STATISTICAL ANALYSIS WITH THE STATE SPACE MODEL

By

Singfat Chu-Chun-Lin

B. Sc.(Hons.) Queen's University at Kingston

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

COMMERCE AND BUSINESS ADMINISTRATION

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1991

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Commerce and Business Administration
The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date: 5 December 1991

Abstract

The State Space Model (SSM) encompasses the class of multivariate linear models, in particular, regression models with fixed, time-varying and random parameters, time series models, unobserved components models and combinations thereof. The well-known Kalman Filter (KF) provides a unifying tool for conducting statistical inferences with the SSM.

A major practical problem with the KF concerns its initialization when either the initial state or the regression parameter (or both) in the SSM are *diffuse*. In these situations, it is common practice to either apply the KF to a transformation of the data which is functionally independent of the diffuse parameters or else initialize the KF with an arbitrarily large error covariance matrix. However neither approach is entirely satisfactory. The data transformation required in the first approach can be computationally tedious and furthermore it may not preserve the state space structure. The second approach is theoretically and numerically unsound. Recently however, De Jong (1991) has developed an extension of the KF, called the Diffuse Kalman Filter (DKF) to handle these diffuse situations. The DKF does not require any data transformation.

The thesis contributes further to the theoretical and computational aspects of conducting statistical inferences using the DKF. First, we demonstrate the appropriate initialization of the DKF for the important class of time-invariant SSM's. This result is useful for maximum likelihood statistical inference with the SSM. Second, we derive and compare alternative pseudo-likelihoods for the diffuse SSM. We uncover some interesting characteristics of the DKF and the *diffuse* likelihood with the class of ARMA models. Third, we propose an efficient implementation of the DKF, labelled the *collapsed* DKF

(CDKF). The latter is derived upon sweeping out some columns of the pertinent matrices in the DKF after an initial number of iterations. The CDKF coincides with the KF in the absence of regression effects in the SSM. We demonstrate that in general the CDKF is superior in practicality and performance to alternative algorithms proposed in the literature. Fourth, we consider maximum likelihood estimation in the SSM using an EM (Expectation-Maximization) approach. Through a judicious choice of the *complete data*, we develop an CDKF-EM algorithm which does not require the evaluation of lag one state error covariance matrices for the most common estimation exercise required for the SSM, namely the estimation of the covariance matrices of the disturbances in the SSM. Last we explore the topic of diagnostic testing in the SSM. We discuss and illustrate the recursive generation of residuals and the usefulness of the latters in pinpointing likely outliers and points of structural change.

Table of Contents

List of Tables

List of Figures

æ

Summary of Notation

v	column vector v
$v^{\#}$	number of components in vector v
$\mathbf{1}$	column vector with all entries equal to 1
M	matrix M
$\mathbf{0}$	matrix with all entries equal to 0
I_p	$p \times p$ identity matrix
$M(i;j)$	(i,j) entry of matrix M
M'	transpose of M
M^*	conjugate transpose of M
M^-	Moore-Penrose generalized inverse of M
M^+	equals $J'(JMJ')^{-1}J$ with $rank(JMJ') = rank(M)$
	if M is symmetric and singular
(M, N)	matrix with column blocks M and N
$(M; N)$	matrix with row blocks M and N
$\|M\|$	determinant of M
$tr(M)$	trace of M
$vec(M)$	stack of columns of M
$M \otimes N$	Kronecker product of M and N
$M^{1/2}$	a square root of positive semidefinite matrix M
$diag(M, N)$	a diagonal matrix with diagonal matrix blocks M and N

$E(x)$	unconditional mean of random variable x
$Cov(x)$	unconditional covariance matrix of random variable x
$x \sim (\mu, V)$	random variable x with mean μ and covariance matrix V
$\hat{x}$	*filtered* estimate of random variable x
$\tilde{x}$	*smoothed* estimate of random variable x
$y\|x$	random variable y conditional on x
$Pred(y\|x)$	random variable $a + bx$ where a and b minimimize diagonal elements of $Cov(y - a - bx)$
$Mse(y\|x)$	equals $Cov\{y - Pred(y\|x)\}$
$\lambda(y)$	$-2\times$ log-likelihood of data y apart from constants
$\lambda^d(y)$	$-2\times$ *diffuse* log-likelihood of data y apart from constants
$\lambda^m(y)$	$-2\times$ *marginal* log-likelihood of data y apart from constants
$\lambda^{\sigma^2}(y)$	$-2\times$ σ^2-concentrated log-likelihood of data y apart from constants
i	positive (complex) square root of -1
t	(integer) time index
n	number of observations in dataset

List of Abbreviations

SSM	State Space Model (as defined in thesis)
ASSM	Augmented State Space Model (as employed in literature)
KF	Kalman Filter
DKF	Diffuse Kalman Filter
IF	Information Filter
AR	autoregressive
MA	moving-average
ARMA	autogressive moving-average
ARIMA	integrated autogressive moving-average
mmse	minimum mean square error
mle	maximum likelihood estimator
log	natural logarithm
gls	generalized least-squares
iff	if and only if

Acknowledgement

I wish to express my deep gratitude to my supervisor, Professor Piet de Jong for his boundless inspiration, guidance, patience, support and advice throughout the course of this work.

Professor Martin L. Puterman provided much encouragement and advice throughout my graduate career. I thank him for making my initial stay at UBC possible.

Professor Jian Liu made valuable comments on this work.

I gratefully acknowledge the financial support of the Faculty of Commerce and Business Administration both for the Dean E.D. MacPhee Fellowship and for the numerous lectureship appointments. Partial support was also provided by the National Sciences and Engineering Research Council under grant 5-88-5-77.

I commend Colleen Colclough and Ruby Visser for their gracious and expert handling of my dealings with the University administration.

I wish to recognise Miss Celine Anime and Mr Chung Chai Tsang, both former teachers of mine, who have been inspiring role models during my younger days in Mauritius.

Thanks, Paul, Melisa, Yuppy and Junior Ho for the happy memories amid the nicest living environment.

To Pa (late), Ma, Vivi and Fifi,

This is yours too.

Chapter 1

Introduction

This thesis deals with the statistical and computational aspects of *prediction*, *model fitting* and *diagnostic testing* in the *State Space Model* (SSM), a model which has become increasingly prominent in the time series literature during the last two decades.

The SSM originates from the systems science and owes much of its theoretical basis to the seminal contributions of Kalman (1960) and Kalman and Bucy (1961). Duncan and Horn (1972) introduced the SSM to the statistical community from the standpoint of a *random parameter regression model* and connected its theory with the fixed parameter regression theory. Harrison and Stevens (1976) refer to the SSM as the *Dynamic Linear Model* in their work on Bayesian forecasting.

The SSM describes an observation process in terms of an underlying unobserved time series known as the *state*. An example of the model is the time series model where the observation (at time t) is specified as the sum of fixed regression effects and the state (at time t) with the latter having components which are interpreted as the unobserved trend and seasonalities. A simple stochastic model for the state may for example stipulate that its trend component follows a random walk model while its seasonal components sum to a white noise process over the span of a year. Observe that this time series model consists of both *fixed* and dynamic *random* effects. The SSM that will be defined in the next Chapter is a generalization of this example.

The SSM has carved a niche in engineering and (more recently) in statistical and socio-economic applications. The flagship application is perhaps the NASA space program which uses the SSM to monitor the progress of its spacecraft. In the field of econometrics, the SSM has been employed in the estimation of unobserved wage rates (Watson and Engle, 1983), the estimation of historical unobserved trend and cycle components of the British industrial production index (Crafts *et al.* , 1989) and in the seasonal adjustment of census data (Burridge and Wallis, 1984). Business applications include inventory control (Downing *et al.* , 1980), short-term forecasting (Mehra, 1979) and statistical quality control (Phadke, 1981). In the area of policy, Harvey and Durbin (1986) report an interesting study of the impact of seat belt legislation on road casualties. Recently in a series of contributions, Harvey (1984, 1989) expounds on the merits, in particular the ease of interpretation, of a class of SSM's called the *structural models* over another class of SSM's, namely the ARMA time series models which have been popularized by Box and Jenkins (1970). Other uses of the SSM in the statistical arena deal with cross-validation (De Jong, 1988b) and spline smoothing (Kohn and Ansley, 1987a). An interesting application of the SSM deals with the prediction of outcomes of National Football League (NFL) games (Sallas and Harville, 1981, 1989).

The SSM is formally defined in Chapter 2. The definition is nonstandard in that fixed and random effects are treated separately. The conceptual and computational advantages derived from this definition will be displayed throughout the thesis. We demonstrate that many practical statistical models are in fact special instances of the SSM. Consequently they can all be treated in a unified fashion upon casting them as SSM's. The major part of the Chapter is devoted to a summary of the technology associated with the SSM. In this respect we cover the statistical and computational aspects of recursive *filtering* (i. e. the Kalman Filter), *smoothing* and *likelihood* evaluation. The concepts discussed therein are central to the contributions presented later in the thesis.

The KF needs to be properly initialized to allow its use for maximum-likelihood-based inference in the SSM. Chapter 3 addresses this issue for an important class of SSM's. The

Chapter first defines the concepts of *time invariance* and *stationarity* in the SSM context. Thereafter, assuming that a time invariant SSM has applied since *time-immemorial* , we derive closed-form expressions for the unconditional mean and covariance matrix of the states. The results hold for both stationary and nonstationary SSM's and are useful for initializing both the KF and the Diffuse Kalman Filter (De Jong, 1991b) when they are applied to these time invariant models.

In Chapter 4, we turn to the problems of the initialization of the KF and the definition of an appropriate likelihood for the general SSM, in particular the *diffuse* SSM. The latter arises when there is uncertainty about the initial state or the regression parameter in the SSM. These diffuse situations can be handled in a unified fashion by including a diffuse random vector (i. e. a random vector with an arbitrarily large covariance matrix) in the SSM. However the effect of this diffuse random vector needs to be factored out prior to any statistical inference. This leads us to the study of the *diffuse* and *marginal* likelihoods which are both suitable pseudo-likelihoods for the diffuse SSM. We establish the exact relationship between these two pseudo-likelihoods. De Jong (1991b) has developed an extension of the KF, called the Diffuse Kalman Filter (DKF), to handle recursive filtering, smoothing, likelihood evaluation and gls estimation of regression effects in the diffuse SSM. Using basic arguments, we demonstrate why the DKF is a natural extension of the KF. The Chapter concludes with the presentation of two interesting characteristics of the DKF with the class of nonstationary ARMA models. These models are often used in socio-economic applications. We demonstrate that when the DKF is applied to nonstationary autoregressive models, it reduces *de facto* to the KF after an initial number of iterations. This "collapse" of the DKF sets the motivation for the work of the following Chapter where we implement a collapsed form of the DKF which, for arbitrary SSM's, is generally not equivalent to the KF. With nonstationary mixed ARMA (p,q) processes, we demonstrate that it is critical from a computational standpoint to restrict ourselves to the *invertible* parametrization of these models.

The implementation aspects of the DKF are discussed in Chapter 5. The starting point is the redefinition of the diffuse SSM in such a way that the diffuse parameter vector γ is partitioned as $\gamma = (\gamma_1; \gamma_2)$ with γ_1 and γ_2 being solely associated with the initial state and the unknown regression parameter. A proper estimate of γ_1 is obtainable (save for collinearity problems) after an initial number of DKF iterations and this in turn can be used to provide *limiting* estimates of the subsequent states. From a computational standpoint, this suggests collapsing the DKF, specifically factoring out those columns of various matrices in the DKF which are associated with γ_1. This collapsed DKF, labelled the *collapsed* DKF (CDKF), coincides with the KF in the absence of a regression effect in the SSM. The smoothing algorithm associated with the DKF can also be collapsed in an analogous fashion as the CDKF. We provide the details of the intricate adjustments required by this smoothing algorithm when it has to be switched back to the smoothing algorithm associated with the DKF in the pre-collapse time period. In the final section, we collate the CDKF and its associated smoothing algorithm with alternative algorithms discussed in the literature. We conclude that the use of the CDKF (and its associated smoothing) algorithm can lead to appreciable computational savings since it employs recursions of state error covariance matrices of lower dimensionalities than its competitors.

Maximum likelihood estimation of parameters in the SSM is covered in Chapter 6. It is well-known that maximum likelihood estimators possess such desirable properties as asymptotic consistency, efficiency, unbiasedness and normality. The estimation method, labelled the *CDKF-EM* method, embeds the CDKF within the EM algorithm, a popular derivative-free likelihood optimization algorithm. We generalize and unify previous works in the literature. We also propose a novel CDKF-EM algorithm specifically designed for the estimation of the error covariance matrices in the SSM. This new algorithm is simpler and computationally more efficient than the general algorithm discussed in the first part of the Chapter since it does not require the evaluation of lag one state error covariance matrices. We illustrate the CDKF-EM algorithm using examples borrowed from the literature. Of interest is the fact that in some cases, the new CDKF-EM algorithm generates solutions with higher log-likelihoods than previously reported.

In Chapter 7, we explore the topic of diagnostic testing in the SSM. The KF generates a sequence of uncorrelated residuals known as the *innovations.* The latters have proved useful in tests of goodness-of-fit (see Harvey (1989), p256-260) but they often fail to distinguish between *outliers* and *structural breaks* in the SSM. In that regard, it is worthwhile to study alternative residuals. The SSM defined in this thesis employs a single disturbance vector u_t with specific components of the latter applying to the observation and state equations (at time t). The Chapter focusses on the study of v_t which is defined as the predictor of u_t conditional on the whole observation set. We demonstrate that the v_t's are more useful than the innovations in the detection of *outliers* and *structural breaks* in the SSM. The v_t's are serially correlated and we therefore consider the idea of orthogonalizing them (in a backward direction). We conclude that these backward orthogonalized versions of v_t merely corresponds to the innovations. This tells us that no advantage is derived from using orthogonalized versions of v_t's in lieu of the innovations in statistical tests of goodness-of-fit in the SSM.

1.1 Preliminaries

For clarity and completeness, we now define the notations employed in this thesis. Matrices are denoted by capital roman or caligraphic characters (e. g. M, $\mathcal{M}$) and vectors by ordinary characters (e. g. v). A matrix with all entries equal to zero is written as $\mathbf{0}$, the identity matrix is denoted by I and a vector with all entries equal to one is denoted by

1. For (appropriate) matrix M, the determinant, the Moore-Penrose generalized inverse, the transpose, the conjugate transpose and a Choleski root are respectively denoted by $|M|$, M^-, M', M^* and $M^{1/2}$. For (appropriate) matrices M and N, $M \otimes N$, (M, N) and $(M; N)$ respectively stand for the Kronecker product of M and N, the matrix with column blocks M and N and the matrix with row blocks M and N.

Time series observations are denoted by y_t, $t = 1, \ldots, n$ with t the time index and n the number of observations in the dataset. We will often use the shorthand notation y to denote the stack of observations $(y_1; \ldots; y_n)$. For a statistical model with parameter vector θ and under the assumption of normally distributed disturbances, -2 times the log-likelihood of y, apart from constant terms which do not depend on θ, is denoted by $\lambda(y|\theta)$ or more compactly by $\lambda(y)$ when the role of θ is unambiguous.

For random variables x and y, $x \sim (\mu, V)$ is shorthand for saying that x has mean $E(x) = \mu$ and covariance matrix $Cov(x) = V$ whereas $Pred(x|y)$ denotes the inhomogeneous linear combination of the components of y which minimize the diagonal elements of $Cov\{x - Pred(x|y)\} \equiv Mse(x|y)$. In this thesis, we often consider the prediction of a random vector x_t conditional on observation vectors $(y_1; \ldots; y_{t-1})$ and $(y_1; \ldots; y_n)$. We use the shorthand notation, $Mse(\hat{x}_t)$ and $Mse(\tilde{x}_t)$ to respectively denote $Mse(x_t|y_1; \ldots; y_{t-1})$ and $Mse(x_t|y_1; \ldots; y_n)$.

Chapter 2

The State Space Model

This Chapter reviews the state of the art in state space technology and in the process it introduces the fundamental concepts behind the contributions presented later in the thesis. The development of the theory associated with these fundamental concepts can be found in the lucid textbooks of Anderson and Moore (1979) and Harvey (1981, 1989).

The programme of this Chapter is as follows. The SSM is formally defined in Section 1. This definition of the SSM differs from that commonly employed in the SSM literature. We spell out the analytic and computational advantages arising from such a definition of the SSM. The next section describes some attractive characteristics of the SSM. In section 3, we demonstrate that familiar statistical models in the linear models literature are in fact special cases of the SSM. Section 4 deals with the prediction aspects associated with the SSM, namely *filtering* and *smoothing*. These operations are conducted via a pair of recursive algorithms known as the *Kalman Filter* (KF) and the *Smoothing* algorithm. Direct implementation of these algorithms can be numerically unsafe, specifically with regards to maintaining the positive semi-definiteness of covariance matrices. An attractive solution is to employ the *square root* forms of the KF and the smoothing algorithm. These propagate covariance matrices in terms of their Choleski square roots. We describe a computationally efficient version of these square root algorithms. We also briefly discuss a variant of the KF which is known as the *Information Filter*.

2.1 Defining the SSM

The SSM stipulates the generation of an observation process, $y = (y_1; y_2; \ldots; y_n)$. In this thesis, the SSM is defined according to the following pair of equations :

$$y_t = X_t\beta + Z_t\alpha_t + G_t u_t \ , \ t = 1, 2, \ldots, n \tag{2.1}$$

$$\alpha_{t+1} = W_t\beta + T_t\alpha_t + H_t u_t \ , \ t = 0, 1, \ldots, n \tag{2.2}$$

The first equation is called the *observation* (or *measurement*) equation while the second equation is known as the *state* (or *transition* or *system*) equation. The latter equation specifies the dynamics of the unobserved *random* vector α_t which is known as the *state* (or *system vector*) at time t. The term β represents a *regression* parameter, while the u_t's are serially uncorrelated disturbances with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I$. The SSM is anchored with $\alpha_0 = 0$ thereby implying an *initial state*, $\alpha_1 = W_0\beta + H_0 u_0$. The *system matrices* Z_t, G_t, T_t and H_t as well as the *regression matrices* X_t, W_t are all assumed known.

Our definition of the SSM differs from the one commonly employed in the literature. We label the latter definition of the SSM, the *augmented* SSM (ASSM) on account of the fact that the state is augmented to accomodate the regression parameter β. The specification of the ASSM which is equivalent to equations (2.1)-(2.2) is,

$$y_t = (Z_t \ X_t)\begin{pmatrix} \alpha_t \\ \beta \end{pmatrix} + G_t u_t \ , \qquad t = 1, 2, \ldots, n \tag{2.3}$$

$$\begin{pmatrix} \alpha_{t+1} \\ \beta \end{pmatrix} = \begin{pmatrix} T_t & W_t \\ \mathbf{0} & I \end{pmatrix}\begin{pmatrix} \alpha_t \\ \beta \end{pmatrix} + \begin{pmatrix} H_t \\ \mathbf{0} \end{pmatrix} u_t, \qquad t = 0, 1, \ldots, n \tag{2.4}$$

The above specification of the ASSM points out that the standard approach in the SSM literature has been to deal with regression effects implicitly rather than explicitly.

We now spell out the advantages associated with using our definition of the SSM as opposed to the ASSM :

1. **Statistical.** Regression effects ($X_t\beta$ and $W_t\beta$) are explicitly introduced in the model. These effects form an integral part of any statistical model and should be handled explicitly and not ignored or removed in an *ad hoc* fashion from the observations. Furthermore this feature is useful (i) conceptually, to separate fixed regression effects from the purely random effects induced by the states, (ii) theoretically, to introduce diffuse parameters in the SSM and (iii) empirically, to describe, for instance, *outliers* and *model shifts* in the SSM.

2. **Computational.** It will be shortly demonstrated that the performance of filtering and smoothing algorithms depends on the size of the state. It is therefore beneficial to keep the dimension of the state in the SSM to a minimum. Furthermore we will argue later in the Chapter that smoothing algorithms based on the ASSM are inefficient since the smoothed estimate of β corresponds to its final estimate in the filtering cycle and is therefore not effectively updated during the smoothing cycle.

3. **Analytic.** Using the same β and u_t in the observation and state equation is not restrictive since through appropriate choices of X_t, W_t, G_t and H_t, different components of β and u_t can be brought into either equation. This parsimony of notation contrasts with the situation in the ASSM which in general employs distinct regression parameters and disturbance vectors in the measurement and state equations.

The technical material introduced in this Chapter assumes that the regression parameter β is either fixed and known or random with known covariance matrix. For the latter case (a rare occurrence in practice), it is necessary to employ the ASSM specification wherein β is included in the state. The more common cases of fixed but unknown β and random but diffuse β require special treatment and will be covered in Chapter 4.

2.2 Characteristics of the SSM

We now list some general characteristics of the SSM. These establish the usefulness of the model especially when viewed in the context of its specializations which are described in the next section.

1. **Dimensionality.** The dimensions of the system matrices are arbitrary at each time point except for conformability constraints. In particular, the SSM covers both univariate and vector observations in a unified framework.

2. **State.** Often the state has a physical meaning. For example, the progress of a spacecraft can be monitored using a SSM with a state whose components describe the velocity, acceleration, coordinates and rate of fuel consumption of the spacecraft. In the structural model introduced in the next section, the components of the state describe economic constructs such as trend and seasonalities. However in many situations, the state can only be given an abstract interpretation.

   Observe that the current state embodies all the information up to the present. Therefore "knowledge" about the state implies the redundance of storing past observations. The latter aspect is a major feature of the filtering and smoothing algorithms described later in the Chapter.

3. **Dynamics.** In many natural or scientific phenomena, the evolution mechanism of the state is known to vary with time. The SSM provides a simple and elegant framework for capturing such knowledge via time-varying system matrices and regression matrices.

4. **Markovian Nature.** The fact that the state equation is Markovian is not restrictive. When the dynamics of the state involve multiple lags, then a suitable augmentation of the state keeps the Markovian feature intact. As an illustration, suppose $\alpha_{t+1} = W_t\beta + T_t\alpha_t + S_t\alpha_{t-1} + H_t u_t$. In this case, an appropriate state equation is,

$$\begin{pmatrix} \alpha_{t+1} \\ \alpha_t \end{pmatrix} = \begin{pmatrix} W_t \\ \mathbf{0} \end{pmatrix} \beta + \begin{pmatrix} T_t & S_t \\ I & \mathbf{0} \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} H_t \\ \mathbf{0} \end{pmatrix} u_t$$

   Kalman (1960) exploited the Markovian nature of the state equation to design the famous recursive filter named after him, namely the Kalman Filter.

5. **Non-Uniqueness.** The SSM specification for a particular process is not unique. For example if U is any orthogonal matrix, then the SSM defined in (2.1)-(2.2) is equivalent to another SSM where Z_t is postmultiplied by U and W_{t-1}, T_{t-1} and H_{t-1} are premultiplied by U'. Furthermore it is possible for "equivalent" SSM's to have states and hence system matrices of different dimensions (equations (2.1)-(2.2) and (2.3)-(2.4) for example). Obviously from a computational standpoint, SSM's with system matrices of minimum dimensions are preferred.

6. **Missing Data.** Missing observations do not require any special handling. Stoffer (1981) has demonstrated a "zeroing-out" strategy whereby missing components of y_t as well as the corresponding rows of X_t, Z_t and G_t are replaced by zeroes and the "revised" data are then processed in the usual manner except for a minor adjustment in the Kalman Filter.

7. **Linearity.** Repeated backsubstitutions of the state equation in the observation equation lead to $y = X\beta + Gu$ where y and u are respectively the stacks of the observations y_t and disturbances u_t and X and G are built up from X_t, Z_t, W_t, T_t

and H_t. This implies that the SSM is a linear model if X and G do not depend on the data or the error terms. As such, the well-known linear model theory applies to the SSM.

8. **Data Irregularities.** The transition matrices X_t and W_t are useful for intervention purposes in the presence of *outliers* and *structural changes* in the SSM.

2.3 Specializations of the SSM

It is now shown that the SSM straddles a wide class of popular multivariate linear models, in particular regression models with fixed, time-varying or random coefficients, time series models and unobserved components models.

Regression Model

Consider the SSM with scalar observations y_t (generalization for vector observations is immediate) of the form,

$$y_t = X_t\beta + \alpha_t, \qquad \alpha_{t+1} = a\ \alpha_t + u_t$$

Through appropriate specifications of X_t and a, we immediately recognise the following familiar statistical models :

- White Noise Model : $y_t = u_t$, (put $X_t = 0$, $a = 0$).
- Fixed Parameter Regression Model : $y_t = X_t\beta + u_t$, (put $a = 0$).
- Autoregressive Model of order 1 : $(y_t - \beta) = a\ (y_{t-1} - \beta) + u_t$, (put $X_t = 1$).
- Regression Model with autoregressive disturbances of order 1 : $y_t = X_t\beta + \alpha_t$; $\alpha_{t+1} = a\ \alpha_t + u_t$, (interpret the state as the autoregressive disturbance term).

- Random Walk with drift : $y_t = \beta + y_{t-1} + u_t$, (put $X_{t+1} - X_t = 1$, $a = 1$).

The random coefficients regression model (Nicholls and Pagan, 1985) can be described by the SSM,

$$y_t \;=\; Z_t\alpha_t + G_t u_t, \qquad \alpha_{t+1} = W_t\beta + T_t\alpha_t + H_t u_t$$

Special cases of this model include (i) the regression model with random-walk parameters (put $W_t = 0$ and $T_t = I$), (ii) the *Return to Normality Model* discussed in Harvey (1981, p202) where the regression parameter α_t evolves according to $(\alpha_{t+1} - \beta) = \phi(\alpha_t - \beta) + u_t$ (put $W_t = 1 - \phi$, $T_t = \phi$ and $H_t = 1$) and (iii) the time-varying (but non-random) coefficient regression model (put $H_t = 0$).

ARMA Model

There are several advantages in casting ARMA models as SSM's. First, scalar and vector ARMA processes are dealt with in a unified framework. Second, their log-likelihood is evaluated in an exact and efficient fashion via the Kalman Filter. This contrasts with the Box-Jenkins methodology which relies on *ad hoc* procedures such as *back-forecasting.* Third, note that through the zeroing-out strategy, missing observations do not require any special handling.

Consider the vector ARMA (p,q) model,

$$y_t \;=\; A_1 y_{t-1} + \ldots + A_p y_{t-p} + \epsilon_t + B_1\epsilon_{t-1} + \ldots + B_q\epsilon_{t-q}$$

which following Gardner *et al.* (1980) can be specified as a SSM of the form,

$$y_t = (I, 0, \ldots, 0)\,\alpha_t\,, \qquad t = 1, 2, \ldots \tag{2.5}$$

$$\alpha_{t+1} = \begin{pmatrix} A_1 & I & 0 & \ldots & 0 \\ A_2 & 0 & I & \ldots & 0 \\ . & . & . & . & . \\ A_{m-1} & 0 & 0 & \ldots & I \\ A_m & 0 & 0 & \ldots & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} I \\ B_1 \\ B_2 \\ . \\ B_{m-1} \end{pmatrix} \epsilon_{t+1}\,,\ t = 0, 1, 2, \ldots \tag{2.6}$$

where $m = max(p, q+1)$, $A_i = 0,\ i > p$ and $B_j = 0, j > q$.

In this specification, the first row block of the state vector is the observation itself. To see this, denote block component j of the state by $\alpha_{t,j}$. Then by repeated backsubstitutions,

$$\begin{aligned} \alpha_{t,1} &= A_1\alpha_{t-1,1} + \alpha_{t-1,2} + \epsilon_t \\ &= A_1\alpha_{t-1,1} + (A_2\alpha_{t-2,1} + \alpha_{t-2,3} + B_1\epsilon_{t-1}) + \epsilon_t \\ &\vdots \\ &= A_1\alpha_{t-1,1} + \ldots + A_m\alpha_{t-m,1} + \epsilon_t + B_1\epsilon_{t-1} + \ldots + B_{m-1}\epsilon_{t+1-m} \end{aligned}$$

As stated earlier, there exist alternative specifications for the SSM. For instance, Akaike (1975) defines a SSM where $\alpha_{t,j}$ $(1 \leq j \leq m)$ is the (j-1)-step ahead predictor of the process.

An immediate extension of the ARMA model is the regression model with ARMA error structure, $y_t = X_t\beta + v_t$ where the disturbance term v_t follows an ARMA process. This model can be written as the SSM above (equations 2.5-2.6) with the regression effect $X_t\beta$ included in the observation equation and the state now interpreted as v_t. Another extension is the *mixed linear model* where the random effects evolve according to an

ARMA process. This is cast as the SSM defined in (2.1)-(2.2) with β and α_t respectively representing the fixed and random effects and the transition equation is as defined for the ARMA process above. This particular SSM has been employed by Sallas and Harville (1981) for the prediction of football scores and dairy cattle breeding values.

Unobserved Components Model

The familiar time series model comprising of trend, seasonal and irregular components is a member of the class of *unobserved components models.* These models arise in many applications. For instance, Watson and Engle (1983) use the SSM to estimate "unobserved" wage rates whereas Downing *et al.* (1980) estimate the shrinkage or loss of materials in an inventory control system. In general, the data is used to suggest the form of the unobserved components model but it is often possible to impose plausible models for the form of the components. In such cases, they are known as *structural models.*

An interesting application of the structural model is described in Harvey and Durbin (1986) who assess the impact of seat belt legislation on British road casualties. One of the structural models used by the authors is of the form,

$$\begin{aligned}
y_t &= X_t\beta + \mu_t + \varsigma_t + \epsilon_t, & \\
\mu_{t+1} &= W_t\beta + \mu_t + \nu_t + \eta_t, & \text{(Level)} \\
\nu_{t+1} &= \nu_t + \xi_t, & \text{(Local Slope)} \\
\varsigma_t &= \mathcal{Z}_t\gamma_t, & \text{(Seasonality)} \\
\gamma_{t+1} &= \Gamma\gamma_t + \omega_t &
\end{aligned}$$

where μ_t is a local linear trend with its level and slope (ν_t) determined by random walks, ς_t describes aspects of the vector of the seasonal components γ_t in effect at time t (the evolution of γ_t being specified by transition matrix Γ and disturbances ω_t), ϵ_t is the irregular component (i. e. measurement error) and X_t and W_t are matrices of explanatory variables such as car traffic index, real price of gasoline and indicator variables reflecting the permanent effects (via X_t) or the transient effects (via W_t) of the seat belt legislation. This structural model can be written as the following SSM,

$$y_t = X_t\beta + (1\ ,\ 0\ , \mathcal{Z}_t)\alpha_t,$$

$$\begin{pmatrix} \mu_{t+1} \\ \nu_{t+1} \\ \gamma_{t+1} \end{pmatrix} = \alpha_{t+1} = \begin{pmatrix} W_t \\ 0 \\ 0 \end{pmatrix}\beta + \begin{pmatrix} \begin{matrix} 1 & 1 \\ 0 & 1 \end{matrix} & \mathbf{0} \\ \mathbf{0} & \Gamma \end{pmatrix}\alpha_t + \begin{pmatrix} \mathcal{H}_1 \\ \mathcal{H}_2 \\ \mathcal{H}_3 \end{pmatrix} u_t$$

The above SSM with $\beta = \mathbf{0}$ and γ_t describing quarterly variations will be often used in this thesis and it will labelled the *Quarterly Basic Structural Model* (QBSM).

SSM's have also been employed in the study of non-linear and non-normal dynamic phenomena, both possibly occurring in a continuous time setting ; see for example Kitagawa (1987,1989) and Pena and Guttman (1988), with the latters employing Bayesian concepts to propose robust recursive algorithms for these situations. A class of nonlinear models of interest to time series practitioners is the *state dependent model* (Priestley, 1988) where the current state depends on its previous realizations and/or past observations. Special cases include the *bilinear* models (Granger and Andersen (1978) and Subba Rao, 1981), the *threshold autoregressive* models (Tong and Lim, 1980), the *exponential autoregressive* model (Haggan and Ozaki, 1981) and the *autoregressive conditional heteroscedasticity* model (ARCH) introduced by Engle (1982) and thereafter generalized by Bollerslev (1986). Finally we also mention the seminal work of Harvey and Fernandes (1989) who extend the SSM technology to deal with count or qualitative time series data. Non-linear models are often of limited practical utility and furthermore their statistical analyses are complicated and lack the elegance of their linear counterparts. This thesis deals exclusively with the class of linear SSM's.

2.4 The Statistics of the SSM

Statistical issues concerning the SSM include the reconstruction of the states $\{\alpha_t\}$ and the evaluation of the likelihood function. Estimation of the states can be achieved in two ways namely *filtering* and *smoothing* whereby the states are respectively estimated conditional on $(y_1; y_2; \ldots; y_{t-1})$ and $(y_1; y_2; \ldots; y_n)$.

2.4.1 Filtering

The famous Kalman Filter (Kalman, 1960) provides a recursive algorithm for the filtering process with the recursiveness ocurring as a result of the Markovian nature of the transition equation in the SSM. Filtering can be viewed as the process of updating a predictor in light of new information. In the Bayesian context, this is equivalent to computing a posterior distribution given a prior distribution and the data. It is therefore appropriate that the KF be also viewed as a Bayesian procedure; see Harrison and Stevens (1976) and Meinhold and Singpurwalla (1983). Jazwinski (1970), Anderson and Moore (1979) and Harvey (1981, 1989) derive the KF using the classical ideas of the *Prediction* or *Projection* Theorem. We now state without proof the following prediction results (where β is assumed known), the equations of which make up the KF.

Theorem 2.1 (Kalman Filter) *Suppose $y_1, \ldots, y_n$ are generated by the SSM. Then $\hat{\alpha}_t$, the predictor of the state α_t conditional on $(y_1, \ldots, y_{t-1})$, $t \leq n+1$, and its associated Mse matrix $\sigma^2 P_t$ are evaluated according to the following recursions,*

$$e_t = y_t - X_t\beta - Z_t\hat{\alpha}_t, \qquad D_t = Z_t P_t Z_t' + G_t G_t', \qquad K_t = (T_t P_t Z_t' + H_t G_t')D_t^{-1},$$

$$\hat{\alpha}_{t+1} = W_t\beta + T_t\hat{\alpha}_t + K_t e_t \qquad \textit{and} \qquad P_{t+1} = T_t P_t T_t' + H_t H_t' - K_t D_t K_t'$$

with $\hat{\alpha}_1 = W_0\beta$ *and* $P_1 = H_0 H_0'$.

The quantities employed in the KF have physical interpretations : e_t is called the *innovation* or the one-step ahead prediction error resulting from the prediction of y_t conditional on $(y_1; \ldots; y_{t-1})$ and its covariance matrix is $\sigma^2 D_t$; K_t is the *Kalman gain* matrix and it is used in updating the estimate of the state in light of observation y_t (or equivalently the innovation e_t).

Some merits of the Kalman Filter are the following :

1. It is suited for *on-line* or *real-time* applications.

2. It is efficient in terms of storage : past data need not be stored ; they manifest themselves in the estimate of the current state vector.

3. It produces *minimum mean square* linear estimators of the states.

4. It provides a recursive scheme for evaluating the likelihood of the SSM. (see section 2.4.3). This contrasts for example with Box-Jenkins evaluation of ARMA processes which employs the technique of iterated back-forecasting.

5. It produces a sequence of residuals namely the innovations, e_t. The latters are generalizations of the *recursive residuals* discussed by Brown *et al.* (1975) in their study of the stability of regression parameters in the fixed regression model. The innovations represent aspects of the observation y_t that cannot be predicted from previous observations and consequently they are serially uncorrelated. The latter property makes the innovations useful in statistical tests of goodness-of-fit in the SSM. This topic is discussed in more detail in Chapter 7.

A major practical problem with the KF concerns the specification of $\hat{\alpha}_1$ and P_1 i. e. $W_0\beta$ and H_0. This will be resolved in the next Chapter for the class of SSM's with a time invariant state equation. More general cases will be dealt with in Chapter 4. Observe that the computation of P_t is the most time-consuming exercise in the KF. This emphasises the importance of keeping the dimension of the state to a minimum. In Chapter 5, we will demonstrate that the CDKF outperforms its competitors on account of the fact that it recurs P_t of lower dimensionality.

