikely that one would 'identify' such a model on the basis of the correlogram in the way Box & Jenkins (1976) advocate. In addition, it is not possible to take account of the fact that the pattern for some days may be the same such that the state vector of the state space form is of a huge dimension ($\#[\alpha_t]=168$). The introduction of a further component, such as a seasonal, is likely to make the selection of a suitable ARIMA model even more difficult, unless, it is derived from a structural time series model in the first place.

An alternative approach is based on treating hourly observations as a collection (vector) of 24 time series of daily observations. A regression model for every specific hour can be specified such that it is based on time functions, explanatory variables (temperature, humidity, etc.) and distributed lags. This approach may be sensible because the dynamic mechanisms driving the observations may be very different during night hours and hours around noon. However, this solution is certainly not a parsimonious way of modelling hourly observations. Therefore, it is not surprising that this alternative approach may give better forecasts if enough observations are available.

Alternatively, a small set of regression models can be formulated as described above for typical periods during the day, say, morning, afternoon, evening and night. The responses of the observations at each hour to these different models can be smoothly distributed through a periodic cubic spline. This may be a viable option in specific cases although it requires still a large set of parameters. Indeed, the number of parameters increases when atypical days do occur such as Saturday and Sunday. In addition, specific models have to be formulated for specific periods during these atypical days.

5. FORECASTING HOURLY ELECTRICITY LOAD DEMAND : THE PUGET POWER CASE

This section illustrates the techniques developed in chapter 5 and the previous sections of this chapter with hourly data from the Puget Sound Power & Light [1] electricity company based in the north west of the USA.

Data description

The available data [1] consists of hourly load and temperature values from 1 January 1985 to 31 August 1991. The load is measured in mega-watts (MWs) and the temperature is measured in degrees Fahrenheit (F). Furthermore, every day is indicated by the day number (1..31), the month (1..12), the year (85..91), the day type (SUN,MON,TUE,WED,THU,FRI,SAT) and a holiday indicator (TRUE/FALSE).

An important feature of the data is the intra-weekly pattern. Figure 6.7 presents the hourly load demand for the 5^{th}, 15^{th}, 25^{th}, 35^{th} and 45^{th} week of 1990. Clearly, the weekly patterns do change throughout the year. However, all weekly patterns show that the Sunday and Saturday are different than the weekdays. The Monday morning (am hours) and the Friday afternoon/evening (pm hours) are, to a lesser extend, also different than these parts of the days during the other weekdays. Note that the Sunday of the 45^{th} week in 1990 was Veteran's day which is an official Holiday in the USA.

It is recognised earlier that the response of temperature on electricity load demand is non-linear, see the illustration of section 5.3. The load demand increases with cold temperatures because heating is switched on. A period with extreme hot temperatures shows an increase in load demand as well because of air-conditioning. Therefore, the response of load demand on temperature is a U-shaped curve. The nature of this non-linear effect depends primarily on climatological factors but also on the technological development in a region. The load demand at some specific hours for all Wednesdays is scattered against the corresponding temperature for the days in 1990 (except for holidays)

in figure 6.8. This shows that the various U-shapes are rather weak which is recognised earlier by Engle et.al. (1986) for the region Seattle (Washington, USA) but there is still a cause for using splines to model the temperature response. Indeed, the climate of the North-West of the USA is rather mild. Also, figure 6.8 shows that the response patterns do not strongly vary within the day.

To get some idea of the long-term properties of the series, the load demand at some specific hours for weekdays is plotted in figure 6.9. These graphs show very clearly a weak upwards trend with a cyclical pattern over the year. The cycles corresponding to the early hours of a day are weaker than the cycles at later hours. The random shocks disturbing the cycle movements are explained by the influence of holidays and extreme weather conditions.

Many other features do also influence the load curve, for example, other weather conditions (humidity, wind speed and cloud coverage of the sky) and various ad-hoc events (national events and TV coverage of sport events). The latter only affect in general a small set of hours.

A preliminary regression model

To get some preliminary insights of how the main features of the time series do interact with each other, a very simple regression model is constructed.. The regression model is concerned with a specific hour of the day in the years of 1986, 1987 and 1988. Thus, 24 different models (for each hour) are estimated using 1096 observations. The regression model is

$$y_{i,t} = \alpha_i + \beta_i x_{i,t} + \mathrm{dsun}_i + \mathrm{ssat}_i + \epsilon_{i,t}$$

where $y_{i,t}$, $x_{i,t}$ and $\epsilon_{i,t}$ are the load demand, the temperature and the Gaussian noise term at hour i and day t, respectively. The parameters dsun and dsat are dummies for Saturday and Sunday, respectively. A Holiday is regarded as a Sunday. The estimation results for the 24 models are reported in table 6.3. The t-tests are not given but almost all coefficients are significant. The coefficients are graphically presented in figure 6.10. These plots

are a good indication of the general, or 'average', behaviour within the day. All coefficients follow, more or less, the same intra-daily pattern as the intra-daily pattern of the observations.

Model specification

Several models of the kind described in the previous sections were fitted to the hourly observations described above. Our preferred model is of the form

$$y_t = \mu_t + \psi_t + \gamma_t + \delta_t + \epsilon_t \qquad (5.1)$$

where μ_t is a random walk, ψ_t is a deterministic cycle of a period of one year (365*24) hours, γ_t is a time-varying weekly spline, δ_t is the response on temperature modelled as a cubic spline and ϵ_t is a random disturbance term with mean zero and variance σ^2.

The random walk μ_t characterizes the long term trend movement of the series. The disturbances, generating this stochastic unobserved level component, do have a constant variance $q_\eta\sigma^2$ where q_η is the unknown signal to noise ratio.

The deterministic cycle is modelled by a single pair of sines and cosines of the form (3.2) so it can be evaluated recursively. Recall that this periodic movement is not primarily due to temperature and other weather variables but arises from other seasonal changes such as the change in the number of hours of daylight.

The time-varying spline is specified as presented in section 5.5, that is $\gamma_t = z_j'\gamma^+_{*t}$ (period j prevails at time t) where

$$\gamma^+_{*t} = \gamma^+_{*(t-1)} + \chi_{*t} \qquad (5.2)$$

and χ_{*t} is a disturbance vector with zero mean and covariance matrix $\sigma^2 q_\chi V_*$ where q_χ is the unknown signal to noise ratio concerning χ_{*t}. The twelve knots for a standard weekday are based on the mesh of hours

3am 6am 7am 8am 9am 11am 2pm 5pm 6pm 8pm 11pm 12pm

and they are restricted to be the same for every standard weekday but the am hours of Monday and the pm hours of Friday (except 2pm) are allowed to be different. The eleven knots for a Sunday and a Saturday are based on the mesh of hours

3am	7am	8am	9am	11am	2pm	6pm	7pm	9pm	10pm	12pm
3am	6am	7am	8am	9am	12am	4pm	6pm	8pm	10pm	12pm

respectively. Thus the total number of free knots (parameters) is 43 and note that the knot at the Saturday midnight hour is restricted for the 'summing-to-zero' constraint.

The temperature response δ_t is modelled by a natural cubic spline based on the mesh

0F	40F	65F	99F

and it is decided to let the parameters be fixed.

The model (5.1) does not explicitly take account of holidays and some days after a holiday. To avoid that these distortions in the series do not influence parameter estimation and forecasting performance, holidays and the two days after holidays are regarded as missing.

The following points have been considered during the process of model specification : (i) transforming y_t by logarithms to give a multiplicative model does not improve the model fit and the forecasting performance; (ii) other explanatory variables, such as humidity and windspeed, were included in preliminary versions of the model, but no specification in which they had a significant influence could be found; (iii) lagged values of temperature were included in the manner suggested earlier in chapter 5, but the main effect seemed to come from current temperature; (iv) some evidence was found of different responses to temperature at different times of the day, but the additional complexity of the model was not justified in terms of model fit and forecasting performance.

The model (5.1) can be placed in state space form straightforwardly. The state vector is the stack of parameters $(\mu_t;\psi_t;\psi_t^*;\gamma_{*t}^+;\delta^x)$. The transition matrix T_t does mainly consist of zero

and unity values. The measurement equation vector Z_t mainly consists of the spline weights. The system covariance matrix $\sigma^2 H_t H_t'$ does contain two hyperparameters, that is q_η and q_χ.

Estimation results

The estimation of the hyperparameters is carried out in the time domain via the prediction error decomposition. Some scoring method is applied to maximize the log-likelihood function. The likelihood evaluation process (Kalman filter) is very time consuming in this case because the dimension of the state vector is relatively large and the number of hourly observations is approximately 57,650. The Kalman filter is initialized by the 'big-κ' method with κ=1000.

The estimates of the hyperparameters are found to be

$$q_\eta = 0.0002 \qquad q_\chi = 0.0012$$

and recall that these values indicate the extent to which the level of the series and the weekly pattern, respectively, are allowed to change over time.

The changing intra-weekly pattern is shown in figure 6.11 for some weeks in 1990. The contrast between a week-pattern in, say, January and May is particularly marked, compare figure 6.7. All the knot points of the periodic spline were significant throughout the year although the intensity of significance (observed via the t-test) does change as well.

The fixed effects in the model are all significant but the temperature effect was highly significant, see table 6.3. The shape of the temperature response is as expected; it shows a weak upturn when temperature exceeds 75F, see figure 6.12. The yearly cyclical pattern ψ_t is plotted in figure 6.13.

The model fitted better in the warmer months than in the winter. One indication may be the aggregation of the squared standardised innovations, the one-step ahead prediction errors, for every week in 1990, see figure 6.14. It shows very clearly that the last two weeks of 1990 have large errors because of the special holidays such as Christmas and New Year.

Forecasting results

The electricity company Puget makes forecasts at 9 am on Monday through Thursday, for the next day, based on information up to one hour earlier, that is 8 am. Thus the maximum length for forecasting is 40 steps ahead. On Friday morning forecasts are made for Saturday, Sunday and Monday, that is forecasting with a maximum of 88 steps ahead.

The load forecasts are conditional on the forecasted temperature values, as supplied by Puget, which are based on forecasts of several weather stations in the Bellevue region. It must be stressed that the accuracy of the forecasted load demand depends heavily on the accuracy of the forecasted temperature. However, latest (unpublished) results indicate that the SHELF forecasts do not depend on the temperature forecasts as much as competitive models do.

The forecasting results for a Tuesday (one-day ahead) and a Friday (three-days ahead) are given in figure 6.15 as examples. It presents (a) the forecasted values against the actual observations and (b) the mean absolute percentage error (MAPE) of forecast errors.

For the period 7 November 1990 until 31 March 1991, the forecasting performance of the model (5.1) is shown in table 6.4. The reported indicators are based on the MAPE of every forecast at a specific hour. The 24 MAPEs of a specific day are aggregated in five different ways :

[i] AVG - average
[ii] MIN - minimum
[iii] MAX - maximum
[iv] AM - average for peak morning hours 7am, 8am & 9am
[v] PM - average for peak afternoon hours 4pm, 5pm & 6pm

These indicators are calculated for the day types Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday, Weekday and Weekend. Their averages for the winter period of 1990/91 are given in table 6.4. The holidays are excluded from this evaluation.

To put the forecast results of table 6.4 in perspective, a simple (naive) forecast equation is applied to the same data and its

forecast performance is evaluated for the same winter period of 1990/91. The forecast function is based on the set 24 dynamic (unit root) regression models

$$y_{i,t} = y_{i,t-7} + \beta_i(x_{i,t}-x_{i,t-7}) + \epsilon_{i,t}$$

where $y_{i,t}$, $x_{i,t}$ and $\epsilon_{i,t}$ are the load demand, the temperature and the Gaussian noise term at hour i and day t, respectively. The regression parameter β_i is estimated by generalised least squares for i=1,...,24. The average hourly MAPEs (indicator [i]) for this forecast function are, with the type of day in parantheses, 7.31 (Sunday), 5.60 (Monday), 5.22 (Tuesday), 5.39 (Wednesday), 5.70 (Thursday), 6.93 (Friday), 6.74 (Saturday), 5.77 (Weekday) and 7.03 (Weekend). The holidays are excluded from this evaluation. It can be concluded that the forecasts of model (5.1) are much more precise than the naive method.

6. THE COMPUTER PROGRAM SHELF

The computer program SHELF - Structural Hourly Electricity Load Forecaster - is the implementation of the model of section 5 and it is primarily developed to forecast hourly load demand. The program is mainly a collection of procedures for cubic splines calculations (see section 5.6) for the Kalman filter (see section 2.8) and for forecasting. The program is structured as follows :

[a] Model specification

The model is placed in SSF as discussed in the previous section. The vector Z_t of the SSF varies over time and it depends on the season and it depends on the level of temperature. Therefore two seperate matrices are defined in the program. One matrix is filled with the weight vectors for the intra-weekly periodic cubic spline, that is z_j for j=1,...,168. Another matrix is filled with the weight vectors for the natural cubic spline of temperature, that is w_j for j=0,...,99. The appropriate vectors are placed in some large vector that acts as the Z_t vector of the SSF.

[b] Reading data

The daily observation is organised in a special record, that is

```
Observation =
Record
  Date    : 1..31;
 Month    : 1..12;
 Year     : 0..99;
 Day      : 1..7;
 Holiday  : Boolean;
 Load     : Array[1..24] Of Integer;
 Temp     : Array[1..24] Of Integer;
End;
```

and a special procedure reads any pre-specified record from a big data-file.

[c] Parameter updating

This part of the program consists mainly of the Kalman filter. The state vector contains the time-varying and fixed parameters. The standardised innovations are stored to construct some simple diagnostics afterwards. The Kalman filter is also used to evaluate the likelihood function and it can be embedded within a maximization procedure (some scoring method) to obtain hyperparameter estimates. However, this is not included in SHELF since it is not worthwhile to re-estimate the hyperparameters every time when new observations are coming available.

[d] Forecasting

After the parameters in the state vector are updated for a sequence of hours, a special procedure is incorporated in the KF procedure that produces the forecasts for the next day or for the next three days. An input procedure reads the forecasted temperature values for the coming days from an external file. An output procedure puts the forecasts of electricity load demand in a special file.

7. CONCLUDING REMARKS

A time-varying periodic cubic spline component appears to provide a good way of modelling the changing electricity load pattern within the week with a relatively small number of parameters. The effect of the non-linear response is captured by a fixed natural cubic spline.

NOTES

[1] The Puget Sound, Power & Light company, Bellevue, Washington State, USA has provided hourly data from 1 January 1985 to 31 August 1991 of load demand in mega-watts and of temperature in Fahrenheit. Casey Brace has been extremely helpful in supplying the data to us.

TABLES & FIGURES

Table 6.1

Regression results of cubic spline for load curve

dependent variable : load 11 March 1988 [1] — t = 1,...,24
number of observations : 24 — season : hourly
model : periodic cubic spline regression — method : OLS

REGRESSION ESTIMATION

parameter	estimate	t-test
Constant	2276.08	2048.74
Knots 4am	-459.08	-135.17
6am	6.33	1.60
7am	621.77	122.97
8am	917.49	203.05
9am	748.49	232.79
3pm	-234.39	-72.37
4pm	-268.43	-59.41
5pm	-219.86	-41.70
6pm	-33.60	-6.34
7pm	214.78	43.05
8pm	226.80	59.82
11pm	-159.34	-40.78

$\hat{\sigma}^2 = 5.148$

Table 6.2

Regression results of preliminary daily model

dependent variable : Load demand [1]
number of observations : 1096
model : linear regression

t = 01/01/86D1 - 31/12/88D366
season : daily
method : OLS

REGRESSION ESTIMATION

Hour	Constant	Temperature	dsat	dsun
1	2449	-19.4	39.1	15.7
2	2403	-19.8	21.4 *	0.0 *
3	2422	-20.6	6.9 *	-13.7 *
4	2487	-21.7	-7.0 *	-29.7
5	2652	-23.6	-49.2	-77.7
6	3150	-28.5	-214.1	-265.2
7	3981	-36.1	-507.5	-594.6
8	4478	-40.5	-531.2	-651.6
9	4443	-39.5	-273.9	-404.5
10	4174	-35.4	-74.7	-200.2
11	3893	-31.1	-12.0 *	-120.0
12	3621	-27.4	-12.2 *	-95.2
13	3414	-24.8	-51.0	-96.7
14	3245	-22.5	-103.3	-124.9
15	3141	-21.4	-134.6	-158.0
16	3192	-22.4	-159.2	-183.5
17	3530	-27.5	-170.8	-191.6
18	4038	-34.7	-200.5	-212.1
19	4209	-37.2	-214.2	-211.6
20	4094	-35.6	-220.1	-179.4
21	3812	-31.2	-218.3	-146.0
22	3387	-25.2	-185.6	-111.8
23	2993	-21.5	-106.5	-108.0
24	2675	-20.0	-34.5	-86.6

* not significant at 95% confidence interval

Table 6.3

Regression results of fixed effects in model (5.1)

dependent variable : Load demand [1]
number of observations : 58368
model : (5.1)
t = 01/01/85H1 - 31/08/91H24
season : hourly
method : KF

REGRESSION ESTIMATION

parameter	estimate	t-test
Cycle		
ψ	-132.05	-15.71
ψ^*	200.01	23.99
Temperature		
00F	3497.71	199.54
40F	2187.06	190.89
65F	1620.39	140.71
99F	1773.52	112.96

$\hat{\sigma}^2 = 5208.68$

Table 6.4

Forecasting performance of SHELF

Day :	Mon	Tue	Wed	Thu	Fri	Sat	Sun	WDay	WEnd
Indication									
Day average	4.3	3.7	3.2	4.1	3.7	3.9	5.1	3.8	4.5
Day minimum	0.5	0.4	0.2	0.8	0.5	0.9	1.0	0.5	0.9
Day maximum	10.5	8.7	8.2	9.8	8.9	8.5	10.8	9.2	9.6
AM average	4.9	3.9	2.7	3.2	4.1	4.6	5.7	3.7	5.2
PM average	4.7	4.6	4.2	4.7	4.2	3.9	5.9	4.5	4.9

Figure 6.1a Load curve on 30 January 1991.