We stated previously that the "zeroing-out" strategy permits one to process missing observations in an automatic way except for a minor adjustment in the KF. We now discuss the details of this adjustment. The zeroing-out strategy implies that D_t are singular and consequently the KF fails. In this situation, it is practical to employ a generalized inverse $D_t^+ = J'(JD_tJ')^{-1}J$ where J is a "selector" matrix (e. g. a permutation of the identity matrix) such that $rank(JD_t) = rank(D_t)$. As De Jong (1991a, section 3) remarks and illustrates with an example, although the choice of J is immaterial for prediction purposes, it can however affect likelihood evaluation and maximum likelihood estimation. To avoid these irregular situations, De Jong (1991a) proposes the consideration of the *regular* SSM.

Definition 2.1 Let y denote the stack of observations y_t generated by the SSM. Then y is said to generated by a *regular SSM* if $y = Fz$ where $Cov(z)$ is nonsingular and has the same rank as $Cov(y)$ and F is functionally independent of any unknown parameter.

The "zeroing-out" strategy coupled with the use of D_t^+ as defined above is implicitly tantamount to dealing with a regular SSM with z interpreted as the stack of the non-missing elements in y. More generally, in the consideration of the regular SSM, $y^{\#}$, $|D_t|$ and the log-likelihood $\lambda(y)$ are to be interpreted respectively as $z^{\#}$, $|JD_tJ'|$ and $\lambda(z)$.

Henceforth in this thesis, we imply the regular SSM whenever we refer to the SSM.

In Chapter 4, we discuss the DKF, a filtering algorithm specifically designed by De Jong (1991b) to handle diffuse parameters in the SSM. It turns out that the DKF is the KF with the vectors e_t and $\hat{\alpha}_t$ turned into matrices E_t and A_t with the same row dimensions. The DKF can be viewed as a generalization of the KF since it uses the KF to update each column of A_t. The DKF therefore enjoys most of the attractive features of the KF.

2.4.2 Forecasting

Forecasting is easily carried out with the KF. For instance, using the concept of *iterated* predictions, the k-step ahead predictor of the state is,

$$\begin{aligned} Pred(\alpha_{t+k}|y_1,\ldots,y_t) &= Pred\{Pred(\alpha_{t+k}|y_1,\ldots,y_{t+k-1})|y_1,\ldots,y_t\} \\ &= W_{t+k-1}\beta + T_{t+k-1}Pred(\alpha_{t+k-1}|y_1,\ldots,y_t) \\ &\vdots \\ &= \{W_{t+k-1} + T_{t+k-1}W_{t+k-2} + \ldots + (T_{t+k-1}\ldots T_{t+2})W_{t+1}\}\beta \\ &+ (T_{t+k-1}\ldots T_{t+1})\hat{\alpha}_{t+1} \end{aligned}$$

The mse matrix of the k-step ahead predictor is evaluated from the prediction error, $\alpha_{t+k} - Pred(\alpha_{t+k}|y_1,\ldots,y_t)$. It follows that,

$$\begin{aligned} \sigma^{-2}Mse(\alpha_{t+k}|y_1,\ldots,y_t) &= H_{t+k-1}H'_{t+k-1} + T_{t+k-1}H_{t+k-2}H'_{t+k-2}T'_{t+k-1} + \ldots \\ &+ (T_{t+k-1}\ldots T_{t+2})H_{t+1}H'_{t+1}(T_{t+k-1}\ldots T_{t+2})' \\ &+ (T_{t+k-1}\ldots T_{t+1})P_{t+1}(T_{t+k-1}\ldots T_{t+1})' \end{aligned}$$

Therefore the k-step ahead observation predictor and its mse matrix are,

$$\begin{aligned} Pred(y_{t+k}|y_1,\ldots,y_t) &= X_{t+k}\beta + Z_{t+k}\ Pred(\alpha_{t+k}|y_1,\ldots,y_t) \qquad \text{and} \\ Mse(y_{t+k}|y_1,\ldots,y_t) &= Z_{t+k}\ Mse(\alpha_{t+k}|y_1,\ldots,y_t)\ Z'_{t+k} + \sigma^2 G_{t+k}G'_{t+k}. \end{aligned}$$

2.4.3 Likelihood Evaluation

Assuming that the disturbances u_t are normally distributed, Schweppe (1965) and Harvey (1981), the latter employing the *Prediction Error Decomposition*, have shown that -2 times the log-likelihood of the SSM, apart from a constant, is

$$\lambda(y|\theta,\sigma^2) = y^{\#}\log\sigma^2 + \sum_{t=1}^{n}\log|D_t| + \sigma^{-2}\sum_{t=1}^{n} e_t' D_t^{-1} e_t \qquad (2.7)$$

Here θ denote the parameters in the SSM (i. e. the system and regression matrices) and $Cov(u_t) = \sigma^2 I$. Interestingly, the likelihood function is expressed as a function of only the innovations e_t and their covariance matrices $\sigma^2 D_t$. This implies that the roles of the regression effects $X_t\beta$ and $W_t\beta$ are implicitly buried within the innovations.

Assuming that σ^2 is known and furthermore noting that e_t and D_t are produced by the KF, it therefore ensues that $\lambda(y|\theta,\sigma^2)$ can be evaluated in a recursive fashion by attaching to the KF the extra recursion $q_{t+1} = q_t + e_t' D_t^{-1} e_t$ with $q_1 = 0$. If σ^2 is unknown then it can be concentrated out (i. e. replaced by its mle) of the above log-likelihood function. The mle of σ^2 is $\hat{\sigma}^2 = q_{n+1}/y^{\#}$ and upon its substitution in (2.7), we obtain the σ^2-concentrated log-likelihood, $\lambda(y|\theta) = y^{\#}\log q_{n+1} + \sum_{t=1}^{n}\log|D_t|$.

In Chapter 4, we derive two pseudo-likelihoods, namely the marginal and diffuse likelihoods. We demonstrate that they are both equal to $\lambda(y|\theta,\sigma^2)$ plus some additional terms. It will also be shown that diffuse log-likelihood can be evaluated in a recursive fashion via the DKF.

2.4.4 Smoothing

The KF constructs an estimate of the state at time t using only the information available at time $t-1$. In many situations however, it is desirable to estimate the state using the entire dataset. For instance, the common exercise of least squares estimation of the parameters of a statistical model is tantamount to a smoothing operation. Smoothing can be a complicated procedure requiring in many cases the inversion of covariance matrices of the order of the data.

The SSM however provides a framework for recursive smoothing. This feature is again due to the Markovian characteristic of the state equation. Smoothing in the SSM context translates to updating the filtered estimate of the state, $\hat{\alpha}_t$, using the observation vector $(y_t; y_{t+1}; \ldots; y_n)$ or equivalently the innovation vector $(e_t; e_{t+1}; \ldots; e_n)$. Therefore smoothing algorithms are designed to run backward using output produced by a forward run of the KF. Smoothing algorithms fall into three categories namely (i) *fixed-interval*, which assume that n, the number of observations, is fixed and t, the time index, varies (ii) *fixed-point*, where t is fixed and n increases and (iii) *fixed-lag*, where both t and n vary. These are all detailed in Jazwinski (1970) and Anderson and Moore (1979, Chapter 7). De Jong (1988b,1989) makes significant contributions towards enhancing the computational efficiency of these smoothing algorithms.

The fixed-interval smoothing algorithm, by far the most commonly used, suffices for the purpose of this thesis. We now state without proof De Jong's (1988b,1989) results concerning fixed-interval smoothing.

Theorem 2.2 (De Jong, 1988b) *Suppose the KF is run and for $t = 1, \ldots, n$, the quantities $Z_t'D_t^{-1}e_t$, $Z_t'D_t^{-1}Z_t$, $\hat{\alpha}_t$, P_t and $L_t = T_t - K_tZ_t$ are stored. The smoothing algorithm proceeds as follows : initialize η_n and R_n respectively as a vector and a matrix of zeroes and for $t = n, \ldots, 1$ run the recursions,*

$$\eta_{t-1} = Z_t'D_t^{-1}e_t + L_t'\eta_t \qquad \text{and} \qquad R_{t-1} = Z_t'D_t^{-1}Z_t + L_t'R_tL_t.$$

Then the smoothed estimator of the state α_t and its associated mse matrix are,

$$\begin{aligned} \tilde{\alpha}_t &= Pred(\alpha_t|y_1, \ldots, y_n) = \hat{\alpha}_t + P_t\eta_{t-1} \qquad \text{and} \\ Mse(\tilde{\alpha}_t) &= Mse(\alpha_t|y_1, \ldots, y_n) = \sigma^2(P_t - P_tR_{t-1}P_t). \end{aligned}$$

Furthermore, for $1 \leq t \leq r \leq n+1$, the mse matrix between smoothed estimators of the states is,

$$Mse(\tilde{\alpha}_t, \tilde{\alpha}_r) \;=\; \sigma^2 P_t L'_{r-1,t}(I - R_{r-1}P_r)$$

where $L'_{r-1,t} = \prod_{i=t}^{r-1} L'_i$ with $L_{t-1,t} = I$.

Kohn and Ansley (1989) independently derive a scalar version of this smoothing algorithm. Furthermore, they compare the efficiency of the algorithm to an alternative one discussed in Anderson and Moore (1979, p187) and report for instance savings of the order of 50% in the number of multiplication and division steps for a SSM with a state consisting of 15 components. This is not surprising since the above smoothing algorithm avoids further matrix inversions following the forward KF pass.

An interesting method to compute $Mse(\tilde{\alpha}_{t-1}, \tilde{\alpha}_t)$ is provided by Watson and Engle (1983). They augment the state α_t by α_{t-1}, redefine the transition equation appropriately and thereafter run the KF and smoothing algorithm. This method automatically yields $\sigma^{-2}Mse(\tilde{\alpha}_{t-1}, \tilde{\alpha}_t)$ as the off-diagonal matrix block of the mse matrix of the smoothed estimate of the augmented state. This approach is however computationally inefficient due to the dimensionality of the augmented state.

Two computational drawbacks of the ASSM are vividly portrayed in the smoothing exercise. First, the smoothed estimate of β corresponds to its final estimate obtained in the KF and hence it is effectively not updated during the smoothing cycle. Second, the augmented state in the ASSM implies that a smoothing algorithm will require more data storage than a smoothing algorithm based on the SSM described by (2.1)-(2.2). These properties make smoothing algorithms based on the ASSM patently inefficient. A numerical illustration attesting for this fact is provided in Chapter 5.

To conclude this subsection, we mention that the smoothing algorithm described in

Theorem 2.2 is easily extended to handle diffuse parameters. This extended smoothing algorithm will be discussed in detail in Chapter 4.

2.4.5 Information Filter

When there is uncertainty regarding either the initial state or the regression parameter in the SSM, it is common practice to initialize the KF with a large P_1. This method, which is commonly known as the "big k" method, is inexact and can furthermore be numerically unstable ; see Chapter 4 for an illustration. The numerical problems can sometimes be avoided by using the *Information Filter* (IF) which differs from the KF in that it recurs the inverse of P_t. The IF is easily derived from the KF using the well-known Matrix Inversion Lemma ; see Anderson and Moore (1979, p139).

The IF has some serious shortcomings vis-a-vis the KF. For example, it does not follow that generalized inverses of the covariance matrices can be readily employed if these matrices are singular. For instance, Ansley and Kohn (1985b) remark that the IF breaks down for a large class of ARMA models. Furthermore, the IF is numerically inefficient, requiring the inversion of large covariance matrices when observations y_t are of a multivariate nature. Applications of the IF can be found in Kitagawa (1981) and Sallas and Harville (1988) which both deal with *nonstationary* time series models.

2.4.6 Computational Aspects

Direct implementation of the KF may lead to asymmetric or even negative definite covariance matrices due to rounding errors. These problems can be circumvented through the propagation of Choleski square roots of these covariance matrices. A survey on square root filtering is provided in Anderson and Moore (1979, p147-162). The following efficient *square root* form of the KF is due to De Jong (1991a).

Theorem 2.3 (Square-Root Kalman Filter) *The Kalman Filter described in Theorem* 2.1 *can be generated as follows : let U be such that $UU' = I$ and*

$$\begin{pmatrix} Z_t P_t^{1/2} & G_t \\ T_t P_t^{1/2} & H_t \end{pmatrix} U = \begin{pmatrix} D & \mathbf{0} \\ K & P \end{pmatrix} \tag{2.8}$$

with D of minimal column dimensions and with the same row dimension as Z_t. Then $D = D_t^{1/2}$, $P = P_{t+1}^{1/2}$ and $K = K_t D$ and $\hat{\alpha}_{t+1} = W_t\beta + T_t\hat{\alpha}_t + KD^{-}(y_t - X_t\beta - Z_t\hat{\alpha}_t)$.

The orthogonal matrix U may be obtained by various means, for example Givens rotation, Householder transformation or the QR algorithm as it is commonly known. All the computations reported in this thesis are based on square-root algorithms with the QR algorithm as their core component. The computer codes were written in the APL language and run on an AT-type microcomputer. The codes for the QR algorithm were taken from Helzer (1983).

The square-root form of the smoothing algorithm follows the same concept as the square root KF. In particular, it only suffices to propagate R_t in square-root form. Thus for $t = n+1, \ldots, 1$, we find an orthogonal matrix U such that $|(D_t^{-1/2} Z_t)'\ ,\ (R_t^{1/2} L_t)'|U = R_{t-1}^{1/2}$.

2.5 Summary

We have demonstrated the versatility and usefulness of the SSM. The importance of treating fixed effects separately from random effects has been emphasised. Statistical inferences in the SSM can be achieved within the unified framework of the KF and the associated smoothing algorithm. A major problem with the KF concerns its initialization. This problem is dealt with in the next Chapter for the special case of SSM's with a time invariant state equation. The issues of how to handle diffuse initial states and regression parameters in the SSM are covered in Chapter 4. In particular, it will be shown that these diffuse situations can be handled by a generalization of the KF technology reported in this Chapter.

Chapter 3

Time invariance and Stationarity in the State Space Model

Most practical time series models have time invariant parameters. ARMA models and structural models are typical examples. In their study of ARMA models, Box and Jenkins (1970) consider the concept of *stationarity* and use it to develop such analytic features as the *unconditional* mean and variance of the ARMA process. The work in this Chapter extends these features to a more general class of SSM's, namely those with a *time invariant* state equation. The results are useful for initializing both the KF and the DKF when they are applied to such SSM's.

This Chapter is divided as follows. The concepts and definitions of time invariance and stationarity within the SSM context which were originally devised by De Jong (1991c) are reported in section 1. This section describes and investigates consequences of these definitions. We derive the necessary and sufficient conditions for stationarity in the SSM. These generalize the well-known conditions for stationarity in ARMA models. Section 2 deals with the evaluation of the unconditional mean and covariance matrix of the states for the time invariant SSM assuming the latter has applied since *time immemorial.* Four applications are reported. The Chapter concludes with an Appendix containing all the technical proofs.

3.1 Preliminaries

Definition 3.1 The SSM is said to be *time invariant* if for $t = 1, \ldots, n$, $W_t = W$, $T_t = T$ and $H_t = H$ in the state equation (2.2).

Remarks

1. The above definition does not say anything about W_0, T_0 and H_0 ; in particular they may differ from the subsequent values of these same matrices.

2. Time invariance does not impose any conditions on the observation equation.

3. Time invariance implies that T is square. This follows since if T_t is a $p \times q$ matrix with $p \neq q$, then T_{t+1} must have p columns thereby implying that T is not time invariant.

Definition 3.2 The SSM is said to be *stationary* if it is time invariant and both $E(\alpha_t)$ and $Cov(\alpha_t)$ are invariant to t.

This definition of stationarity parallels the definition of *second-order* or *weak* stationarity used in the ARMA model literature. Weak stationarity assumes that the autocovariance function is a function of the lag between the arguments. In the case of stationary SSM's, this assumption translates to the following result.

Lemma 3.1 *For a stationary SSM,* $Cov(\alpha_t, \alpha_{t-s}) = T^s Cov(\alpha_t) \quad , \quad 0 \leq s \leq t.$

Proof. The result is obvious for $s = 0$. For $s > 0$, observe that

$$\alpha_t \;=\; (I + T + T^2 + \ldots + T^{s-1})W\beta + T^s\alpha_{t-s} + (Hu_{t-1} + THu_{t-2} + \ldots + T^{s-1}Hu_{t-s})$$

and hence $Cov(\alpha_t, \alpha_{t-s}) = T^s Cov(\alpha_{t-s}) = T^s Cov(\alpha_t)$. •

The following Theorem states the necessary and sufficient conditions for stationarity in the SSM.

Theorem 3.1 *The SSM is stationary iff W_0 and H_0 are such that $W_0\beta = W\beta + TW_0\beta$ and $H_0H_0' = TH_0H_0T' + HH'$.*

Proof. Suppose W_0 and H_0 are as specified. Then $E(\alpha_2) = W\beta + TW_0\beta = W_0\beta$ and $\sigma^{-2}Cov(\alpha_2) = TH_0H_0T' + HH' = H_0H_0'$. Clearly for $t > 2$, $E(\alpha_t) = W_0\beta$ and $Cov(\alpha_t) = \sigma^2 H_0H_0'$ which are both invariant to t. Conversely if W_0 and H_0 do not satisfy the given relations then $E(\alpha_2) \neq W_0\beta = E(\alpha_1)$ and $Cov(\alpha_2) \neq \sigma^2 H_0H_0' = Cov(\alpha_1)$ and hence the SSM is not stationary. •

The equations $W_0\beta = W\beta + TW_0\beta$ and $H_0H_0' = TH_0H_0T' + HH'$ need not have proper solutions for W_0 or H_0. This is the case, for example when T, H and $W\beta$ are all equal to one. Furthermore note that these equations do not appear to bear any connection to the Box-Jenkins approach where stationarity conditions are couched in terms of roots of polynomials. However a consequence of the next result, which gives necessary and sufficient conditions for the existence of proper solutions and hence for stationarity, is that these two approaches do in fact coincide for the ARMA model. The result is stated in terms of the eigenvalues of T. An eigenvalue of T is called *stationary* if it has modulus less than one ; otherwise it is called *nonstationary.*

Theorem 3.2 *The equation $W_0\beta = W\beta + TW_0\beta$ has a solution for W_0 iff $(W\beta)'x = \mathbf{0}$ whenever $T'x = x$. The equation $H_0H_0' = TH_0H_0T' + HH'$ has a solution for H_0 iff $H'x = \mathbf{0}$ whenever x is an eigenvector of T corresponding to a nonstationary eigenvalue.*

The proof is given in the Appendix at the end of the Chapter. A sufficient condition for both conditions in the Theorem to hold is that T only have stationary eigenvalues. Then, as suggested by Gardner *et al.* (1980), H_0 may be solved as $vec(H_0H_0') = \{I -$

$(T \otimes T)\}^{-1} vec(HH')$. Note that $I - (T \otimes T)$ is nonsingular since none of the eigenvalues of T equals ± 1.

Application to ARMA(p,q) model. The SSM specification of an ARMA (p,q) model is given in equations (2.5)-(2.6). Recall that the determinant of a matrix does not change if a multiple of one row is added to another row. Consider $T - zI$ and add zI times the first block of rows to the second block of rows leading to a second block of rows of $(A_2 + zA_1 - z^2 I, \mathbf{0}, I, \mathbf{0}, \ldots, \mathbf{0})$. Repeat this procedure by multiplying the resulting second block of rows by zI and adding to the third block of rows and so on for subsequent row blocks. The resulting matrix has a determinant equal to

$$(-1)^{p+1} det\ (A_p + zA_{p-1} + \ldots + z^{p-1}A_1 - z^p I)$$

$$\text{or equivalently} \qquad (-1)^{p+1} det\ \{(-z^p)\ (I - z^{-1}A_1 - \ldots - z^{-p}A_p)\}$$

The Box-Jenkins approach states that the model is stationary iff the roots (z^{-1}) of the polynomial $(I - z^{-1}A_1 - \ldots - z^{-p}A_p)$ all lie outside the unit circle. This is equivalent to z lying inside the unit circle or in other words T has all stationary eigenvalues.

3.2 Automatic Initialization of the Kalman Filter

The KF recursion for the SSM must be initialized with the unconditional mean and covariance matrix of the initial state. In this section, explicit expressions for these quantities are developed for the time invariant SSM. The expressions hold for both the stationary and nonstationary cases. In related work, Ansley and Kohn (1985a) show how to initialize the KF for the special case of ARIMA models.

The previous section dealt with the assignment of W_0 and H_0 in order to induce stationarity in the SSM. A conceptual tool to get around the explicit assignments of W_0 and H_0 is to suppose that the SSM has applied since *time immemorial.*

Definition 3.3 The SSM is said to have applied since time immemorial if it is time invariant and the state equation $\alpha_{t+1} = W\beta + T\alpha_t + Hu_t$ is assumed to hold for $t = r, \ldots, -1, 0, 1, \ldots, n$ where $r \longrightarrow -\infty$ and $\alpha_r = 0$.

The concept of a model having applied since time immemorial is exploited in the Box-Jenkins methodology to evaluate the unconditional mean and variance of an ARMA process. This time immemorial assumption can also be used to find the unconditional mean and covariance matrix of the states in a time invariant SSM. We first consider two simple classes of time invariant SSM's.

Theorem 3.3 *Suppose the SSM has applied since time immemorial. If T has all stationary eigenvalues then for $t = \ldots, -1, 0, 1, \ldots, n+1$,*

$$E(\alpha_t) = (I - T)^{-1} W\beta \qquad \textit{and} \qquad Cov(\alpha_t) = \sigma^2 M$$

where M is such that $M = TMT' + HH'$. If T has all nonstationary eigenvalues then $\{Cov(\alpha_t)\}^- = \mathbf{0}$.

The proof of this Theorem is also given in the Appendix. An immediate consequence of the Theorem is that for stationary SSM's which are assumed to have applied since time immemorial, the KF may be initialized with $E(\alpha_1) = (I - T)^{-1} W\beta$ and $P_1 = M$. These generalize the expressions derived by Gardner *et al.* (1980) who assume $\beta = \mathbf{0}$.

In general, T can have both stationary and nonstationary eigenvalues. In this case, the time immemorial argument leads to an arbitrarily large covariance matrix for the states. This has led to the suggestion that the KF should then be initialized either explicitly or implicitly with a state covariance matrix of the form kI or more generally $kC + D$ where k is large. For example, both Burridge and Wallis (1985) and Burmeister, Wall and Hamilton (1986) propose taking $Cov(\alpha_1) = kI$ with k large. In their theoretical works on filtering, smoothing and signal extraction, Ansley and Kohn (1985b,1987) and Kohn and Ansley (1986) use $Cov(\alpha_1) = kC + D$ with $k \to \infty$. De Jong (1988b,1991b) also employs this specification. There appears to be no literature explicitly justifying the choice of $E(\alpha_1)$ or the use of $Cov(\alpha_1) = kC+D$ with k large or how to actually determine the matrices C and D. The next Theorem provides expressions for these quantities under the time immemorial assumption.

Theorem 3.4 *Suppose the SSM is assumed to have applied since time immemorial. Then for* $t = \ldots, -1, 0, 1, \ldots, n+1$,

$$\sigma^2\{Cov(\alpha_t)\}^- \quad = \quad S'M^-S = \lim_{k\to\infty}(kU_1U_1' + U_2MU_2')^-$$

where M *is such that* $M = QMQ' + SHH'S'$. *Furthermore if for given* x, $x'U_1 = 0$ *then*

$$x'E(\alpha_t) \quad = \quad x'U_2(I-Q)^{-1}SW\beta$$

The matrices Q, S, U_1 *and* U_2 *are as follows :* U *is any matrix such that* $UTU^{-1} = diag(P,Q)$ *where* P *has all nonstationary eigenvalues and* Q *has all stationary eigenvalues ;* $U = (R;S)$ *with* S *having the same row dimension as* Q *and* $U^{-1} = (U_1, U_2)$ *where* U_2 *has the same column dimension as the row dimension of* S.

Proof. Consider $U\alpha_{t+1} = UW\beta + UTU^{-1}(U\alpha_t) + UHu_t$. This can be written as the following system of equations,

$$\begin{aligned} n_{t+1} &= RW\beta + Pn_t + RHu_t, \\ m_{t+1} &= SW\beta + Qm_t + SHu_t \end{aligned}$$

with $n_t = R\alpha_t$ and $m_t = S\alpha_t$.

Since P has all nonstationary eigenvalues, it follows that $\{Cov(n_t)\}^- = \mathbf{0}$. The matrix Q has all stationary eigenvalues thereby implying $E(m_t) = (I - Q)^{-1}SW\beta$ and $Cov(m_t) = \sigma^2 M$ where M is such that $M = QMQ' + SHH'S'$. These results are now used to show

$$\begin{aligned}
\sigma^2\{Cov(\alpha_t)\}^- &= (U^{-1}\{Cov(U\alpha_t)\}U^{-1'})^- = U'\{Cov(U\alpha_t)\}^- U = S'M^-S \\
&= \lim_{k\to\infty} U'diag(k^{-1}I, M^-)U = \lim_{k\to\infty}\{U^{-1}diag(kI, M)U^{-1'}\}^- \\
&= \lim_{k\to\infty}(kU_1U_1' + U_2MU_2')^-
\end{aligned}$$

Finally, if $x'U_1 = \mathbf{0}$, then $x'E(\alpha_t) = x'U^{-1}E(U\alpha_t) = x'\{U_1E(n_t) + U_2E(m_t)\} = x'U_2(I - Q)^{-1}SW\beta$. This concludes the proof of the Theorem. •

Noteworthy features of Theorem 3.4 are :

1. It encompasses Theorem 3.3. If all the roots of T are stationary, then $U = S = I$ and U_1 is null and the Theorem implies the well known results (Gardner *at al.*, 1980), $Cov(\alpha_t) = \sigma^2 M$ where $M = TMT' + HH'$ and $E(\alpha_t) = (I - T)^{-1}W\beta$. If all the roots of T are nonstationary then U_2 is null, $U_1 = I$ and thus $\{Cov(\alpha_0)\}^- = \lim_{k\to\infty} k^{-1}I = \mathbf{0}$.

2. A general construction for U is as follows. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the roots of T with algebraic multiplicities $n_1, n_2, \ldots, n_p$ and where conjugate pairs of roots are included just once and $|\lambda_1| \geq |\lambda_2| \geq \ldots |\lambda_p|$. For i=1,2, ...,p, define matrices N_i as follows. If λ_i is real then N_i has n_i columns spanning the null space of $(T - \lambda_i I)^{n_i}$. If λ_i is complex then N_i has $2n_i$ columns spanning the null space of $(T^2 - 2r_iT + |\lambda_i|^2 I)^{n_i}$ where r_i is the real part of λ_i. Put $U^{-1} = (N_1, N_2, \ldots, N_p)$. The decomposition $UTU^{-1} = diag(P, Q)$ is called a Real Jordan Canonical form of T (see Brown (1988), p 141-150).

3. With U as constructed in 2., the matrix Q is block diagonal with blocks of size $n_i \times n_i$ or $(2n_i) \times (2n_i)$ depending on whether the corresponding root is real or complex. Accordingly, the inversions of $I - Q \otimes Q$ and $I - Q$ are reduced to inverting relatively small diagonal blocks.

4. If $x = S'\gamma$ for some γ, then $x'E(\alpha_t) \rightarrow \gamma'(I-Q)^{-1}SW\beta$. This result is obtained upon noting that $SU_2 = I$ and $SU_1 = \mathbf{0}$.

5. The Theorem suggests that for nonstationary SSM's, α_0 may be specified as $\alpha_0 = U_1\gamma + \alpha_0^\dagger$ where γ and $\alpha_0^\dagger$ are uncorrelated and $\sigma^{-2}\ Cov(\gamma) = kI$ with k large and $\sigma^{-2}\ Cov(\alpha_0^\dagger) = U_2MU_2'$. This specification in turn implies the initialization of the KF with $P_1 = kTU_1U_1'T' + TU_2MU_2'T' + HH'$, where $k \rightarrow \infty$. This however is not a satisfactory solution for it induces numerical instability in the KF. The latter situation arises since the update of P_t in the KF (Theorem 2.1) may possibly involve the difference of two large quantities to yield a required small quantity. These numerical difficulties will be circumvented via the use of an extended KF in the next Chapter.

6. The Information Filter sometimes serves as a useful alternative to the KF, in particular when the latter is to be initiated with a large state error covariance matrix (P_1), as for instance in the case of nonstationary SSM's. For these cases, the Theorem suggests that a suitable initialization for the IF is $P_1^- = S'M^-S$. This contrasts with the common practice of employing $P_1^- = \mathbf{0}$.

We now illustrate the results of Theorem 3.4 in four applications. These provide expressions for the unconditional mean and covariance matrix of the states in the AR (2) model and three other empirical models borrowed from the literature. As stated in point 5. above, the derived expressions are useful for initializing the DKF when it is applied to these models. The results for the second and third applications were obtained with the use of software written in APL.

Application to the AR(2) model. The AR(2) model $y_t = \beta + ay_{t-1} + by_{t-2} + \epsilon_t$ has a SSM representation,

$$y_t \;=\; (1,0)\;\alpha_t, \qquad \alpha_{t+1} = \begin{pmatrix} \beta \\ 0 \end{pmatrix} + \begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \epsilon_{t+1}$$

The roots of T are $\delta = (a + \sqrt{d})/2$ and $\gamma = (a - \sqrt{d})/2$ where $d = a^2 + 4b$. If both roots are stationary then $M = \sigma^{-2}Cov(\alpha_t)$ satisfies $M = TMT' + diag(1,0)$ and $E(\alpha_t) = \beta(1-a-b)^{-1}(1;b)$. When both roots are nonstationary, $\{Cov(\alpha_t)\}^{-} \to \mathbf{0}$ and both components of $E(\alpha_t)$ diverge. Now suppose δ is stationary and γ is nonstationary and hence the roots are real. Define the matrix U^{-1} with columns $U_1 = (1;-\delta)$ and $U_2 = (1;-\gamma)$. Put $Q = \delta$. Then

$$\lim_{k\to\infty} \frac{1}{k} \left\{ E(\alpha_0) + \frac{\delta\beta}{(1-\delta)(\gamma-\delta)} \begin{pmatrix} 1 \\ -\gamma \end{pmatrix} \right\} \;\to\; \beta \begin{pmatrix} 1 \\ -\delta \end{pmatrix} \qquad \text{and}$$

$$\lim_{k\to\infty} \frac{1}{k} \left\{ \sigma^{-2}Cov(\alpha_0) - \frac{\delta}{(1-\delta^2)(\gamma-\delta)^2} \begin{pmatrix} 1 & -\gamma \\ -\gamma & \gamma^2 \end{pmatrix} \right\} \;\to\; \begin{pmatrix} 1 & -\delta \\ -\delta & \delta^2 \end{pmatrix}$$

Application to the seasonal adjustment of data. Burridge and Wallis (1985) deal with the seasonal adjustment of U.S. employment data. Part of one of the models relates to a cyclical component specified by

$$W\beta = \mathbf{0}, \qquad T = \begin{pmatrix} 2.26 & 1 & 0 & 0 \\ -1.52 & 0 & 1 & 0 \\ 0.26 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad H = \begin{pmatrix} 1 \\ -.989 \\ .00686 \\ .00001 \end{pmatrix}, \qquad \sigma^2 = 14409$$

The matrix T has a nonstationary root 1 repeated twice and stationary roots of 0 and .26. Assuming the model has applied since time immemorial then $\sigma^{-2}Cov(\alpha_0)$ is of the form

$$k\begin{pmatrix} .99 & -.02 & -.06 & 0 \\ & .94 & -.24 & 0 \\ & & .07 & 0 \\ & & & 0 \end{pmatrix} + \begin{pmatrix} .12 & .24 & .12 & -.33\times 10^{-5} \\ & .48 & -.24 & .67\times 10^{-5} \\ & & .12 & -.33\times 10^{-5} \\ & & & 1\times 10^{-10} \end{pmatrix}$$

where $k \to \infty$. Burridge and Wallis (1985) employed $Cov(\alpha_0) = 10^{12}I$.

Application to the unobserved components model. Burmeister, Wall and Hamilton (1986) apply the Kalman filter to estimate unobserved monthly inflation rate in economic time series. One of their SSM's has parameters

$$W\beta = \mathbf{0}, \qquad T = \begin{pmatrix} \gamma' & 0 \\ I & \mathbf{0} \end{pmatrix}, \qquad H' = (1,0,0,0), \qquad \sigma^2 = 1.9537\times 10^{-3}$$

where $\gamma' = (.14135, .89635, -.3817, .11173)$. The roots of T are $0, .1905 \pm .2905i, .8497$ and -1.0894. Under the time immemorial assumption $\sigma^{-2}Cov(\alpha_0)$ is of the form

$$k\begin{pmatrix} .27 & -.25 & .23 & -.21 & .19 \\ & .23 & -.21 & .19 & -.18 \\ & & .19 & -.18 & .16 \\ & & & .16 & -.15 \\ & & & & .14 \end{pmatrix} + \begin{pmatrix} 1.40 & 1.42 & 1.00 & 1.00 & .70 \\ & 1.53 & 1.30 & 1.11 & .90 \\ & & 1.64 & 1.20 & 1.19 \\ & & & 1.73 & 1.12 \\ & & & & 1.80 \end{pmatrix}$$

where $k \to \infty$. Burmeister, Wall and Hamilton (1986) used $Cov(\alpha_0) = 20I$.

Application to panel survey data. In many applications, it is possible to directly partition the state into a nonstationary component and a stationary component. For instance, Pfeffermann (1991) employs the following state-space model for the estimation and seasonal adjustment of population means based on rotating panel surveys carried out on a quarterly basis :

$$\begin{aligned}
y_t &= (1\ 0\ 1\ 0\ 0\ 1/4\ 1/4\ 1/4\ 1/4\ 0\ 0)\alpha_t\ , \qquad t = 1,2,\ldots \\
\alpha_{t+1} &= diag(T1, T2)\alpha_t + \epsilon_t\ , \qquad t = 0,1,2,\ldots \text{ where} \\
T1 &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \qquad T2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho^3 & 0 \\ 0 & 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \\
Var(\epsilon_t) &= diag(V_1, V_2), \qquad V_1 = diag(\sigma_T^2, \sigma_L^2, \sigma_S^2, 0, 0) \qquad \text{and} \\
V_2 &= \sigma_v^2\ diag\{(1-\rho^2)^{-1}, 1, (\rho^4 + \rho^2 + 1), 1, 0, 0\}
\end{aligned}$$

Here $T1$ describes the *quarterly basic structural model* reported in Chapter 2 while $T2$ describes the rotation patterns of the units of the panel data under the assumption that observations from the same unit follow an AR(1) model with autoregressive coefficient ρ ($|\rho| < 1$). Observe that T1 has eigenvalues 1,1,i,$-i$ and 0 while T2 has all its eigenvalues equal to 0.

To initialize the KF, Pfeffermann (1991) following Harvey and Peters (1990) uses a diagonal covariance matrix of the form $diag(\mathcal{P}_1, \mathcal{P}_2)$ where $\mathcal{P}_1$ is diagonal with a zero diagonal element plus four arbitrarily large diagonal elements and $\mathcal{P}_2 = \sigma_v^2(1-\rho^2)^{-1}I_6$. These specifications are in line with the results of Theorem 3.4. In particular, observe that the transition matrix is already in the required "nonstationary-stationary" diagonal form thereby implying that $U = I$. Thus $\mathcal{P}_1 \equiv diag(kI_4, 0)$, $k \to \infty$ and $\mathcal{P}_2$ satisfies $\mathcal{P}_2 = T_2\mathcal{P}_2T_2' + V_2$.

3.3 Summary

We have derived expressions for the unconditional mean and covariance matrix of the states in time invariant SSM's under the assumption that the latters have applied since time immemorial. These expressions are useful for initializating the KF. However if the covariance matrix of the states is arbitrarily large as in the case of nonstationary SSM's, then the KF will fail numerically. The next Chapter demonstrates an extension of the KF to deal with this problem. The results of this Chapter will prove useful for initializing this extended KF for the class of time invariant nonstationary SSM's.

3.4 Appendix

The following Lemma and its consequences will be useful for the various proofs. Proof and details can be found in Brown (1988).