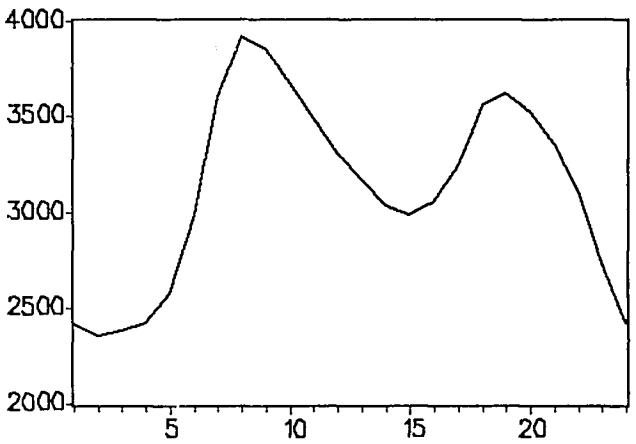


Figure 6.1b Load curve on 10 April 1991

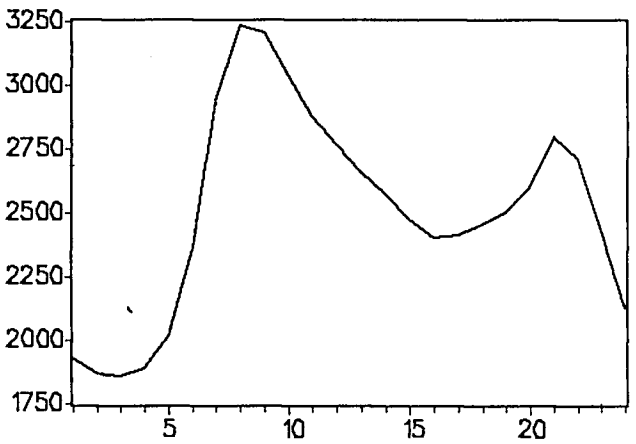


Figure 6.1c Load curve on 19 June 1991

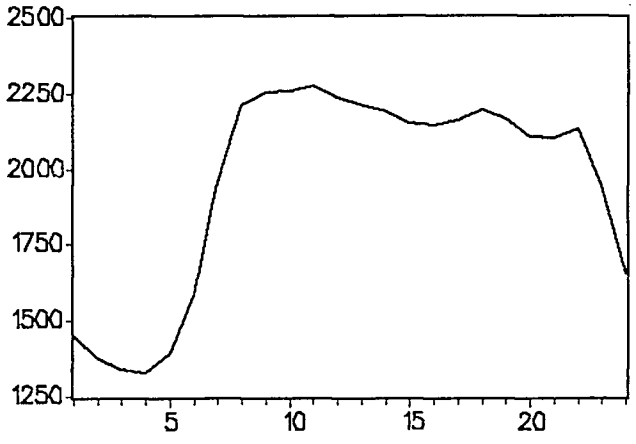


Figure 6.1d Load curve on 28 August 1991

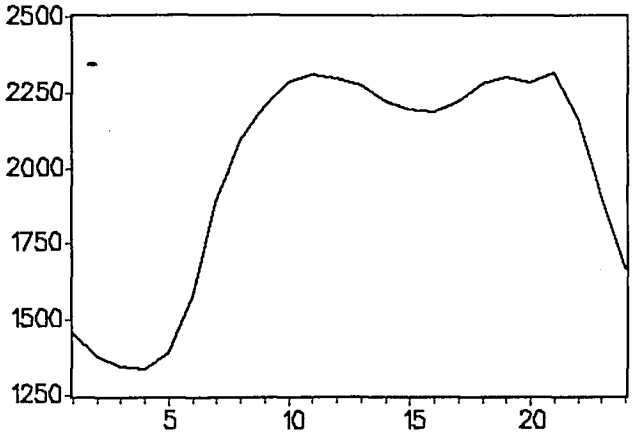


Figure 6.2 Load curve on 11 March 1988

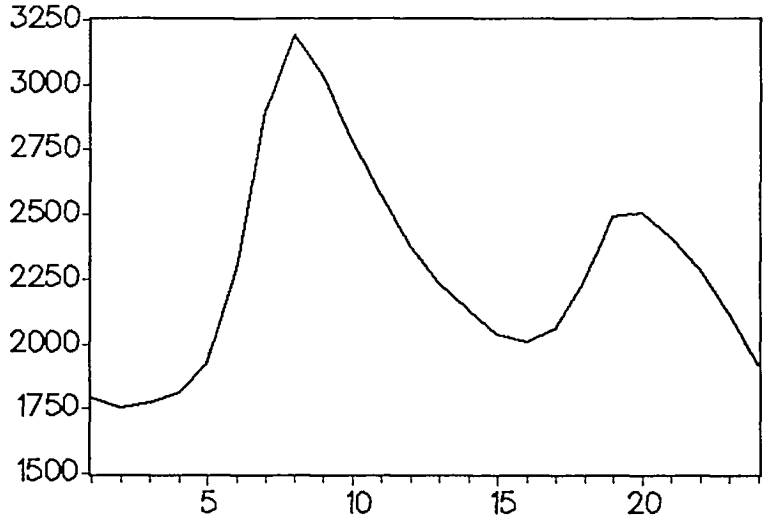


Figure 6.3 Interpolation errors with spline and trigonometrics

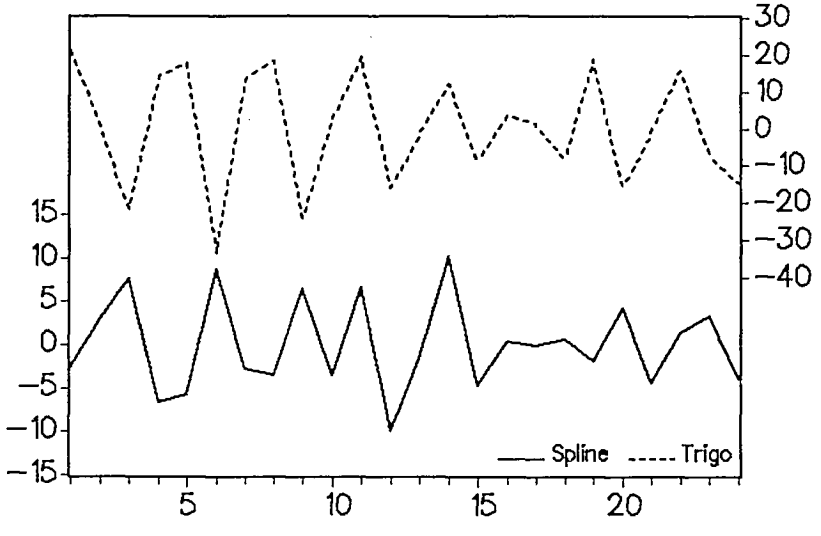


Figure 6.4 Load demand pattern of first week in March 1988

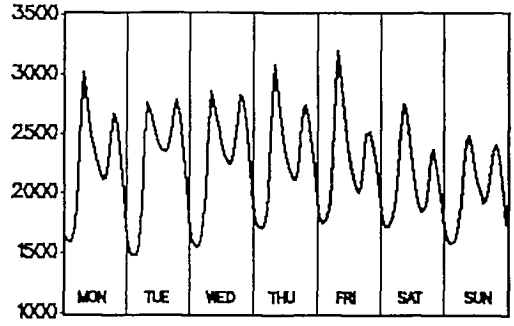


Figure 6.5a Interpolation errors after standard day spline

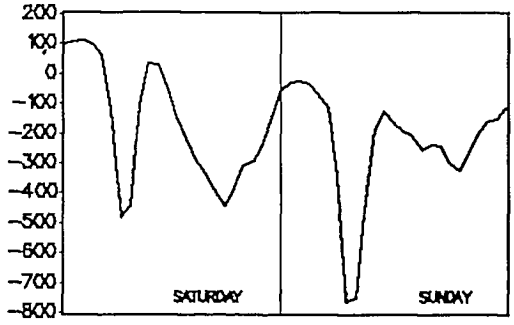


Figure 6.5b Interpolation errors after Saturday correction

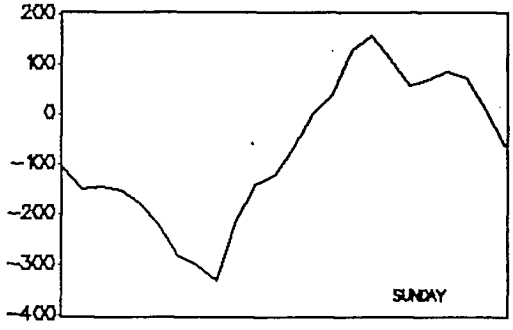


Figure 6.5a Estimated weekly load pattern by standard day spline

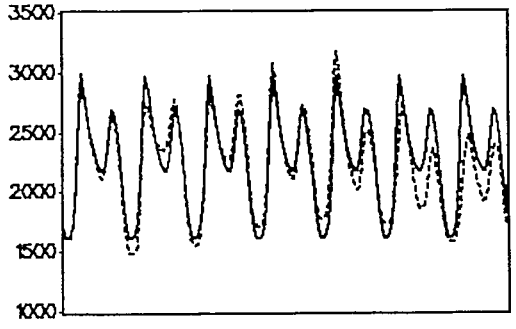


Figure 6.5b Estimated weekly load pattern with Saturday correction

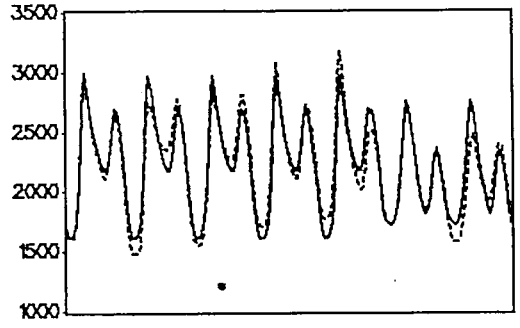


Figure 6.5c Estimated weekly load pattern with Weekend correction

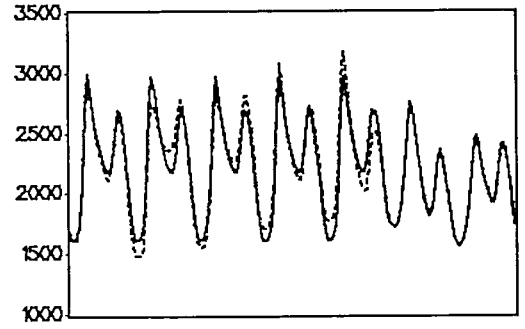


Figure 6.7a Load pattern of week 5 of 1990

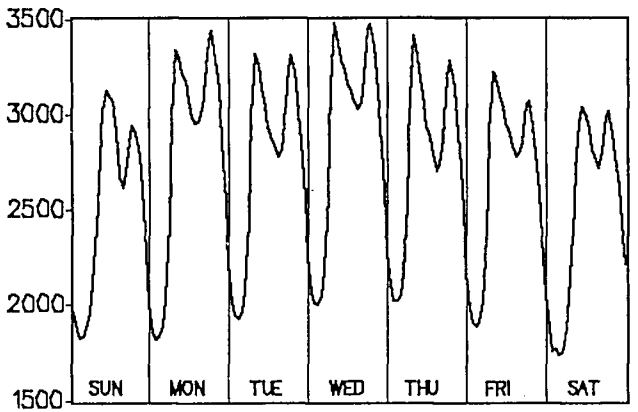


Figure 6.7b Load pattern of week 15 of 1990

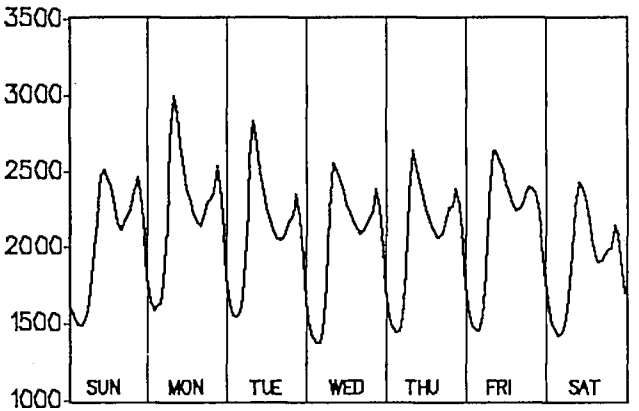


Figure 6.7c Load pattern of week 25 of 1990

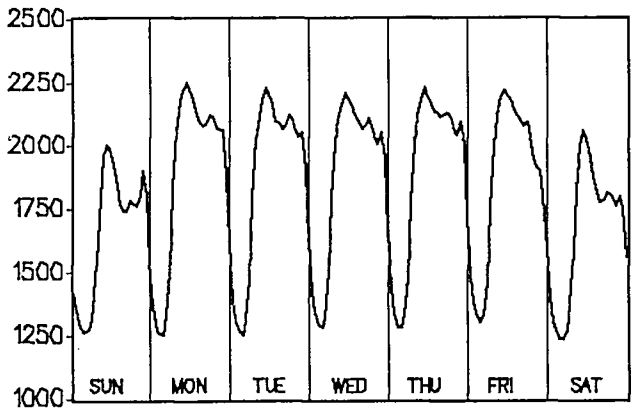


Figure 6.7d Load pattern of week 35 of 1990

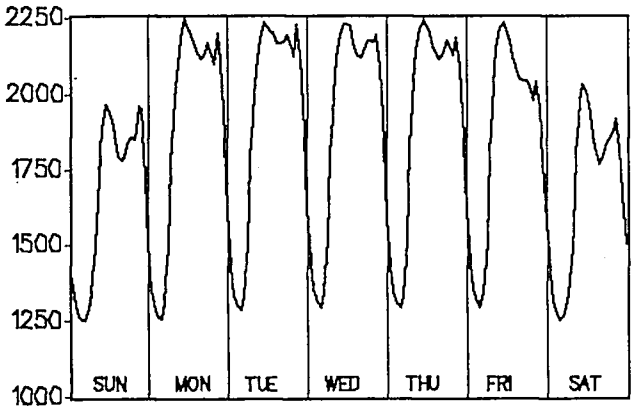
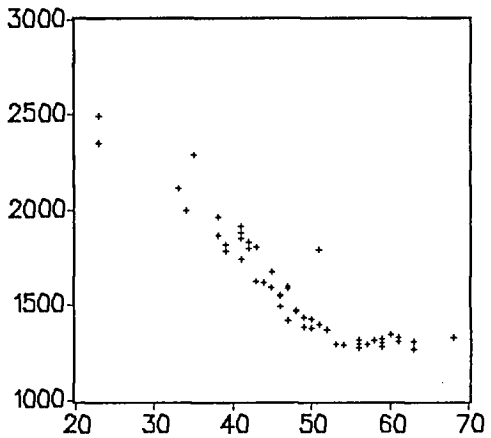