Lemma 3.2 *Every square matrix T has a Jordan Decomposition such that $T' = U(D + K)U^{-1}$, where D is a diagonal matrix with the eigenvalues of T as its diagonal entries and K is a matrix with zeroes everywhere except on its superdiagonal where there may be one or more ones. The matrices D and K are related as follows : if $K(j, j+1) = 1$ then $D(j,j) = D(j+1, j+1)$. The matrices D and U may be complex with the columns of U being the generalized eigenvectors of T'.*

Furthermore D and K have the following useful properties, (i) $DK = KD$, (ii) $K^m = \mathbf{0}$ where m is the order of T and (iii) $(D+K)^t = D^t \sum_{j=0}^{m-1} \begin{pmatrix} t \\ j \end{pmatrix} (D^{-}K)^j$.

Proof of Theorem 3.2 The equation $W_0\beta = W\beta + TW_0\beta$ is equivalent to

$$\{\beta' \otimes (I - T)\}vec(W_0) = W\beta$$

Clearly the equation is consistent if $\beta = \mathbf{0}$. Suppose $\beta \neq \mathbf{0}$. Then consistency results iff for $x \neq \mathbf{0}$, $x'\{\beta' \otimes (I-T)\} = 0$ implies $x'W\beta = 0$. Since $\beta \neq \mathbf{0}$, $x'\{\beta' \otimes (I-T)\} = 0$ implies $x'T = x'$. Thus for given W, T and β, there is a W_0 such that $W_0\beta = W\beta + TW_0\beta$ iff $T'x = x$ implies $(W\beta)'x = 0$.

Next consider $w^*H_0H_0'w = w^*TH_0H_0'T'w + w^*HH'w$ for w arbitrary. Then for $t \geq 0$,

$$w^*H_0H_0'w \quad = \quad w^*T^{t+1}H_0H_0'T^{t+1'}w + w^*T^tHH'T^{t'}w + \ldots + w^*HH'w$$

Using Lemma 3.2, write $w^*T^tHH'T^{t'}w$ as $w^*U^{-1*}(D+K)^{*t}(U^*HH'U)(D+K)^tU^{-1}w$. A typical entry of $(D+K)^tU^{-1}w$ is,

$$\lambda^t\{r_0 + tr_1/\lambda + \ldots + \begin{pmatrix} t \\ k \end{pmatrix} r_k/\lambda^k\}$$

where λ is an eigenvalue of T and $(r_0, \ldots, r_k)$ are consecutive elements in $U^{-1}w$, with $r_k \neq 0$ if $k > 0$. Thus for large t, $w^*T^tHH'T^{t'}w$ is dominated by a term of the form

$$x^*HH'x|\lambda|^{2(t-k)} \begin{pmatrix} t \\ k \end{pmatrix}^2 |r_k|^2$$

where x is a generalized eigenvector of T' associated with eigenvalue λ. If λ is nonstationary and $x^*HH'x \neq 0$ then the dominating term diverges and hence the equation cannot hold for finite H_0. Thus the stated condition is necessary for stationarity.

To show sufficiency, suppose the condition holds. Premultiply and postmultiply both sides of $H_0H_0' = TH_0H_0'T' + HH'$ by U^* and U to yield

$$C \quad = \quad (D+K)^*C(D+K) + U^*HH'U$$

where $C = U^*H_0H_0'U$. Without any loss in generality assume $D+K$ and $U^*HH'U$ are arranged so that $D+K$ has two seperate blocks each containing all the nonstationary and all the stationary eigenvalues of T'. A solution for C is obtained by setting the diagonal block of C corresponding to the "nonstationary" block in $D+K$ to zero which is possible since the condition states that the "nonstationary" block in $U^*HH'U$ is a zero matrix. The vec of the "stationary" diagonal block of C (denoted by $\mathcal{C}$) can be solved as $vec(\mathcal{C}) = \{I-(L\otimes L^*)\}^{-1}vec(R)$ where L and R are respectively the "stationary" blocks of $(D+K)$ and $U^*HH'U$. Thus $U^{-1*}diag(0,\mathcal{C})U^{-1}$ is a solution to $H_0H_0' = TH_0H_0'T'+HH'$.

•

Proof of Theorem 3.3. Consider the state equation $\alpha_{t+1} = W_t\beta + T_t\alpha_t + H_t u_t$ where for $t > 0$, $W_t = W$, $T_t = T$ and $H_t = H$. Backsubstitution shows

$$\begin{aligned} \alpha_t &= (I + T + \ldots + T^{t-2})W\beta + T^{t-1}W_0\beta \\ &+ Hu_{t-1} + THu_{t-2} + \ldots + T^{t-2}Hu_1 + T^{t-1}Hu_0 \end{aligned}$$

and in turn this implies,

$$\begin{aligned} E(\alpha_t) &= (I + T + \ldots + T^{t-2})W\beta + T^{t-1}W_0\beta \qquad \text{and} \\ \sigma^{-2}Cov(\alpha_t) &= HH' + THH'T' + \ldots + T^{t-2}HH'(T^{t-2})' + T^{t-1}H_0H_0'T^{t-1'} \end{aligned}$$

Clearly the behaviour of T^t dictates the convergence of $E(\alpha_t)$ and $Cov(\alpha_t)$. Observe that

$$T^{t'} = U(D+K)^t U^{-1} = U\{D^t \sum_{j=0}^{m-1} \begin{pmatrix} t \\ j \end{pmatrix} (D^- K)^j\} U^{-1}$$

Suppose T has all stationary eigenvalues. Then $D^t \rightarrow \mathbf{0}$ implying $T^t = \mathbf{0}$. Accordingly, using a geometric sum argument, $(I + T + T^2 + \ldots + T^{t-1}) \rightarrow (I - T)^{-1}$. Hence as $t \rightarrow \infty$, $E(\alpha_t) = (I + T + T^2 + \ldots + T^{t-1})W\beta + T^{t-1}W_0\beta \rightarrow (I - T)^{-1}W\beta$. The limiting covariance matrix of the state vector is such that

$$vec\{Cov(\alpha_t)\} = \sigma^2(I + (T \otimes T) + \ldots + (T \otimes T)^{t-1})vec(HH')$$

As $t \rightarrow \infty$, $vec\{Cov(\alpha_t)\} \rightarrow \sigma^2\{I - (T \otimes T)\}^{-1}vec(HH')$ which is the solution $\sigma^2 vec(M)$ where M is such that $M = TMT' + HH'$. Gardner, Harvey and Philips (1980) obtain a similar solution. Note that $(I - T)$ and $\{I - (T \otimes T)\}$ are both nonsingular since T does not have any root with unit modulus.

Now suppose T has all nonstationary roots. Clearly $(I + T + T^2 + \ldots + T^{t-1})$ diverges and hence $x'E(\alpha_t)$ diverges for every $x \neq \mathbf{0}$. Finally it is shown that for any w, $w^*T^tHH'T'^t w$ converges to zero or diverges to $+\infty$. Thus every root of $Cov(\alpha_t)$

converges to zero or diverges to $+\infty$ which implies that $\{Cov(\alpha_t)\}^- \rightarrow \mathbf{0}$. Recall from the proof of Theorem (3.2), $w^*T^tHH'T'^tw$ is dominated by a term of the form $x^*HH'x|\lambda|^{2(t-k)}\begin{pmatrix} t \\ k \end{pmatrix}^2 |r_k|^2$ where x is an eigenvector of T' associated with the nonstationary eigenvalue λ. As $t \rightarrow \infty$ the dominating term diverges unless $x^*HH'x = 0$ in which case $U^*HH'U = \mathbf{0}$ which in turn implies $H = \mathbf{0}$ and therefore $w^*T^tHH'T'^tw = 0$.

•

Chapter 4

The Diffuse State Space Model

This Chapter is concerned with the problem of diffuseness in the SSM. In this thesis, a *diffuse* random variable is viewed as one with an arbitrarily large covariance matrix. Diffuseness arises in three situations within the SSM context : (i) when the SSM is nonstationary and is assumed to have applied since time-immemorial, a situation we encountered in the last Chapter, (ii) when there is uncertainty about the initial state in a time-varying SSM and (iii) when the regression parameter vector β is unknown ; this covers both the cases of fixed but unknown β and random β with unknown covariance matrix. To reflect the lack of knowledge about the parameters in the last two situations, it is convenient to regard them as diffuse random variables. It will be shown in this Chapter that these three diffuse situations can be addressed in a unified fashion using a transparent generalization of the KF technology introduced in Chapter 2. In the next Chapter, we will demonstrate that this approach, when properly implemented, is superior in practicality and computational performance to alternative approaches discussed in the literature.

These alternative approaches can be categorised into four methods, all of which apply to the ASSM, which as described in Chapter 2, employs an augmented state to accomodate the regression parameter β. The first one, commonly known as the "big k" method, initiates the KF with an arbitrarily large covariance matrix in order to reflect the diffuseness in the initial state. The "big k" method is popular in empirical works (see Burridge and Wallis (1985) and Den Butter and Mourik (1990) for example) since it makes use of readily-available KF software. However as we will illustrate graphically later in this Chapter, it is inexact and numerically unstable. The second method employs the Information Filter (IF) ; see Kitagawa (1981), Sallas and Harville (1988) and Pole and West (1989). The drawbacks of the IF have already been discussed in Chapter 2. The third method, due to Harvey and Pierse (1984), is best introduced within the sphere of the ARIMA model. Here the state vector used is the one associated with the differenced model (i. e. stationary ARMA model) but augmented with the first d (the order of differencing) raw observations. More generally, the augmented part of the state corresponds to regression-type estimates based on an initial stretch of the raw observations. This technique is tantamount to producing an estimate of the state in effect at time $t = d$ and allows one to initiate the KF at $t = d$ where d is equal to the number of regression estimates. The method has two drawbacks : (i) the evaluation of the regression-type estimates can be difficult and potentially complicated and messy (e. g. with missing data) and (ii) the excessive augmentation of the state, when the SSM also includes an unknown regression parameter vector, makes it computationally unattractive. The fourth method, devised by Ansley and Kohn (1985b), applies in more generality than the three methods discussed above. Conceptually, the method amounts to removing the diffuseness in the SSM through a data transformation and thereafter applying the KF to the transformed data. This data transformation is achieved in an implicit fashion by a "modified" KF which is hereafter referred to as the AKKF. However the implementation of the AKKF can be complicated ; in particular, "existing Kalman Filter software cannot be used" (Bell and Hillmer, 1991).

The method used in this Chapter to treat the diffuseness problem in the SSM is due to De Jong (1991b). It expresses diffuse aspects of the SSM via a parameter vector γ, extends the KF technology discussed in Chapter 2 to estimate γ in parallel with the non-diffuse aspects of the state and finally indicates the appropriate adjustments required for factoring out the diffuse effects. This extended KF, called the *Diffuse Kalman Filter* (DKF), operates by applying the KF technology to each column of a *column-augmented* state. The latter consists of $\gamma^{\#}+1$ columns with the first $\gamma^{\#}$ columns each corresponding to a particular aspect of γ and the last column corresponding to the non-diffuse aspects of the state. The DKF is therefore a transparent generalization of the KF ; it turns a couple of vector iterations in the KF into matrix iterations. The DKF and the AKKF are conceptually similar since they are designed to factor out the effects of the diffuse parameter γ. However they differ in approach : γ is factored out explicitly in the DKF but implicitly in the AKKF. Although matrix recursions are employed in the DKF, it does not follow that the DKF is computationally expensive since its performance, like that of the KF, is significantly more dependent on the number of rows than the number of columns in the matrices that it computes. The issue of computational comparisons is covered in detail in the next Chapter.

In this Chapter, we address two original topics associated with diffuseness in the SSM. First, we derive, compare and relate two alternative pseudo-likelihoods called the *diffuse* and *marginal* likelihoods which are suitable normalized likelihoods for a SSM with diffuse parameters. This study is useful since it allows us to relate likelihoods evaluated from different transformations of the data. In the absence of diffuseness in the SSM, both the diffuse and marginal likelihoods reduce to the "ordinary" likelihood defined in Chapter 2. Second, we report some interesting properties of the DKF when it is applied to the often-used class of nonstationary ARMA models. We show that when it is applied to autoregressive processes, the DKF collapses *de facto* to the KF after an initial run. With mixed ARMA processes, we demonstrate the prudence of restricting grid-searching of the diffuse likelihood function to the invertibility region. This avoids numerical roundoff and overflow problems in the DKF.

The contents of this Chapter are as follows. In the first section, the SSM is redefined to incorporate diffuse parameters. Section 2 is concerned with the derivation and comparison of the marginal and diffuse likelihoods. Section 3 deals the DKF. We start with an intuitive explanation of the concepts behind the derivation of the DKF and thereafter we summarize the results of De Jong (1991b) concerning filtering, smoothing, gls estimation of the regression parameter and evaluation of the diffuse likelihood with the DKF. The results of Chapter 3 are then used to initialize the DKF for use with the class of time-invariant SSM's. We also report the square-root forms of the DKF and its associated smoothing algorithm. The section concludes with a graphical illustration of the pitfalls of employing the inexact "big k" method as opposed to an exact method like the DKF. Section 4 illustrates the DKF with the class of nonstationary ARMA models.

4.1 Anchoring the Diffuse SSM

In order to accomodate the three diffuse situations described in the preamble, it is necessary to redefine the anchoring of the SSM. This will be achieved via the device of a diffuse random variable.

Definition 4.1 A sequence of random variables $\{\gamma^1, \gamma^2, \ldots\}$ is said to be *diffuse* if $Cov(\gamma^k)$ is nonsingular and the sequence of inverses of $Cov(\gamma^k)$ converges, as $k \to \infty$, to the zero matrix in the Euclidean norm.

Remarks

1. The above definition of diffuseness translates to the assumption of a noninformative prior in Bayesian analysis.

2. This thesis henceforth uses the expression "diffuse random vector" to refer to the above sequence of random variables.

Definition 4.2 The *diffuse* SSM (DSSM) is the SSM defined by equations (2.1)-(2.2) with α_0 and β now specified as

$$\alpha_0 = a + A\gamma \qquad \text{and} \qquad \beta = b + B\gamma \tag{4.1}$$

where a and b are known, $(A; B)$ is of full column rank, $\gamma \sim (c, \sigma^2 C)$ with $C^{-1} \to \mathbf{0}$ (i. e. γ is diffuse) and γ uncorrelated with $(u_0, \ldots, u_n)$.

The above specifications are flexible : (i) if α_0 is totally diffuse then $rank\ (A) = \alpha_0^{\#}$, i. e. the number of components in α_0, (ii) if α_0 is partially diffuse then $rank\ (A) < \alpha_0^{\#}$ and (iii) if α_0 is not diffuse then A is null. Similar statements apply to β and $rank\ (B)$.

Remark

In line with Definition 4.1, a diffuse SSM is a sequence of SSM's with each term in the sequence corresponding to a term in the sequence $\{\gamma^k\}$.

4.2 Pseudo-Likelihoods for the Diffuse SSM

The presence of a diffuse parameter in the DSSM implies that the likelihood function of the latter converges pointwise to zero at every possible set of values of the parameter and it is therefore uninformative. One approach suggested to deal with this problem is to consider a likelihood based on a SSM where γ and consequently the initial conditions are fixed. However, it has been established in the literature (Tunnicliffe Wilson (1989) and Shepard and Harvey, 1990) that this approach may lead to erroneous statistical inferences with a high probability. This has led researchers to instead study pseudo-likelihood functions based on particular *normalizations* of the probability distribution of the model. An instance of normalization is differenced data which is employed by Box and Jenkins (1970) in the context of scalar ARIMA models. In this section, we generalize the differencing technique and thereafter construct normalized likelihoods based on implicit differencing of the data.

The differencing operation effects a data transformation which results in the transformed data being functionally independent of the diffuse parameter. The likelihood based on the transformed data is called the *marginal* (or *restricted* or *residual* or *invariant*) likelihood. The marginal likelihood was introduced by Kalbfleish and Sprott (1970) who were interested in defining a likelihood for those aspects of the data which are invariant to a set of *nuisance* parameters. These can be viewed in an analogous fashion to diffuse parameters. Thus if the "exact" likelihood is a product of two terms each supplying exclusive information on two different sets of parameters, one of which is considered a nuisance for the estimation process, then that part of the exact likelihood which relates to the parameter set of interest is called the marginal likelihood. Kalbfleish and Sprott (1970) "expounds marginal likelihood as a means for treating nuisance parameters and reducing bias in the parameters of interest" (Tunnicliffe Wilson, 1989). Marginal likelihood has been employed in the modelling of variance components (Patterson and Thompson, 1975) and in the estimation of ARMA parameters (Cooper and Thompson, 1977). Parameter estimation based on the marginal likelihood is commonly known as *restricted maximum likelihood* (REML) estimation. Both Ansley and Kohn (1985b) and Sallas and Harville (1988) employ a marginal likelihood for estimation purposes within the context of the SSM.

Two practical problems confront the use of the marginal likelihood. First, there is the issue of data transformation. This can be computationally tedious and intricate (e. g. in the case of missing observations). Furthermore although differencing is a popular data transformation in the scalar time series arena, its application to vector time series is still a moot point. For instance, Tsay and Tiao (1990) report that for the case of vector ARMA processes, identification of the "genuine" nonstationary components of the vector process is not currently feasible and furthermore differencing of the individual components of the observation vector is not justified and may even induce noninvertibility in the resulting stationary model. Second, we are faced with the issue of evaluating the marginal likelihood in an efficient manner. From our discussion in Chapter 2, we would expect that this could be done recursively. However this is not the case since the SSM structure is not maintained after a data transformation except in special circumstances. For instance, a necessary condition for the implementation of the AKKF of Ansley and Kohn (1985b) is that the matrix A (defined in equation 4.1) have a canonical form.

De Jong (1991b) has proposed an alternate pseudo-likelihood called the *diffuse likelihood* which does not have the shortcomings of the marginal likelihood for it is based on the untransformed data. Furthermore he demonstrates the evaluation of the diffuse log-likelihood via the DKF. In the next two subsections we review the derivation of the diffuse likelihood and establish its connection with the marginal likelihood.

4.2.1 The Diffuse Likelihood

An expression for the diffuse likelihood is best derived upon regarding the DSSM as a linear model. It is a straightforward exercise to show that upon repeated substitutions of the transition equation into the measurement equation and taking into account the anchoring of the DSSM defined in equation (4.1) that the DSSM can be written as the linear model $y = X(a;b) + X(A;B)\gamma + Gu$ where y and u are respectively the stacks of the observations and the error terms and X and G are built up from the regression and system matrices in the model. We will find it convenient to employ the following shorthand notations : $GG' = \Sigma$, $X(a;b) = x$ and $X(A;B) = \mathcal{X}$ and the linear model itself will be denoted by $(y, \{x, \mathcal{X}\}, G)$. Furthermore we remind the reader that the various log-likelihoods ($\lambda(y)$, $\lambda^d(y)$ and $\lambda^m(y)$) employed in the thesis are to be interpreted as -2 times the "exact" log-likelihood minus all additive constants which are independent of the parameters. These preliminaries set the stage for the derivation of the diffuse likelihood.

Theorem 4.1 (De Jong, 1991b) *Suppose y is generated by a DSSM with $(\gamma; u)$ normally distributed. Then as $C^{-1} \to \mathbf{0}$, $\lambda(y) - \log|\sigma^2 C|$ converges to the diffuse log-likelihood,*

$$\begin{aligned} \lambda^d(y) &= (y^\# - \gamma^\#)\log\sigma^2 + \log|S| + \log|\Sigma| + \sigma^{-2}(q - s'S^{-1}s) \qquad \textit{where} \\ S &= \mathcal{X}'\Sigma^{-1}\mathcal{X}, \qquad s = \mathcal{X}'\Sigma^{-1}(y-x) \qquad \textit{and} \qquad q = (y-x)'\Sigma^{-1}(y-x). \end{aligned}$$

The proof of the Theorem makes use of the following Lemma.

Lemma 4.1 *Suppose y is generated by a SSM with $(\gamma; u)$ normally distributed. Then $\gamma|y \sim N(\hat{\gamma}, \sigma^2(C^{-1}+S)^{-1})$ where $\hat{\gamma} = (C^{-1}+S)^{-1}(C^{-1}c+s)$.*

Proof. That $\gamma|y$ is normally distributed is a well known property of the normal distribution. Two well-known identities are needed to prove the result. Suppose P and Q are nonsingular matrices and R is conformable with P. Then

$$PR'(RPR'+Q)^{-1} = (R'Q^{-1}R + P^{-1})^{-1}R'Q^{-1} \tag{4.2}$$

$$(R'Q^{-1}R + P^{-1})^{-1} = P - PR'(RPR'+Q)^{-1}RP \tag{4.3}$$

These identities are now used to derive $E(\gamma|y)$ and $Cov(\gamma|y)$. Under the normality assumption, it follows that

$$\begin{aligned} E(\gamma|y) &= E(\gamma) + Cov(\gamma, y)\{Cov(y)\}^{-1}[y - E(y)] \\ &= c + C\mathcal{X}'(\mathcal{X}C\mathcal{X}' + \Sigma)^{-1}\{y - x - \mathcal{X}c\} \\ &= c + (\mathcal{X}'\Sigma^{-1}\mathcal{X} + C^{-1})^{-1}\mathcal{X}'\Sigma^{-1}(y - x - \mathcal{X}c) \qquad \text{by (4.2)} \\ &= c + (S + C^{-1})^{-1}(s - Sc) \\ &= (I - (C^{-1}+S)^{-1}S)c + (C^{-1}+S)^{-1}s \\ &= (C^{-1}+S)^{-1}(C^{-1}c+s) = \hat{\gamma} \end{aligned}$$

The last equality is obtained upon noting that $(C^{-1}+S)(I-(C^{-1}+S)^{-1}S) = C^{-1}$. Finally,

$$\begin{aligned}
\sigma^{-2}Cov(\gamma|y) &= Cov(\gamma) - Cov(\gamma,y)\{Cov(y)\}^{-1}\{Cov(\gamma,y)\}' \\
&= C - C\mathcal{X}'\{\mathcal{X}C\mathcal{X}'+\Sigma\}^{-1}\mathcal{X}C \\
&= (C^{-1}+S)^{-1} \text{ by (4.3)}
\end{aligned}$$

This asserts the Lemma. •

Proof of Theorem 4.1. Using Bayes' Theorem, it follows that $\lambda(y) = \lambda(\gamma)+\lambda(y|\gamma)-\lambda(\gamma|y)$ with

$$\begin{aligned}
\lambda(\gamma) &= \gamma^{\#}\log\ \sigma^2 + \log|C| + \sigma^{-2}(\gamma-c)'C^{-1}(\gamma-c) \\
\lambda(y|\gamma) &= y^{\#}\log\ \sigma^2 + \log|\Sigma| + \sigma^{-2}(y-x-\mathcal{X}\gamma)'\Sigma^{-1}(y-x-\mathcal{X}\gamma) \\
&= y^{\#}\log\ \sigma^2 + \log|\Sigma| + \sigma^{-2}(q-2s'\gamma+\gamma'S\gamma) \\
\lambda(\gamma|y) &= \gamma^{\#}\log\sigma^2 - \log|C^{-1}+S| + \sigma^{-2}(\gamma-\hat{\gamma})'(C^{-1}+S)(\gamma-\hat{\gamma})
\end{aligned}$$

Following direct simplification, $\lambda(y) = (y^{\#}-\gamma^{\#})\log\ \sigma^2 + \log|\sigma^2C| + \log|\Sigma| + \log|C^{-1}+S| + \sigma^{-2}(q+c'C^{-1}c-(s+C^{-1}c)'(C^{-1}+S)^{-1}(s+C^{-1}c)$. Now let $C\to\infty$. Then $\lambda^d(y) = \lambda(y) - \log|\sigma^2C|$ is as stated in the Theorem. •

The mle's of γ and σ^2 are respectively $\hat{\gamma} = S^{-1}s$ and $\hat{\sigma}^2 = (q-s'S^{-1}s)/(y^{\#}-\gamma^{\#})$. Subtracting $\log|\sigma^2C|$ from $\lambda(y)$ is tantamount to a normalization of the exact log-likelihood to ensure nondegeneracy. The "ordinary" log-likelihood described in Chapter 2 is a special instance of the diffuse likelihood ; it is obtained upon regarding $\gamma = \mathbf{0}$.

4.2.2 Connection between the Diffuse and the Marginal Likelihoods

In this subsection, we derive the marginal likelihood and establish its connection to the diffuse likelihood. As previously stated, the marginal likelihood is based on a transformation of $(y, \{x, \mathcal{X}\}, G)$ which is functionally independent of γ. This data transformation is achieved by a class of well-known linear maps whose interesting properties are described in the following Lemma.

Lemma 4.2 *Let* $M = I - \mathcal{X}[\mathcal{X}'\Sigma^{-1}\mathcal{X}]^{-1}\mathcal{X}'\Sigma^{-1}$. *Then (i)* M *is idempotent with rank* $y^\# - \gamma^\#$, *(ii)* $M\mathcal{X} = \mathbf{0}$ *and (iii)* $\Sigma^{-1}M = M'\Sigma^{-1}$.

Furthermore suppose N *is a* $(y^\# - \gamma^\#) \times y^\#$ *matrix spanning the rowspace of* M. *Then (iv)* $N\mathcal{X} = \mathbf{0}$ *and (v)* $N'(N\Sigma N')^{-1}N = \Sigma^{-1}M$.

Proof. Results (i)-(iii) are direct. For the second part of the Lemma, write $N = JM$ where J is of full-row rank. Then (iv) immediately follows from (ii). Now let $V = N'(N\Sigma N')^{-1}N$. Then

$$\begin{aligned}
N\Sigma V &= N\Sigma N'(N\Sigma N')^{-1}N = N \\
&\Rightarrow M\Sigma V = M \qquad \text{since } N \text{ spans the row space of } M \\
&\Rightarrow \Sigma^{-1}M\Sigma V = \Sigma^{-1}M \\
&\Rightarrow M'V = \Sigma^{-1}M \qquad \text{by result (iii)} \\
&\Rightarrow V = \Sigma^{-1}M \qquad \text{by results (iii) and (i)}
\end{aligned}$$

This asserts result (v) of the Lemma. •

Corollary 4.1 *For* q, s *and* S *as defined in Theorem* 4.1 *and* N *as defined in Lemma* 4.2,

$$q - s'S^{-1}s = \{N(y - x)\}'(N\Sigma N')^{-1}\{N(y - x)\}.$$

Proof. Direct manipulation of the pertinent quantities leads to

$$\begin{aligned} q - s'S^{-1}s &= \{(y-x)'\Sigma^{-1}(y-x)\} - \{(y-x)'\Sigma^{-1}\mathcal{X}\}\{\mathcal{X}'\Sigma^{-1}\mathcal{X}\}^{-1}\{\mathcal{X}'\Sigma^{-1}(y-x)\} \\ &= (y-x)'\Sigma^{-1}M(y-x) \\ &= \{N(y-x)\}'(N\Sigma N')^{-1}\{N(y-x)\} \end{aligned}$$

with the final equality following from result (v) of Lemma 4.2. •

The results of the Lemma and the Corollary are now used to establish the marginal likelihood and thereafter connect it with the diffuse likelihood. It is critical to recall at this stage that we are considering the *regular* SSM whereby y is interpreted as the stack of non-missing elements of the observations y_t. This ensures that every aspect of the non-missing data can manifest themselves in all possible linear transformations of y, in particular Ny.

Definition 4.3 The marginal likelihood of data y is the likelihood based on the transformed data Ny where N is as defined in Lemma 4.2.

Theorem 4.2 *The marginal log-likelihood apart from an additive constant equals,*

$$\lambda^m(y) \quad = \quad \lambda(Ny) = (y^\# - \gamma^\#)\sigma^2 + \log|N\Sigma N'| + (q - s'S^{-1}s)/\sigma^2.$$

Proof. The result follows from corollary 4.1. •

An immediate consequence of the Theorem is the following result.

Corollary 4.2 *The mle of σ^2 is the same when it is calculated from either the diffuse or marginal likelihoods.*

We now establish the connection between the diffuse and marginal likelihoods.

Theorem 4.3 *For the model* $(y, \{x, \mathcal{X}\}, G)$*, the diffuse and marginal log-likelihoods differ by* $\log|NN'| - \log|\mathcal{X}'\mathcal{X}|$.

Proof. We first express $|N\Sigma N'|$ as follows,

$$\begin{aligned}
|N\Sigma N'| &= |(N\Sigma N')^{-1}|^{-1} \\
&= |(NN')^{-1}(NN')(N\Sigma N')^{-1}(NN)'(NN')^{-1}|^{-1} \\
&= |(NN')^{-1}N\{N'(N\Sigma N')^{-1}N\}N'(NN')^{-1}|^{-1} \\
&= |(NN')|^2|N\Sigma^{-1}MN'|^{-1} \\
&= |NN'|^2|N\{\Sigma^{-1} - \Sigma^{-1}\mathcal{X}(\mathcal{X}'\Sigma^{-1}\mathcal{X})^{-1}\mathcal{X}'\Sigma^{-1}\}N'| \\
&= |NN'|^2 \begin{vmatrix} N\Sigma^{-1}N' & N\Sigma^{-1}\mathcal{X} \\ \mathcal{X}'\Sigma^{-1}N' & \mathcal{X}'\Sigma^{-1}\mathcal{X} \end{vmatrix}^{-1} |\mathcal{X}'\Sigma^{-1}\mathcal{X}| \\
&= |NN'|^2 \left| \begin{pmatrix} N \\ \mathcal{X}' \end{pmatrix} \Sigma^{-1}(N' \ \mathcal{X}) \right|^{-1} |\mathcal{X}'\Sigma^{-1}\mathcal{X}| \\
&= |NN'|^2|\Sigma| \left| \begin{pmatrix} N \\ \mathcal{X}' \end{pmatrix} (N' \ \mathcal{X}) \right|^{-1} |\mathcal{X}'\Sigma^{-1}\mathcal{X}| \\
&= |NN'|^2|\Sigma| \begin{vmatrix} NN' & \mathbf{0} \\ \mathbf{0} & \mathcal{X}'\mathcal{X} \end{vmatrix}^{-1} |\mathcal{X}'\Sigma^{-1}\mathcal{X}| \\
&= |NN'| \ |\mathcal{X}'\mathcal{X}|^{-1} \ |\Sigma| \ |S|
\end{aligned}$$

The second equality follows from result (v) of Lemma 4.2 ; the sixth equality from the well known formula for the determinant of a covariance matrix and the penultimate equality uses $N\mathcal{X} = \mathbf{0}$. Upon substitution of the above expression for $|N\Sigma N'|$, we obtain

$$\lambda^m(y) \quad = \quad (y^{\#} - \gamma^{\#})\sigma^2 + \log|N\Sigma N'| + (q - s'S^{-1}s)/\sigma^2$$

$$
\begin{aligned}
&= (y^{\#} - \gamma^{\#})\sigma^2 + \log|\Sigma| + \log|S| + (q - s'S^{-1}s)/\sigma^2 + \log|NN'| - \log|\mathcal{X}'\mathcal{X}| \\
&= \lambda^d(y) + \log|NN'| - \log|\mathcal{X}'\mathcal{X}|
\end{aligned}
$$

This asserts the Theorem. •

It follows from Theorem 4.3 that the difference between the diffuse and marginal log-likelihoods, namely $\log|NN'| - \log|\mathcal{X}'\mathcal{X}|$, can be interpreted as a penalty term resulting from the non-removal of diffuse effects in the SSM. The diffuse likelihood coincides with the marginal likelihood when the SSM is non-diffuse. Hence it can be used to discriminate between diffuse and non-diffuse SSM's. The penalty term contrasts with its counterpart in the Akaike's Information Criterion where it is expressed as an *ad hoc* function of the number of parameters in the model. The next two results illustrate the significance of this penalty term in the case of time invariant SSM's.

Theorem 4.4 *If matrix N represents ordinary differencing then $\lambda^m(y) \equiv \lambda^d(y)$.*

Proof. Consider the scalar model $(1-L)y_t = v_t$, where v_t is an arbitrary disturbance. This can be written as $y = \mathbf{1}y_0 + v$ where y and v are respectively the stacks of y_t's and v_t's. Here $\mathcal{X} = \mathbf{1}$ thereby implying that $|\mathcal{X}'\mathcal{X}| = n$. The matrix N has dimensions $(n-1) \times n$ with $N(i,i) = -1$, $N(i,i+1) = 1$ and zero elsewhere. It follows that NN' is tridiagonal with diagonal entries of 2 and subdiagonal and superdiagonal entries of -1. It is easy to show that $|NN'| = n$. Therefore $\lambda^m(y) \equiv \lambda^d(y)$ as asserted. •

Remarks

1. The Theorem is easily extended to higher order differencing i. e. $(1-L)^d y_t$, $d \geq 1$ and to seasonal differencing i. e. $(1-L^s)y_t$, $s \geq 1$. It also extends to vector observations y_t.

2. The Theorem indicates that differencing of the data, whereby the differenced data are regarded as noise, is in fact unnecessary since statistical inferences based on the diffuse and marginal likelihoods will in fact coincide. This is verified in the following application which uses the DKF in the context of an ARIMA (0,1,1) model.

Application to IBM stock prices. Box and Jenkins (1970) fit the model $z_t = e_t + 0.09e_{t-1}$, $t = 1, 2, \ldots, 369$, where $z_t = y_t - y_{t-1}$ with y_t representing the closing prices of IBM stock for the period 17^{th} May 1961 to 2^{nd} November 1962 (Series B). In this situation, the diffuse and marginal log-likelihoods are respectively based on data y_t and its differenced form z_t. Using the square-root DKF algorithm (presented in the next section), we obtain $\lambda^d(y) = 1814.9$ and $\hat{\sigma}^2 = 52.2$. These results coincide with those provided by Box and Jenkins.

The next example links Theorem 3.4 and Theorem 4.3. It is shown that in the context of time invariant SSM's with state transition matrix T, the penalty term $\log|NN'| - \log|\mathcal{X}'\mathcal{X}|$ is a function of the nonstationary eigenvalues of T.

Theorem 4.5 *Consider the DSSM,*

$$y_t = Z_t\alpha_t + G_t u_t, \qquad \alpha_{t+1} = T\alpha_t + Hu_t, \qquad t = 0, 1, 2, \ldots$$

Then the difference between the diffuse and marginal log-likelihoods of this DSSM is a function of the nonstationary eigenvalues of T.

Proof. As a consequence of Theorem 3.4 one can write $\alpha_0 = U_1\gamma + \alpha_0^\dagger$. Repeated substitution of the transition equation into the measurement equation allows one to write $y_t = Z_t T^t U_1 \gamma + v_t$ where v_t is a linear function of $\alpha_0^\dagger$ and $(u_0, \ldots, u_t)$ and U_1 is as defined in Theorem 3.3. Therefore the t^{th} row of $\mathcal{X} = Z_t T^t U_1 = Z_t U_1 P^t$. Matrix $\mathcal{X}$ also affects the determination of N (see Lemma 4.2). Therefore we conclude that the difference between the diffuse and marginal log-likelihoods is explained by the nonstationary eigenvalues of T. •

Application to scalar explosive autoregressive processes. To assess the relevance of Theorem 4.5, consider the scalar autogressive model $y_t = ay_{t-1} + \epsilon_t$ where $a > 1$. In this case, the i^{th} component of $\mathcal{X}$ is a^i while an appropriate $(n-1) \times n$ matrix N has entries $N(i,i) = -a$, $N(i,i+1) = 1$ and zero elsewhere. It is easy to show through direct algebra that $|\mathcal{X}'\mathcal{X}| = a^2|NN'|$ thereby implying that the diffuse and marginal likelihoods differ by the logarithm of the square modulus of the nonstationary eigenvalue, a.

4.3 Statistical Inference with the Diffuse SSM

The estimation of the states and regression parameters in the DSSM requires dealing with the diffuse parameter γ. In the initial part of this section, we demonstrate using ideas borrowed from Rosenberg (1973) that an efficient filtering algorithm for the DSSM is tantamount to a modified KF which estimates γ in parallel with the nondiffuse aspects of the states. We then show that the DKF of De Jong (1991b) immediately follows from these ideas. The section ends with a summary of the results of De Jong (1991a, 1991b) concerning diffuse filtering, smoothing, likelihood evaluation and generalized least squares estimation of regression parameters with the DKF. We also briefly discuss the implementation of the DKF via its square-root form.

For this informal introduction to the ideas behind the DKF, we assume without any loss in generality $(a; b) = \mathbf{0}$. The following result is useful for subsequent discussions.