Figure 6.8a LOAD vs TEMP for all Wednesdays in 1990 at 3am



6.8b LOAD vs TEMP for all Wednesdays in 1990 at 8am

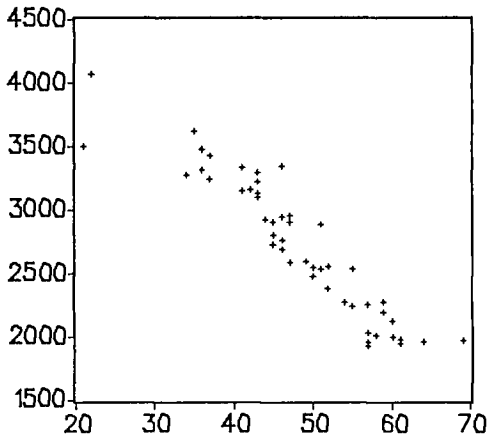
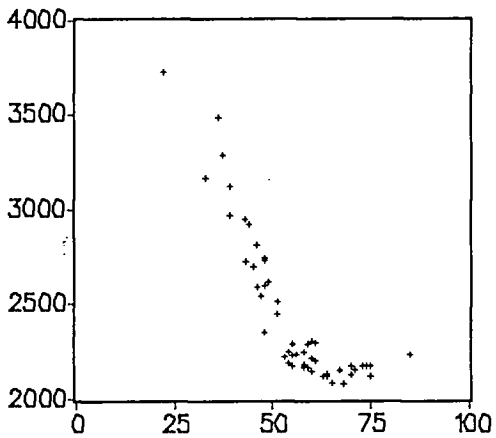


Figure 6.8c LOAD vs TEMP for all Wednesdays in 1990 at 1pm



6.8d LOAD vs TEMP for all Wednesdays in 1990 at 8pm

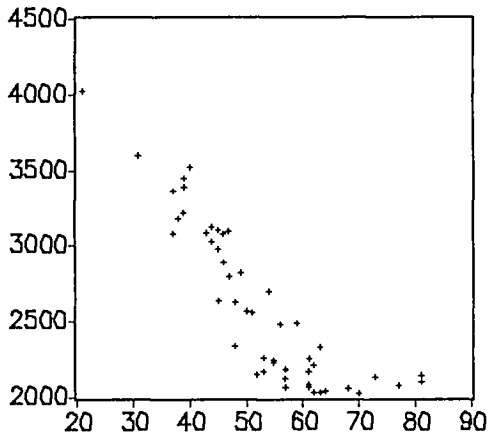
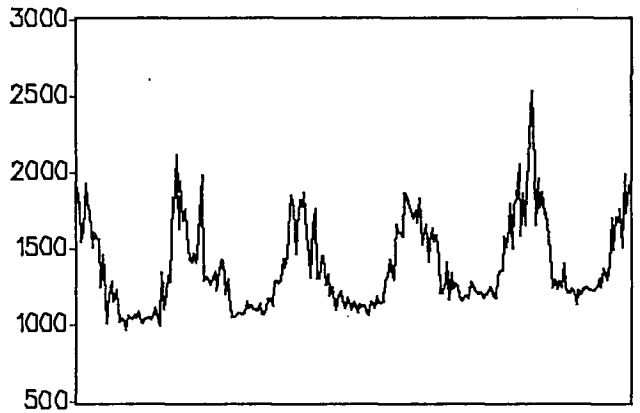


Figure 6.9a Load demand on Wednesdays in 1985–1989 at 3am



6.9b Load demand on Wednesdays in 1985–1989 at 8am

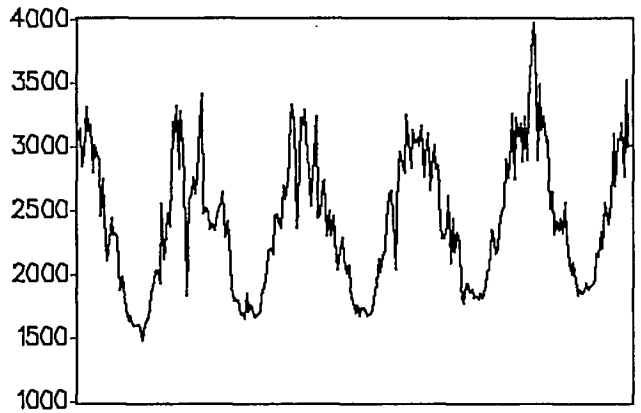
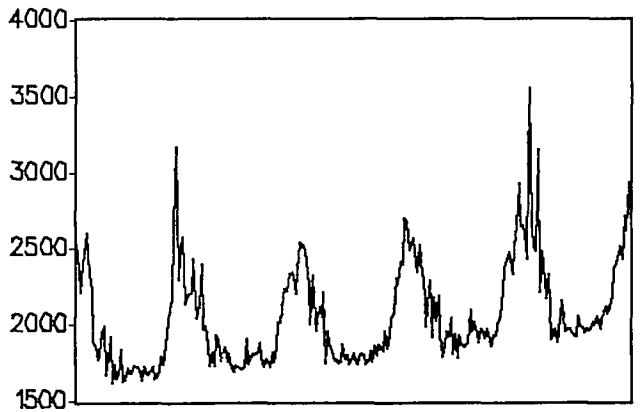