Lemma 4.3 *The DSSM can be expressed as* $y_t = \mathcal{X}_t\gamma + v_t$ *with* $\mathcal{X}_t$ *built up from the system and regression matrices and where* v_t *is generated as follows,*

$$v_t \;=\; Z_t\alpha_t^\dagger + G_t u_t \;, \qquad \alpha_{t+1}^\dagger = T_t\alpha_t^\dagger + H_t u_t \qquad \textit{where}$$

$$\alpha_0^\dagger = \mathbf{0}, \qquad \sigma^{-2}Cov(\alpha_0^\dagger) = H_0H_0'.$$

Proof. The proof easily follows upon repeated substitutions of the transition equation into the observation equation. •

Now suppose the KF is applied to $\{v_t\}$. This is possible since the initial state of this process is non-diffuse. Denote the innovations and their covariance matrices by $e_t = v_t - Pred(v_t|v_0;\ldots;v_{t-1})$ and $\sigma^2 D_t$. Put $v = (v_1;\ldots;v_n)$, $e = (e_1;\ldots;e_n)$ and $D = Diag(D_1,\ldots,D_n)$. Then the log-likelihood of v is given by,

$$\lambda(v) = n\log\sigma^2 + \log|D| + \sigma^{-2}e'D^{-1}e$$

Observe that $e = Kv = K(y - \mathcal{X}\gamma)$ where y is the stack of the $y_t's$ and K is lower triangular with ones on the diagonal thereby implying $|K| = 1$. It is crucial to recognise that K is orthogonal and is implicitly produced by the KF. Furthermore,

$$\begin{aligned} e'D^{-1}e &= (Kv)'D^{-1}(Kv) \\ &= (y-\mathcal{X}\gamma)'K'D^{-1}K(y-\mathcal{X}\gamma) \\ &= \{\mathcal{K}(y-\mathcal{X}\gamma)\}'\{\mathcal{K}(y-\mathcal{X}\gamma)\} \end{aligned}$$

where $\mathcal{K} = D^{-1/2}K$. Therefore the gls of γ is obtained upon regressing $\mathcal{K}y$ on $\mathcal{K}\mathcal{X}$, a process than can be done in parallel with the filtering of v. This suggests applying a KF-like algorithm to the *augmented* observation $(\mathcal{X}_t, y_t)$ instead of y_t alone. This is the concept behind the DKF.

4.3.1 Filtering and Likelihood Evaluation with the DKF

This subsection summarises the work of De Jong (1991b) with regards to the use of the DKF in filtering, smoothing, gls estimation of regression parameters and evaluation of the diffuse log-likelihood. First consider how the diffuse log-likelihood defined in

Theorem 4.1 can be evaluated in a recursive fashion. As Schweppe (1965) did earlier in the derivation of the likelihood of a SSM (see Chapter 2), De Jong (1991b) exploits the fact that $\Sigma^{-1} = K'D^{-1}K$ to express S and s as,

$$S = \mathcal{X}'\Sigma^{-1}\mathcal{X} = (K\mathcal{X})'D^{-1}(K\mathcal{X}), \qquad s = \mathcal{X}'\Sigma^{-1}(y-x) = (K\mathcal{X})'D^{-1}K(y-x)$$

In Chapter 2, we saw that the stack of innovations produced by the KF is $e = K(y - \mathcal{X}\beta)$, where $\mathcal{X}\beta$ is a known mean effect. Therefore replacing β by $(a; b)$ allows the computation of $f = K(y-x)$. Furthermore each column of $K\mathcal{X}$ may be computed by substituting y by a column of zeroes and β by the relevant column of $-(A; B)$. These ideas are used in the DKF for the recursive evaluation of S and s and ultimately the diffuse log-likelihood. We now formally define the DKF.

Definition 4.4 The DKF is the KF (see Theorem 2.1) with the equations for e_t and $\hat{\alpha}_{t+1}$ respectively replaced by

$$E_t = (X_tB, y_t - X_tb) - Z_tA_t \qquad \text{and} \qquad A_{t+1} = W_t(-B, b) + T_tA_t + K_tE_t$$

with $A_1 = W_0(-B, b) + T_0(-A, a)$ and P_1 unchanged.
Furthermore attach the recursion $Q_{t+1} = Q_t + E_t'D_t^{-1}E_t$ with $Q_1 = \mathbf{0}$ to the DKF. The matrix Q_t is of the form,

$$Q_t = \begin{pmatrix} S_t & s_t \\ s_t' & q_t \end{pmatrix}$$

and hence the diffuse log-likelihood as stated in Theorem 4.1 is given by,

$$\lambda^d(y) = (y^{\#} - \gamma^{\#})\log\sigma^2 + \log|S_{n+1}| + \sum_{t=1}^{n}\log|D_t| + \sigma^{-2}(q_{n+1} - s_{n+1}'S_{n+1}^{-1}s_{n+1})$$

Remarks

1. The DKF turns two vector recursions in the KF into matrix recursions. Furthermore evaluation of the diffuse log-likelihood is made possible upon appending the recursion of the matrix Q_t to the DKF. This contrasts with the KF where Q_t is a scalar.

2. The matrices A and B are the same matrices used in defining α_0 and β. We will address the appropriate specification of A for the nonstationary time-invariant SSM in a future subsection.

3. The last column of A_t and E_t are interpreted as the nondiffuse aspects of the state and the innovation at time t.

The above ideas put us in a position to appreciate De Jong's (1991b) results on filtering, smoothing with the DSSM. The predictors therein are interpreted as *limiting predictors* since they assume that the diffuse parameter γ is such that $\sigma^2\{Cov(\gamma)\}^{-1} = C^{-1} \to 0$.

Theorem 4.6 (De Jong , 1991b) *Suppose $y = (y_1; \ldots; y_n)$ is generated by the DSSM. Let (i) $\hat{x}_t$ and $\tilde{x}_t$ respectively denote the limiting predictors of the random variable x conditional on $(y_1; \ldots; y_{t-1})$ and $(y_1; \ldots; y_n)$ and (ii) $M_{t\gamma}$ denote all but the last column of matrix M_t. Then*

$$\begin{aligned}
\hat{\gamma}_t &= S_t^{-1} s_t, \qquad \sigma^{-2} Mse(\hat{\gamma}_t) = S_t^{-1} \\
\hat{\alpha}_t &= A_t(-\hat{\gamma}_t; 1), \qquad \sigma^{-2}\ Mse(\hat{\alpha}_t) = P_t + A_{t\gamma} S_t^{-1} A'_{t\gamma} \\
y_t - \hat{y}_t &= E_{t\gamma}(-\hat{\gamma}_t; 1), \qquad \sigma^{-2} Mse(\hat{y}_t) = D_t + E_{t\gamma} S_t^{-1} E'_{t\gamma} \\
\hat{\beta}_t &= b + B\hat{\gamma}_t, \qquad \sigma^{-2} Mse(\hat{\beta}_t) = B S_t^{-1} B'
\end{aligned}$$

Furthermore, for $1 \leq t \leq r \leq n+1$,

$$\tilde{\alpha}_t = F_t(-\hat{\gamma}_{n+1}; 1) \qquad \textit{and} \qquad \sigma^{-2} Mse(\tilde{\alpha}_t, \tilde{\alpha}_r) = P_t L'_{r-1,t}(I - R_{r-1}P_r) + F_{t\gamma} S_{n+1}^{-1} F_{r\gamma}$$

$$\textit{where} \qquad N_{t-1} = Z_t' D_t^{-1} E_t + L_t' N_t \qquad \textit{and} \qquad R_{t-1} = Z_t' D_t^{-1} Z_t + L_t' R_t L_t$$

with N_n *and* R_n *equal to zero matrices,* $F_t = A_t + P_t N_{t-1}$, $L_t = T_t - K_t Z_t$ *and* $L'_{r-1,t} = \prod_{i=t}^{r-1} L'_i$ *with* $L_{t-1,t} \equiv I$.

If S_t is singular then a generalized inverse (e. g. S_t^+) can be employed. The results of the Theorem clearly generalize those described in Chapter 2 for the non-diffuse SSM. Thus the DKF is a transparent generalization of the KF. The conceptual elegance and the computational simplicity of the DKF makes it an attractive proposition vis a vis a competitor like the AKKF of Ansley and Kohn (1985b).

4.3.2 Square Root DKF

The connection between the KF and the DKF makes it obvious that their square-root forms should be similar except for differences in the dimensionalities of various matrices. This subsection presents a slightly modified version of the square root DKF originally devised by De Jong (1991a). The algorithm proceeds as follows,

- Step 0. Initialize $A = W_0(-B, b) + T_0(-A, a)$, $P = H_0$, $\lambda = m = 0$ and set Q to a null matrix.

  For steps $t = 1, 2, \ldots, n$ do

- Step 1. Postmultiply matrix on the left with an orthogonal matrix U such that

$$\begin{vmatrix} Z_t P_t^{1/2} & G_t \\ T_t P_t^{1/2} & H_t \end{vmatrix} U = \begin{vmatrix} D & \mathbf{0} \\ K & P \end{vmatrix}$$

  and D is row-echelon with the same number of rows as Z_t.

- Step 2. $E = D^{-}\{(\mathcal{X}_t B, y_t - \mathcal{X}_t b) - Z_t A\}$, $\quad A = W_t(-B, b) + T_t A + KE$
  $\lambda = \lambda + \log|D|$, $\quad m = m +$ column-rank (D)
- Step 3. Update Q via an orthogonal transformation U,

$$(Q; E)U \;=\; Q$$

De Jong (1991a) shows that at the end of each iteration, $D = D_t^{1/2}$, $P = P_{t+1}^{1/2}$, $A = A_{t+1}$ and $Q = Q_t^{1/2}$. The matrix Q' has canonical form $Q' = \{(\mathcal{Q}, w); (0, r)\}$, with r, a scalar. Upon multiplying Q by its transpose, we immediately recognise $S_t = \mathcal{Q}'\mathcal{Q}$, $s_t = \mathcal{Q}'w$ and $q_t = r^2 + w'w$. Therefore it follows that (i) $\hat{\gamma}_t = S_t^{-1}s_t = (\mathcal{Q}'\mathcal{Q})^{-1}\mathcal{Q}'w = \mathcal{Q}^{-1}w$ and (ii) $m\hat{\sigma}_t^2 = q_t - s_t' S_t^{-1} s_t = (r^2 + w'w) - w'\mathcal{Q}(\mathcal{Q}'\mathcal{Q})^{-1}\mathcal{Q}'w = r^2$.

Step 3 in the square-root DKF devised by De Jong (1991a) is more elaborate in the sense that the update of the (generalized) inverse of Q is achieved with an orthogonal matrix U such that,

$$\begin{pmatrix} Q & E \\ Q'^{-} & \mathbf{0} \end{pmatrix} U \;=\; \begin{pmatrix} Q & \mathbf{0} \\ R & W \end{pmatrix}$$

De Jong shows that R automatically holds $Q^{-1'}$ if Q was square on the previous iteration. If Q is not square, then its generalized inverse must be explicitly computed. This is necessary for only a few initial iterations unless a multicollinearity problem exists. The advantage of this approach is that $S_{t+1}^{-1/2}$ is immediately available in R. This "luxury" however comes at the price of carrying out an orthogonal transformation on an augmented matrix at each iteration.

The use of either version of step 3 depends on the purpose under consideration. If the square-root DKF is used for likelihood evaluation and smoothing, then our step 3 is more appropriate since only $\hat{\gamma}_{n+1}$ is required and therefore only one matrix inversion is needed. However if we are interested in monitoring the behaviour of $\hat{\gamma}_t$, for example in a study of stability of regression relationships over time, then De Jong's version is more appropriate.

On a final note, we mention that a square-root form of the diffuse smoothing algorithm is identical to the one used in the non-diffuse context except for the diffference in the dimensions of the pertinent quantities.

4.3.3 Automatic Initialization of the DKF

We now demonstrate, using the results of Chapter 3, the appropriate initialization of the DKF for the class of nonstationary time-invariant SSM's. From Theorem 3.4, it follows that the initial state can be written as $\alpha_1 = T(U_1\gamma + \alpha_0^\dagger) + Hu_0$ where $\gamma \sim (0, kI)$ with $k \to \infty$ and $\sigma^{-2}Cov(\alpha_0^\dagger) = U_2MU_2'$. Therefore for this class of SSM's, an appropriate initialization of the DKF is $A_1 = (-TU_1, a) + W(-B, b)$ and $P_1 = TU_2MU_2'T' + HH'$.

4.3.4 Pitfall of employing the "big k" method

We have emphasized the fact that the "big k" method is inexact. As previously noted, it is nevertheless a popular approach adopted in empirical works since it employs readily-available KF software. We now demonstrate that employing the "big k" method can have serious consequences for all aspects of statistical inference. This can be seen by considering the behaviour of the time series $\sigma^2 D_t$, the variance of the innovations, as evaluated in the first instance from the KF using the "big k" method and in the second instance using the DKF. For the purpose of illustration, we consider the QBSM discussed in Chapter 2. Recall that the QBSM is a nonstationary SSM with its state transition matrix having four nonstationary eigenvalues each with unit modulus. Therefore for the "big k" method, the KF is initialized with the estimate of the initial state covariance matrix equal to kI where k is large.

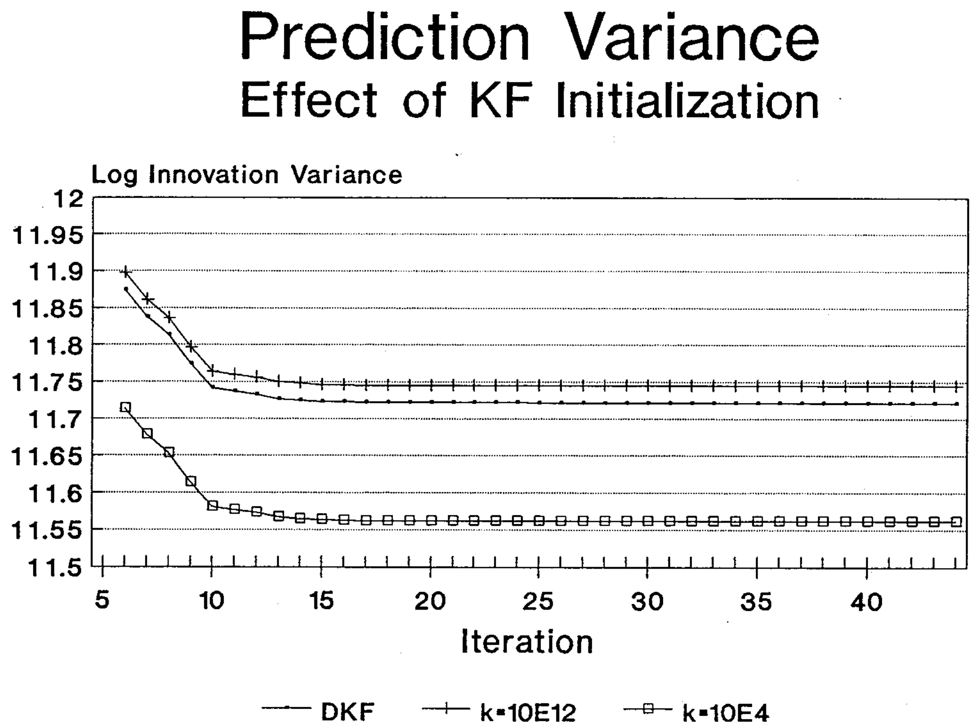

Figure 4.1: $\sigma^2 D_t$ evaluated vith DKF and "big k" methods.

Since the DKF can be viewed as equivalent to the "big k" method with k arbitrarily large, one would expect $\sigma^2 D_t$ to be highest when the DKF is employed. However this is not the case in Figure 4.1. This is explained by the fact that the use of the "big k" method, with k large, is numerically unstable. Bell and Hillmer (1991) report a similar phenomenon upon applying a modified AKKF to a seasonal ARMA model.

4.4 Characteristics of the DKF with ARMA Models

The ARMA model is a popular tool in Time Series Analysis. Box and Jenkins (1970) have built up a complete set of statistical techniques around stationary ARMA models. However many applications, especially those arising in the socio-economic areas, require the use of nonstationary ARMA models such as ARIMA models. In these situations, statistical inferences have been traditionally conducted with the stationary models resulting from iterated differencing of these nonstationary models. This practice has however been questioned in the case of vector data by Tsay and Tiao (1990) who ultimately go on to recommend that statistical analysis be carried using the raw data. The material in the last Chapter and the current Chapter indicate that the DKF provides a means for statistical inference in nonstationary ARMA models at all levels of generality. In particular, vector ARMA processes and data irregularity problems such as missing data are covered.

In the next subsection, we display an interesting collapsing property of the DKF when it is applied to nonstationary autoregressive processes. The subsection thereafter focusses on the consequences of noninvertibility on the diffuse likelihood. For ease of presentation, we have only considered scalar ARMA models ; however generalization for the vector models is direct.

4.4.1 Autoregressive Processes

We demonstrate in this subsection that when the DKF is applied to a nonstationary AR(p) process, the estimate of the diffuse parameter γ associated with the state α_0 is in fact obtained from the first p observations (we are assuming a regular SSM i. e. without any missing observations). This has the interesting implication that after the p^{th} iteration, the DKF self-collapses to the KF.

Theorem 4.7 *Suppose the DKF is applied to a nonstationary AR(p) process. Partition $A_t = (A_{t\gamma}, a_t)$ and $E_t = (E_{t\gamma}, e_t)$ where a_t and e_t are vectors. Then for $t > p$, $E_{t\gamma} = A_{t\gamma} = \mathbf{0}$ and a_t and e_t corresponds to the limiting predictors of the state and the innovation at time t.*

Proof. Consider the AR (p) model, $y_t = a_1 y_{t-1} + \ldots + a_p y_{t-p} + \epsilon_t, t \geq 1$. Upon repeated substitutions, one can write $y = (y_1; \ldots; y_n)$ as $y = Ay^\dagger + B\epsilon$ where $y^\dagger = (y_1; \ldots; y_p)$ and $\epsilon = (\epsilon_1; \ldots; \epsilon_n)$. Hence all the diffuse aspects of the process can be efficiently inferred from $y^\dagger$ only and this in turn implies that S and s attain their final values after iteration p of the DKF. Since S is a positive semi-definite matrix, it then follows that for $t > p$, $E_{t\gamma}$ and in turn $A_{t\gamma}$ are matrices of zeroes. Consequently, for $t > p$, a_t and e_t must correspond to the limiting predictors of the state and the innovation at time t. •

Remark

The Theorem asserts when the DKF is applied to nonstationary AR (p) processes, it collapses *de facto* to the KF i. e. at $t = p + 1$, we can reinitialize $A_t = (0, a_t)$ and Q_{p+1} by a_t and $q_{p+1} - s'_{p+1} S_{p+1}^{-1} s_{p+1}$, update the diffuse log-likelihood by $\log |S_{p+1}|$ and then run the KF for $t \geq p + 1$. Hence this collapsed DKF is as computationally efficient as any alternative algorithm proposed in the literature (such as those described in the next Chapter).

The *de facto* collapse of the DKF to the KF occurs whenever the diffuse aspects of the SSM are completely determined by an initial stretch of the observations $(y_1; \ldots; y_m)$ (say) in which case $A_t = (A_{t\gamma}, a_t) = (0, a_t)$ for $t > m$. Whether A_t exhibits such a behaviour in other SSM's is a moot point. It is an easier task to find SSM's which do not lead to such A_t's. Consider the following two examples. First, suppose the SSM contains a diffuse regression parameter β. In that case, the optimal estimator of β is based on the entire observation set and hence $A_{t\gamma}$ does not necessarily stay a zero matrix. Second, consider a nonstationary mixed ARMA (p,q) process. This can be written as $y = Ay^\dagger + B\epsilon + C\epsilon^\dagger$ where y, $y^\dagger$, and ϵ are as described in the proof of Theorem 4.7 and $\epsilon^\dagger = (\epsilon_{1-q}; \ldots; \epsilon_0)$. In this case, the diffuse aspects of the process are captured in both $y^\dagger$ and $\epsilon^\dagger$. The optimal estimator of $\epsilon^{\dagger}$ requires the entire observation set y and consequently $A_{t\gamma} \neq \mathbf{0}$.

Remark

Since it is often possible to approximate a mixed ARMA process by a relatively low order AR process, we would expect, in view of Theorem 4.7, the entries of the matrices $E_{t\gamma}$ to be close to zero after an initial number of DKF iterations. Therefore we expect S_t and s_t to substantially attain their their final values in the early iterations of the DKF.

The next Chapter focusses on the computational aspects of the DKF. The work therein is motivated by the results in this subsection. In particular, it is shown that after an initial run, the DKF can always be switched to a KF based on the ASSM (i. e. the SSM with augmented states). This KF is however not computationally efficient due to the dimensions of the states. This leads us to consider a collapsing strategy which consists of reducing the column dimensions of pertinent matrices in the DKF. This collapsed DKF is shown to outperform the competition.

4.4.2 Mixed ARMA Processes

Box and Jenkins (1970, p198-199) observe that a stationary ARMA (p,q) model may have up to 2^q representations thereby implying that the processes described by these representations have the same autovariance function. Therefore these processes must also share the same likelihood function and consequently they also generate one-step-ahead prediction errors with identical means and variances. Osborn (1976) argues that due to roundoff errors, grid searching of likelihood values across the parameter space must be restricted to the *invertibility* region. This invertibility property is satisfied by only one of these 2^q parametrizations of the ARMA model. For completeness, we now define the concept of invertibility.

Definition 4.5 The ARMA (p,q) process $a(L)y_t = b(L)\epsilon_t$, where L is the lag operator (i. e. $Lx_t = x_{t-1}$) and $a(.)$ and $b(.)$ are polynomials of order p and q in L, is said to be *invertible* if the roots of $b(L) = 0$ lie outside the unit circle.

In this subsection, we demonstrate that the above remarks transcends to nonstationary ARMA processes with the diffuse likelihood used instead of the "exact" likelihood. We show how roundoff errors arise in the DKF and therefore stress the prudence of keeping to the invertible region while grid-searching the diffuse likelihood function.

Theorem 4.8 *Up to 2^q parametrizations of a nonstationary ARMA(p,q) process share the same diffuse likelihood function.*

Proof. It suffices to consider an MA (q) process, since the nonstationary ARMA (p,q) process, $a(L)y_t = b(L)\epsilon_t$ can be viewed as $z_t = b(L)\epsilon_t$ with $z_t = a(L)y_t$. Suppose the model is invertible with the roots of $b(L)$ (possibly complex) denoted by $\theta_j, j = 1, \ldots, q$. The spectrum of $z = (z_1; \ldots; z_n)$ is given by,

$$\begin{aligned} f_z(\lambda) &= \sigma^2 |1 + \sum_{j=1}^{q} b_j \exp^{-i\lambda j}|^2 \\ &= \sigma^2 \prod_{j=1}^{q} |1 + \theta_j \exp^{-i\lambda j}|^2 \\ &= \sigma^2 \prod_{j=1}^{q} (1 + \theta_j^2 + 2\theta_j \cos \lambda j) \end{aligned}$$

When θ_j is real, it follows that

$$1 + \theta_j^2 + 2\theta_j \cos \lambda j \equiv \theta_j^2 (1 + (1/\theta_j)^2 + 2(1/\theta_j) \cos \lambda j)$$

This asserts that the spectrum is invariant to root flippings. Since the spectrum stands a 1:1 relationship with the autocovariance function, this implies that the latter and hence by extension the diffuse log-likelihood is invariant to possibly 2^q different parametrizations of the ARMA (p,q) process. •

Theorem 4.8 implies that an identification problem is likely during grid-searching of the diffuse likelihood function. Restricting the grid-searching to the invertibility region avoids this identification problem. However a more consequential argument to keeping to the invertibility region is that the evaluation of the diffuse likelihood of nonstationary, noninvertible ARMA processes is prone to acute roundoff errors. To see this, write the ARMA model as, in the previous subsection, $y = Ay^\dagger + B\epsilon + C\epsilon^\dagger$. Then it is easy to see that in this context, roundoff errors are likely to arise since when t is sufficiently large, the entries of the t^{th} row of A, B and C diverge.

We now illustrate with a simple example how the evaluation of the diffuse likelihood of noninvertible processes is plagued by overflow problems when the DKF is employed. Consider the ARIMA (0,1,1) process $y_t = y_{t-1} + \epsilon_t + b\epsilon_{t-1}$, where the disturbances ϵ_t's are serially independent with mean zero and variance v. Assign $b = \theta$ ($|\theta| < 1$) and $v = \sigma^2$ for the invertible model and $b = 1/\theta$ and $v = \sigma^2\sigma^2$ for the noninvertible model. Upon assuming that this process has applied since time-immemorial, it follows from Theorem 3.4 that the stack of observations can be written as $y = \mathbf{1}\gamma + (\mathbf{1},\mathbf{1})\alpha_0^\dagger + Gu$, where $\sigma^{-2}Cov(\alpha_0^\dagger) = \{(1,-1);(-1,1)\}$ and G is lower-triangular with diagonal entries of one and subdiagonal entries of $(1+b)$. The "non-diffuse" portion of $\sigma^{-2}Cov(y)$ is $\Sigma = \sigma^{-2}(\mathbf{1},\mathbf{1})Cov(\alpha_0^\dagger)(\mathbf{1},\mathbf{1})' + GG' = GG'$. This implies $|\Sigma| = 1$. Upon direct manipulation, it can be shown that $S(b) = \mathbf{1}'(GG')^{-1}\mathbf{1}' = \sum_{t=0}^{y^\# - \gamma^\#} b^{2t}$ (here $\gamma^\# = 1$) where $S(b)$ is S evaluated with MA coefficient b. Let $C(b,v)$ denote the sum of square of errors $q - s'S^{-1}s$ evaluated with parameters b and v. Then it is easy to deduce that $C(\theta,\sigma^2) = \theta^2 C(\theta^{-1}, \theta^2\sigma^2)$. Therefore, after noting that $\log|\Sigma| = 0$ independantly of the value of b, we obtain

$$\begin{aligned}
\lambda^d(y|\theta^{-1},\theta^2\sigma^2) &= (y^{\#}-1)\log(\theta^2\sigma^2)+\log\sum_{t=0}^{y^{\#}-1}\theta^{-2t}+(\theta\sigma)^{-2}C(\theta^{-1},\theta^2\sigma^2) \\
&= 2(y^{\#}-1)\log\theta+(y^{\#}-1)\log\sigma^2+\log\theta^{-2(y^{\#}-1)}\sum_{t=0}^{y^{\#}-1}\theta^{2t} \\
&+ (\theta\sigma)^{-2}\theta^2 C(\theta,\sigma^2) \\
&= (y^{\#}-1)\log\theta^2+\log\sum_{t=0}^{y^{\#}-1}\theta^{2t}+\sigma^{-2}C(\theta,\sigma^2) \\
&= \lambda^d(y|\theta,\sigma^2)
\end{aligned}$$

as asserted in Theorem 4.8. Observe however that with the noninvertible process, S blows up and this makes computations in the DKF unsound. This demonstrates the prudence of restricting ourselves to invertible ARMA models.

4.5 Summary

In this Chapter, we have treated the problem of statistical inference in the DSSM. We have demonstrated that recursive filtering, smoothing, evaluation of the log-likelihood and the gls estimation of regression parameters can be carried out with a transparent extension of the KF labelled the DKF. We have considered the merits of the diffuse and marginal likelihoods as suitable pseudo-likelihoods for the DSSM. We displayed two interesting characteristics of the DKF when it is applied to nonstationary ARMA models. First with autoregressive models, the DKF is shown to collapse *de facto* to the KF after an initial number of iterations. Second we have stressed the prudence of restricting the grid-searching of the diffuse likelihood function to the invertibility region. This is necessary to avoid numerical roundoff and overflow problems. The work in the next Chapter is motivated by the collapsibility property of the DKF. There we show a means of forcing the collapse of the DKF, which for arbitrary SSM's, is not necessarily equal to the KF.

Chapter 5

Efficient Algorithms for the State Space Model

In this Chapter we show that the DKF, when properly implemented, is superior in performance to alternative algorithms proposed in the literature for the purposes of filtering, smoothing, likelihood evaluation, gls estimation of regression effects and diagnostics generation in the DSSM. This may appear surprising since the DKF has two apparent shortcomings : (i) the vector recursions for e_t, $\hat{\alpha}_t$ and the scalar recursion for q_t in the KF are replaced by matrix recursions and (ii) it does not immediately provide limiting predictors of the state or estimates of the regression effects.

The alternative approaches to the DKF do not suffer from these shortcomings since they apply the KF to all but an initial stretch of the observations. Ansley and Kohn (1985b,1990) switch from their modified KF (hereafter called AKKF) to the standard KF after an initial startup period whereas Sallas and Harville (1988) and Pole and West (1989) both initially use the Information Filter (IF) and thereafter switch to the KF. In both cases, the switch to the KF is explained by the fact that once a proper estimate of the diffuse parameter γ can be constructed (from an initial stretch of the observations) then it can subsequently be used to construct limiting predictors of the states via the KF. This concept is evident in Harvey and Pierse (1984) and Bell and Hillmer (1991) : they both deal directly with the diffuseness problem by using an initial stretch of the observations to construct regression type estimates for initializing the KF used for the subsequent stretch of the observations.

It therefore appears that the usefulness of either the AKKF or the IF is confined to providing estimates for initializing the KF used thereafter. Since the DKF achieves the same objectives as the AKKF and IF, it is of interest to study the merits of switching from the DKF to the KF after a sufficient number of iterations. De Jong (1991a, 1991b) provides the implementation details and discusses the utility of switching from the DKF to the KF. The latter is based on a SSM with states augmented by the diffuse parameter γ.

We make several contributions in this Chapter. In section 1, we demonstrate that without any loss in generality, the SSM can be defined with the diffuse parameter γ partitioned as $\gamma = (\gamma_1; \gamma_2)$ where γ_1 and γ_2 are respectively the diffuse effects associated with the initial state and the regression parameter. With this redefined SSM, we argue in the following section, that from the standpoint of likelihood evaluation, De Jong's collapse of the DKF to the KF only necessicates the augmentation of the states by γ_2. This KF, which we label the *Augmented* KF (AKF), coincides with the alternative algorithms to the DKF (discussed above) since these are based on the ASSM wherein the states are augmented by the regression parameter β. We next show in section 3 that an analogue to the AKF is a *column-reduced* DKF, labelled the *collapsed* DKF (CDKF), where the appropriate submatrices of A_t, E_t and Q_t associated with γ_1 are *partialled-out* after an initial stretch of the observations has been processed. Both the AKF and CDKF coincide with the KF in the absence of a regression parameter in the SSM. Square root forms of the AKF and CDKF and their associated smoothing algorithms are also described in these two sections. Section 4 is devoted to the comparison of the computational aspects of the DKF, AKF and CDKF. We conclude that the CDKF is generally more efficient than either the DKF or the AKF since (i) it employ matrices A_t, E_t and Q_t with lower dimensionalities than the DKF and (ii) it recurs mse matrices P_t of lower dimensionalities than the AKF. This tells us that the performance of filtering and smoothing algorithms is significantly more affected by the number of rows than the number of columns in their pertinent matrices and thereby validates our reservation (see Chapter 2) on incorporating a regression parameter within the state, as in the ASSM specification.

5.1 The Canonical Form of the Diffuse State Space Model

In the previous Chapter, the DSSM was anchored with $\alpha_0 = a + A\gamma$ and $\beta = b + B\gamma$ with $(A; B)$ of full column rank. The following result shows that without any loss of generality, $(A; B)$ may assume a canonical structure.

Lemma 5.1 *The DSSM (see Definition* 4.2*) can be transformed such that* $(A; B)$ *has the canonical structure* $\{(\mathcal{A}_1\ \mathcal{A}_2); (\mathbf{0}\ \mathcal{B})\}$ *where* $\mathcal{A}_1$ *and* $\mathcal{B}$ *respectively have the same row dimensions as* A *and* B*.*

Proof. There exists a nonsingular matrix Q such that $(A; B)Q$ has the stated canonical structure. The effect of Q is undone upon reinterpreting γ as $Q^{-1}\gamma$. •

The canonical structure on $(A; B)$ often arises naturally ; therefore it is rarely necessary to determine the transformation matrix Q. In most applications, $\mathcal{A}_2 = \mathbf{0}$ but this does not necessarily entail further simplifications of the results reported in this Chapter. Also note that the theoretical results developed in the previous Chapter are unaffected by nonsingular transformations of the diffuse parameter γ. Henceforth the DSSM will always be anchored with,

$$\alpha_0 = a + (\mathcal{A}_1, \mathcal{A}_2)(\gamma_1; \gamma_2) \qquad \text{and} \qquad \beta = b + (\mathbf{0}, \mathcal{B})(\gamma_1; \gamma_2)$$

Consequently, γ_1 and γ_2 can then be viewed as the diffuse parameters associated respectively with the initial state and the regression parameter. The next two sections discuss two ramifications of partitioning γ in such a fashion.

5.2 Switching from the DKF to the KF

De Jong (1991a) rewrites the DSSM as,

$$y_t = X_t b + (Z_t, X_t B)\begin{pmatrix}\alpha_t \\ \gamma\end{pmatrix} + G_t u_t \tag{5.1}$$

$$\begin{pmatrix}\alpha_{t+1} \\ \gamma\end{pmatrix} = \begin{pmatrix}W_t \\ \mathbf{0}\end{pmatrix} b + \begin{pmatrix}T_t & W_t B \\ \mathbf{0} & I\end{pmatrix}\begin{pmatrix}\alpha_t \\ \gamma\end{pmatrix} + \begin{pmatrix}H_t \\ \mathbf{0}\end{pmatrix} u_t \tag{5.2}$$

This is a SSM with states augmented by the diffuse parameter γ. Clearly if $(\hat{\alpha}_t; \hat{\gamma}_t)$ is available and corresponds to the limiting predictor of $(\alpha_t; \gamma)$ using $(y_1, \ldots, y_{t-1})$, then subsequent iterations of the KF (applied to (5.1)-(5.2)) yield $(\hat{\alpha}_r; \hat{\gamma}_r)$, $r > t$ and the associated error covariance matrix.

Since $B = (\mathbf{0}, \mathcal{B})$ where the zero matrix consists of $\gamma_1^{\#}$ columns, it follows that the first $\gamma_1^{\#}$ columns of both $X_t B$ and $W_t B$ only contain zero entries. Consequently γ_1, apart from being captured in the initial state α_1, does not figure in either y_t or α_t and hence it need not be accomodated in the augmented state in (5.2). In essence then, the omission of γ_1 from the augmented state has no repercussion with regards to likelihood evaluation or prediction in the SSM. This observation leads us to the definition of the *Augmented* KF (AKF).

Definition 5.1 The AKF is the KF applied to the SSM,

$$y_t = X_t b + \mathcal{Z}_t \delta_t + G_t u_t, \qquad \delta_{t+1} = \mathcal{W}_t b + \mathcal{T}_t \delta_t + \mathcal{H}_t u_t \qquad \text{where} \tag{5.3}$$

$$\mathcal{Z}_t = (Z_t, X_t \mathcal{B}),\ \delta_t = \begin{pmatrix}\alpha_t \\ \gamma_2\end{pmatrix},\ \mathcal{W}_t = \begin{pmatrix}W_t \\ \mathbf{0}\end{pmatrix},\ \mathcal{T}_t = \begin{pmatrix}T_t & W_t \mathcal{B} \\ \mathbf{0} & I\end{pmatrix} \text{ and } \mathcal{H}_t = \begin{pmatrix}H_t \\ \mathbf{0}\end{pmatrix}.$$

Remarks

1. The AKF coincides with the alternative KF-based algorithms listed in the preamble since these are based on the ASSM wherein the regression parameter is incorporated within the state.

2. The AKF coincides with the KF when β is null.

3. When $\gamma_1 \neq \mathbf{0}$, it ensues that the state in (5.3) has $\gamma_1^{\#}$ less components than the state in (5.2). Consequently the AKF will outperform the KF based on (5.1)-(5.2). The computational savings can be appreciable, as for instance with monthly seasonal data when $\gamma_1^{\#} \geq 11$.

4. The AKF does not update the estimate of γ_1. However $Pred(\gamma_1|y_1;\ldots;y_n)$ is sometimes required for smoothing purposes. The reconstruction of this estimator is dealt with in the proof of Theorem 5.2.

We now turn to the problem of the appropriate initialization of the AKF. This requires the construction of the limiting predictor of the augmented state and its associated error covariance matrix at the point of collapse.

Theorem 5.1 *Suppose y is generated by the DSSM. Apply the DKF and partition,*

$$A_m = (A_{m\gamma}, a_m) \qquad \textit{and} \qquad Q_m = \begin{pmatrix} S_m & s_m \\ s'_m & q_m \end{pmatrix}$$

Suppose S_m is nonsingular and for $t \geq m$, replace $B = (\mathbf{0}, \mathcal{B})$ by $\mathcal{B}$ and run the AKF initialized with

$$\hat{\delta}_m = \begin{pmatrix} \hat{\alpha}_m \\ \hat{\gamma}_{m,2} \end{pmatrix} = \begin{pmatrix} a_m \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} A_{m\gamma} \\ \{\mathbf{0}, -I_{\gamma_2^{\#}}\} \end{pmatrix} S_m^{-1} s_m \qquad \textit{and}$$

$$\mathcal{P}_m = \sigma^{-2}Mse(\hat{\delta}_m) = \begin{pmatrix} P_m + A_{m\gamma}S_m^{-1}A'_{m\gamma} & -A_{m\gamma}S_m^{*2} \\ -S_m^{*2\prime}A'_{m\gamma} & S_m^{22} \end{pmatrix}$$

where S_m^{*2} *denotes the last* $\gamma_2^{\#}$ *columns of* S_m^{-1} *and* $S_m^{22} = \sigma^{-2}Mse(\hat{\gamma}_{m,2})$ *is the bottom diagonal block of order* $\gamma_2^{\#}$ *of* S_m^{-1}. *Then subsequent recursions of the AKF yield the limiting predictor of* δ_t, $t > m$ *and its error covariance matrix.*

Furthermore if the recursion $q_{t+1}^* = q_t^* + e_t'D_t^{-1}e_t$ *where* e_t *is the innovation at time* t, $Cov(e_t) = \sigma^2 D_t$ *and* $q_m^* = q_m - s_m'S_m^{-1}s_m$ *is attached to the KF, then* $-2\times$ *the diffuse log-likelihood, apart from a constant, is given by,*

$$\lambda^d(y) = (y^{\#} - \gamma^{\#})\log\sigma^2 + \log|S_m| + \sum_{t=1}^{n}\log|D_t| + \sigma^{-2}q_{n+1}^*.$$

Proof. It follows from Theorem 4.6 that $\hat{\delta}_m$ is the limiting predictor of δ_m and that $\mathcal{P}_m = \sigma^{-2}Mse(\hat{\delta}_m)$. Therefore subsequent iterations of the AKF yield the limiting predictor of δ_t, $t > m$ and its error covariance matrix. The second result is obtained as follows. Let $C = \sigma^{-2}Cov(\gamma) \to \infty$. Then

$$\begin{aligned}
\lambda^d(y) &= \lambda(y) - \log|\sigma^2 C| \\
&= \{\lambda(y_1,\ldots,y_{m-1}) - \log|\sigma^2 C|\} + \lambda(y_m,\ldots,y_n|y_1,\ldots,y_{m-1}) \\
&\to [\{(y_1;\ldots;y_{m-1})^{\#} - \gamma^{\#}\}\log\sigma^2 + \log|S_m| + \sum_{t=1}^{m-1}\log|D_t| + \sigma^{-2}q_m^*] \\
&+ (y_m;\ldots;y_n)^{\#}\log\sigma^2 + \sum_{t=m}^{n}\log|D_t| + \sigma^{-2}\sum_{t=m}^{n}e_t'D_t^{-1}e_t \\
&= (y^{\#} - \gamma^{\#})\log\sigma^2 + \log|S_m| + \sum_{t=1}^{n}\log|D_t| + \sigma^{-2}q_{n+1}^*
\end{aligned}$$

This completes the proof of the Theorem. •

With scalar observations, the switch from the DKF to the AKF can take place at the earliest when $m = 1 + \gamma^{\#}$ in which case $q_m^* = 0$. This will be the case when each iteration of the DKF leads to the identification of a separate component of γ. This however is not the case in general. For instance if for $1 \leq t \leq \gamma^{\#}$, either of Z_t or X_t or W_t is equal to a zero matrix then more than $\gamma^{\#}$ iterations of the DKF will be required to identify γ and in this case $q_m^* \neq 0$. With vector observations, fewer than $\gamma^{\#}$ iterations may be needed to identify γ.

The AKF has three attractive features vis-a-vis the DKF. First, it automatically generates (i) limiting estimates of the states and the innovations and (ii) gls estimates of the regression parameter ; with the DKF, such estimates are only obtained after further computations. Second, it employs a scalar recursion for q_t as opposed to the matrix recursion for Q_t in the DKF. This implies that $y^{\#}\hat{\sigma}^2 = q_{n+1}^*$ can be read off from the AKF ; this compares with the DKF where extra computations are required to evaluate the same estimate. Third, it is numerically more stable than the DKF since it does not recurs S_t. The latter was shown to explode when the DKF was applied to nonstationary-noninvertible ARMA models. These advantages are however overriden by the fact that the high dimensionality of the augmented state implies time-consuming computations of crucial matrices in the AKF, in particular the computation of the state error covariance matrix $\sigma^2 P_t$. This fact will be highlighted during the discussion on the efficiencies of collapsing strategies in section 5.4.

The square root form of the AKF is as described in Theorem 2.3 except for the modified system matrices which are specified in (5.3) . The next Lemma describes the safe computation of its initializing quantities (i. e. $\hat{\delta}_m$, $\mathcal{P}_m^{1/2}$ and λ) using output generated by the square root DKF.

Lemma 5.2 *Suppose the square root form DKF described in section* 4.3.2 *is run until say* $t = m$ *when* $rank(Q) = \gamma_1^{\#} + \gamma_2^{\#}$. *Write (i)* $Q' = \{(\mathcal{Q}, w); (0, r)$ *where* $\mathcal{Q} = \{(\mathcal{Q}_{11}, \mathcal{Q}_{12}); (0, \mathcal{Q}_{22})\}$ *with* $\mathcal{Q}_{11}$ *and* $\mathcal{Q}_{22}$ *both square matrices with respective order* $\gamma_1^{\#}$ *and* $\gamma_2^{\#}$, *w is a vector and r is a scalar and (ii)* $A_m = (A_{m\gamma}, a_m) = (A_{m1}, A_{m2}, a_m)$ *where*

a_m is the final column of A_m while A_{m1} consists of the first $\gamma_1^\#$ columns of A_m. Then the reinitializations described in Theorem 5.1 *can be evaluated as follows :*

$$\begin{aligned}
\hat{\delta}_m &= \begin{pmatrix} \hat{\alpha}_m \\ \hat{\gamma}_{m,2} \end{pmatrix} = \begin{pmatrix} a_m \\ 0 \end{pmatrix} - \begin{pmatrix} A_{m\gamma} \\ \{0, -I_{\gamma_2^\#}\} \end{pmatrix} \mathcal{Q}^{-1} w \ , \\
\mathcal{P}_m^{1/2} &= \begin{pmatrix} (P_m^{1/2}, A_{m1}\mathcal{Q}_{11}^{-1})U & A_{m\gamma}\mathcal{S}^{*2} \\ 0 & \mathcal{Q}_{22}^{-1} \end{pmatrix}, \\
\sqrt{q_m^*} &= r \qquad \text{and} \qquad \lambda \leftarrow \lambda + \log|\mathcal{Q}|
\end{aligned}$$

where U is an orthogonal matrix, $(P_m^{1/2}, A_{m1}\mathcal{Q}_{11}^{-1})U$ has no trailing zero columns and $\mathcal{S}^{*2} = (-\mathcal{Q}_{11}^{-1}\mathcal{Q}_{12}\mathcal{Q}_{22}^{-1}; \mathcal{Q}_{22}^{-1})$.

Proof. With matrix Q' as stated, it follows that

$$\begin{pmatrix} S_m & s_m \\ s_m' & q_m \end{pmatrix} = QQ' = \begin{pmatrix} \mathcal{Q}' & 0 \\ w' & r \end{pmatrix} \begin{pmatrix} \mathcal{Q} & w \\ 0 & r \end{pmatrix} = \begin{pmatrix} \mathcal{Q}'\mathcal{Q} & \mathcal{Q}'w \\ w'\mathcal{Q} & w'w + r^2 \end{pmatrix}$$

Therefore $|S_m| = |\mathcal{Q}|^2$ and $q_m - s_m' S_m^{-1} s_m = r^2$ and these assert the reinitialization of λ and $\sqrt{q_m^*}$. The expression for $\hat{\delta}_m$ ensues upon noting that $S_m^{-1}s_m = (\mathcal{Q}'\mathcal{Q})^{-1}\mathcal{Q}'w = \mathcal{Q}^{-1}w$. Using the canonical expression for $\mathcal{Q}$ it follows that,

$$\begin{aligned}
S_m^{-1} &= \mathcal{Q}^{-1}\mathcal{Q}^{-1\prime} = \begin{pmatrix} \mathcal{Q}_{11}^{-1} & -\mathcal{Q}_{11}^{-1}\mathcal{Q}_{12}\mathcal{Q}_{22}^{-1} \\ 0 & \mathcal{Q}_{22}^{-1} \end{pmatrix} \begin{pmatrix} \mathcal{Q}_{11}^{-1\prime} & 0 \\ -\mathcal{Q}_{22}^{-1\prime}\mathcal{Q}_{12}\mathcal{Q}_{11}^{-1} & \mathcal{Q}_{22}^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \mathcal{Q}_{11}^{-1}\mathcal{Q}_{11}^{-1\prime} + \mathcal{Q}_{11}^{-1}\mathcal{Q}_{12}\mathcal{Q}_{22}^{-1}\mathcal{Q}_{22}^{-1\prime}\mathcal{Q}_{12}'\mathcal{Q}_{11}^{-1\prime} & -\mathcal{Q}_{11}^{-1}\mathcal{Q}_{12}\mathcal{Q}_{22}^{-1}\mathcal{Q}_{22}^{-1\prime} \\ -\mathcal{Q}_{22}^{-1}\mathcal{Q}_{22}^{-1\prime}\mathcal{Q}_{12}'\mathcal{Q}_{11}^{-1\prime} & \mathcal{Q}_{22}^{-1}\mathcal{Q}_{22}^{-1\prime} \end{pmatrix}
\end{aligned}$$

It is then a direct task, using this expression, to verify that $\mathcal{P}_m^{1/2}$ when post-multiplied by its transpose equals $\mathcal{P}_m$ as specified in Theorem 5.1. •

It is clear that for $t \geq m$, the smoothing algorithm associated with the KF (see Theorem 2.2) can also be used in conjunction with the AKF provided that the pertinent matrices are appropriately redefined. This smoothing procedure is however not efficient since it entails the redundant smoothing of γ_2. This is explained by the fact that γ_2 can be viewed as a (diffuse) regression parameter and hence its smoothed estimate coincides with the final estimate provided by the AKF. An extra drawback arises when $t < m$ since $\hat{\alpha}_t$ is not available then. Thus for $t < m$, we need to revert to a diffuse smoothing algorithm. The following Theorem specifies the necessary adjustments for extending the smoothing algorithm associated with the AKF to a diffuse smoothing algorithm.

Theorem 5.2 *Suppose the DKF is switched to the AKF at $t = m$ and for $t \geq m$, run the smoothing algorithm described in Theorem 2.2 using AKF-derived data. Partition $\eta_{m-1} = (n_1; n_2)$ and $R_{m-1} = \{(R^{11}, R^{12}); (R^{21}, R^{22})\}$, with n_1 and $R^{11} = \sigma^{-2}Cov(n_1)$ of the same order as the unaugmented state α_m. For $t < m$, carry the modified recursions*

$$N_{t-1} \;=\; Z_t'D_t^{-1}E_t + L_t'N_t \qquad \text{and} \qquad \mathcal{R}_{t-1} = Z_t'D_t^{-1}Z_t + L_t'\mathcal{R}_tL_t$$

where $N_{m-1} = (\mathbf{0}, n_1)$ with matrix $\mathbf{0}$ having $\gamma_1^\# + \gamma_2^\#$ columns and $\mathcal{R}_{m-1} = R^{11}$. Then for $1 \leq t \leq r \leq m$, $\tilde{\alpha}_t$ and $\sigma^{-2}Mse(\tilde{\alpha}_t, \tilde{\alpha}_r)$ are respectively given by,

$$F_t(w;1) \text{ and } P_tL_{r-1,t}'(I - \mathcal{R}_{r-1}P_r) + F_{t\gamma}WF_{r\gamma}' - P_tL_{m,t}'JF_{r\gamma}' - F_{t\gamma}J'L_{m,r}P_r$$

where $F_t = A_t + P_tN_{t-1}$, $w = H\eta_{m-1} - S_m^{-1}s_m$, $H = S_m^{-1}\{A_{m\gamma}', (\mathbf{0}; -I)\}$ with matrix $\mathbf{0}$ having dimensions $\gamma_1^\# \times \gamma_2^\#$, $W = \sigma^{-2}Mse(w) = S_m^{-1} - HR_{m-1}H'$ and $J = (R^{11}, R^{12})H'$.

The proof of this Theorem (and Theorems 5.4 and 6.1) requires the following result.

Lemma 5.3 *Consider the KF recursion given in Theorem 2.1. Then for $t \geq 1$,*

$$e_t \;=\; Z_t(\alpha_t - \hat{\alpha}_t) + G_tu_t \qquad \text{and} \qquad \alpha_{t+1} - \hat{\alpha}_{t+1} = L_t(\alpha_t - \hat{\alpha}_t) + J_tu_t$$

where $L_t = T_t - K_tZ_t$ and $J_t = H_t - K_tG_t$.

Proof. For the first result, observe that

$$\begin{aligned} e_t &= y_t - X_t\beta - Z_t\hat{\alpha}_t \\ &= (X_t\beta + Z_t\alpha_t + G_tu_t) - X_t\beta - Z_t\hat{\alpha}_t \\ &= Z_t(\alpha_t - \hat{\alpha}_t) + G_tu_t \end{aligned}$$

This result is now used to prove the second result :

$$\begin{aligned} \alpha_{t+1} - \hat{\alpha}_{t+1} &= (W_t\beta + T_t\alpha_t + H_tu_t) - (W_t\beta + T_t\hat{\alpha}_t + K_te_t) \\ &= T_t(\alpha_t - \hat{\alpha}_t) + H_tu_t - K_t\{Z_t(\alpha_t - \hat{\alpha}_t) + G_tu_t\} \\ &= (T_t - K_tZ_t)(\alpha_t - \hat{\alpha}_t) + (H_t - K_tG_t)u_t \\ &= L_t(\alpha_t - \hat{\alpha}_t) + J_tu_t \end{aligned}$$

This concludes the proof of the Lemma. •

Remark

The results of Lemma 5.3 still hold when the AKF is applied to the SSM defined by (5.3) provided that α_t is interpreted as the augmented state $\delta_t = (\alpha_t; \gamma_2)$ and all the matrices are appropriately redefined.

Proof of Theorem 5.2. We first show that $-w = \hat{\gamma}_{n+1} = Pred(\gamma|y_1; \ldots; y_n)$. For $t \geq m$, let $\delta_t = (\alpha_t; \gamma_2)$, $\mathcal{Z}_t = (Z_t, X_t\mathcal{B})$, $\mathcal{L}_t = \mathcal{T}_t - \mathcal{K}_t\mathcal{Z}_t$ where $\mathcal{T}_t$ is the transition matrix in (5.4) and $\mathcal{K}_t$ is Kalman gain matrix generated by the AKF at time t. Then

$$\begin{aligned} Pred(\gamma|y_1; \ldots; y_n) &= Pred(\gamma|y_1; \ldots; y_{m-1}; e_m; \ldots; e_n) \\ &= Pred(\gamma|y_1; \ldots; y_{m-1}) + \sigma^{-2}\sum_{t=m}^{n} Cov(\gamma, e_t)D_t^{-1}e_t \\ &= S_m^{-1}s_m + \sigma^{-2}\sum_{t=m}^{n} Cov(\gamma, \delta_t - \hat{\delta}_t)\mathcal{Z}_t'D_t^{-1}e_t \\ &= S_m^{-1}s_m + \sigma^{-2}Cov(\gamma, \delta_m - \hat{\delta}_m)\sum_{t=m}^{n} \mathcal{L}_{t-1,m}'\mathcal{Z}_t'D_t^{-1}e_t \end{aligned}$$

$$
\begin{aligned}
&= S_m^{-1}s_m + \sigma^{-2}Cov\{\gamma, -H'S_m(\gamma - \hat{\gamma}_m)\}\eta_{m-1} \\
&= S_m^{-1}s_m - \sigma^{-2}Mse(\hat{\gamma}_m)S_mH\eta_{m-1} \\
&= S_m^{-1}s_m - H\eta_{m-1}
\end{aligned}
$$

where $\mathcal{L}'_{t-1,m} = \prod_{j=t-1}^{m} \mathcal{L}'_j$ with $\mathcal{L}_{m-1,m} = I$. The second equality follows since the stack of observations $(y_1; \ldots; y_{m-1})$ is uncorrelated with $(e_m; \ldots; e_n)$, which is the stack of innovations produced by the AKF for $t \geq m$. The third and fourth equality uses Lemma 5.3 and the fact that γ and u_t are uncorrelated for all t. In the fifth equality, the expression $\hat{\delta}_m$ follows from Theorem 5.1. Furthermore repeated backsubstitutions of η_t as defined in Theorem 2.2 shows that η_{m-1} is as asserted.

We now show that $Mse(\gamma|y) = \sigma^2 W$:

$$
\begin{aligned}
Mse(\gamma|y) &= Cov\{\gamma, \gamma - Pred(\gamma|y)\} \\
&= Cov\{\gamma, \gamma - Pred(\gamma|y_1; \ldots; y_{m-1})\} + Cov(\gamma, \eta_{m-1})H' \\
&= Mse(\gamma|y_1; \ldots; y_{m-1}) + Cov(S_m^{-1}s_m - H\eta_{m-1}, \eta_{m-1})H' \\
&= \sigma^2(S_m^{-1} - HR_{m-1}H') = \sigma^2 W
\end{aligned}
$$

The above expression for $\hat{\gamma}_{n+1}$ implies that,

$$
\begin{aligned}
\tilde{\alpha}_t &= (F_{t\gamma}, f_t)(-\hat{\gamma}_{n+1}; 1) \;=\; (F_{t\gamma}, f_t)(H\eta_{m-1} - S_m^{-1}s_m; 1) \\
&= \{F_{t\gamma}(-S_m^{-1}s_m) + f_t\} + F_{t\gamma}H\eta_{m-1} \\
&= \tilde{\alpha}_{t|m} + F_{t\gamma}H\eta_{m-1}
\end{aligned}
$$

where $F_t = (F_{t\gamma}, f_t)$ and $\tilde{\alpha}_{t|m} = Pred(\alpha_t|y_1; \ldots; y_m)$.

Let $P_{i,j} = P_iL'_{j,i}$. Then for $1 \leq t \leq r \leq m$,

$$
\begin{aligned}
Mse(\tilde{\alpha}_t, \tilde{\alpha}_r) &= Cov(\alpha_t - \tilde{\alpha}_t, \alpha_r - \tilde{\alpha}_r) \\
&= Cov(\alpha_t - \tilde{\alpha}_{t|m} - F_{t\gamma}H\eta_{m-1}, \alpha_r - \tilde{\alpha}_{r|m} - F_{r\gamma}H\eta_{m-1})
\end{aligned}
$$

$$
\begin{aligned}
&= Mse(\tilde{\alpha}_{t|m}, \tilde{\alpha}_{r|m}) + Cov(F_{t\gamma}H\eta_{m-1}, F_{r\gamma}H\eta_{m-1}) \\
&- Cov(\alpha_t - \tilde{\alpha}_t + F_{t\gamma}H\eta_{m-1}, F_{r\gamma}H\eta_{m-1}) \\
&- Cov(F_{t\gamma}H\eta_{m-1}, \alpha_r - \tilde{\alpha}_r + F_{r\gamma}H\eta_{m-1}) \\
&= Mse(\tilde{\alpha}_{t|m}, \tilde{\alpha}_{r|m}) - Cov(F_{t\gamma}H\eta_{m-1}, F_{r\gamma}H\eta_{m-1}) \\
&- Cov(\alpha_t - \tilde{\alpha}_t, F_{r\gamma}H\eta_{m-1}) - Cov(F_{t\gamma}H\eta_{m-1}, \alpha_r - \tilde{\alpha}_r) \\
&= \{P_{t,r-1}(I - \mathcal{R}^*_{r-1}P_r) + F_{t\gamma}S_m^{-1}F'_{r\gamma} - P_{t,m}R^{11}P'_{r,m}\} - F_{t\gamma}HR^{-1}_{m-1}H'F'_{r\gamma} \\
&- Cov(P_{t,m}n_1, \eta_{m-1})H'F'_{r\gamma} - F_{t\gamma}H\ Cov(\eta_{m-1}, n_1)P'_{r,m} \\
&= P_{t,r-1}\{I - (\mathcal{R}_{r-1} - L'_{m,r}R^{11}L_{m,r})P_r\} - P_{t,m}R^{11}P'_{r,m} \\
&+ F_{t\gamma}WF'_{r\gamma} - P_{t,m}(R^{11}, R^{12})H'F'_{r\gamma} - F_{t\gamma}H(R^{11}; R^{21})P'_{r,m} \\
&= P_{t,r-1}(I - \mathcal{R}_{r-1}P_r) + P_{t,m}R^{11}P'_{r,m} - P_{t,m}R^{11}P'_{r,m} \\
&+ F_{t\gamma}WF'_{r\gamma} - P_{t,m}JF'_{r\gamma} - F_{t\gamma}J'P'_{r,m} \\
&= P_{t,r-1}(I - \mathcal{R}_{r-1}P_r) + F_{t\gamma}WF'_{r\gamma} - P_{t,m}JF'_{r\gamma} - F_{t\gamma}J'P'_{r,m}
\end{aligned}
$$

where the expression for $Mse(\tilde{\alpha}_{t|m}, \tilde{\alpha}_{r|m})$ is obtained upon noting that

$$
\tilde{\alpha}_{t|m} \;=\; (A_t + P_tN^*_{t-1})(-\hat{\gamma}_m; 1) \;=\; (A_t + P_tN^*_{t-1})(-S_m^{-1}s_m; 1)
$$

with $N^*_{m-1} = \mathbf{0}$ and hence $\mathcal{R}^*_t$ follows the same recursion as $\mathcal{R}_t$ except that $\mathcal{R}^*_{m-1} = \mathbf{0}$. This concludes the proof of the Theorem. •

Remarks

1. For $1 \le t \le r \le m$, the expression for $Mse(\tilde{\alpha}_t, \tilde{\alpha}_r)$ is, except for a couple of adjustment terms induced by the switch from the smoothing algorithm associated with the KF to the diffuse smoothing algorithm, equal to the mse expression given in Theorem 4.6 for an uncollapsed DKF.

2. For $1 \le t < m < r \le n+1$, $Mse(\tilde{\alpha}_t, \tilde{\alpha}_r)$ can be derived in an analogous fashion as

in Theorem 5.2. This is however not required in this thesis since we only use lag zero and lag one mse matrices of predictors of the states.

With regards to square-root smoothing, observe that the square-root smoothing algorithm associated with the KF (see section 2.4.6) can be employed using the AKF-derived data in the post-switch time period (i. e. $t \geq m$). In the pre-switch time period however, square-root propagation of the lag zero mse matrices of the states does not appear to be possible in light of the adjustments arising upon the switch from the AKF to the DKF. Ansley and Kohn (1990) make the same comments with regards to the smoothing algorithm associated with the AKKF.

The final section of this Chapter makes clear that the switch from the DKF to the AKF is not advantageous save for the case of a null β. This conclusion is explained by the fact that the AKF is based on a SSM with augmented states and as we have previously noted in Chapter 2, it is the dimensions of the states which determine the performances of filtering and smoothing algorithms. The next section discusses a collapse of the DKF. This actually reduces the dimensions of pertinent matrices in the DKF and hence it is of major computational interest.

5.3 The Collapsed DKF

The preceding section suggests the idea of partialling out the effect of γ_1 in the DKF itself after an initial run of the latter. Specifically this entails the partialling out of those columns and rows of E_t, A_t and Q_t which relate to γ_1. This modified DKF, which we label the *collapsed* DKF (CDKF), can be viewed as an analogue of the AKF since the only diffuse effects that it estimates are those associated with the regression parameter. We now state the main result of this section.

Theorem 5.3 *Suppose the DKF is applied to observations y generated by the DSSM until $t = m$. Partition A_m and Q_m conformably with γ_1 and γ_2,*

$$A_m = (A_{m1}, A_{m2}, a_m) \qquad \textit{and} \qquad Q_m = \begin{pmatrix} S_m & s_m \\ s'_m & q_m \end{pmatrix} = \left(\begin{array}{cc|c} S_{m11} & S_{m12} & s_{m1} \\ S_{m21} & S_{m22} & s_{m2} \\ \hline s'_{m1} & s'_{m2} & q_m \end{array}\right)$$

and suppose that S_{m11} is nonsingular. For $t \geq m$, replace $B = (0, \mathcal{B})$ by $\mathcal{B}$ where the latter is defined in Lemma 5.1 *and reinitialize A_m, P_m and Q_m as follows :*

$$\mathcal{A}_m = (A_{m2} - A_{m1}S_{m11}^{-1}S_{m12}, a_m - A_{m1}S_{m11}^{-1}s_{m1}), \qquad \mathcal{P}_m = P_m + A_{m1}S_{m11}^{-1}A'_{m1} \qquad \textit{and}$$

$$\mathcal{Q}_m = \begin{pmatrix} S_{m22.1} & S_{m21}S_{m11}^{-1}s_{m1} - s_{m2} \\ s'_{m1}S_{m11}^{-1}S_{m12} - s'_{m2} & q_m - s'_{m1}S_{m11}^{-1}s_{m1} \end{pmatrix}$$

where $S_{m22.1} = S_{m22} - S_{m21}S_{m11}^{-1}S_{m12}$.
Then this collapsed DKF can be employed in lieu of the standard DKF for limiting prediction of the states and gls estimation of β. Furthermore $-2 \times$ the diffuse log-likelihood, apart from a constant, is

$$\begin{aligned} \lambda^d(y) &= (y^{\#} - \gamma^{\#})\log\sigma^2 + \sum_{t=1}^{n}\log|D_t| + \log|S_{m11}| + \log|S_{n+1}| \\ &+ \sigma^{-2}(q_{n+1} - s'_{n+1}S_{n+1}^{-1}s_{n+1}). \end{aligned}$$

Proof. Without any loss in generality, assume $\gamma = (\gamma_1; \gamma_2) \sim N\{0, diag\ \sigma^2(C_1, C_2)\}$. Let $C_1 \to \infty$. Then from Lemma 4.1,

$$\begin{aligned} Pred(\gamma|y_1; \ldots; y_{m-1}) &= \{S_m + diag(0, C_2^{-1})\}^{-1}s_m \qquad \text{and} \\ \sigma^{-2}Mse(\gamma|y_1; \ldots; y_{m-1}) &= \{S_m + diag(0, C_2^{-1})\}^{-1} \end{aligned}$$

Observe that $\{S_m + diag(0, C_2^{-1})\}^{-1}$ can be written as,

$$\left\{\begin{pmatrix} I & 0 \\ S_{m21}S_{m11}^{-1} & I \end{pmatrix}\begin{pmatrix} S_{m11} & 0 \\ 0 & S_{m22.1} + C_2^{-1} \end{pmatrix}\begin{pmatrix} I & S_{m11}^{-1}S_{m12} \\ 0 & I \end{pmatrix}\right\}^{-1}$$

$$\begin{aligned} &= \begin{pmatrix} I & -S_{m11}^{-1}S_{m12} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} S_{m11}^{-1} & \mathbf{0} \\ \mathbf{0} & \{S_{m22.1} + C_2^{-1}\}^{-1} \end{pmatrix} \begin{pmatrix} I & \mathbf{0} \\ -S_{m21}S_{m11}^{-1} & I \end{pmatrix} \\ &= UD^{-1}U' \end{aligned}$$

Let $\Omega = (S_{m22.1} + C_2^{-1})^{-1}$ and $\mu = \Omega(s_{m2} - S_{m21}S_{m11}^{-1}s_{m1})$. Then using the above identity, the predictor of α_m conditional on $(y_1; \ldots; y_{m-1})$ is

$$\begin{aligned} \hat{\alpha}_m &= a_m - (A_{m1}, A_{m2})UD^{-1}U'(s_{m1}; s_{m2}) \\ &= (A_{m2} - A_{m1}S_{m11}^{-1}S_{m12}, a_m - A_{m1}S_{m11}^{-1}s_{m1})(\mu; 1) \qquad \text{and} \\ \sigma^{-2}Mse(\hat{\alpha}_m) &= P_m + (A_{m1}, A_{m2})UD^{-1}U'(A_{m1}, A_{m2})' \\ &= P_m + A_{m1}S_{m11}^{-1}A'_{m1} + (A_{m2} - A_{m1}S_{m11}^{-1}S_{m12})\Omega(A_{m2} - A_{m1}S_{m11}^{-1}S_{m12})' \end{aligned}$$

Therefore $(y_m; \ldots; y_n)$ can be envisaged as being generated by a DSSM anchored with,

$$\alpha_m = (a_m - A_{m1}S_{m11}^{-1}s_{m1}) + (A_{m2} - A_{m1}S_{m11}^{-1}S_{m12})\omega \qquad \text{where } \omega \sim \{\mu, \sigma^2\Omega)$$

Thus the DKF, reinitialized as stated in the Theorem, yields the same limiting predictors as the standard DKF.

Next, express the diffuse log-likelihood as

$$\lambda^d(y) = \lambda(y_1; \ldots; y_{m-1}) + \lambda(y_m; \ldots; y_n | y_1; \ldots; y_{m-1}) - \log|\sigma^2 C_1| - \log|\sigma^2 C_2|$$

The two likelihoods on the right can be evaluated by appealing to Theorem 4.1 and Definition 4.4. These assert that the log-likelihood of a diffuse SSM with the diffuse parameter $\gamma \sim (c, \sigma^2 C)$ is given by,

$$\begin{aligned} \lambda(y) &= (y^{\#} - \gamma^{\#})\log\sigma^2 + \log|\sigma^2 C| + \sum_{t=1}^{n}\log|D_t| \\ &+ \log|C^{-1} + S| + \sigma^{-2}\{q + c'C^{-1}c - (s + C^{-1}c)'(C^{-1} + S)^{-1}(s + C^{-1}c)\} \end{aligned}$$

Recall $\gamma_1 \sim (\mathbf{0}, \sigma^2 C_1)$. Thus as $C_1 \to \infty$,

$$\begin{aligned}
\lambda(y_1;\ldots;y_{m-1}) - \log|\sigma^2 C_1| &= \{(y_1;\ldots;y_{m-1})^{\#} - \gamma_1^{\#}\}\log\sigma^2 + \sum_{t=1}^{m-1}\log|D_t| \\
&+ \log|S_m + diag(\mathbf{0}, C_2^{-1})| + \log|\sigma^2 C_2| \\
&+ [q_m - s_m'\{S_m + diag(\mathbf{0}, C_2^{-1})\}^{-1}s_m]/\sigma^2 \\
&= \{(y_1;\ldots;y_{m-1})^{\#} - \gamma_1^{\#}\}\log\sigma^2 + \sum_{t=1}^{m-1}\log|D_t| \\
&+ \log|S_{m11}| - \log|\Omega| + \log|\sigma^2 C_2| \\
&+ \sigma^{-2}(q_m - s_{m1}'S_{m11}^{-1}s_{m1} - \mu'\Omega^{-1}\mu)
\end{aligned}$$

Furthermore $\lambda(y_m;\ldots;y_n|y_1;\ldots;y_{m-1})$ equals

$$\begin{aligned}
&(y_m;\ldots;y_n)^{\#}\log\sigma^2 + \log|\Omega| + \sum_{t=m}^{n}\log|D_t| + \log|\Omega^{-1} + S_{n+1}^{\dagger}| \\
+ \;& \sigma^{-2}\{q_{n+1}^{\dagger} + \mu'\Omega^{-1}\mu - (s_{n+1}^{\dagger} + \Omega^{-1}\mu)'(\Omega^{-1} + S_{n+1}^{\dagger})^{-1}(s_{n+1}^{\dagger} + \Omega^{-1}\mu)\}
\end{aligned}$$

where $Q_t^{\dagger} = \{(S_t^{\dagger}, s_t^{\dagger}); (s_t^{\dagger'}, q_t^{\dagger})\}$ follows the same recursion as $\mathcal{Q}_t$ except that it is initialized with $Q_m^{\dagger} = \mathbf{0}$. Now let $C_2 \to \infty$. Then in view of the initialization of $\mathcal{Q}_m$, it is a simple task to ascertain the stated expression for $\lambda^d(y)$. This concludes the proof of the Theorem.

•

Remarks

1. With scalar observations, the earliest time that the DKF can be collapsed to the CDKF is when $m = 1 + \gamma_1^{\#}$, in which case $\mathcal{Q}_m = \mathbf{0}$. This contrasts to the switch from the DKF to the AKF which can take place at $m = 1 + \gamma^{\#}$ at the earliest.

2. When β is null, both the CDKF and the AKF coincide with the KF.

3. The matrix U performs a sweep operation : it factors out the effect of γ_1 in the DKF. As such, the matrices $\mathcal{A}_t$, $\mathcal{E}_t$ and $\mathcal{Q}_t$ generated by the CDKF respectively correspond to the γ_1-partialled-out versions of A_t, E_t and Q_t in the DKF.

4. The matrix $\mathcal{A}_1$ defined in Lemma 5.1 can be in row-echelon form. Then using the ideas in Theorem 5.3, it is possible to implement a progressive collapse of the DKF. This amounts to partialling out the columns related to those distinct elements of γ_1 as soon as the latters are identified. Ansley and Kohn (1990) employ this strategy in turning their AKKF to the AKF. However in our opinion, progressive collapsing is not an attractive proposition in view of the intricate bookeeping that is required when smoothing (see Theorems 5.2 and 5.4) is called for.

We now consider the square root form of the CDKF. It follows the line of the square root form of the DKF described in section 3.2.3 except for the modification of the pertinent matrices for $t \geq m$. The following Lemma indicates the safe computations of the reinitialized quantities.