Figure 6.9c Load demand on Wednesdays in 1985–1989 at 4pm



6.9d Load demand on Wednesdays in 1985–1989 at 9pm

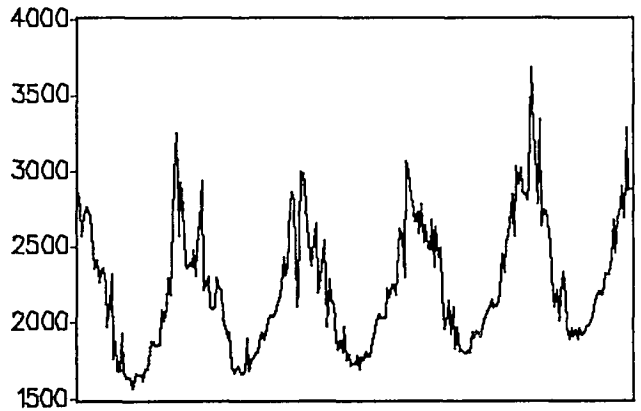


Figure 6.10a Estimated constant in preliminary regression

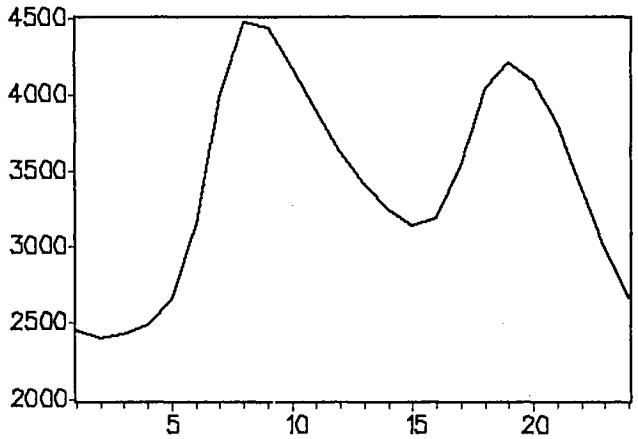


Figure 6.10b Estimated beta in preliminary regression

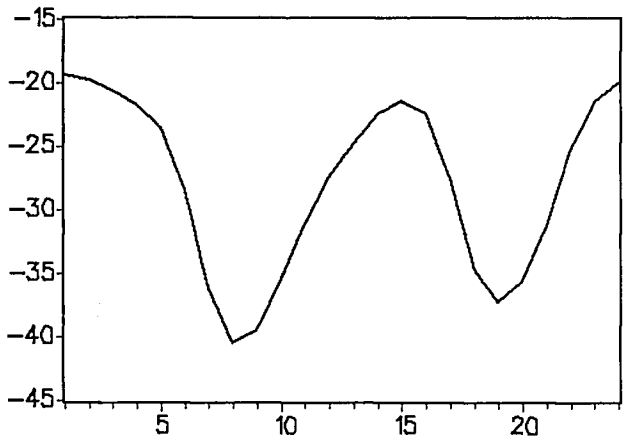


Figure 6.10c Estimated DSAT in preliminary regression

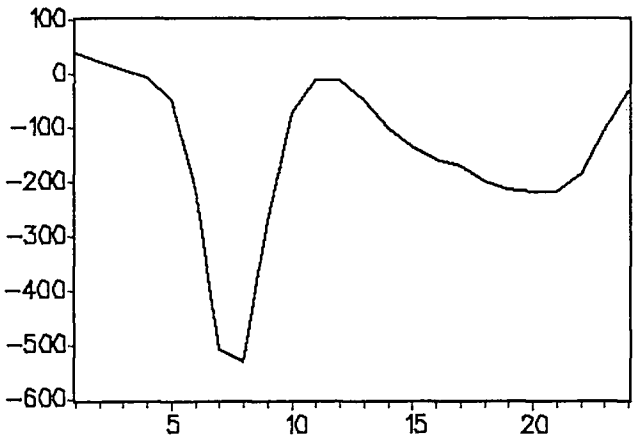


Figure 6.10d Estimated DSUN in preliminary regression

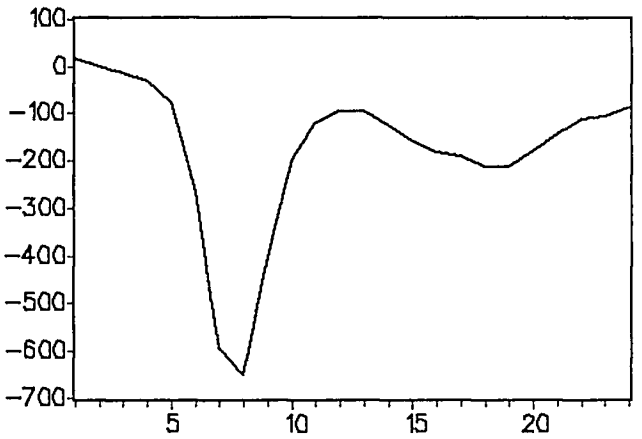


Figure 6.11a Estimated weekly load pattern during FEB

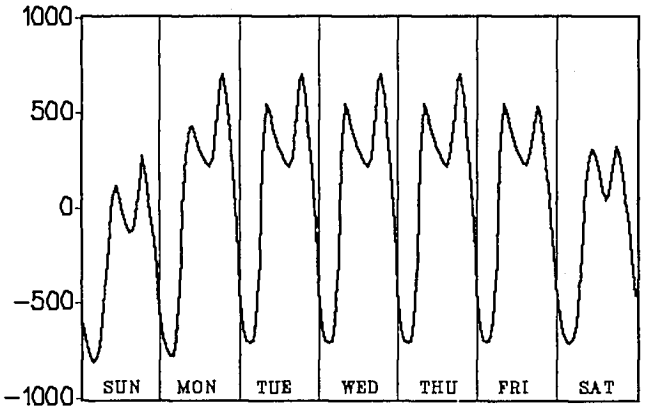


Figure 6.11b Estimated weekly load pattern during MAY

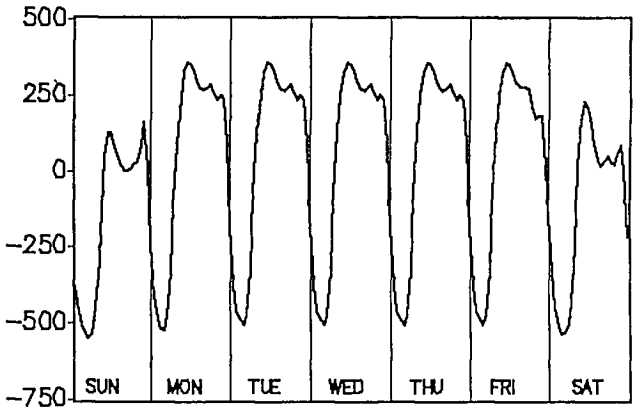


Figure 6.11c Estimated weekly load pattern during AUG

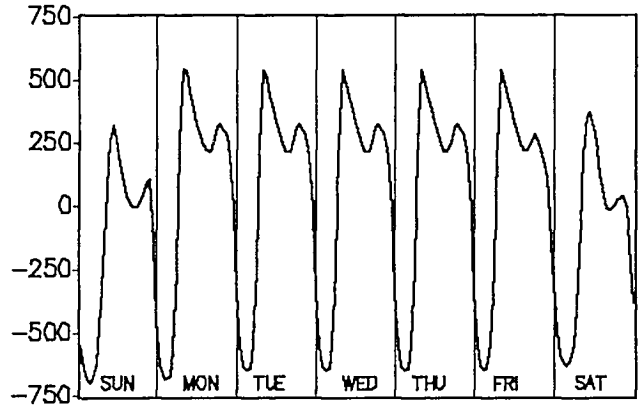


Figure 6.11d Estimated weekly load pattern during NOV

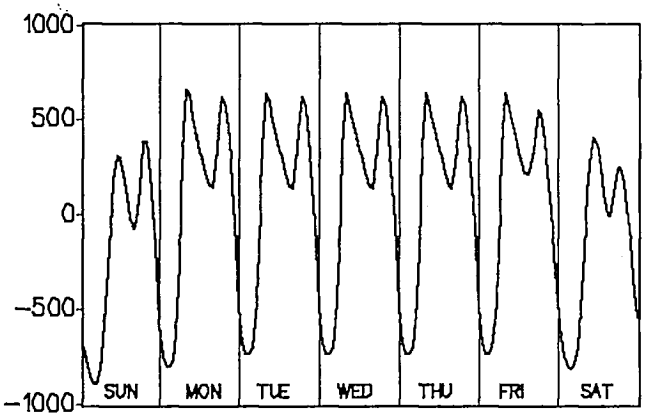


Figure 6.12 Estimated non-linear response to temperature

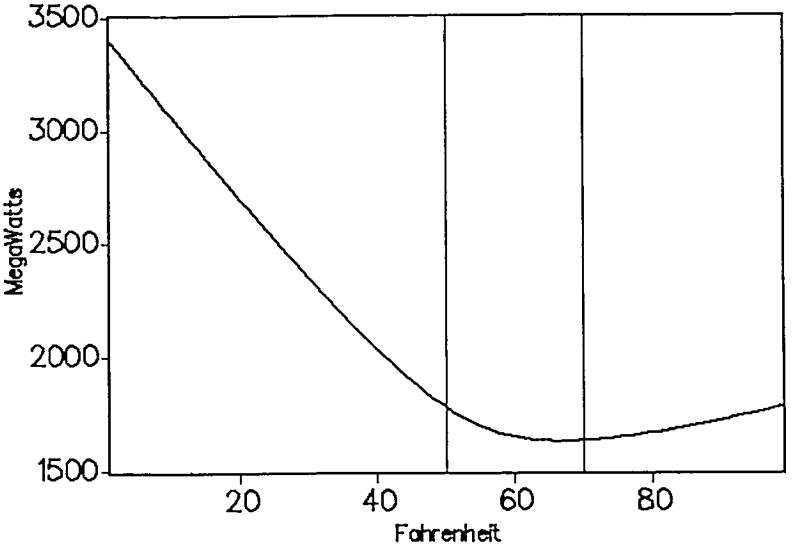


Figure 6.13 Estimated yearly cyclical pattern

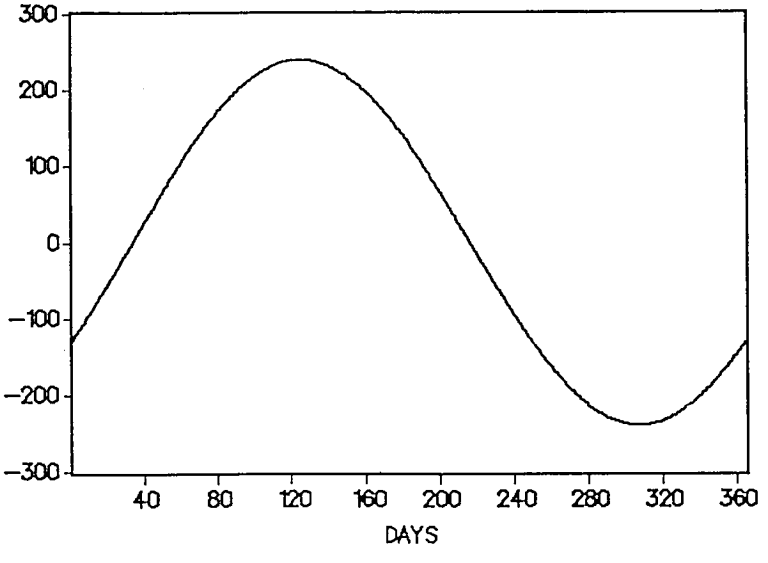


Figure 6.14 Weekly aggregated standard prediction error in 1990

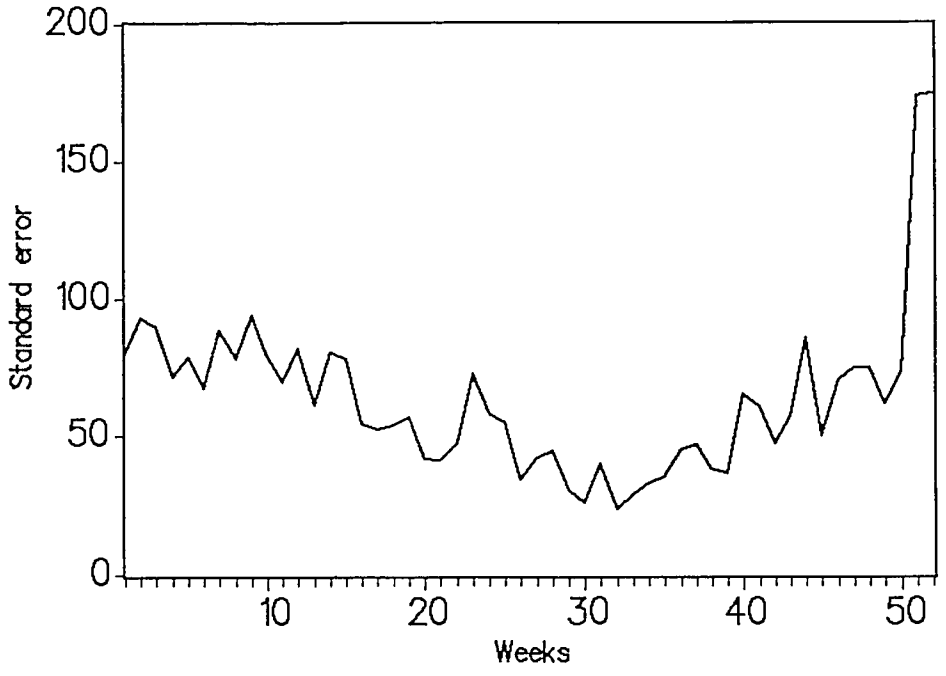


Figure 6.15a Actual vs Forecast Load on a Wednesday

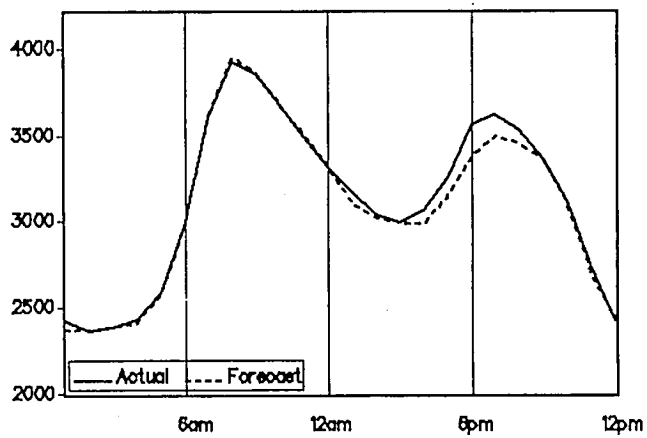


Figure 6.15b Absolute % Forecast Error for a Wednesday

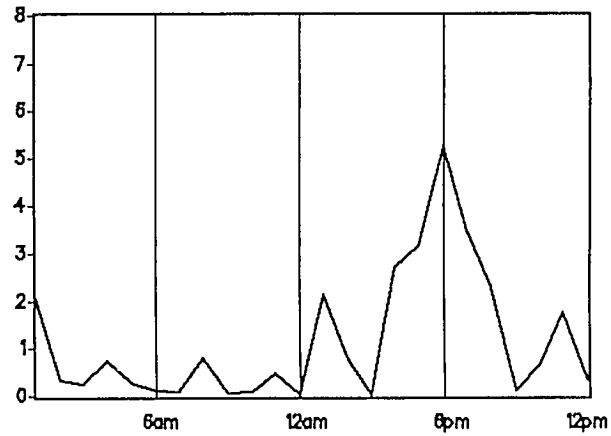


Figure 6.15c Actual vs Forecast Load for Weekend & Monday

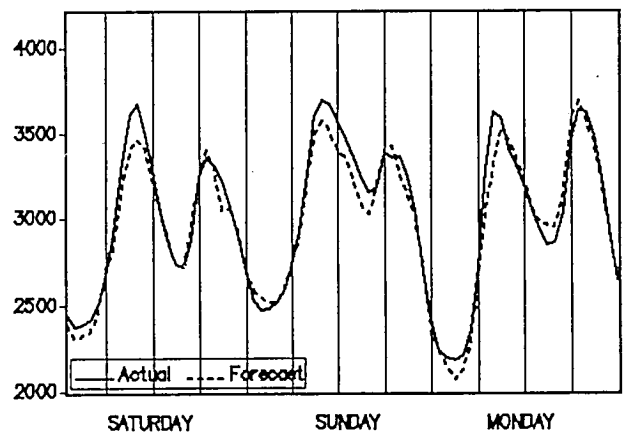
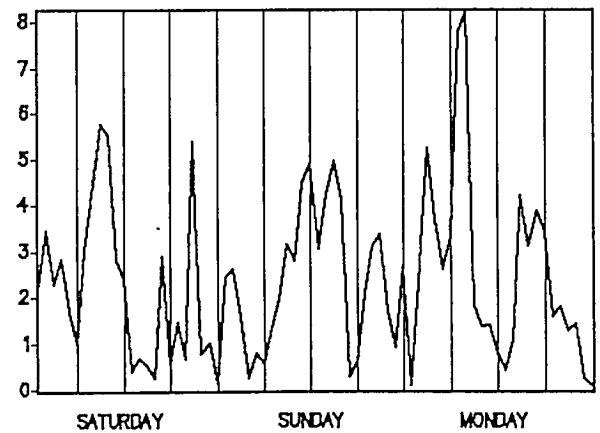