Lemma 5.4 *Suppose the square root form DKF described in section* 4.3.2 *is run for* $t < m$ *where* m *is as described in Theorem* 5.2. *Write*

$$Q' = \begin{pmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} & w_1 \\ \mathbf{0} & \mathcal{Q}_{22} & w_2 \\ \mathbf{0} & \mathbf{0} & r \end{pmatrix}$$

where for $i = 1, 2$, *square matrix* $\mathcal{Q}_{ii}$ *and vector* w_i *both conform with* γ_i *and* r *is a scalar. Then the reinitializations described in Theorem* 5.3 *can be obtained as follows :*

$$\mathcal{A}_m = (A_{m2} - A_{m1}\mathcal{Q}_{11}^{-1}\mathcal{Q}_{12}, a_m - A_{m1}\mathcal{Q}_{11}^{-1}w_1) \ , \qquad \mathcal{P}_m^{1/2} = (P_m^{1/2}, A_{m1}\mathcal{Q}_{11}^{-1})U \ ,$$

$$\mathcal{Q}_m^{1/2'} = \begin{pmatrix} \mathcal{Q}_{22} & w_2 \\ \mathbf{0} & r \end{pmatrix} \qquad \textit{and} \qquad \lambda = \lambda + \log|\mathcal{Q}_{11}|$$

where U *is any orthogonal matrix and* $\mathcal{P}_m^{1/2}$ *is without any trailing zero columns.*

Proof. With Q' as specified in the Lemma, we obtain

$$\begin{aligned}\begin{pmatrix} S_{m11} & S_{m12} & s_{m1} \\ S_{m21} & S_{m22} & s_{m_2} \\ s'_{m1} & s'_{m2} & q \end{pmatrix} &= QQ' = \begin{pmatrix} Q'_{11} & \mathbf{0} & \mathbf{0} \\ Q'_{12} & Q'_{22} & \mathbf{0} \\ w'_1 & w'_2 & r \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} & w_1 \\ \mathbf{0} & Q_{22} & w_2 \\ \mathbf{0} & \mathbf{0} & r \end{pmatrix} \\ &= \begin{pmatrix} Q'_{11}Q_{11} & Q'_{11}Q_{12} & Q'_{11}w_1 \\ Q'_{12}Q_{11} & Q'_{12}Q_{12} + Q'_{22}Q_{22} & Q'_{12}w_1 + Q'_{22}w_2 \\ w'_1Q_{11} & w'_1Q_{12} + w'_2Q_{22} & w'_1w_1 + w'_2w_2 + r^2 \end{pmatrix}\end{aligned}$$

Therefore, $S_{m11}^{1/2} = Q'_{11}$ and

$$\begin{aligned} S_{m11}^{-1}S_{m12} &= (Q'_{11}Q_{11})^{-1}Q'_{11}Q_{12} = Q_{11}^{-1}Q_{12}, \\ S_{m11}^{-1}s_{m1} &= (Q'_{11}Q_{11})^{-1}Q'_{11}w_1 = Q_{11})^{-1}w_1, \\ S_{m22.1} &= (Q'_{12}Q_{12} + Q'_{22}Q_{22}) - Q'_{12}Q_{11}(Q'_{11}Q_{11})^{-1}Q'_{11}Q_{12} = Q'_{22}Q_{22}, \\ S_{m21}S_{m11}^{-1}s_{m1} - s_{m2} &= (Q'_{12}Q_{11}^{-1'})Q'_{11}w_1 - (Q'_{12}w_1 + Q'_{22}w_2) = -Q'_{22}w_2, \\ q - s'_{m1}S_{m11}^{-1}s_{m1} &= (w'_1w_1 + w'_2w_2 + r^2) - w'_1Q_{11}((Q'_{11}Q_{11})^{-1}Q'_{11}w_1 = w'_2w'_2 + r^2 \end{aligned}$$

We then obtain upon using these expressions, the asserted reinitializations of λ, $\mathcal{A}_m$, $\mathcal{P}^{1/2}$ and $Q^{1/2}$. •

It is clear that the collapse the DKF to the CDKF also permeates to the related smoothing algorithm. Note that the smoothing algorithm related to the CDKF makes use of $Pred(\gamma_2|y_1;\ldots;y_n)$. Hovever for $t < m$, this smoothing algorithm additionally requires $Pred(\gamma_1|y_1;\ldots;y_n)$. The next Theorem details this reconstruction as well as other necessary adjustments to this smoothing algorithm for the pre-collapse time period.

Theorem 5.4 *Suppose the DKF is collapsed to the CDKF at $t = m$. Put $N_n = \mathbf{0}$, a zero matrix with dimensions $\alpha_{n+1} \times (\gamma_2^{\#} + 1)$, and $R_n = \mathbf{0}$, a square matrix of order α_{n+1}. Iterate*

$$N_{t-1} = Z'_tD_t^{-1}E_t + L'_tN_t \qquad \textit{and} \qquad R_{t-1} = Z'_tD_t^{-1}Z_t + L'_tR_tL_t$$

except that for $t < m$, N_t is replaced by $\mathcal{N}_t$ with $\mathcal{N}_{m-1} = (\mathbf{0}, N_{m-1})$ where $\mathbf{0}$ consists of $\gamma_1^{\#}$ columns. Then for $1 \leq m \leq t \leq r$, $\tilde{\alpha}_t$ and $\sigma^{-2}Mse(\tilde{\alpha}_t, \tilde{\alpha}_r)$ are obtained as in Theorem 4.6 provided $\hat{\gamma}_{n+1}$ is replaced by $\hat{\gamma}_{n+1,2}$.
Furthermore for $1 \leq t \leq r \leq m$, $\tilde{\alpha}_t$ and $\sigma^{-2}Mse(\tilde{\alpha}_t, \tilde{\alpha}_r)$ are respectively,

$$F_t(w;1)\ ,\ P_tL'_{r-1,t}(I - R_{r-1}P_r) + F_{t,\gamma}WF'_{r,\gamma} - P_tL'_{m,t}JF'_{r1} - F_{t1}J'L_{m,r}P_r$$

where

$$\begin{aligned} w &= \begin{pmatrix} -\hat{\gamma}_{n+1,1} \\ -\hat{\gamma}_{n+1,2} \end{pmatrix} = \begin{pmatrix} Hn_{m-1} - S_{m11}^{-1}s_{m1} - G\hat{\gamma}_2 \\ -\hat{\gamma}_{n+1,2} \end{pmatrix}, \\ W &= \sigma^{-2}Mse(w) = \begin{pmatrix} S_{m11}^{-1} - HR_{m-1}H' + G\Gamma G' & G\Gamma \\ \Gamma G' & \Gamma \end{pmatrix}, \end{aligned}$$

F_{t1} and $F_{t,\gamma}$ respectively denote the first $\gamma_1^{\#}$ and the first $\gamma^{\#}$ columns of $F_t = A_t + P_t\mathcal{N}_{t-1}$, $H = S_{m11}^{-1}A'_{m1}$, $G = HN_{m-1,2} - S_{m11}^{-1}S_{m12}$ where $N_{m-1} = (N_{m-1,2}, n_{m-1})$ with n_{m-1} as its final column, $\Gamma = \sigma^{-2}Mse(\hat{\gamma}_{n+1,2})$ and $J = R_{m-1}H'$.

Proof. From Theorems 5.3 and 4.6, $Pred\{\alpha_m|(\gamma_1;\gamma_2)\}$ is given by either

$$(A_{m1}, A_{m2}, a_m)\begin{pmatrix} -\gamma_1 \\ -\gamma_2 \\ 1 \end{pmatrix} \qquad \text{or} \qquad (A_{m2} - A_{m1}S_{m11}^{-1}S_{m12}, a_m - A_{m1}S_{m11}^{-1}s_{m1})\begin{pmatrix} -\gamma_2 \\ 1 \end{pmatrix}$$

This implies after direct algebra that $\gamma_1 = S_{m11}^{-1}s_{m1} - S_{m11}^{-1}S_{m12}\gamma_2$. Hence

$$\begin{aligned} Pred(\gamma_1|y) &= Pred(S_{m11}^{-1}s_{m1}|y) - Pred(S_{m11}^{-1}S_{m12}\gamma_2|y) \\ &= Pred(S_{m11}^{-1}s_{m1}|y_1;\ldots;y_{m-1};e_m;\ldots;e_n) - S_{m11}^{-1}S_{m12}\hat{\gamma}_2 \\ &= S_{m11}^{-1}s_{m1} + \sigma^{-2}\sum_{t=m}^{n} Cov(S_{m11}^{-1}s_{m1}, \alpha_t - \hat{\alpha}_t)Z'_tD_t^{-1}e_t - S_{m11}^{-1}S_{m12}\hat{\gamma}_{n+1,2} \\ &= S_{m11}^{-1}(s_{m1} - S_{m12}\hat{\gamma}_{n+1,2}) + \sigma^{-2}Cov(S_{m11}^{-1}s_{m1}, \alpha_m - \hat{\alpha}_m)N_{m-1}(-\hat{\gamma}_{n+1,2};1) \end{aligned}$$

$$
\begin{aligned}
&= S_{m11}^{-1}(s_{m1} - S_{m12}\hat{\gamma}_{n+1,2}) \\
&+ \sigma^{-2}Cov(S_{m11}^{-1}s_{m1}, -A_{m1}S_{m11}^{-1}s_{m1})N_{m-1}(-\hat{\gamma}_{n+1,2};1) \\
&= S_{m11}^{-1}(s_{m1} - S_{m12}\hat{\gamma}_{n+1,2}) - S_{m11}^{-1}A'_{m1}N_{m-1}(-\hat{\gamma}_{n+1,2};1) \\
&= S_{m11}^{-1}s_{m1} + G\hat{\gamma}_{n+1,2} - Hn_{m-1}
\end{aligned}
$$

where $e_t = E_t(-\hat{\gamma}_{2,t};1)$ is the limiting innovation and $N_{m-1}(-\hat{\gamma}_{n+1,2};1)$ coincides with n_1 in Theorem 5.2. The third and fourth equalities as in the proof of Theorem 5.2, follow from Lemma 5.3. Therefore $w = Pred(-\gamma|y)$ can be written as,

$$
w = \begin{pmatrix} I & G \\ 0 & I \end{pmatrix} \begin{pmatrix} Hn_{m-1} - S_{m11}^{-1}s_{m1} \\ -\hat{\gamma}_{n+1,2} \end{pmatrix}
$$

Observe that $\sigma^{-2}Mse(H\eta - S_{m11}^{-1}s_{m1}) = S_{m11}^{-1} - HR_{m-1}H'$ (as in Theorem 5.2) and $H\eta - S_{m11}^{-1}s_{m1}$ is uncorrelated with $\hat{\gamma}_2$. Therefore

$$
\begin{aligned}
\sigma^{-2}Mse(\gamma|y) &= \begin{pmatrix} I & G \\ 0 & I \end{pmatrix} \begin{pmatrix} S_{m11}^{-1} - HR_{m-1}H' & 0 \\ 0 & \Gamma \end{pmatrix} \begin{pmatrix} I & 0 \\ G' & I \end{pmatrix} \\
&= \begin{pmatrix} S_{11}^{-1} - HR_{m-1}H' + G\Gamma G' & G\Gamma \\ \Gamma G' & \Gamma \end{pmatrix} = W
\end{aligned}
$$

Now observe that in both the pre-switch (from DKF to AKF) and the pre-collapse (from DKF to CDKF), we employ the same DKF quantities. Furthermore the adjustments to N_{m-1} in both Theorem 5.2 and the present Theorem coincide since N_{m-1} as described in this Theorem can always be written as $N_{m-1} = (\mathbf{0}, n_1)$ upon factoring out the effect of γ_2 and this in turn leads to a reinitialization $\mathcal{N}_{m-1} = (\mathbf{0}, n_1)$ as in Theorem 5.2. Thus the arguments developed in the proof of the cross state mse expressions in Theorem 5.2 apply *verbatim* here. This concludes the proof of the Theorem. •

A square-root form for the smoothing algorithm associated with the CDKF is similar to one associated with the DKF (section 4.3.2) except that it only applies for $t \geq m$, the post-collapse period. Due to the adjustments at $t = m$, square-root smoothing algorithm for $t \leq m$ appears intricate at best.

5.4 Efficiency of Collapsing Strategies

Ansley and Kohn (1990, p282) and Bell and Hillmer (1991, p284) have raised concerns about the computational efficiency of the DKF, specifically the fact that it employs recursions of matrices A_t and E_t in lieu of the vector recursions for $\hat{\alpha}_t$ and e_t in the AKF. However we counter through the observation that the efficiency of the DKF and the AKF is significantly more dependent on the dimensions of the mse matrices of the predictors of the states. Thus the AKF, since it is based on a SSM with augmented states, is not necessarily more computationally efficient than the DKF.

In the present section, we demonstrate that the CDKF, the collapsed version of the DKF which was derived in the last section, is computationally superior than both the DKF and the AKF-type algorithms (unless β is null in which case they all coincide with the KF) proposed by Ansley and Kohn (1990), Bell and Hillmer (1991) and Harvey and Pierse (1984). As discussed previously, these alternative algorithms are all based on the ASSM wherein the states are augmented to incorporate the regression parameter β. These algorithms therefore construct larger error covariance matrices than the CDKF and this time-consuming activity explains the computational superiority of the latter algorithm. The same explanation permeates to the associated smoothing algorithm. Additionally, the latter employs, among other quantities, the state error covariance matrices generated by the CDKF and therefore has less data storage requirements than the smoothing algorithm associated with the AKF.

A minor reason behind the computational superiority of the CDKF over the AKF is due to the fact that the switch from the DKF to the AKF generally takes place at a later

Strategy	A_t	P_t	Q_t
DKF	$\tau \times (\gamma_1^\# + \gamma_2^\# + 1)$	$\tau \times \tau$	$(\gamma_1^\# + \gamma_2^\# + 1) \times (\gamma_1^\# + \gamma_2^\# + 1)$
AKF	$(\tau + \gamma_2^\#) \times 1$	$(\tau + \gamma_2^\#) \times (\tau + \gamma_2^\#)$	1×1
CDKF	$\tau \times (\gamma_2^\# + 1)$	$\tau \times \tau$	$(\gamma_2^\# + 1) \times (\gamma_2^\# + 1)$

Table 5.1: Dimensionalities in filtering algorithms

stage than the collapse of the DKF to the CDKF. This follows since the switch can only occur when S_t is nonsingular while the collapse requires that only the topmost diagonal block of order $\gamma_1^\#$ of S_t be nonsingular. Therefore the use of the CDKF as opposed to the AKF implies gains in the areas of computational efficiency and data storage requirements when smoothing is called for.

The DKF, CDKF and AKF (and their associated smoothing algorithms) share the same recursions except that they employ matrices of different dimensions. Hence the difference in their computational performances is solely accountable to the dimensions of pertinent matrices in these algorithms. As stated in Chapter 2, the most time-consuming recursion in the KF, and by extension the DKF, CDKF and AKF, is the one concerning the state error covariance matrix i. e. $\sigma^2 P_t$. In the same vein, the smoothing algorithm (see Theorems 2.2 and 4.6) evaluates a covariance matrix R_t with the same dimension as P_t. The dimensions of these matrices, as well as other pertinent matrices, subsequent to the switch or collapse of the DKF are given in Table 5.1 for the filtering cycle and in Table 5.2 for the smoothing cycle. These tables assume that the transition matrix T_t is time-invariant with dimensions $\tau \times \tau$. For the time-varying case, the relative differences will be exactly the same.

Since the computations of P_t, Q_t and R_t are the most time-consuming in the filtering and smoothing algorithms especially when their square-root forms are employed, we can immediately infer that the CDKF should be the most efficient of these three strategies

Strategy	N_t	R_t
DKF	$\tau \times (\gamma_1^{\#} + \gamma_2^{\#} + 1)$	$\tau \times \tau$
AKF	$(\tau + \gamma_2^{\#}) \times 1$	$(\tau + \gamma_2^{\#}) \times (\tau + \gamma_2^{\#})$
CDKF	$\tau \times (\gamma_2^{\#} + 1)$	$\tau \times \tau$

Table 5.2: Dimensionalities in smoothing algorithms

for two reasons namely, (i) it is based on a SSM with states of minimal dimensionality and (ii) its associated smoothing algorithm require less data storage.

5.4.1 Numerical Illustration

In order that the DKF and the CDKF can be compared on an equal footing with the AKF, we must estimate the state, its error covariance matrix and also the gls estimate of β at each iteration of the algorithms. The same benchmark applies with regards to smoothing. For illustration purposes, we have chosen a regression model with $\gamma_2^{\#} = 0, 1, 2, 3$ regressors and a quarterly seasonal error term. This can be expressed as the following SSM,

$$y_t = X_t\beta + (1\ 0)\alpha_t, \qquad \alpha_{t+1} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} u_t$$

The transition matrix has nonstationary roots 1 and $\pm i$. Hence following the collapse stage, the quantities E_t, A_t and Q_t in the CDKF have $\gamma_1^{\#} = 3$ fewer columns than their counterparts in the DKF. The AKF is based on a SSM with augmented states $(\alpha_t; \beta)$ (i. e. the ASSM). Therefore the matrices P_t in the AKF and R_t in the associated smoothing algorithm consists of $\gamma_2^{\#}$ more rows and columns than their counterparts in the DKF and its associated smoothing algorithm.

The computations were carried out on an AT-type microcomputer using the square root forms of the DKF, AKF, and CDKF and their associated smoothing algorithms.

$\gamma_2^{\#}$	0	1	2	3
DKF	23.34	30.92	40.53	53.99
AKF	15.26	19.33	26.75	37.29
CDKF	16.04	19.28	23.07	28.95

Table 5.3: Run times (seconds) for state prediction, regression parameter estimation and likelihood evaluation

$\gamma_2^{\#}$	0	1	2	3
DKF	38.34	45.98	55.86	69.86
AKF	31.03	42.67	60.15	77.28
CDKF	31.46	35.15	38.50	43.55

Table 5.4: Run times (seconds) for smoothing

Recall that these square root algorithms are identical except that they employ matrices of different dimensions. Tables 5.3 and 5.4 display the run times of these square root algorithms when the same 20 randomly generated observations are employed to construct limiting estimates (filtered and smoothed) of the state and the related error covariance matrix and the gls estimate of the regression parameter at each iteration of these algorithms as well as evaluating the diffuse log-likelihood at the final iteration.

These run times confirm our assertion that the CDKF is computationally more efficient than its competitors in all facets of statistical inference with the DSSM. The results also clearly indicate that the inclusion of the regression parameters within the state leads to patently inefficient smoothing algorithms.

5.5 Summary

We have demonstrated that a properly implemented DKF, labelled the CDKF, is computationally superior to alternatives discussed in the literature. The next Chapter discusses maximum likelihood estimation of parameters in the SSM. This is conducted by embedding the DKF within the iterative EM algorithm. Chapter 7 deals with the recursive generation of residuals using the DKF. The results of this Chapter indicate that computational efficiency will be enhanced if the algorithms developed in these two Chapters employ the CDKF in lieu of the DKF.

Chapter 6

Maximum Likelihood Estimation in the State Space Model

In many applications of the SSM, the focus in on the estimation of the unknown parameters in its system matrices on account of their practical interpretation. In this Chapter, *maximum likelihood estimates* (mle's) are derived for these unknown parameters under the assumption that they are time invariant. The estimation method, labelled the *DKF-EM* method, consists of embedding the DKF within the EM (Expectation-Maximization) algorithm. The latter is a popular derivative-free likelihood optimization procedure.

The time series literature reports several applications employing an EM approach for the estimation of unknown parameters in the SSM : Harvey and Peters (1990) for the estimation of the error covariance matrices $\sigma^2 GG'$ and $\sigma^2 HH'$ in the *structural models*, Shumway and Stoffer (1981) for the estimation of a nonstationary scalar transition matrix T which is interpreted as an inflation rate and Watson and Engle (1983) for the estimation of the observation matrix Z with the components of the latter interpreted as the unobserved wage rates.

All these works employ a KF-EM estimation method with the KF initiated according to the big "k" method. Initializing the KF in such a fashion is, as argued in previous chapters, both theoretically and computationally unsound. The DKF is more appropriate in diffuse situations and therefore in this Chapter, we focus on maximum likelihood estimation in the SSM via the DKF-EM technology. The DKF (like the KF) serves two purposes : (i) evaluate the diffuse log-likelihood and (ii) provide the required data for the smoothing algorithm ; the latter is required by the EM algorithm. In view of the results of the previous chapter, it is computationally more efficient to employ the CDKF in lieu of the DKF. The results reported in this Chapter are obtained via the CDKF-EM estimation method.

This Chapter is organized along two main sections. Section 1 reviews the concepts behind the EM algorithm and discusses its merits relative to other likelihood maximizing procedures. A general CDKF-EM algorithm for estimating system matrices in the SSM is developed. It generalizes and unifies the works of Shumway and Stoffer (1981), Watson and Engle (1983) and Harvey and Peters (1990). The CDKF-EM estimation technology is illustrated via two financial applications. The first application requires the estimation of the exponential growth rate of a time series of quarterly earnings of a company. This is tantamount to estimating the growth coefficient of the trend in a trend-seasonal transition matrix. The second application employs the *Capital Asset Pricing Model* under the assumption that the market premium (equivalent to the state) follows a random walk to estimate the risk-free rate of return (equivalent to an unknown regression parameter) and the "betas" (equivalent to the observation matrix) of three stocks traded on the New York Stock Exchange (NYSE).

The majority of estimation applications in the SSM deal with the estimation of the covariance matrices of the disturbances in the SSM, namely $\sigma^2 GG'$ and $\sigma^2 HH'$. The second half of the Chapter focusses on this estimation problem. Following the judicious choice of the *complete data* (concept is explained in the next section), we investigate a new and computationally more efficient CDKF-EM estimation method which avoids the time-consuming computation of lag one state error covariance matrices. In a recent discussion paper, Koopman (1991) also suggests a similar estimation strategy. The section concludes with a tabulation of the results obtained on employing this novel version of the CDKF-EM algorithm to structural models previously illustrated by Harvey (1989), Harvey and Peters (1990) and West and Harrison (1989). Interestingly, we obtain solutions with higher log-likelihoods than previously found.

6.1 The EM approach

The EM algorithm is a popular tool for likelihood maximization in statistical models which either explicitly involve missing data or which can be formulated in terms of missing information. As such, it is often employed for maximum likelihood estimation purposes in a variety of applications. Instances of its applications can be traced quite far in statistical history. For example, Lauritzen (1981) reports a paper by Thiele (1880) who formulated a time series model consisting of the sum of a regression component, a Brownian motion and a white noise. Thiele employed a variant of the Kalman Filter to estimate the regression component and evaluated the variances of the Brownian motion and the white noise via an iterative process which is akin to the EM algorithm. A recent application of the EM algorithm deals with the reconstruction of data images in the field of positron emission tomography (PET) (Vardi *et al.*, 1985). Dempster, Laird and Rubin (1977) building on previous work by Orchard and Woodbury (1972) and Sundberg (1974) generalize and unify the theory behind the EM algorithm and demonstrate its usefulness to such applications as variance components estimation, hyperparameter estimation, finite mixture models and factor analysis.

The SSM provides a suitable framework for explaining the concepts behind the EM algorithm. In the SSM, we observe the *incomplete data* y_t as a function of an unobserved time series α_t. Call $(\{y_t\}, \{\alpha_t\})$ the *complete data.* Suppose the mle of a parameter vector ψ is required. One possible estimation method is to maximize the likelihood based on the incomplete data. However this is usually a complicated task requiring the optimization of a nonlinear likelihood function. An attractive alternative is to employ the EM algorithm. This consists of the repeated iterations of the following two steps :

1. *Expectation* or *E*-step : ***complete data sufficient statistics*** for ψ are estimated conditional on the *current* estimate of ψ and the observed data y.

2. *Maximimization* or *M*-step : a *new* estimate for ψ is evaluated using the expectation of the complete data sufficient statistics computed in the E-step.

Dempster, Laird and Rubin (1977) prove that repeated iterations of the E and M steps eventually lead to a stationary point of the likelihood function. The starting estimate of ψ is arbitrarily chosen ; however from the viewpoint of accelerating the convergence of the EM algorithm, it is often possible to employ an easily-derived consistent estimate of ψ.

The appealing features of the EM algorithm are : (i) the E and M steps are often very simple and interpretable, amounting in many cases to regression-like equations, (ii) the sequence of log-likelihoods is monotone nondecreasing (Boyles (1983) and Wu, 1983) (iii) a neighbourhood of the stationary point of the likelihood function is usually found within a few initial iterations even when the EM algorithm is initiated with poor estimates of the unknown parameters and (iv) the parameter estimates are mle's if a global maximum of the likelihood function is attained. Consequently, in the PET application, subsequent iterations of the EM algorithm yield sharper images. The EM algorithm does have some deficiencies : (i) it does not provide covariance matrices for the parameter estimates and (ii) its convergence rate may be linear or sublinear in the neighbourhood of the stationary point and thus unbearably slow for many applications. Louis (1982) and Meilijson (1989) have proposed extensions to the EM algorithm for computing mse matrices for the parameter estimates and speeding up the convergence of the EM algorithm using pseudo-Aitken's acceleration methods. These extensions are not covered in this thesis.

Alternative likelihood maximization methods include Newton-Raphson (for example the Gill-Murray-Pitfield algorithm) and quasi-Newton methods such as *scoring* which require the use of complex software to solve the nonlinear equations which arise from differentiating the likelihood function based on the incomplete data y. These estimation methods have the merits of (i) requiring fewer (but more involved) steps than the EM algorithm to achieve convergence and (ii) providing mse matrices of the parameter estimates. However they are not as stable as the EM algorithm in the sense that their repeated applications do not generally yield a monotone sequence of log-likelihoods. Some researchers like Watson and Engle (1983) have used a combination of both the EM and scoring algorithms in order to exploit their desirable features. The EM algorithm is first used to locate a neighbourhood of the stationary point of the likelihood function and the scoring algorithm is subsequently applied to achieve the convergence of the parameter estimates and the estimation of an approximate information matrix.

A serious flaw to the EM algorithm and most optimization routines is that they converge to a stationary point and not necessarily to the global maximum of the likelihood function. It is therefore always a wise strategy to initiate these algorithms with different starting points and then compare the results. For instance, in the applications reported in section 6.1.2, we have used two strategies to initiate the EM algorithm. The first strategy (labelled **A**) initiates the EM algorithm with solutions provided by previous researchers while the second strategy (labelled **B**) uses "naive" starting points where for example the relative variances of disturbances in the SSM are set to one.

6.1.1 The general DKF-EM algorithm

This subsection customizes the EM algorithm for use with the SSM. We generalize and unify the works of Shumway and Stoffer (1981), Watson and Engle (1983) and Harvey and Peters (1990) dealing with the estimation of unknown parameters in the system matrices (i. e. Z, T, G or H) in the SSM. Consider the SSM,

$$y_t = X_t\beta + Z\alpha_t + Gu_t, \qquad \alpha_{t+1} = W_t\beta + T\alpha_t + Hu_t$$

where $\alpha_0 \sim N(a_0, \sigma^2 P_0)$ with a_0 and P_0 known, X_t and W_t are known and $u_t \sim N(\mathbf{0}, \sigma^2 I)$. Suppose $\psi = (\beta; Z, G, T, H, \sigma^2)$ and furthermore assume without any loss in generality that y_t and α_t respectively have p and k components. Then *assuming* $\{\alpha_t\}$ *is known*, the SSM can be expressed as,

$$\begin{pmatrix} y_t \\ \alpha_{t+1} \end{pmatrix} = \begin{pmatrix} X_t & \alpha_t' \otimes I_p & \mathbf{0} \\ W_t & \mathbf{0} & \alpha_t' \otimes I_k \end{pmatrix} \begin{pmatrix} \beta \\ z \\ \tau \end{pmatrix} + \begin{pmatrix} G \\ H \end{pmatrix} u_t$$

where $z = vec\ (Z)$ and $\tau = vec\ (T)$.

It then follows from well-known linear models theory that the mle of $(\beta; z; \tau)$ is,

$$\begin{pmatrix} \hat{\beta} \\ \hat{z} \\ \hat{\tau} \end{pmatrix} = \left\{ \sum \begin{pmatrix} X_t' & W_t' \\ \alpha_t' \otimes I_p & \mathbf{0} \\ \mathbf{0} & \alpha_t' \otimes I_k \end{pmatrix} \begin{pmatrix} GG' & GH' \\ HG' & HH' \end{pmatrix}^{-1} \begin{pmatrix} X_t & \alpha_t' \otimes I_p & \mathbf{0} \\ W_t & \mathbf{0} & \alpha_t' \otimes I_k \end{pmatrix} \right\}^{-1}$$

$$\times \sum \begin{pmatrix} X_t' & W_t' \\ \alpha_t' \otimes I_p & \mathbf{0} \\ \mathbf{0} & \alpha_t' \otimes I_k \end{pmatrix} \begin{pmatrix} GG' & GH' \\ HG' & HH' \end{pmatrix}^{-1} \begin{pmatrix} y_t \\ \alpha_{t+1} \end{pmatrix}$$

This general mle expression is easily modified when some of the parameters (possibly time-varying) are known. Furthermore,

$$\hat{\sigma}^2 = tr \left\{ \begin{pmatrix} GG' & GH' \\ HG' & HH' \end{pmatrix}^{-1} M \right\} / (y; \alpha)^{\#}$$

$$\sigma \begin{pmatrix} \hat{G} \\ \hat{H} \end{pmatrix} = (M/n)^{1/2} \qquad \text{where}$$

$$M = \sum_t m_t m_t' \qquad \text{with } m_t = \begin{pmatrix} y_t \\ \alpha_{t+1} \end{pmatrix} - \begin{pmatrix} X_t & \alpha_t' \otimes I_p & 0 \\ W_t & 0 & \alpha_t' \otimes I_k \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{z} \\ \hat{\tau} \end{pmatrix}$$

It is easily ascertained that the sufficient statistics for the above mle's are $\sum_{t=1}^n y_t$, $\sum_{t=1}^n y_t y_t'$, $\sum_{t=1}^n y_t \alpha_t'$, $\sum_{t=1}^n \alpha_t$ and $\sum_{t=1}^n \alpha_{t+i}\alpha_{t+j}'$, $i, j = 0, 1$. Therefore the E-step of the EM algorithm evaluates $\tilde{\alpha}_t = E(\alpha_t|y)$ and $E(\alpha_{t+i}\alpha_{t+j}'|y) = Mse(\tilde{\alpha}_{t+i}, \tilde{\alpha}_{t+j}) + \tilde{\alpha}_{t+i}\tilde{\alpha}_{t+j}'$. In essence then, the E-step constructs the *complete data sufficient statistics* by running the smoothing algorithm associated with the CDKF. The M-step uses these complete data sufficient statistics to evaluate a *new* estimate for ψ.

We conclude this subsection by pointing out a problem which affects the estimation of system matrices in the SSM. In Chapter 2, we discussed the non-uniqueness of the SSM specification. This property implies that parameter estimation may be only feasible up to a scale and/or orthogonal transformation and consequently the EM algorithm converges to a ridge of stationary points of the likelihood function. This characteristic is evidenced by the multiple solutions obtained in the applications illustrated in this Chapter.

6.1.2 Illustrations

Estimation of growth rate of earnings. Shumway (1988, p186-192) describes an exponential trend plus seasonal variation model for a time series of quarterly earnings (from the fourth quarter, 1970 to the first quarter, 1980) of the U.S. conglomerate Johnson & Johnson with the SSM,

$$y_t = (1\ 1\ 0\ 0)\alpha_t + (0\ 0\ 1)u_t, \qquad \alpha_{t+1} = \begin{pmatrix} \phi & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} h_1 & 0 & 0 \\ 0 & h_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} u_t$$

The mle's of ϕ, h_1, h_2 and σ^2 conditional on y are easily derived using the results of the previous subsection. In particular, the mle of ϕ conditional on y is $S_{10}[1;1]/S_{11}[1;1]$ where $S_{ij} = \sum E(\alpha_{t+i}, \alpha_{t+j}|y) = \sum Mse(\tilde{\alpha}_{t+i}, \tilde{\alpha}_{t+j}) + \tilde{\alpha}_{t+i}\tilde{\alpha}'_{t+j}$, $i, j = 0, 1$. Table 6.1 lists the results obtained upon initiating the EM algorithm with (i) the final results obtained by Shumway (strategy **A**) and (ii) naive estimates of the parameters (strategy **B**). The second and third columns report the σ^2-concentrated log-likelihoods (denoted by λ^{σ^2}) respectively computed from the starting and final estimates generated by the EM algorithm.

Starting Points $\phi, h_1, h_2,$	Start. λ^{σ^2}	Final λ^{σ^2}	Solutions $\tilde{\phi}$, $\tilde{h}_1$, $\tilde{h}_2$, $\tilde{\sigma}^2$	Number of Iterations
A : 1.037, 0.53, 1.66	-93.99	-93.83	1.037, 0.5853, 1.528, 0.0394	2
B : 1, 1, 1	-136.67	-94.04	1.0368, 1.0583, 1.8851, 0.0234	295

Table 6.1: Estimation results with Johnson & Johnson data

The results tell us that during the study period, Johnson & Johnson experienced an average 3.7% quarterly increase in earnings. Furthermore the high value of $\tilde{h}_2$ is evidence of seasonal fluctuations in the earnings figures. The EM algorithm does not provide the standard errors of its estimates. However, in the present context, hypothesis testing on a particular value of ϕ can be conducted in an indirect fashion by analysing the innovations generated upon employing the stated value of ϕ in the transition matrix.

It is clear from the results of strategy **A** that Shumway's solution is indeed very close to a stationary point of the likelihood function. His results and ours may possibly differ due to the initialization of the KF : Shumway initializes the KF according to the "big k" method whereas we employ an exact method, namely the CDKF. Interestingly, the estimate of ϕ converges very quickly (within the first 5 iterations) as opposed to the other estimates. The other applications reported in this Chapter which also employ seasonal

SSM's also experience slow rates of convergence of the error covariance matrices.

Finally note that in this application, the CDKF coincides with the KF and thus A_t and E_t are vectors. This contrasts with employing the DKF when A_t and E_t would then consist of 4 columns. Thus employing the CDKF in lieu of the DKF leads to substantial computational savings.

Estimation of asset betas. The *Capital Asset Pricing Model* (CAPM), devised by Sharpe (1965) and which earned him a share of the 1990 Nobel prize in Economics, is used in Finance to estimate the "beta" (a measure of riskiness or volatility) of financial assets vis-a-vis a *market portfolio.* The latter is defined as the basket of all assets in a financial market, such as the New York Stock Exchange. In its simplest form, the CAPM can be viewed as a simple linear regression model with the intercept and regressor being respectively interpreted as the *risk-free rate of return* and the *market premium*, which is a premium offered as compensation for investing in a risky asset as opposed to a risk-free asset. By design, risk-free assets like government treasury bills are assigned a "beta" of zero while the market portfolio has a "beta" of one. Over the years, several modifications have been proposed to the CAPM. One such modification posits a random walk model for the unobserved market premium. This modified CAPM can be written as the SSM,

$$y_t = \mathbf{1}_p\beta + Z\alpha_t + (G, \mathbf{0})u_t, \qquad \alpha_{t+1} = \alpha_t + (\mathbf{0}, 1)u_t$$

where y_t is a vector of the rates of return of p assets, β is the unknown risk-free rate of interest, α_t is the market premium thereby implying that Z is a vector of the "betas" of the assets and G is diagonal with unknown diagonal entries g_i, $i = 1, \ldots, p$. More complicated asset-pricing models exist. For instance in the *Arbitrage Pricing Model*, the state α_t has components which are thought to be inflation rate, growth in industrial production, difference between long-term and short-term treasury bond yields etc. (Chen *et al.*, 1986). The estimation of these components is of obvious practical interest.

We now derive mle's for the parameters of interest in the above CAPM. Denote $z = vec(Z)$ and rewrite the CAPM as,

$$y_t = (\mathbf{1}_p \ , \ \alpha_t \otimes I_p) \begin{pmatrix} \beta \\ z \end{pmatrix} + (G, \mathbf{0})u_t$$

Then the mle of $(\beta; z)$ is given by,

$$\begin{aligned}
\begin{pmatrix} \hat{\beta} \\ \hat{z} \end{pmatrix} &= \{\sum_{t=1}^{n}(\mathbf{1}_p' \ ; \ \alpha_t \otimes I_p) \ Diag(g_1^{-2}, \ldots, g_p^{-2}) \ (\mathbf{1}_p \ , \ \alpha_t \otimes I_p)\}^{-1} \\
&\times \sum_{t=1}^{n}(\mathbf{1}_p' \ ; \ \alpha_t \otimes I_p) \ Diag(g_1^{-2}, \ldots, g_p^{-2}) \ y_t \\
&= \left(\begin{array}{c|cccccc}
n\sum_{i=1}^{p} g_i^{-2} & g_1^{-2}\sum\alpha_t & g_2^{-2}\sum\alpha_t & g_3^{-2}\sum\alpha_t & \ldots & g_{p-1}^{-2}\sum\alpha_t & g_p^{-2}\sum\alpha_t \\
\hline
g_1^{-2}\sum\alpha_t & g_1^{-2}\sum\alpha_t^2 & 0 & 0 & \ldots & 0 & 0 \\
g_2^{-2}\sum\alpha_t & 0 & g_2^{-2}\sum\alpha_t^2 & 0 & \ldots & 0 & 0 \\
. & . & . & . & . & . & . \\
g_p^{-2}\sum\alpha_t & 0 & 0 & 0 & \ldots & 0 & g_p^{-2}\sum\alpha_t^2
\end{array}\right)^{-1} \\
&\times \begin{pmatrix} \sum_i\sum_t g_i^{-2} y_t[i] \\ g_1^{-2}\sum\alpha_t y_t[1] \\ g_2^{-2}\sum\alpha_t y_t[2] \\ . \\ g_p^{-2}\sum\alpha_t y_t[p] \end{pmatrix}
\end{aligned}$$

where $y_t[i]$ is the i^{th} component of y_t. Therefore the E-step of the EM algorithm constructs the estimates of the sufficient statistics $\sum\alpha_t$ and $\sum\alpha_t^2$ conditional on y and feeds them to the M-step to update the current estimate of $(\beta; z)$. The mle's of σ^2 and G conditional on y are derived according to the expressions given in the previous subsection.