Figure 6.15d Absolute % Forecast Error for Weekend & Monday



CHAPTER 7
OVERVIEW AND DISCUSSION

This thesis deals with smoothing algorithms, diagnostic checking of unobserved components time series models, time-varying cubic splines and short term forecasting of time series with the periodic movements being intra-daily or intra-weekly. All topics are discussed extensively with emphasis on theory, algorithms and practical implications. The novel contributions to the time series literature are

[1] a smoothing algorithm for the disturbance vector and its (lagged) mean square error matrix

[2] an efficient smoothing algorithm for the state vector

[3] an alternative (more efficient) EM algorithm for estimating hyperparameters in the covariance structure of the SSF

[4] auxiliary residuals are proposed as an supplement to innovations

[5] diagnostic test statistics for auxiliary residuals are developed to detect and distinguish between outliers and structural change

[6] cubic splines are allowed to vary over time by letting the parameters follow stochastic processes

[7] a structural time series model is proposed for observations with intra-daily or intra-weekly periodic movements with an emphasis on short term forecasting

These novel contributions are illustrated with various examples which can be found especially in chapters 4 and 6. More elaborate

applications are presented on the quarterly coal production in the UK between 1960 and 1983 (Chapters 2 and 3), on the yearly consumption of spirits in the UK between 1870 and 1930 (Chapter 4) and on the hourly load demand of electricity at the Puget Sound Power & Light company in Bellevue, Washington, USA, between 1985 and August 1991 (Chapter 6). The following conclusions can be drawn from our analyses :

[a] the quarterly production of coal in the UK can be satisfactory described by a cumulation of three uncorrelated stochastic processes which can be directly interpreted as the unobserved components trend, seasonal and irregular; no significant distortions can be detected from the estimated model as observed from the innovations and the auxiliary residuals.

[b] the yearly consumption of spirits in the UK between 1870 and 1930 follow a time-varying trend and can be further explained by relative price and income per capita; a shift in the level of about 9 % is observed in 1909 from the auxiliary residuals and can be explained by the Government's policy on social issues and taxation; outliers are distinguished in 1915 and 1918 but there is also a cause for outliers in the remaining years of World War I.

[c] the hourly electricity demand at the Puget Power company in the period from 7 November 1990 to 31 March 1991 are well forecasted by the SHELF model (6.5.1) using the hourly data from 1 January 1985 onwards.

The necessary calculations are implemented in computer programs which are presented at the end of this thesis. The computer programs are computationally efficient.