Using the CDKF in lieu of the DKF in this application implies that (i) the matrices

A_t and E_t consist of 2 as opposed to 3 columns and (ii) Q_t has order 2 as opposed to order 3. Thus the use of the CDKF attracts both computational and storage savings.

In order to obtain proper "betas", i. e. the "betas" are such that the market portfolio is assigned a "beta" equal to one, it is necessary to augment y_t by a proxy representing the market rate of returns. The choice of an appropriate market index has been a subject of debate in the financial research community. Two common choices are the CRSP (Center for Research in Security Prices) *equally-weighted* and *value-weighted* market indices which are respectively the arithmetic and price-weighted averages of the rates of return of all the assets in the NYSE. Therefore an interesting side result of the current application is to produce evidence if any in favour of these two market indices. This is achieved by comparing the relative "betas" estimated when no market index is used to the "betas" estimated by augmenting y_t with the two CRSP market indices.

The dataset considered in this application (courtesy of Dr. Dilip Madan of the University of Maryland) consists of 336 monthly rates of returns of $p = 3$ assets for the period 1959-1986 inclusive (see Appendix at the end of the dissertation). Listed in Table 6.2 are the summary statistics of the CRSP equally-weighted and value-weighted market indices (denoted by EW and VW) and these 3 assets.

Asset	EW	VW	1	2	3
Mean ($\times 10^{-4}$)	87	117	95	90	79
Variance ($\times 10^{-4}$)	18	28	22	27	21

Table 6.2: Summaries for financial data

Table 6.3 summarises our results (with assets in the same order as above) obtained under 3 strategies namely (I) without the use of a market proxy (therefore yielding relative "betas", which are denoted with † sign), (II) using the equally-weighted CRSP index and (III) using the value-weighted CRSP index. The average monthly market premium $\bar{\alpha} = n^{-1} \sum \tilde{\alpha}_t$ is also reported. In all instances, the EM algorithm was initiated with estimates of the "betas" and the diagonal elements of G (i. e. the g_i's) all set to one.

	I	II	III
λ^{σ^2}	6767.74	9880.62	9880.64
$\tilde{\beta} \times 10^{-3}$	5.4160	7.5432	7.5433
$\tilde{Z}$	-	1	-
	-	-	1
	$1^{\dagger}$	1.0480	1.0480
	$1.1256^{\dagger}$	1.1538	1.1538
	$1.0034^{\dagger}$	1.0519	1.0519
$10^{-3} \times \tilde{\sigma}^2$	2.48	2.59	2.59
$10^{-3} \times diag\ \tilde{\sigma}^2 \tilde{G}\tilde{G}'$	-	0.1904	-
	-	-	0.0938
	0.4422	0.3428	0.3041
	0.4814	0.3886	0.3908
	0.3540	0.2869	0.2130
# iterations	56	51	57
$10^{-3} \times \bar{\alpha}$	3.1556	1.1553	1.1552

Table 6.3: Estimates with financial data

From Table 6.3, we infer the following :

1. The CAPM explains about 80% of the variations in the assets since the residual variances (i. e. the diagonal elements of $\tilde{\sigma}^2 \tilde{G}\tilde{G}'$) each account for less than 20% of the unconditional variances listed in Table 6.1.

2. The similarity in the results produced by strategies **II** and **III** suggest that the value-weighted and the equally-weighted CRSP market indices are equally good market proxies.

3. The relative betas of assets 1 through 3 in the three strategies are about the same (since 1.04 : 1.15 : 1.05 translates to 1 : 1.1 : 1.003). In strategy **I**, the estimates

of the beta were rescaled so that the first asset had a beta of one. This has the drawback of subsequently confounding the estimates of β and $\bar{\alpha}$.

4. The results of strategies **II** and **III** indicate average risk-free and market premium annualized rates of return of 9.4% (= $1.00754^{12} - 1$) and 1.4%. Therefore the average annualized rate of return in the NYSE for the period 1959-1986 is 10.8% , a figure which is in line with commonly held estimates.

5. The higher beta for the second asset tells us that the returns of this asset are more sensitive to market fluctuations than the returns of the other two assets. The latters have "betas" close to one and therefore can be categorised as conservative assets.

6.2 Estimation of covariance matrices in the SSM

The most common parameter estimation required in the SSM deals with the estimation of the covariance matrices of its disturbances. This section considers a novel and more efficient CDKF-EM algorithm for this specific application. Koopman (1991) recently suggested a similar approach in the case of structural models. As a motivation, consider the *quarterly basic structural model* (QBSM),

$$
\begin{aligned}
y_t &= (1\ 0\ 1\ 0\ 0)\ \alpha_t + (0\ 0\ 0\ 1)\ u_t, \\
\begin{pmatrix} \mu_{t+1} \\ \nu_{t+1} \\ \gamma_{t+1} \\ \gamma_t \\ \gamma_{t-1} \end{pmatrix} = \alpha_{t+1} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} h_1 & 0 & 0 & 0 \\ 0 & h_2 & 0 & 0 \\ 0 & 0 & h_3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} u_t
\end{aligned}
$$

The QBSM is a seasonal structural model. Non-seasonal structural models include (i) the *basic structural model* (BSM) where the state consists of only a *level* (μ) and a *slope* (β) components and (ii) the *random walk plus noise* (RWM) model where both the observation matrix and transition matrix are equal to 1. The illustrations described later in this section employ the RWM, BSM and QBSM.

The unknown parameters in these structural models are σ^2 and a subset of $\{h_1, h_2, h_3\}$. In the case of the QBSM, their mle's are derived upon exploiting the following relationships,

$$
\begin{aligned}
\epsilon_{0,t} &= (0\ 0\ 0\ 1)\ u_t = y_t - z_t\ \alpha_t, \text{ where } z_t = (1\ 0\ 1\ 0\ 0) \\
\epsilon_{1,t} &= (h_1\ 0\ 0\ 0)\ u_t = (1\ 0\ 0\ 0\ 0)\ \alpha_{t+1} - (1\ 1\ 0\ 0\ 0)\ \alpha_t \\
\epsilon_{2,t} &= (0\ h_2\ 0\ 0)\ u_t = (0\ 1\ 0\ 0\ 0)(\alpha_{t+1} - \alpha_t) \\
\epsilon_{3,t} &= (0\ 0\ h_3\ 0)\ u_t = (0\ 0\ 1\ 1\ 1)\ \alpha_{t+1} + (0\ 0\ 0\ 0\ 1)\ \alpha_t
\end{aligned}
$$

Denote the variance of $\epsilon_{i,t}$ by σ_i^2, $i = 0, \ldots, 3$ with $\sigma_0^2 = \sigma^2$. Clearly then, the mle of σ_i^2 is $\hat{\sigma}_i^2 = 1/n \sum_{t=1}^n \epsilon_{i,t}^2$. Using the general CDKF-EM algorithm described in section 6.1.1, we therefore obtain, conditional on y,

$$
\begin{aligned}
\tilde{\sigma}_0^2 &= 1/n \sum_{t=1}^{n} \{(y_t - z_t\tilde{\alpha}_t)(y_t - z_t\tilde{\alpha}_t)' + z_t Mse(\tilde{\alpha}_t) z_t'\} \\
\tilde{\sigma}_1^2 &= 1/(n\sigma^2)\{S_{11}(1;1) + \sum_{i=1}^{2}\sum_{j=1}^{2} S_{00}(i;j) - 2 \times [S_{10}(1;1) + S_{10}(1;2)]\} \\
\tilde{\sigma}_2^2 &= 1/(n\sigma^2)\{S_{11}(2;2) + S_{00}(2;2) - 2 \times S_{10}(2;2)\} \\
\tilde{\sigma}_3^2 &= 1/(n\sigma^2)\{\sum_{i=3}^{5}\sum_{j=3}^{5} S_{11}(i;j) + S_{00}(5;5) + 2 \times \sum_{i=3}^{5} S_{10}(i;5)\} \qquad \text{where} \\
S_{ij} &= \sum_{t=0}^{n-1} Mse(\tilde{\alpha}_{t+i}, \tilde{\alpha}_{t+j}) + \tilde{\alpha}_{t+i}\tilde{\alpha}_{t+j}', \qquad i, j = 0, 1
\end{aligned}
$$

These mle's require the computation of $Mse(\tilde{\alpha}_{t-1}, \tilde{\alpha}_t)$. To obtain this quantity, Harvey and Peters (1990), while working with the QBSM, innovate an idea originally devised by Watson and Engle (1983) by augmenting α_t as defined above with $(\gamma_{t-3}; \mu_{t-1}; \nu_{t-1})$ and thereafter write $\epsilon_{i,t}$, $t = 1, 2, 3$ as contrast functions of the augmented state. The mle's of

σ_i^2 conditional on y are then evaluated using smoothed estimates of the augmented states and their mse matrices. From the discussions and illustrations in Chapters 2 and 5, it is clear that the performance of filtering and smoothing algorithms are dependent on the size of the state. This may explain why Harvey and Peters (1990) found the performance of the EM algorithm to be unsatisfactory. The next subsection proposes a means for avoiding the computation of lag one state error covariance matrices.

6.2.1 A new CDKF-EM algorithm

In this subsection, we consider a version of the CDKF-EM algorithm for the efficient estimation of G and H in the SSM. Its derivation is based on the observation that it is more natural to regard the disturbances in the structural model or in the general SSM as linear functions of u_t and *not* the state α_t. This suggests the consideration of $(y; u)$ as the complete data in lieu of $(y; \alpha)$. Consequently we are led to devise an CDKF-EM algorithm which does not require any lag 1 error covariance terms since the components of u (as opposed to α) are serially uncorrelated. Koopman (1991) has suggested a similar implementation of this version of the EM algorithm for structural models.

The E-step of this new DKF-EM algorithm requires the evaluation of $E(u_t|y)$ and $Mse(u_t|y)$. The following Theorem indicates the recursive evaluation of these quantities.

Theorem 6.1 (De Jong, 1991c) *Suppose $y = (y_1; y_2; \ldots; y_n)$ is generated by the SSM. Then $v_t = Pred(u_t|y)$ and $V_t = I - \sigma^{-2}Mse(u_t|y)$ are computed as,*

$$v_t = G_t' D_t^{-1} e_t + J_t' \eta_t, \qquad V_t = G_t' D_t^{-1} G_t + J_t' R_t J_t$$

where $J_t = H_t - K_t G_t$ and all the quantities are as defined in the KF and in the smoothing algorithm presented in Chapter 2.

The proof is given in the Appendix at the end of the Chapter. Koopman (1991) also proves a similar result. The above Theorem will also prove useful in the next Chapter where we consider diagnostic testing in the SSM. We now state the main result of this section.

Theorem 6.2 *Consider the SSM,*

$$y_t = X_t\beta + Z_t\alpha_t + Gu_t, \qquad \alpha_{t+1} = W_t\beta + T_t\alpha_t + Hu_t$$

where $\alpha_0 = \mathbf{0}$, X_t, Z_t, W_t *and* T_t *are known and* $u_t \sim N(\mathbf{0}, \sigma^2 I)$. *Then the complete data sufficient statistics in the mle's of* β, σ^2, G *and* H *are constructed from*

$$b_t = \begin{pmatrix} X_t \\ W_t \end{pmatrix}\beta + \begin{pmatrix} G \\ H \end{pmatrix} v_t \qquad \textit{and} \qquad B_t = \sigma^2 \begin{pmatrix} G \\ H \end{pmatrix}(I - V_t)(G' \; H').$$

Proof. Define for $t = 0, 1, \ldots, n$,

$$f_t = \begin{pmatrix} X_t \\ W_t \end{pmatrix}\beta + \begin{pmatrix} G \\ H \end{pmatrix} u_t, \qquad F = \begin{pmatrix} G \\ H \end{pmatrix}(G' \; H')$$

Observe that $\alpha_{t+1} = T_t\alpha_t + (\mathbf{0}, I)f_t$ and $y_t = Z_t\alpha_t + (I, \mathbf{0})f_t$. Therefore $f = (f_0; f_1; \ldots; f_n)$ is a complete dataset for the SSM. Furthermore $-2 \times$ the log-likelihood of f,apart from a constant is,

$$\lambda(\psi|f) \;=\; f^{\#}\log(\sigma^2) + n\log|F| + \sigma^{-2}\sum_{t=0}^{n}\{f_t - (X_t; W_t)\beta\}'F^{-1}\{f_t - (X_t; W_t)\beta\}$$

Thus $E(\lambda(\psi|f) \mid y)$ requires the evaluation of

$$E(f_t|y) = b_t \qquad \text{and} \qquad E(f_t f_t'|y) = Mse(f_t|y) + E(f_t|y)E'(f_t|y) = B_t + b_t b_t'$$

and hence the Theorem is asserted. •

Therefore the Theorem indicates that the E-step of the CDKF-EM algorithm uses the iterations described in Theorem 6.1 to evaluate $\sum_{t=0}^{n} v_t$ and $\sum_{t=0}^{n} V_t$. The M-step is described in the following result.

Theorem 6.3 (M-step) *For the SSM described in Theorem 6.2, the mles of σ^2, F and β conditional on the observed data y are*

$$
\begin{aligned}
\tilde{\sigma}^2 &= 1/f^{\#}\ tr[\tilde{F}^{-1}\sum_{t=0}^{n}\{B_t + b_t b_t' + (X_t;W_t)\tilde{\beta}\tilde{\beta}'(X_t;W_t)' - (X_t;W_t)\tilde{\beta}b_t - b_t'\tilde{\beta}'(X_t;W_t)'\}] \\
\tilde{F} &= 1/(n\tilde{\sigma}^2)\sum_{t=0}^{n}\{B_t + b_t b_t' + (X_t;W_t)\tilde{\beta}\tilde{\beta}'(X_t;W_t)' - (X_t;W_t)\tilde{\beta}b_t - b_t'\tilde{\beta}'(X_t;W_t)'\} \\
\tilde{\beta} &= [\sum_{t=0}^{n}(X_t;W_t)'\tilde{F}^{-1}(X_t;W_t)]^{-1}\sum_{t=0}^{n}(X_t;W_t)'\tilde{F}^{-1}b_t
\end{aligned}
$$

Proof. Differentiating $\lambda(\psi|f)$ in turn with respect to σ^2, F and β and equating each normal equation to zero respectively yields,

$$
\begin{aligned}
\hat{\sigma}^2 &= 1/f^{\#}\ tr[F^{-1}\sum_{t=0}^{n}\{f_t f_t' + (X_t;W_t)\beta\beta'(X_t;W_t)' - (X_t;W_t)\beta f_t - f_t'\beta'(X_t;W_t)'\}] \\
\hat{F} &= 1/(n\sigma^2)\sum_{t=0}^{n}\{f_t f_t' + (X_t;W_t)\beta\beta'(X_t;W_t)' - (X_t;W_t)\beta f_t - f_t'\beta'(X_t;W_t)'\} \\
\hat{\beta} &= [\sum_{t=0}^{n}(X_t;W_t)'F^{-1}(X_t;W_t)]^{-1}\sum_{t=0}^{n}(X_t;W_t)'F^{-1}f_t
\end{aligned}
$$

Taking the expectation of these mle's conditional on y immediately leads to the results listed in the Theorem. •

If $\beta = \mathbf{0}$, as for instance in the structural models, then the expressions given in the Theorem for $\tilde{\sigma}^2$ and $\tilde{F}$ simplify considerably.

Corollary 6.1 *Suppose $\beta = \mathbf{0}$ in Theorem 6.3. Then the new estimate of σ^2 and $(G;H)$ are*

$$
tr(\sum_{t=0}^{n} v_t v_t')/tr(\sum_{t=0}^{n} V_t) \qquad and \qquad \begin{pmatrix} \tilde{G} \\ \tilde{H} \end{pmatrix}\{I - 1/n(\sum_{t=0}^{n} V_t - v_t v_t'/\tilde{\sigma}^2)\}^{1/2}
$$

Proof. Put $\beta = \mathbf{0}$ in Theorem 6.3. Then

$$
\tilde{\sigma}^2 = 1/f^{\#}\ tr\{\sum_{t=0}^{n}\tilde{\sigma}^2(I - V_t) + v_t v_t'\} \Rightarrow \tilde{\sigma}^2 = tr\ (\sum_{t=0}^{n} v_t v_t')/tr(\sum_{t=0}^{n} V_t)
$$

Furthermore the new estimate of $(G; H)$ is

$$(\tilde{G}; \tilde{H})\{\sum_{t=0}^{n}[\tilde{\sigma}^2(I - V_t) + v_t v_t']/n\tilde{\sigma}^2\}^{1/2} = (\tilde{G}; \tilde{H})\{I - 1/n(\sum_{t=0}^{n} V_t - v_t v_t'/\tilde{\sigma}^2)\}^{1/2}$$

This completes the proof of the Corollary. •

In order to derive consistent estimates of σ^2, it may be necessary to impose some constraints on G and H. For instance in the QBSM, the zero constraints must be enforced in the updated estimates of G and H. Furthermore these estimates should also be rescaled such that the updated estimate of G has the form $\tilde{G} = (0\ 0\ 0\ 1)$.

This new version of the CDKF-EM algorithm is clearly simpler and more efficient than the one discussed in section 6.1. It provides gains in storage requirements whenever u_t is of lower dimensionality than α_t. This occurs for instance in the seasonal structural model when the number of seasons is greater than 3. Furthermore, as previously noted, it does not require the evaluation of lag one error covariance matrices.

6.2.2 Estimation of Structural Models

This subsection illustrates the improved CDKF-EM algorithm with some structural models. A factor influencing the convergence of the EM algorithm is the initial estimate of the parameter. It is desirable that a consistent and easily computable estimate be employed. Harvey and Todd (1983) and Harvey (1989, p56) have suggested the construction of consistent estimates of the unknown parameters in structural models through the consideration of the autocovariance function. They state that the autocovariance function of $(1 - L)(1 - L^s)y_t$, where L is the usual lag operator, s is the number of seasons and $\sigma_j^2, j = 0, 1, 2, 3$ are as defined previously in this Chapter, is

$$\begin{aligned}
\gamma(0) &= 4\sigma_0^2 + 2\sigma_1^2 + s\sigma_2^2 + 6\sigma_3^2 \\
\gamma(1) &= -2\sigma_0^2 + (s-1)\sigma_2^2 - 4\sigma_3^2
\end{aligned}$$

$$
\begin{aligned}
\gamma(2) &= (s-2)\sigma_2^2 + \sigma_3^2 \\
\gamma(i) &= (s-i)\sigma_2^2, \qquad i = 3, \ldots, s-2 \\
\gamma(s-1) &= \sigma_0^2 + \sigma_2^2 \\
\gamma(s) &= -2\sigma_0^2 - \sigma_1^2 \\
\gamma(s+1) &= \sigma_0^2 \\
\gamma(i) &= 0, \qquad i \geq s+2
\end{aligned}
$$

Therefore particular estimates of σ_j^2 are obtained upon solving any set of four autocovariance equations listed above. A serious flaw is that there is no guarantee of non-negative solutions. To get around this problem, we may substitute a small positive number for each negative estimate of σ_j^2. We will employ this strategy (labelled **C**) in addition to stategies **A** (which uses the reported solutions) and **B** (which uses naive estimates) to initiate the EM algorithm in the next 3 applications.

We now report the results of applying this new version of the CDKF-EM algorithm in the first instance to structural models that have been estimated by Harvey (1989) and Harvey and Peters (1990). In the second instance, we apply the CDKF-EM algorithm to structural models considered (but not estimated) by West and Harrison (1989). In the applications below, the EM algorithm is stopped when the increase in the σ^2-concentrated log-likelihood is less than 10^{-4}.

Estimation of variability in purse snatchings. Harvey (1989, p89) uses a random walk plus noise model (RWM) for a time series of reported purse snatchings in the Hyde Park area of Chicago. He reports estimates $\tilde{\sigma}^2 = 24.79$ and $\tilde{h}_1 = 0.4557$. With these as starting points in the new EM algorithm (strategy **A**) we could not attain a higher log-likelihood (=-554.264). For strategy **B** we evaluated the lag 0 and lag 1 covariances of $(1-L)y_t$ as 57.23 and -31 thereby implying an estimate of σ_1^2 of -4.77 ; strategy **B**

therefore initiates the EM algorithm with $h_1 = 0.01$. After 130 iterations, we obtain a log-likelihood of -554.264, $\tilde{\sigma}^2 = 24.8347$ and $\tilde{h}_1 = 0.4537$. Finally strategy **C** initiates the EM algorithm with $h_1 = 1$. After 26 iterations, we obtain a similar log-likelihood with $\tilde{\sigma}^2 = 24.7763$ and $\tilde{h}_1 = 0.4557$. For this dataset, the CDKF-EM algorithm convergences in all instances to a ridge of local maxima.

Estimation of seasonality in air travel. Box and Jenkins (1970, p531) provide a dataset containing the number of monthly international airline departures for the period January 1949 to December 1960. This dataset is a popular benchmark test in time series analysis. Harvey (1990,p93-94) and Harvey and Peters (1990) aggregate the data into 48 quarterly observations and thereafter apply the log transformation to them.

In the tables below, the first column lists the starting points employed by strategies **A**, **B** and **C**. The second column and third column respectively list the σ^2-concentrated log-likelihoods based on the initial and final estimates of the parameters ; the fourth column contains the final parameter estimates and the last column reports the number of iterations required for convergence.

The results in Table 6.4 relate relate to Harvey (1990) who estimates the QBSM using all the 48 observations. Harvey and Peters (1990), on the other hand, only employ the first

Start. Points h_1, h_2, h_3	Start. λ^{σ^2}	Final λ^{σ^2}	Solutions $\tilde{\sigma}^2, \tilde{h}_1, \tilde{h}_2, \tilde{h}_3$	Number of Iterations
A : 22.19, 1, 11	121.88	123.22	6.88×10^{-7}, 29.9946, 0.8138, 10.7035	110
B : 15, 15, 15	115.74	123.21	1.37×10^{-7}, 66.9838, 2.0305, 23.9407	185
C : 3.52, 1, 1.24	119.76	123.21	2.83×10^{-6}, 14.7877, 0.3258, 5.2631	301

Table 6.4: Estimation results for airline departures data (I)

40 observations for estimation puposes. They report solutions obtained upon using four estimation methods namely (1) TD, maximization of the time domain prediction error

decomposition form of the likelihood function using the Gill-Murray-Pitfield algorithm (2) EM, the EM algorithm as discussed in section 1 but modified to incorporate a line search in order to speed up convergence and using a stopping criterion based on differences in log-likelihoods (3) EM*, same as (2) but using a different stopping criterion which is based on differences between prediction error variances and (4) TD, maximization of the frequency domain form of the likelihood function using the Gill-Murray-Pitfield algorithm.

Start. Points h_1, h_2, h_3	Start. λ^{σ^2}	Final λ^{σ^2}	Solutions $\tilde{\sigma}^2$, $\tilde{h}_1$, $\tilde{h}_2$, $\tilde{h}_3$	Number of Iterations
A (TD) : 13, 1, 5.77	101.71	102.31	2.17×10^{-6}, 18.24, 0.84, 6.20	99
A (EM) : 14.35, 1, 4.47	102.16	102.31	2.41×10^{-6}, 17.31, 0.82, 5.89	96
A (EM*) : 9.8, 4.11, 1	90.43	102.32	2.22×10^{-6}, 18.07, 0.74, 6.11	236
A (FD) : 16.63, 1, 5.48	102.27	102.31	1.95×10^{-6}, 19.26, 0.88, 6.54	88
B : 10, 10, 10	97.79	102.32	4.89×10^{-7}, 38.32, 2.03, 13.09	140
C : 9.1, 1.24, 1	95.16	102.31	3.17×10^{-6}, 15.14, 0.59, 5.11	240

Table 6.5: Estimation results for airline departures data (II)

In both Tables 6.4 and 6.5, the estimate of h_2 is relatively small compared the estimates to h_1 and h_3. The same situation occurs in the next two applications (see Tables 6.6 and 6.7). Ledolter explains this phenomenom in the discussion of Harvey (1984) as follows : structural models (apart from the random walk model) can be expressed as ARIMA models with MA coefficients lying on the boundary of the invertibility region and it is "structural components with small variances that introduce moving operators in the equivalent ARIMA model that are close to the invertibility boundary".

West and Harrison (1989) and Ng and Young (1990) discuss subjective interventions in state-space models, specifically structural models. Following their Bayesian approach, the first set of authors fix the unknown parameters of the basic structural model *a priori* rather than estimate them. The second set of authors assume zero variance for the level and slope components of the state except at intervention points. Both papers illustrate their methods with two applications. We have a two-pronged interest in the work of these researchers : (i) provide mle's for the unknown parameters of these structural models under the assumption of no data irregularities and (ii) provide diagnostics based on the models estimated in (i) and thereafter incorporating the necessary interventions. The latter area of work is covered in the next Chapter.

Estimation of trend in tobacco sales. West and Harrison (1989) employ a BSM to model standardized monthly total sales of tobacco products by a major company in the UK for the period 1955-1959. In this application, we ignore the effects of possible outliers and structural breaks in the model. Table 6.6 lists the results obtained upon applying the CDKF-EM algorithm to this dataset.

Start. Points h_1, h_2	Start. λ^{σ^2}	Final λ^{σ^2}	Solutions $\tilde{\sigma}^2$, $\tilde{h}_1$, $\tilde{h}_2$	Number of Iterations
B : 0.2, 0.2	-665.42	-659.93	540.80, 0.9357, 0.0626	124
C : 1, 1	-678.31	-659.92	539.67, 0.9402, 0.0600	105

Table 6.6: Estimation results for tobacco products sales data

These results tell us that the level component in the state account for more variability in the observations than the slope components. In fact in our work in diagnostic-testing in Chapter 7, we attribute the cause of data irregularities in this model to shifts in the mean level.

Estimation of seasonality in UK weddings. West and Harrison (1989) posit a QBSM for the quarterly number of UK weddings for the period 1965-1970. Ignoring possible data iregularities, we obtain the following estimates for the parameters of this model in Table 6.7.

Start. Points h_1, h_2, h_3	Start. λ^{σ^2}	Final λ^{σ^2}	Solutions $\tilde{\sigma}^2, \tilde{h}_1, \tilde{h}_2, \tilde{h}_3$	Number of Iterations
B : 1, 1, 1	-174.25	-150	0.0582, 0.5035, 0.0933, 38.1495	181
C : 0.01, 0.03, 0.052	-166.08	-150	0.1865, 0.0888, 0.5, 21.30	287

Table 6.7: Estimation results for UK weddings data

The interesting finding in the UK weddings dataset is the wide variability of the seasonal effects. West and Harrison (1989) attributes this to the abolishment of a tax incentive which used to affect the timing of weddings. This will be considered in more detail in the next Chapter.

Remarks

1. The results from the three seasonal models considered in this section suggest that the likelihood function surface is flat. This is attested for by the multiplicity of solutions. These findings are in line with those observed by Laird *et al.* (1987) who employ the EM approach to estimate variance components models of which the structural model is one.

2. The relatively high number of iterations required for convergence of the CDKF-EM algorithm does not necessarily constitute a drawback. For instance, scoring methods require fewer iterations in the neighbourhood of a stationary point of the likelihood function but each of these iterations are very involved usually requiring several passes of the KF to compute first and second derivatives of the likelihood with respect to the unknown parameters. For instance Watson and Engle(1983) find the performance of the EM algorithm and the scoring method to be comparable in an application involving vector observations.

6.3 Summary

This Chapter considered maximum likelihood estimation in the SSM using an EM approach. We showed a general method for estimating unknown time invariant system matrices in the SSM. We also developed an efficient CDKF-EM estimation method for the estimation of error covariance matrices in the SSM. This novel approach does not require the computation of lag one state error covariance matrix. The preceding chapters dealt with the prediction aspects of the SSM. Therefore following the guidelines set forward by Box and Jenkins (1970), it remains to cover the topic of model-fitting or diagnostic testing in order to complete the statistical analysis of the SSM. This is the focus of the following Chapter.

6.4 Appendix

Proof of Theorem 6.1 The proof makes use of the fact that the innovation vector $e = (e_1; \ldots; e_n)$ generated by the KF has the same information content as the observation vector $y = (y_1; y_2; \ldots; y_n)$. Therefore,

$$\begin{aligned} Pred(u_t|y) &= Pred(u_t|e) = Pred(u_t|e_t; e_{t+1}; \ldots; e_n) \\ &= \sum_{j=t}^{n} Cov(u_t, e_j)(\sigma^2 D_j)^{-1} e_j \end{aligned}$$

Using Lemma 5.3, $e_t = Z_t(\alpha_t - \hat{\alpha}_t) + G_t u_t$ and for $j = t+1, \ldots, n$,

$$\begin{aligned} e_j &= G_j u_j + Z_j\{J_{j-1}u_{j-1} + L_{j-1}(\alpha_{j-1} - \hat{\alpha}_{j-1})\} \\ &= G_j u_j + Z_j\{J_{j-1}u_{j-1} + L_{j-1}J_{j-2}u_{j-2} + L_{j-1}L_{j-2}(\alpha_{j-2} - \hat{\alpha}_{j-2})\} \\ &= G_j u_j + Z_j\{J_{j-1}u_{j-1} + L_{j-1}J_{j-2}u_{j-2} + \ldots \\ &+ (L_{j-1}L_{j-2}\ldots L_{t+1})J_t u_t + (L_{j-1}L_{j-2}\ldots L_t)(\alpha_t - \hat{\alpha}_t)\} \end{aligned}$$

After noting that $Cov(u_t, \alpha_j - \hat{\alpha}_j) = \mathbf{0}$ for all t and j, we obtain

$$\begin{aligned} Pred(u_t \mid y) &= G_t' D_t^{-1} e_t + \sum_{j=t+1}^{n} (Z_j L_{j-1} L_{j-2} \ldots L_{t+1} J_t)' D_j^{-1} e_j \\ &= G_t' D_t^{-1} e_t + J_t' r_t \\ &= v_t, \text{ as asserted} \end{aligned}$$

Finally,

$$\begin{aligned} Mse(u_t|y) &= Cov(u_t) - \sum_{j=t}^{n} Cov(u_t, e_j)(\sigma^2 D_j)^{-1}\{Cov(u_t, e_j)\}' \\ &= \sigma^2(I - G_t' D_t^{-1} G_t \\ &- J_t'\{\sum_{j=t+1}^{n} (Z_j L_{j-1} L_{j-2} \ldots L_{t+1} J_t)' D_j^{-1} (Z_j L_{j-1} L_{j-2} \ldots L_{t+1} J_t)\} J_t') \\ &= \sigma^2(I - G_t' D_t^{-1} G_t - J_t' R_t J_t) \\ &= \sigma^2(I - V_t) \end{aligned}$$

This asserts the Theorem. •

Chapter 7

Residual Analysis in the State Space Model

Time series models, especially those arising in socioeconomic applications, are prone to data irregularities such as discordant observations, or *outliers* as they are commonly called, and *structural breaks.* The exercise of detecting these unanticipated or extraordinary events, generally dubbed as *residual analysis* or *diagnostic testing,* is now firmly entrenched as an essential and integral part of any statistical modelling. Residual analysis allows us to revise the statistical model under consideration and consequently it enhances the various facets of statistical inference namely parameter estimation, prediction and tests for goodness of fit.

The residual analysis literature is extensive in the case of the linear regression model where the observations are mutually independent. Foremost publications are the established textbooks of Belsey *et al.* (1980) and Cook and Weisberg (1982). The latter authors introduce three types of residuals and discuss their uses in the detection of *outliers* and *influential* observations. For the SSM defined by equations (2.1)-(2.2), these residuals are defined as follows,

1. *Ordinary* or *Signal* residuals :

$$\begin{aligned}\tilde{e}_t &= y_t - Pred(X_t\beta + Z_t\alpha_t \mid y_1, \ldots, y_n) \\ &= Z_t\{\alpha_t - Pred(\alpha_t \mid y_1, \ldots, y_n)\} + G_t u_t\end{aligned}$$

2. *Deleted* or *Leave-one-out* residuals :

$$\bar{e}_t = y_t - Pred(y_t \mid y_1, \ldots, y_{t-1}, y_{t+1}, \ldots, y_n)$$

$$\begin{aligned} &= Z_t\{\alpha_t - Pred(\alpha_t \mid y_1, \ldots, y_{t-1}, y_{t+1}, \ldots, y_n)\} \\ &+ G_t\{u_t - Pred(u_t \mid y_1, \ldots, y_{t-1}, y_{t+1}, \ldots, y_n)\} \end{aligned}$$

3. *Recursive* or *Innovation* residuals :

$$\begin{aligned} e_t &= y_t - Pred(y_t | y_1, \ldots, y_{t-1}) \\ &= Z_t\{\alpha_t - Pred(\alpha_t \mid y_1, \ldots, y_{t-1})\} + G_t\{u_t - Pred(u_t \mid y_1, \ldots, y_{t-1})\} \end{aligned}$$

The ordinary and deleted residuals are often outputted by standard regression packages. Brown *et al.* (1975) demonstrate the usefulness of the recursive residuals in assessing the constancy of the regression parameter in the context of cross-sectional regression analysis.

The ideas pertaining to the above residuals clearly extend to dynamic linear models but they become more intricate due to the dependent nature of the observations (or more precisely the states). The latter characteristic also makes the leave-one-out residual less useful. This point is emphasised in a recent paper dealing with diagnostics for ARIMA fitting of time series data where Bruce and Martin (1989) emphasise that "the dependency aspect of time series data gives rise to a *smearing* effect, which confounds the diagnostics for the coefficients ..." and thereafter propose a "*leave-k-out*" diagnostics approach to deal with patches of outliers.

Carrying out residual analysis via a linear regression model approach has the drawback of being computationally demanding since the evaluation of the residuals requires the inversion of error covariance matrices with dimensions equal to the size of the data. The execution of this exercise via a SSM approach is however more attractive from a computational standpoint. For instance, innovations are automatically generated by the KF or the DKF after further algebraic manipulation. Early works in residual analysis within the SSM context made exclusive use of the innovations ; see Harvey (1990, p256-260) who survey their uses in tests of misspecifactions for serial correlation, non-linearity, heteroscedasticity and normality in the SSM. Innovations have the nice statistical property of uncorrelatedness but suffer from the fact that they may not convey as much information content as the signal and deleted residuals. These alternative residuals can be generated in an efficient fashion using for example the recursive algorithms derived independently by De Jong (1988b,1989) in the vector data case and Kohn and Ansley (1989) in the scalar data case.

The residuals described above confound the observation errors $(G_t u_t)$ and the errors incurred in the estimation of the states. Therefore they are unlikely to distinguish between outliers and structural breaks. The detection of these data irregularities is more satisfactorily addressed via the separate studies of estimators of $G_t u_t$ and the state or transition errors $H_t u_t$. Harvey and Koopman (1991) advocate and study the use of these residuals for goodness-of-fit tests and diagnostic checks.

The SSM specification employed in this thesis attracts two benefits for residual analysis. First, estimates of $G_t u_t$ and $H_t u_t$ can be generated in a unified fashion from estimators of the disturbance vector u_t. Second, interventions in the SSM to incorporate data irregularities is easily carried out via the use of the regression matrices X_t and W_t.

This Chapter emphasises the use of the *predicted residuals* $v_t = Pred(u_t \mid y_1, \ldots, y_n)$ for exploratory residual analysis. Recursive formulae for the generation of v_t and $Mse(v_t)$ have already been provided in the previous Chapter (Theorem 6.1). Other predictors of u_t are less useful. For instance, observe that the random variable u_t conditional on $y_1, \ldots, y_{t-1}$ has mean $\mathbf{0}$ and covariance matrix $\sigma^2 I$ and therefore $Pred(u_t \mid y_1, \ldots, y_{t-1})$ is uninformative from the standpoint of diagnostics. Furthermore, as argued above, leave-one-out residuals are almost similar in characteristics to signal residuals in the presence of dependent data. These will be apparent in the illustrations presented in the final section of the Chapter. A recursive algorithm for the generation of the leave-one-out residuals is provided by De Jong (1988b, 1989).