COMPUTER PROGRAMS

```
Program Filtering;
Const
       DimVec =  4D;
       DimObs = 5DD;
Type
       Vector  = Array[D..DimVec] of Float;
       Matrix  = Array[D..DimVec] of Vector;
       SysFlt  = Record
                   Code     : D..9;
                   Positive : Boolean;
                   Value    : Real;
                   Pos      : Integer;
                 End;
       SysVec  = Array[D..DimVec] Of SysFlt^;
       SysMat  = Array[D..DimVec] Of SysVec;
       Data    = Array[D..DimObs] Df Float;
Var
   { SSF       yt = z_t α_t + x_t β + g_t ε_t
               α_t+1 = T_t α_t + W_t β + H_t ε_t
               α1 = W_D β + H_D ε_D        β = b + Bδ    }
    TSM {z_t;T_t},
    RGM {x_t b,x_t B;W_t b,W_t B},
    CSM {g_t g_t',g_t H_t';H_t g_t',H_t H_t'},
    ICM {W_D(b,B)},
    CIM {H_0 H_0'}                  : SysMat;
    DimSt {#α_t},
    DimBt {#β},
    DimEp {#ε_t},
    DimRg {#δ}                      : Integer;

    Y                               : Data;
    X                               : Array[1..DimVec] Of Data;
    t,n,fobs,lobs                   : Integer;

    { fobs and lobs allow analyzing subset of Y[1..n] : Y[fobs..lobs] }

Function Mult(t : Integer; Fl : Float; SFl : SysFlt^) : Float;
Begin
 With SFl^ do
 Case Code Of
  D : Mult:=D.D;
  1 : Mult:=Fl;
  2 : Mult:=Fl*Value;
  3 : Mult:=Fl*X[Pos,t];
  4 : Mult:=Fl*Sin(2*pi/Value);
  5 : Mult:=Fl*Cos(2*pi/Value);
  6 : Mult:=Fl*Tan(2*pi/Value);
  7 : Mult:=Fl*Sqr(Value);
  8 : Mult:=Fl*Sqrt(Value);
  9 : Mult:=Fl*Exp(Value);
 End;
 If Not SFl^.Positive then Mult:=-Mult;
End;
Function Add(t : Integer; SFl : SysFlt^) : Float;
Begin
 With SFl^ do
 Case Code Of
  D : Add:=D.D;
  1 : Add:=1.D;
  2 : Add:=Value;
  3 : Add:=X[Pos,t];
  4 : Add:=Sin(Value);
  5 : Add:=Cos(Value);
  6 : Add:=Tan(Value);
  7 : Add:=Sqr(Value);
  8 : Add:=Sqrt(Value);
  9 : Add:=Exp(Value);
 End;
 If Not SFl^.Positive then Add:=-Add;
End;
```

```
Procedure InitialF1( JobCode : Byte; Var St,CSt : Matrix );
Var i,j : Byte;
Begin
 Case JobCode of
 1 : { KF diffuse starting condition : St=0,CSt=∞ }
     For i:=1 to DimSt do
     Begin
      St[i,0]:=0.0;ZeroVec(CSt[i]);CSt[i,i]:=999999.99
     End;
 2 : { KF zero starting condition : St=0,CSt=0 }
     For i:=1 to DimSt do
     Begin
      St[i,0]:=0.0;ZeroVec(CSt[i]);
     End;
 3 : { Modified KF : St=W0(b,B);CSt=H0H0' }
     For i:=1 to DimSt do
     Begin
      For j:=0 to DimRg do  St[i,j]:=Add(0,ICM[i,j]);
      For j:=1 to DimSt do CSt[i,j]:=Add(0,CIM[i,j]);
     End
 End
End;

Procedure KFUpdate(Y : Real; DimKF : Integer; Var A,P : Matrix);
Var   M1,M2  : Matrix;
      i,j,k  : Byte;
Begin
{Step 1}
    For i:=0 to DimSt do
    Begin
     ZeroVec(M1[i]);
{!}  ZeroVec(M2[i]);
     For j:=1 to DimSt do
     If TSM[i,j]<>NIL then
     Begin
      For k:=0 to DimKF do M1[i,k]:=M1[i,k]+Mult(t,A[j,k],TSM[i,j]);
{!}   For k:=1 to DimSt do M2[i,k]:=M2[i,k]+Mult(t,P[j,k],TSM[i,j]);
     End
    End;

{Step 2}
{!} For i:=0 to DimSt do
    Begin
     ZeroVec(P[i]);
     For j:=1 to DimSt do
     If TSM[i,j]<>NIL then
     For k:=0 to DimSt do P[i,k]:=P[i,k]+Mult(t,M2[k,j],TSM[i,j]);
    End;

{Step 3}
    For i:=0 to DimSt do
    Begin
     EqualVec(A[i],M1[i]);
     For j:=0 to DimKF do A[i,j]:=A[i,j]+Add(t,RGM[i,j]);
{!}  For j:=0 to DimSt do P[i,j]:=P[i,j]+Add(t,CSM[i,j]);
    End;

{Step 4 and 5}
    A[0,0]:=Y-A[0,0];
    For i:=1 to DimKF do A[0,i]:=-A[0,i];
{!} For i:=1 to DimSt do P[0,i]:=P[0,i]/P[0,0];

{Step 6}
    For i:=1 to DimSt do
    Begin
     For j:=0 to DimKF A[i,j]:=A[i,j]-(P[0,i]*A[0,j]);
{!}  For j:=1 to DimSt do P[i,j]:=P[i,j]-(P[0,i]*P[0,0]*P[0,j]);
    End;
End;

{ the sign ! means that the statement can be dropped in a steady state
  the procedure ZeroVec(x) puts zeroes in each element of x
  the procedure EqualVec(x,y) puts subsequently the elements of y into x}
```

```
Procedure Estimation
( Var Q : Matrix; Var L : Float; Var d : Integer; Obs : Data );

Var   A,P,S     : Matrix;
      k         : Vector;
      v,f,sigma : Float;
      i,j       : Integer;
Begin
 For i:=0 to DimRg do
 Begin
  ZeroVec(S[i]);ZeroVec(Q[i]);
 End;
 L:=0.0;
 d:=0
 InitialKF(3,A,P);
 For t:=fobs to lobs do
 Begin
  KFUpdate(y[t],DimRg,A,P);
  { Check p.d. of F_t and P_{t+1|t} and other numerical unstabilities
    Store vector A[0,0..DimRg]
(!) Store vector P[0,0..DimRg] }
  L:=L+Ln(P[0,0]);

  If d=0 then
  Begin
   For i:=1 to DimRg do
   For j:=1 to DimRg do S[i,j]:=S[i,j]+(A[0,i]*A[0,j]/P[0,0]);
   If Not (t<DimRg) then
   Begin
    Det:=Determinant(1,DimRg,S);
    If Not (Det=0) then
    Begin
     Inverse(1,DimRg,Det,S,Q);
     Q[0,0]:=S[0,0];
     For i:=1 to DimRg do
     Begin
      Q[0,i]:=S[0,i];
      Q[i,0]:=0.0;
      For j:=1 to DimRg do Q[i,0]:=Q[i,0]-(S[i,j]*S[0,j]);
     End;
     d:=t;
    End;
   End;
  End
  Else  { If d is not zero }
  Begin
   For j:=0 to DimRg do Q[0,j]:=Q[0,j]+(A[0,j]*A[0,0]/P[0,0]);
   v:=A[0,0];f:=P[0,0];
   For i:=1 to DimRg do
   Begin
    v:=v+(A[0,i]*Q[i,0]);
    k[i]:=0.0;
    For j:=1 to DimRg do
    Begin
     k[i]:=k[i]+(A[0,j]*Q[j,i]);
    End;
    f:=f+(k[i]*A[0,i])
   End;
   For i:=1 to DimRg do
   Begin
    Q[i,0]:=Q[i,0]-(k[i]*v/f);
    For j:=1 to DimRg do Q[i,j]:=Q[i,j]-(k[i]*k[j]/f)
   End;
   sigma:=Q[0,0];
   For i:=1 to DimRg do sigma:=sigma-(Q[i,0]*Q[0,i]);
   sigma:=sigma/(t-fobs+1);
  End;
 End;

 Q[0,0]:=sigma;
 Det:=Determinant(1,DimRg,Q);
 L:=L-Det+((lobs-fobs+1)*(1+ln(2*pi*Q[0,0])));
End;
```

```
Procedure SmoothUpdate
( DimKF : Integer; Var DSE,DSD,DSU,DSC : Matrix );
Var M1,M2  : Matrix;
    i,j,k  : Integer;
Begin
{Step 1}
    For i:=0 to DimKF do DSE[0,i]:=DSE[0,i]/DSD[0,0];
    DSD[0,0]:=1/DSD[0,0];

    For i:=1 to DimSt do
    Begin
     For j:=0 to DimKF do DSE[0,j]:=DSE[0,j]-(DSD[0,i]*DSE[i,j]);
     DSD[i,0]:=0;
     For j:=1 to DimSt do DSD[i,0]:=DSD[i,0]-(DSD[i,j]*DSD[0,j]);
     DSD[0,0]:=DSD[0,0]-(DSD[i,0]*DSD[0,i])
    End;
    For i:=1 to DimSt do DSD[0,i]:=DSD[i,0];

{Step 2}
    For i:=0 to DimSt do
    Begin
     ZeroVec(DSU[i]);ZeroVec(M2[i]);
     For j:=0 to DimSt do
     If CSM[j,i]<>NIL then
     Begin
      For k:=0 to DimKF do DSU[i,k]:=DSU[i,k]+Mult(t,DSE[j,k],CSM[j,i]);
      For k:=0 to DimSt do M2[i,k]:=M2[i,k]+Mult(t,DSD[j,k],CSM[j,i])
     End;
    End;

{Step 3}
    For i:=0 to DimSt do
    Begin
     For j:=0 to DimSt do DSC[i,j]:=0.0;
     For j:=0 to DimSt do
     If CSM[j,i]<>NIL then
     For k:=1 to DimSt do DSC[i,k]:=DSC[i,k]+Mult(t,M2[k,j],CSM[j,i]);
    End;

{Step 4}
    For i:=1 to DimSt do
    Begin
     ZeroVec(M1[i]);ZeroVec(M2[i]);
     For j:=0 to DimSt do
     If TSM[j,i]<>NIL then
     Begin
      For k:=0 to DimKF do M1[i,k]:=M1[i,k]+Mult(t,DSE[j,k],TSM[j,i]);
      For k:=0 to DimSt do M2[i,k]:=M2[i,k]+Mult(t,DSD[j,k],TSM[j,i])
     End;
    End;

{Step 5}
    For i:=1 to DimSt do
    Begin
     For j:=1 to DimSt do DSD[i,j]:=0.0;
     For j:=0 to DimKF do DSE[i,j]:=M1[i,j];

     For j:=0 to DimSt do
     If TSM[j,i]<>NIL then
     For k:=1 to DimSt do DSD[i,k]:=DSD[i,k]+Mult(t,M2[k,j],TSM[j,i]);
    End;
End;
```

```
Procedure AuxResiduals
( Nr : Integer; Q : Matrix; Var r,vr : Data );
Var E,D,U,C      : Matrix;
    rl           : Real;
    i,j,k        : Integer;
Begin
   { Open memory space for KF quantities }
   { Read (F_n,K'_n) and put it into D[0,..] }
   { R_n=0 N_n=0 }
   For i:=0 to DimSt do Begin ZeroVec(E[i]);ZeroVec(D[i]) End;

   For t:=lobs downto fobs do
   Begin
    { Read V_t and put it into E[0,..] }
{!} { Read F_t,K_t }
    { Put (F_t,K'_t) into D[0,..] }
    SmoothUpdate(DimRg,E,D,U,C);

    r[t]:=U[Nr,0];
    For j:=1 to DimRg do r[t]:=r[t]+(U[Nr,j]*Q[j,0]);
    vr[t]:=Add(t,CSM[Nr,Nr])-C[Nr,Nr];
    For j:=1 to DimRg do
    Begin
     rl:=0.0;
     For k:=1 to DimRg do rl:=rl+(U[Nr,k]*Q[k,j]);
     vr[t]:=vr[t]+(U[Nr,j]*rl);
    End;
   End;
End;

Procedure AuxCrossMse
( Nr1,Nr2 : Integer; Q : Matrix; Var cc : Data );
Var E,D,U,C           : Matrix;
    rl                : Real;
    i,j,k             : Integer;
Begin
   { Open memory space for KF quantities }
   { Read (F_n,K'_n) and put it into D[0,..] }
   { R_n=0 N_n=0 }
   For i:=0 to DimSt do Begin ZeroVec(E[i]);ZeroVec(D[i]) End;

   For t:=lobs downto fobs do
   Begin
    { Read V_t and put it into E[0,..] }
{!} { Read F_t,K_t }
    { Put (F_t,K'_t) into D[0,..] }
    SmoothUpdate(DimRg,E,D,U,C);

    cc[t]:=Add(t,CSM[Nr1,Nr2])-C[Nr1,Nr2];
    For j:=1 to DimRg do
    Begin
     rl:=0.0;
     For k:=1 to DimRg do rl:=rl+(U[Nr1,k]*Q[k,j]);
     cc[t]:=cc[t]+(U[Nr2,j]*rl);
    End;
   End;
End;
```

```
Program Splines;
Const
       DimVec  =  40;
       DimObs  = 500;
Type
       Vector  = Array[D..DimVec] of Float;
       Matrix  = Array[0..DimVec] of Vector;
       Data    = Array[0..DimObs] of Real;

Procedure NatPinvQ
( k : Integer; xk : Vector; Var M : Matrix );
Var  q,v,z,P1,P2,P3 : Vector;
     rl,rk          : Real;
     i,j,l          : Integer;
Begin
   { k=number of knots xk=mesh M=P^-1Q }

   { distances between knots }
   For i:=1 to k do z[i]:=xk[i]-xk[i-1];

   { tridiagonal structure of P put in three vectors P1 P2 P3 }
   For i:=1 to k-1 do
   Begin
    rl:=z[i]+z[i+1];
    P1[i]:=z[i]/rl;
    P2[i]:=2.0;
    P3[i]:=z[i+1]/rl;
   End;
   P1[k]:=0.0;P2[0]:=1.0;P2[k]:=1.0;P3[0]:=0.0;

   { sequence of solving k sets of linear restrictions to get PinvQ }
   For i:=0 to k do
   Begin
    { i-th column of Q is put in q }
    For j:=0 to k do q[j]:=0.0;
    If (i>1) then q[i-1]:=6/(z[i]*(z[i-1]+z[i]));
    If ( (i>0) AND (i<k) ) then q[i]:=-6/(z[i]*z[i+1]);
    If (i<k-1) then q[i+1]:=6/(z[i+1]*(z[i+1]+z[i+2]));

    { solving m in Pm=q where m is i-th column of M, see Press et.al (1988) }
    rl:=P2[0];
    M[0,i]:=q[0]/rl;
    For j:=1 to k do
    Begin
     v[j]:=P3[j-1]/rl;
     rl:=P2[j]-(P1[j]*v[j]);
     M[j,i]:=(q[j]-(P1[j]*M[j-1,i]))/rl;
    End;
    For j:=(k-1) downto 0 do
    M[j,i]:=M[j,i]-(v[j+1]*M[j+1,i]);
   End;
End;
```

```
Procedure PerPinvQ
( k : Integer; xk : Vector; Var M : Matrix );
Var q,v,z,P1,P2,P3 : Vector;
    W,bM           : Matrix;
    rl,rk          : Real;
    i,j,l          : Integer;

Begin
   { k=number of knots xk=mesh M=P^-1Q }

   { distances between knots }
   For i:=1 to k do z[i]:=xk[i]-xk[i-1];
   z[0]:=z[k];
   z[k+1]:=z[1];

   { tridiagonal structure of P* put in three vectors P1 P2 P3 }
   For i:=1 to k do
   Begin
    rl:=z[i]+z[i+1];
    P1[i]:=z[i]/rl;
    P2[i]:=2.0;
    P3[i]:=z[i+1]/rl;
   End;

   For i:=0 to k+1 do
   Begin
    { i-th column of Q is put in q for i=1,...,k }
    { If i=0   then the first  column of U' is put into q }
    { If i=k+1 then the second column of U' is put into q }

    For j:=1 to k do q[j]:=0.0;
    If i=0   then q[1]:=P1[1];
    If i=k+1 then q[k]:=P3[k];
    If ((i<k+1) AND (i>0)) then
    Begin
     q[i-1]:=6/(z[i]*(z[i-1]+z[i]));
     q[i]:= -6/(z[i]*z[i+1]);
     q[i+1]:=6/(z[i+1]*(z[i+1]+z[i+2]));
     If i=1 then q[k]:=6/(z[1]*(z[1]+z[k]));
     If i=k then q[1]:=6/(z[1]*(z[1]+z[2]));
    End;

    { solving m in Pm=q where m is i-th column of M, see Press et.al (1988) }
    rl:=P2[1];
    M[1,i]:=q[1]/rl;
    For j:=2 to k do
    Begin
     v[j]:=P3[j-1]/rl;
     rl:=P2[j]-(P1[j]*v[j]);
     M[j,i]:=(q[j]-(P1[j]*M[j-1,i]))/rl;
    End;
    For j:=k-1 downto 1 do
    M[j,i]:=M[j,i]-(v[j+1]*M[j+1,i]);
   End;

   { rl is determinant of (VP*^-1U'+I) }
   rl:=((1.0+M[k,0])*(1.0+M[1,k+1]))-(M[1,0]*M[k,k+1]);

   { calculate P*^-1U'(VP*^-1U'+I)^-1 }
   For i:=1 to k do
   Begin
    P1[i]:=( ((1.0+M[1,k+1])*M[k,i])-(M[k,k+1]*M[1,i]) )/rl;
    P2[i]:=( ((1.0+M[k,0])*M[1,i])-(M[1,0]*M[k,i]) )/rl;
   End;

   { calculate P^-1Q }
   For i:=1 to k do
   For j:=1 to k do
   M[i,j]:=M[i,j]-(M[i,0]*P1[j])-(M[i,k+1]*P2[j]);

End;
```

```
Procedure NatSplineWeight
( k : Integer; kn : Vector; M : Matrix; x : Real; Var w : Vector );
Var h,j     : Integer;
    z,rl,rk : Real;
Begin
 j:=1;
 z:=kn[1]-kn[0];
 While (x>kn[j]) do
 Begin
  j:=j+1;
  z:=kn[j]-kn[j-1];
 End;
 rl:=kn[j]-x;
 rk:=x-kn[j-1];

 For h:=0 to k do
 w[h]:=((rl/(6*z))*(sqr(rl)-sqr(z))*M[j-1,h]) +
       ((rk/(6*z))*(sqr(rk)-sqr(z))*M[j,h]);
 w[j-1]:=w[j-1]+(rl/z);
 w[j]  :=w[j]  +(rk/z);
End;

Procedure PerSplineWeight
( k : Integer; kn : Vector; M : Matrix; x : Real; Var w : Vector );
Var h,j     : Integer;
    z,rl,rk : Real;
Begin
 j:=1;
 z:=kn[1];
 While (x>kn[j]) do
 Begin
  j:=j+1;
  z:=kn[j]-kn[j-1];
 End;
 rl:=kn[j]-x;
 rk:=x-kn[j-1];

 If j=1 then
 Begin
  For h:=1 to k do
  w[h]:=((rl/(6*z))*(sqr(rl)-sqr(z))*M[k,h]) +
        ((rk/(6*z))*(sqr(rk)-sqr(z))*M[j,h]);
  w[k]:=w[k]+(rl/z);
  w[j]:=w[j]+(rk/z);
 End
 Else
 Begin
  For h:=1 to k do
  w[h]:=((rl/(6*z))*(sqr(rl)-sqr(z))*M[j-1,h]) +
        ((rk/(6*z))*(sqr(rk)-sqr(z))*M[j,h]);
  w[j-1]:=w[j-1]+(rl/z);
  w[j]  :=w[j]  +(rk/z);
 End;
End;

Procedure SeasSpline
( p,k : Integer; Var W : Matrix );
Var t,i : Integer;
    sW  : Vector;
Begin
 For i:=1 to k do sW[i]:=0.0;
 For t:=1 to p do
 For i:=1 to k do sW[i]:=sW[i]+W[t,i];
 For t:=1 to p do
 Begin
  For i:=1 to k-1 do W[t,i]:=W[t,i]-((W[t,k]/sW[k])*sW[i]);
  W[t,k]:=1.0;
 End;
End;
```

REFERENCES

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. Econometrica 33, 178-196.

Anderson, B.D.O. and Moore, J.B. (1979). Optimal Filtering. Englewood Cliffs, New Jersey: Prentice Hall.

Ansley, C.F. and Kohn, R. (1982). A geometrical derivation of the fixed interval smoothing algorithm. Biometrika 69, 486-487.

Ansley, C.F. and Kohn, R. (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. Annals of Statistics 13, 1286-1316.

Bell, W. and Hillmer, S. (1991). Initializing the Kalman filter for nonstationary time series models. Journal of Time Series Analysis 12, 283-300.

Boneva, L.I, Kendall, D. and Stefanov, I. (1971). Spline transformations : three new diagnostic aids for the statistical data analysis (with discussion). Journal of the Royal Statistical Society (series B) 33, 1-70.

Bowman, K.O. and Shenton, L.R. (1975). Omnibus test contours for departures from normality based on $b_1^{\frac{1}{4}}$ and b_2. Biometrika 62, 243-250.

Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis: Forecasting and control (revised edition). San Francisco: Holden-Day.

Box, G.E.P. and Pierce, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated-moving average time series models. Journal of the American Statistical Association 65, 1509-1526.

Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques of testing the constancy of regression relationships over time. Journal of the Royal Statistical Society (series B) 37, 141-192.

Bruce, A.G. and Martin, R.D. (1989). Leave-k-out diagnostics for time series. Journal of the Royal Statistical Society (series B) 51, 363-424.

Bunn, D. and Falmer, E.D. (1985). Comparitive models for electrical load forecasting. New York: John Wiley & Sons.

Burmeister, E., Wall, K.D. and Hamilton J.D. (1986). Estimation of unobserved expected monthly inflation using Kalman filtering. Journal of Business and Economic Statistics 4, 147-160.

Burridge, P. and Wallis, K.F. (1985). Calculating the variance of seasonally adjusted series. Journal of the American Statistical Association 80, 541-552.

Casteren, P.H.F.M. van (1991). An improved frequency polygon. Kwantitatieve Methoden 38, 5-16.

Chang, I., Tiao, G.C. and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. Technometrics 30, 193-204.

Chow, G.C. (1984). Random and changing coefficient models. In Griliches, Z. and Intriligator, M. (editors) Handbook of Econometrics (volume 2), 1213-1245. Amsterdam: North Holland.

Chu-Chun-Lin, S. (1991). Statistical analysis with the state space model. Ph.D thesis, University of British Columbia, Vancouver.

Cochrane, D. and Orcutt, G.H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. Journal of the American Statistical Association 44, 32-61.

De Jong, P. (1988a). The likelihood for a state space model. Biometrika 75, 165-169.

De Jong, P. (1988b). A cross-validation filter for time series models. Biometrika 75, 594-600.

De Jong, P. (1989). Smoothing and interpolation with the state space model. Journal of the American Statistical Association 84, 1085-1088.

De Jong, P. (1991a). Stable algorithms for the state space model. Journal of Time Series Analysis 12, 143-157.

De Jong, P. (1991b). The diffuse Kalman filter. Annals of Statistics 19, 1073-1083.

De Jong, P. and Chu-Chun-Lin, S. (1991). Stationary and nonstationary state space models. Mimeo.

De Jong, P. and MacKinnon M.J. (1988). Covariances for smoothed estimates in state space models. Biometrika 75, 601-602.

Duncan, D.B. and Horn, S.D. (1972). Linear dynamic regression from the viewpoint of regression analysis. Journal of the American Statistical Association 67, 815-821.

Durbin, J. and Watson, G.S. (1951). Testing for serial correlation in least squares regression II. Biometrika 56, 1-15.

Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. Journal of the American Statistical Association 81, 310-320.

Engle, R.F. and Watson M.W. (1981). A one-factor multivariate time series model of metropolitan wage rates. Journal of the American Statistical Association 76, 744-781.

Fuller, W.A. (1976). Introduction to statistical time series. New York: John Wiley & Sons.

Garcia-Ferrer, A. and del Hoyo, J. (1991). On trend extraction models: interpretation, empirical evidence and forecasting performance (with discussion). Journal of Forecasting (forthcoming).

Gardner, G., Harvey, A.C. and Phillips, G.D.A. (1980). An algorithm for exact maximum likelihood estimation by means of Kalman filtering. Applied Statistics 29, 311-322.

Golub, G.H. and Van Loan, C.F. (1989). Matrix computations (second edition). Baltimore: The Johns Hopkins University Press

Granger, C.W.J. and Newbold, P. (1977). Forecasting economic time series. New York: Academic Press.

Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting. Journal of the Royal Statistical Society (series B) 38, 205-247.

Harvey, A.C. (1981). Time series models. London: Philip Allen.

Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press.

Harvey, A.C. and Durbin, J. (1986). The effects of seat belts legislation on british road casualties: A case study in Structural time series modelling. Journal of the Royal Statistical Society (series A) 149, 187-227.

Harvey, A.C. and Peters, S. (1990). Estimation procedures for structural time series models. Journal of Forecasting 9, 89-108.

Harvey, A.C. and Phillips, G.D.A. (1979). The estimation of regression models with autoregressive-moving average disturbances. Biometrika 66, 49-58.

Harvey, A.C. and Pierse, R.G. (1984). Estimating missing observations in economic time series. Journal of the American Statistical Association 79, 125-131.

Harvey, A.C. and Shephard, N.G. (1992). Stuctural time series models. In Rao, C.R. Econometrics (forthcoming).

Harvey, A.C. and Todd, P.H.J. (1983). Forecasting economic time series with structural and Box-Jenkins models (with discussion). Journal of Business and Economic Statistics 1, 299-315.

Hendricks, W., Koenker, R. and Poirier, D.J. (1979). Residential demand for electricity: an econometric approach. Journal of econometrics 9, 33-57.

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.

Johnston, J. (1984). Econometric methods (third edition). Singapore: McGraw-Hill.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82, 35-45.

Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction theory. Journal of Basic Engineering 83, 95-107.

Kitagawa, G. and Gersch, W. (1984). A smoothness priors-state space modeling of time series with trend and seasonality. Journal of the American Statistical Association 79, 378-389.

Knuth, D.E. (1981). The art of computer programming (vol II). Reading, Massachusetts: Addison-Wesley.

Kohn, R. and Ansley, C.F. (1985). Efficient estimation and prediction in time series regression models. Biometrika 72, 694-697.

Kohn, R. and Ansley, C.F. (1989). A fast algorithm for signal extraction, influence and cross-validation in state space models. Biometrika 76, 65-79.

Koopman, S.J. (1991). Efficient smoothing algorithms for time series models. Mimeo.

Koopman, S.J. (1992). On density approximation by fitting a spline through a histogram. Mimeo.

Ljung G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. Biometrika 66, 67-72.

Lomnicki, Z.A. (1961). Tests for departure from normality in the base of linear stochastic processes. Metrika 4, 37-62.

Maddala, G.S. (1988). Introduction to econometrics. New York: Macmillan.

Magnus, J.R. and Neudecker, H. (1988). Matrix differential calculus with applications in statistics and econometrics. Chichester: John Wiley & Sons.

Maravall, A. (1987). Minimum mean squared errors estimation of the noise in unobserved components models. Journal of Business and Economic Statistics 5, 115-120.

Marshall, P. (1992). State space models with diffuse initial conditions. Journal of Time Series Analysis (forthcoming).

McCullagh, P. and Nelder, J.A. (1989). Generalized linear models (second edition). London: Chapman and Hall.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974). Introduction to the theory of statistics (third edition). Singapore: McGraw-Hill.

Morf, M. and Kailath, T. (1975). Square root algorithms for least squares estimation. IEEE Transactions on Automatic Control AC-20, 487-497.

Pierce, D.A. (1979). Signal extraction error in nonstationary time series. Annals of Statistics 7, 1303-1320.

Poirier, D.J. (1973). Piecewise regression using cubic splines. Journal of the American Statistical Association 68, 515-524.

Poirier, D.J. (1976). The econometrics of structural change with special emphasis on spline functions. Amsterdam: North Holland.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1989). Numerical recipes in Pascal. Cambridge: Cambridge University Press.

Prest, A.R. (1949). Some experiments in demand analysis. Review of Economics and Statistics 31, 33-49.

Rao, C.R. (1973). Linear statistical inference and its applications (second edition). New York: John Wiley & Sons.

Rosenberg, B. (1973). Random coefficient models : the analysis of a cross-section of time series by stochastically convergent parameter regression. Annals of Economic and Social Measurement 2, 399-428.

Shephard, N.G. and Harvey, A.C. (1990). On the probability of estimating a deterministic component in the local level model. Journal of Time Series Analysis 11, 339-347.

Shumway, R.H. & Stoffer D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. Journal of Time Series Analysis 3, 253-264.

Tsay, R.S. (1986). Time series model specification in the presence of outliers. Journal of the American Statistical Association 81, 132-141.

Tsay, R.S. (1988). Outliers, level shifts and variance changes in time series. Journal of Forecasting 7, 1-20.

Watson, M.W. and Engle, R.F. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression. Journal of Econometrics 23, 385-400.

West, M. and Harrison, J. (1989). Bayesian forecasting and dynamic models. New York: Springer Verlag.

Whittle, P. (1991). Likelihood and cost as path integrals (with discussion). Journal of the Royal Statistical Society (series B) 53, 505-538.