In the first section of this Chapter, we demonstrate that the results of De Jong (1988b,1989) and Kohn and Ansley (1989) concerning signal residuals are in fact consequences of Theorem 6.1. Since the prediction residuals v_t are serially correlated, we consider, in section 2, whitening or orthogonalizing $\{v_t\}$ in a backward direction and conclude that this orthogonalized sequence corresponds, up to a scaled factor, to the innovations. This tells us that innovations are as statistically efficient as the backward orthogonalized prediction residuals in statistical tests for goodness-of-fit of the SSM. In section 3, we apply residual analysis to both the tobacco sales dataset and the the UK weddings dataset which were discussed in the last Chapter. We illustrate the detection of outliers and points of structural breaks via simple graphical devices.

7.1 Connection with the Literature

In many applications, it is only necessary to consider specific aspects of v_t, for example $G_t v_t$ and $H_t v_t$. These have lower dimensionalities than v_t and hence it is worthwhile to specialize Theorem 6.1 to these cases.

We now demonstrate how the results of De Jong (1988b,1989) and hence Kohn and Ansley (1989) concerning the signal residuals follow from Theorem 6.1. The following Theorem due to De Jong (1989) is reexpressed with notation consistent with this thesis ; see also Koopman (1991) for a closely connected result.

Theorem 7.1 (De Jong, 1989) *Consider the SSM defined by* $y_t = X_t\beta + Z_t\alpha_t + G_t u_t$ *and* $\alpha_{t+1} = W_t\beta + T_t\alpha_t + H_t u_t$ *where* $G_t H_t' = \mathbf{0}$ *and the* u_t*'s are mutually and serially uncorrelated zero-mean disturbance vectors with covariance matrix* $\sigma^2 I$. *Let* $y = (y_1, \ldots, y_n)$.

Then

$$\begin{aligned} Pred(G_t u_t|y) &= G_t G_t' m_t \quad \text{and} \quad Mse(G_t u_t|y) = \sigma^2(G_t G_t' - G_t G_t' M_t G_t G_t') \text{ where} \\ m_t &= D_t^{-1} e_t - K_t' \eta_t \quad \text{and} \quad M_t = D_t^{-1} + K_t' R_t K_t \end{aligned}$$

with η_t and R_t as defined in Theorem 2.2.

Proof. We first establish a useful identity, namely $Z_t P_t L_t' = -G_t J_t'$, where $L_t = T_t - K_t Z_t$ and $J_t = H_t - K_t G_t$. To see this, manipulate a couple of equations making up the KF to obtain

$$\begin{aligned} Z_t P_t L_t' &= Z_t P_t (T_t - K_t Z_t)' \\ &= (D_t K_t' - G_t H_t') - (D_t - G_t G_t') K_t' \\ &= -G_t (H_t - K_t G_t)' \\ &= -G_t J_t' \end{aligned}$$

Using the expression given in Theorem 6.1 for v_t, it follows that

$$\begin{aligned} G_t v_t &= G_t G_t' D_t^{-1} e_t + G_t J_t' \eta_t \\ &= G_t G_t' D_t^{-1} e_t - Z_t P_t (T_t - K_t Z_t)' \eta_t \\ &= G_t G_t' D_t^{-1} e_t - \{D_t^{-1} K' - (D_t - G_t G_t') K_t'\} \eta_t \\ &= G_t G_t' (D_t^{-1} e_t - K_t' \eta_t) \\ &= G_t G_t' m_t \end{aligned}$$

The third equality follows from the KF (see Chapter 2), taking into account that $G_t H_t' = 0$. Finally $\sigma^{-2} Mse(G_t v_t) = G_t(I - V_t) G_t'$ equals,

$$\begin{aligned} G_t(I - V_t) G_t' &= G_t(I - G_t' D_t^{-1} G_t - J_t' R_t J_t) G_t' \\ &= G_t G_t' - G_t G_t' D_t^{-1} G_t G_t' - G_t G_t' K_t' R_t K_t G_t G_t' \end{aligned}$$

$$= G_tG_t' - G_tG_t'(D_t^{-1} + K_t'R_tK_t)G_tG_t'$$
$$= G_tG_t' - G_tG_t'M_tG_tG_t'$$

These assert the Theorem. •

Theorem 6.1 can also be specialized for the efficient generation of H_tv_t and their associated error covariance matrices. These residuals, which are sometimes known as the smoothed *auxilliary* residuals, estimate the errors associated with components of the state and convey information which is usually not apparent in the innovations.

Theorem 7.2 *Consider the SSM described in Theorem 7.1. Then*

$$Pred(H_tu_t|y) = H_tH_t'\eta_t \text{ and } Mse(H_tu_t|y) = \sigma^2(H_tH_t' - H_tH_t'R_tH_tH_t')$$

Proof. Observe that $Pred(H_tu_t|y) = H_tv_t = H_t\{G_t'D_t^{-1}e_t + (H_t' - G_t'K_t')\eta_t\} = H_tH_t'\eta_t$ upon noting that $H_tG_t' = \mathbf{0}$. Finally, $Mse(H_tu_t|y) = Cov(H_tu_t) - Cov(H_tu_t|y) = \sigma^2(H_tH_t' - H_tH_t'R_tH_tH_t')$. •

7.2 Backward Orthogonalization of Predicted Residuals

The predicted residuals v_t's are serially correlated since they are inhomogeneous linear combination of $(y_1, \ldots, y_n)$ or equivalently $e_1, \ldots, e_n$. This property makes the use of the v_t's in statistical tests for goodness-of-fit in the SSM a complicated task. This contrasts with the ease with which the (uncorrelated) innovation lend themselves to in the same tests. Hence we are led to consider the idea of *whitening* or orthogonalising the v_t's.

Theorem 7.3 *Suppose that for $1 \leq t \leq n$, the space spanned by $(v_t; v_{t+1}; \ldots; v_n)$ coincides with the space spanned by $(e_t; e_{t+1}; \ldots; e_n)$. Then a backward orthogonalization of v_t corresponds, up to a weighting matrix, to the innovations e_t generated in the KF.*

Proof. The assumption in the Theorem implies

$$\begin{aligned} v_t - Pred(v_t|v_{t+1};\ldots;v_n) &= v_t - Pred(v_t|e_{t+1};\ldots;e_n) \\ &= (G_t'D_t^{-1}e_t + J_t'r_t) - J_t'r_t \\ &= G_t'D_t^{-1}e_t \end{aligned}$$

Hence the backward orthogonalized version of v_t, corresponds up to a scale factor, to the innovations. •

Remarks

1. A sufficient condition for the Theorem to hold is that G_t has full rank for all t.

2. The Theorem implies that innovations are as efficient as backward whitened versions of v_t's in statistical tests of goodness-of-fit.

7.3 Illustrations

Theorem 6.1 was useful for maximum likelihood estimation of parameters in the SSM. We now illustrate the Theorem in a different setting, namely exploratory residual analysis. This consists of assessing time series plots of the studentized observation residuals, $\{\sigma^2 G_t(I-V_t)G_t'\}^{-1}G_tv_t$ and the studentized auxilliary residuals, $\{\sigma^2 H_t(I-V_t)H_t'\}^{-1}H_tv_t$. We have chosen the final two datasets covered in the previous Chapter for the purpose of illustration.

Diagnostics for tobacco sales. The sales figures are graphed in Figure 7.1. The observations clearly suggest the presence of outliers and possibly structural breaks. The various residuals, displayed in the top two graphs of Figure 7.2 were obtained, from the structural model estimated in the previous Chapter. A cursory examination of these diagnostic plots (especially the smoothed observation residuals and the auxilliary residuals) suggest the presence of one-time mean effects or outliers at Dec'55, Jan'57 and Jan'58. In particular, note the statistically significant departures of the innovations and the residuals associated with the level component from their expected value of zero at these points.

The basic structural model (model **C**) employed in the previous Chapter was modified to incorporate interventions at these points. This consists of defining X_t as indicator variables at Dec'55, Jan'57 and Jan'58. The matrix H has revised parameters $\tilde{h}_1 = 1.30$ and $\tilde{h}_2 = .08$ whereas $\tilde{\sigma}^2 = 223.8$ (compared respectively to 0.94, 0.06 and 540.8 in the pre-intervention model). The residuals (innovations, signal and level residuals) produced by this revised model (displayed in the bottom half of Figure 7.2) look reasonable and furthermore reflect the larger variability in the data for the period Jan'58 - Dec'59.

Diagnostics for UK Weddings. West and Harrison (1989) attribute the unanticipated seasonal variations in the observations (see Figure 7.3), particularly in the first quarter of each year to the tax benefits enjoyed upon matrimony. These benefits were abolished at the end of 1967. The diagnostic plots in the top half of Figure 7.4 clearly indicates this fact. Specifically observe the huge residual associated with the seasonal component of the state at the first quarter of 1968.

We intervened in the model and associated a dummy regression variable with the seasonal component of the state for the first quarter of 1968. The revised model has parameters $\tilde{h}_1 = .11$, $\tilde{h}_2 = .014$, $\tilde{h}_3 = 1.32$ and $\tilde{\sigma}^2 = 30.69$ (compared respectively to 0.50, 0.09, 0.06 and 38.15 in the model without any intervention). The residuals in the revised model (see bottom half of Figure 7.4) especially those arising after 1968 appear to conform to expectations. In the absence of any financial incentives, we would expect a seasonal low during the first quarter since it coincides with the winter months and seasonal highs during the two middle quarters. From the revised model, we inferred that the abolishment of the tax benefits caused a decrease of about 34,200 weddings (or a 28% drop) from the expected number in the first quarter of 1968.

7.4 Summary

We have demonstrated that the specification of the SSM employed in this thesis allows us to generate, in a unified fashion, residuals which are useful for pinpointing likely outliers and point of structural change in the SSM. These residuals, unlike the innovations, are serially correlated and should therefore be interpreted with care. We have shown that the backward whitened versions of these residuals correspond (up to weighting matrix) to the innovations generated by the KF. This implies that innovations are as efficient as these whitened residuals in statistical tests of goodness-of-fit.

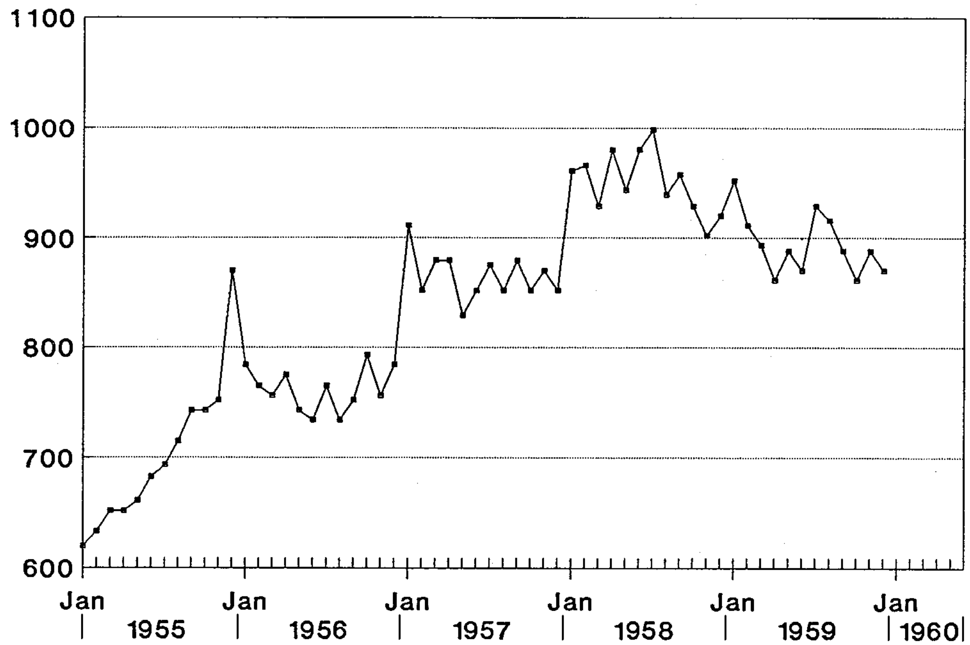

Figure 7.1: Tobacco sales data

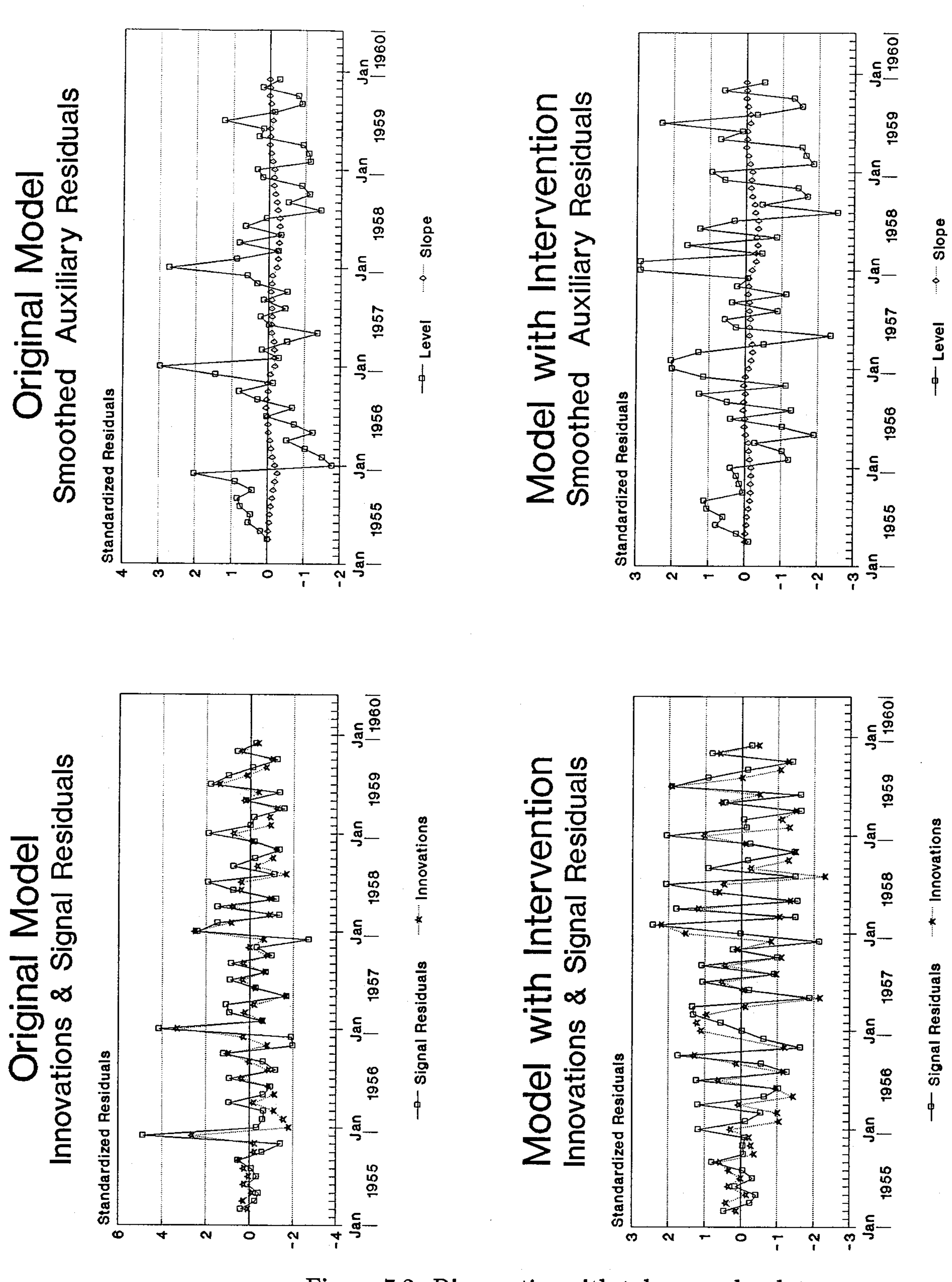

Figure 7.2: Diagnostics with tobacco sales data

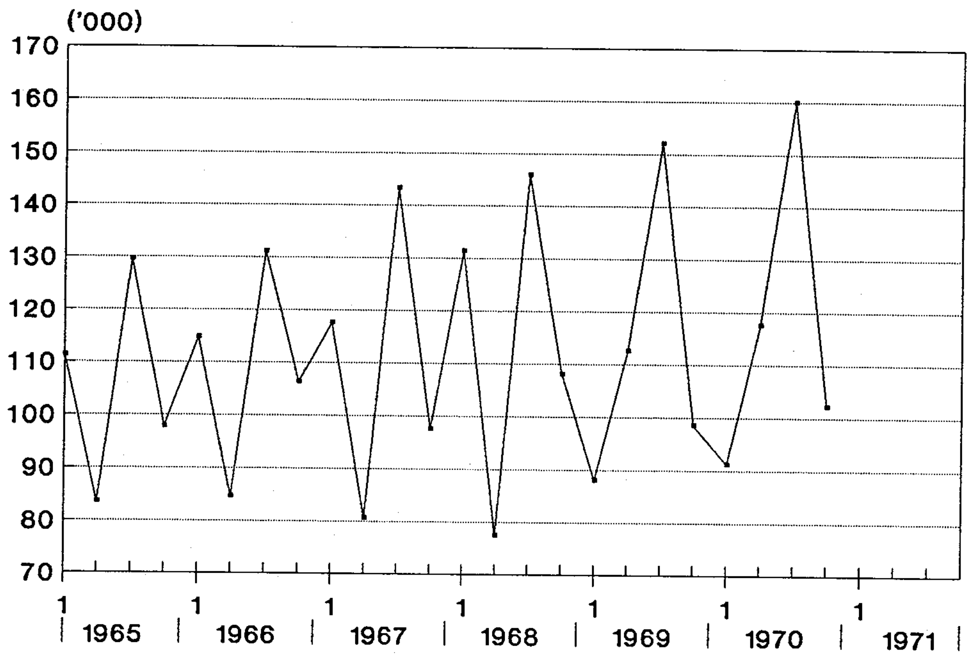

Figure 7.3: UK weddings data

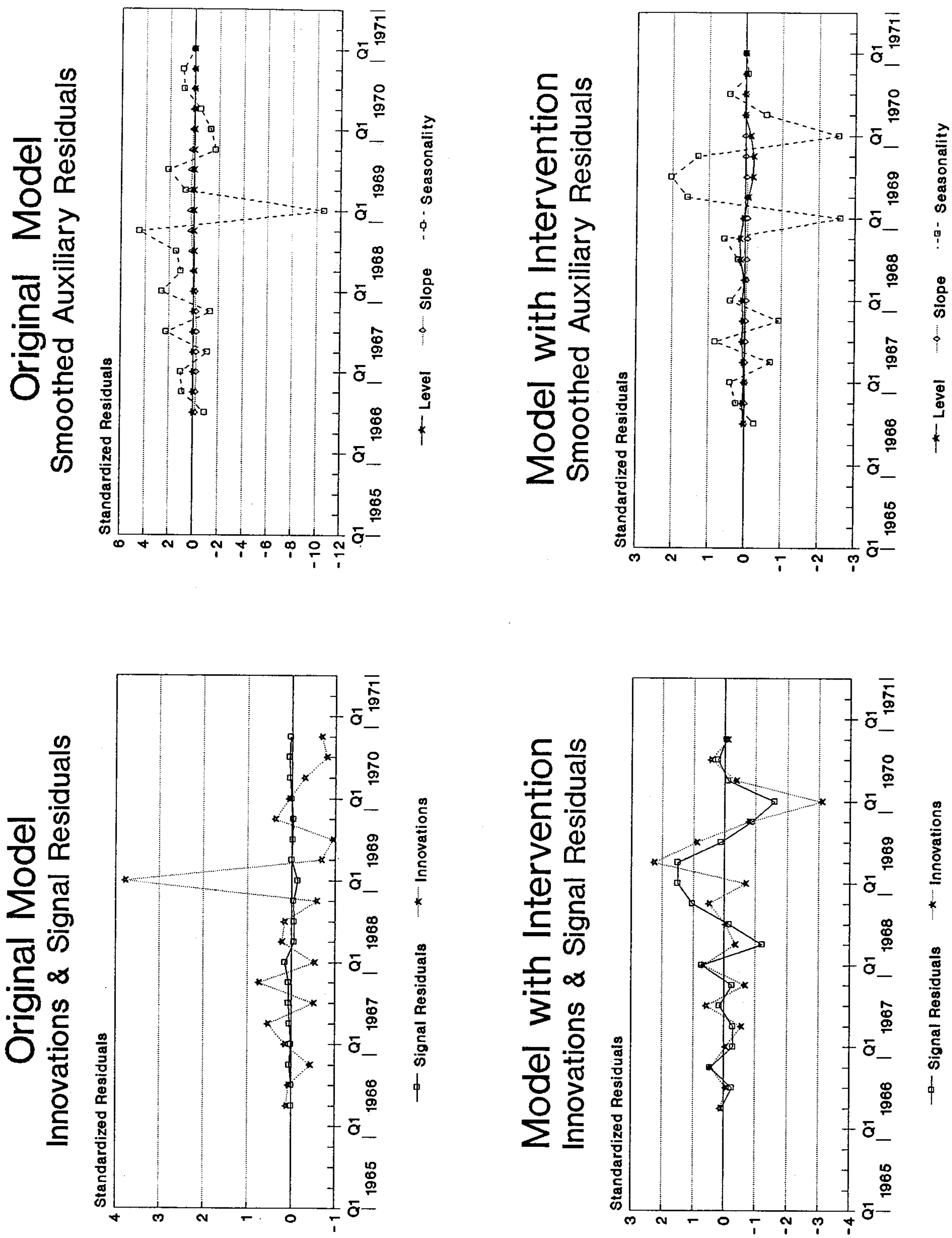

Figure 7.4: Diagnostics with UK weddings data

Chapter 8

Epilogue

In this thesis, we have shown that the three facets of statistical analysis with the SSM, namely prediction, model-fitting and residual analysis can be conducted in a computationally sound and efficient manner using the DKF, which is the Kalman Filter extended in order to handle diffuse effects in the SSM.

We have illustrated throughout the thesis the conceptual and computational advantages of our definition of the SSM over the standard definition employed in the literature, namely the ASSM. Recursive algorithms based on the latter have been shown to be inefficient when the states in the ASSM accomodate a regression parameter and/or initial diffuse effects since they then require recursions of larger error covariance matrices than the DKF. Another disadvantage of including a regression parameter in the state arises in the smoothing cycle wherein the smoothed estimator of the regression parameter is effectively not updated since it coincides with its final estimator in the filtering cycle.

We have discussed practical issues concerning the use of the DKF namely its initialization and its efficient implementation. We have shown how the DKF could be initialized in the general case as well as in particular instances such as when the SSM is time invariant. We have implemented a collapsing strategy in the DKF whereby columns of pertinent matrices related to diffuse initial conditions were factored out. This revised DKF, labelled the CDKF, which coincides with the KF when regression effects are absent from the SSM, has been shown to be computationally more efficient than alternative algorithms considered in the literature.

We have displayed two beneficial consequences of employing a single source of error in the SSM. First, it allowed us to implement a novel and efficient CDKF-EM approach which does not require the estimation of lag-one state error covariance matrices for the maximum likelihood estimation of the covariance matrices of the error vector in the SSM. Second, we showed how the predictor of the error vector conditional on the entire dataset can be obtained recursively. These estimates were then used to detect likely outliers and structural breaks in the SSM.

Appendix : CAPM Dataset

Dataset is kindly provided by Dr. Dilip Madan, University of Maryland. Read across and down. Entries in each column are monthly rate of returns for the three assets. Example : (.00488 ; .00153 ; .00509), (-.01604 ; .02413 ; .03227) etc.

.00488 .00153 .00509	-.01604 .02413 .03227	-.00586 .00869 -.00231
.03378 .07872 .04534	-.00876 .03910 .03543	-.00329 .00237 .01892
.03052 .03892 .04550	.00753 -.00653 -.01043	-.05926 -.04643 -.04989
.02093 .02631 .00987	.01396 .04359 .02384	.03798 .04033 .02503
-.05316 -.10617 -.08731	-.00287 .00868 -.00263	.01118 -.03933 -.02236
-.01749 -.01572 -.03292	.03306 .04078 .04443	-.00200 .00302 .01749
-.01202 -.02909 -.03807	.03044 .01588 .01471	-.03827 -.07757 -.07178
-.00321 -.01668 -.01467	.05385 .02855 .04241	.03924 .00750 .04233
.07804 .05111 .06694	.03832 .03008 .04395	.05504 .03143 .01655
.01214 -.01640 -.00129	-.00229 .03824 .04389	-.02598 -.02390 -.03444
.02313 .05060 .02535	.03695 .01951 .02989	-.01909 .01390 -.04414
.03810 -.00859 .00185	.06536 .06916 .02560	-.00498 .00770 .00199
-.04063 -.00709 -.04625	.04792 -.01064 .01806	-.02873 -.00471 .00071
-.05668 -.07087 -.08011	-.07937 -.06558 -.09000	-.07169 -.08016 -.09779
.05346 .08264 .06045	.02504 .03517 .00977	-.03258 -.04971 -.05365
-.01239 .02177 .02360	.12416 .10097 .13546	.01665 .01728 -.01003
.05362 .05869 .06336	-.00766 -.03700 -.03178	.03491 .04262 .03144
.03980 .05834 .05317	-.00132 .05658 .01680	-.01304 -.02987 -.02801
.00026 -.00849 -.01286	.05478 .06585 .07177	-.03505 -.00905 -.00808
.01528 .09369 .02447	-.00694 -.03556 .00030	.03485 -.00061 .02842
-.00065 .01083 .03446	.02394 .01915 .01883	.01734 .02048 .02921
.01836 -.00259 -.01343	.01041 .00977 .01194	-.00302 .01548 .00254
.01987 .05097 .02117	-.00663 .01181 -.01976	.02140 .04828 .04824
.01233 .00143 .00728	-.00252 -.00407 -.00405	-.01540 .01553 .00163
.03547 .03677 .05415	-.00468 .00467 .01670	-.00486 .01195 -.01284
.02181 .06228 .02978	-.01176 -.02218 -.00591	-.02801 -.07065 -.06289
.02140 .02766 .01519	.03324 .04000 .02734	.01984 .07106 .02728
.02561 .05777 .03724	.00288 -.00590 -.00740	.00375 .01148 .03698
-.00022 .02053 .03021	-.02186 -.00538 -.01263	-.04000 -.02890 -.02684
.01958 .02128 .01063	-.03393 -.08579 -.05524	-.00681 -.00701 -.01778
-.00644 -.02633 -.01106	-.08353 -.07456 -.07760	.03477 -.01601 -.03141
.08750 .00799 .03519	.01378 -.00466 .02407	.01034 -.01518 .01248

.08576	.11789	.09994	-.00356	.00238	-.00381	.02187	.04189	.06165
.02738	.08751	.03871	-.04347	-.05871	-.03825	.03894	.03532	.01706
.06433	.09556	.06151	.00057	-.00641	-.00260	.01146	.04519	.03042
-.04199	-.03762	-.05762	.00738	-.00395	-.00186	.04100	.01968	.04843
.00410	-.07368	-.05400	-.03117	-.03256	-.03703	.00167	.00429	.00699
.07802	.12163	.09577	.01883	.00341	.02061	.05541	-.00814	.00648
-.01269	-.03754	-.03446	.02933	.01418	.02022	.07839	.05467	.04417
.04046	.02523	.02390	.04321	.02152	.05921	-.04785	-.04135	-.01939
.01004	-.02060	-.00440	-.07823	-.04737	-.05577	.02161	.03755	.02940
.01957	.00660	.02581	.00013	-.00600	-.00227	-.08081	-.06860	-.08060
-.08161	-.07171	-.03729	.08002	.03335	.05140	-.01448	.00070	-.01293
.07821	.05345	.07560	-.05321	-.04181	-.01976	-.02572	-.03004	-.00413
-.08882	-.08982	-.06666	.10130	.05973	.07073	-.00295	.01695	-.01166
-.12047	-.11071	-.08238	-.06444	-.07420	-.05777	-.03412	-.04836	-.03248
.08884	.08788	.05401	.05562	.07688	.02984	.03772	.04152	.04197
-.04746	-.03388	-.01826	.06190	.06513	.03061	.08108	.06549	.06844
.05034	.05716	.04601	.01776	.02397	.00731	.06776	.04626	.03572
.02161	.07846	.02923	-.04848	-.03633	-.02824	.00132	-.00480	.00073
-.01385	-.06055	-.04150	.05350	.08663	.04744	-.01083	-.00079	-.00814
-.02845	-.06168	-.05308	.01049	.00316	-.00681	.06075	.08147	.08777
.01587	.03794	.02355	.00295	.02031	.05592	.03640	.02434	.00923
.01766	.00258	.01202	-.00446	-.00223	.02148	-.03220	-.03823	-.01816
-.00968	-.02122	.01487	.06204	.03414	.03575	-.00719	-.01242	-.01785
.02468	-.01991	.01639	.05275	.07311	.04180	-.00634	.01638	.01743
-.05566	-.06866	-.01700	-.05875	-.05895	-.00979	.00247	-.03338	.01001
-.06642	-.06723	-.03791	-.02465	-.02966	-.00280	.01495	-.03433	-.00290
.07761	.09048	.07028	-.01715	-.03953	-.03121	.07600	.08756	.04532
-.01144	-.04287	.02047	-.11437	-.16418	-.11602	.04143	-.03505	.01387
-.00687	.05354	.00202	.01101	-.00359	.01046	-.04428	-.02153	-.00342
-.07467	-.03801	-.01158	-.08675	-.00969	-.02994	-.05663	-.03825	.00311
-.05316	-.10298	-.06862	-.10610	-.06650	-.09100	-.03900	-.12378	-.13364
.20065	.06390	.21557	-.01773	-.03795	-.04117	-.01950	-.03489	-.04007
.15402	.19543	.08092	-.01292	.06492	.09249	.01155	.07801	.04573
.01600	.05121	.08181	.07187	.04990	.03656	.07908	.09545	.02840
-.07351	-.04895	-.07783	-.05480	-.03188	-.00615	-.05120	-.03301	-.06751
.04024	.07696	.06146	.04050	.03322	.03347	-.00022	-.00808	-.02513
.11204	.17004	.14620	.00994	.03371	-.00557	.03755	.03176	.02528
-.01916	.00492	-.01265	-.01188	-.01690	-.01819	.05937	.06495	.02745
.00578	-.01767	-.01495	.00278	-.01268	-.01229	.01784	.03969	.02155
-.00123	-.03328	-.02646	.02182	.00406	-.03885	.08856	.08683	.06774

-.04467	-.04224	-.07186		-.02559	-.02601	-.00299		-.00703	-.01835	-.00879
.02260	.02803	-.01452		-.00744	-.00192	-.04461		.05775	.05226	.03151
.00137	-.02982	-.03608		-.02434	-.01408	-.00400		-.00203	.00103	-.01689
-.04581	-.04159	-.03653		.04768	.03491	.04258		-.00286	-.00536	.00713
-.06845	-.05918	-.04970		-.00163	-.00023	-.03876		.05136	.06597	.03842
.08680	.10346	.09831		.00535	.02346	.02888		-.01164	-.01443	-.01531
.04755	.06836	.07846		.04396	.02872	.02753		.00451	-.01772	-.01335
-.10496	-.11840	-.09411		.01901	.02536	.01811		.01491	.00304	.01715
.04939	.04795	.05707		-.03785	-.05004	-.03385		.07242	.06531	.06813
.01981	.01795	-.00873		.00690	-.02123	-.02298		.07885	.04290	.03175
.03206	.00990	.01048		.04549	.06756	.08332		-.00381	.00176	.00860
-.07246	-.08609	-.08062		.06894	.02200	.04872		.00871	.04504	.04777
.03404	.10487	.04552		.00016	-.05645	-.00775		-.08970	-.09564	-.10789
.06077	-.00218	.03377		.06792	.05224	.07461		.06707	.01880	.01899
.02749	.12732	.09354		.01182	.02858	.00888		.04016	.03769	.02494
.03299	-.00331	-.01642		.10915	.09398	.06943		-.02087	-.01321	-.00811
-.05670	-.04634	-.00859		.00780	.03459	.03477		.06955	.09137	.07477
-.04553	.02893	-.01914		.01606	.04328	-.00204		.01715	-.06166	-.01682
-.00292	-.03099	-.02681		-.05553	-.07644	-.05209		-.02737	-.05894	-.06538
.06408	.02972	.03664		.04925	.02293	.05434		-.04413	-.00628	-.02378
-.03266	-.00285	-.02933		-.04998	-.05334	-.03777		-.01197	-.00045	-.03072
.04914	.05548	.04331		-.05096	-.03713	-.04824		-.03189	.00852	-.02857
-.04256	-.00453	-.01856		.12889	.14373	.14028		.04524	-.01120	-.00262
.14788	.16851	.11336		.04919	.10020	.03990		-.02211	.02439	.00840
.00280	.04122	.04850		.04167	.03691	.02399		.06484	.01443	.02524
.09146	.08260	.06611		-.00614	.00897	.00906		.01555	.06979	.01632
-.03127	-.04357	-.02216		.01568	-.02049	.02952		.00582	.04651	.00502
-.02396	-.01691	-.01642		.03691	.05742	.03401		-.00883	.00552	-.02902
.00616	-.04841	-.02854		-.03884	-.06501	-.03690		.01581	.01272	.03733
-.01160	.00939	-.00634		-.05965	-.07030	-.06563		.01397	.03906	.01896
-.01098	-.00468	-.03757		.11213	.12325	.10835		.03535	-.01558	-.03242
.00518	.01440	.00969		-.00930	-.03926	-.00139		.03756	.03901	.02083
.08983	.10779	.06663		.02262	-.00023	.00862		.00009	-.04152	.00087
.02094	-.03234	-.01243		.05626	.05513	.06738		.02621	.01885	.01066
-.02228	.01388	.02802		-.00309	-.01907	-.00914		-.04015	-.04467	-.03129
.07113	.02856	.03426		.06854	.07975	.08742		.04794	.06467	.04023
.02569	.01368	.01490		.07963	.08708	.09132		.06504	.07057	.07876
-.03602	-.00694	-.01509		.05076	.01318	.06446		.01509	-.00566	.02751
-.04740	-.08905	-.05804		.09072	.07577	.06013		-.09890	-.06565	-.08089
.04393	.05273	.07335		.00046	.03024	.02335		-.01890	-.01764	-.01138

Read across and down. Entries are monthly rate of returns for the CRSP equally-weighted and value-weighted market indices. Example : (.00958 ; .04149), (.01120 ; .02747) etc.

.00958 .04149	.01120 .02747	.00490 .01280	.03917 .02559
.02004 .00758	.00121 .00450	.03427 .03015	-.01160 -.01508
-.04475 -.04589	.01514 .02322	.01822 .01486	.02902 .02124
-.06726 -.03910	.01246 .00813	-.01212 -.02447	-.01548 -.01919
.03328 .02533	.02313 .02136	-.02077 -.01807	.03004 .03495
-.05811 -.05880	-.00457 -.02299	.04793 .04659	.04811 .03740
.06446 .08239	.03709 .05986	.03096 .05042	.00586 .00941
.02589 .04176	-.02848 -.04239	.03070 .01154	.02716 .02112
-.01881 -.02979	.02714 .02110	.04609 .04630	.00067 -.00351
-.03626 -.00789	.01904 .01544	-.00564 -.00594	-.06273 -.06806
-.08452 -.09804	-.08265 -.08497	.06608 .06378	.02287 .02800
-.05007 -.06058	.00418 -.02184	.11182 .13783	.01306 -.00914
.05129 .07819	-.02253 -.01477	.03421 .02083	.04789 .03845
.02038 .03268	-.01801 -.01577	-.00185 -.00968	.05408 .05094
-.01264 -.01950	.02937 .01633	-.00508 -.00766	.02273 .00772
.02590 .02024	.01766 .02661	.01767 .03179	.00432 -.00348
.01690 .01230	.01624 .01530	.01975 .02772	-.01140 -.00903
.03052 .03699	.00953 .01720	.00247 .00074	.00384 -.00693
.03785 .05926	.00710 .02795	-.01069 .00512	.03407 .03590
-.00434 -.00784	-.05035 -.07439	.01711 .02900	.03009 .04497
.03223 .03212	.02878 .04744	.00165 .02972	.01233 .03212
.01005 .04228	-.01024 .01108	-.02131 -.02182	.02371 .03372
-.05109 -.07242	-.01112 -.04999	-.01208 -.01208	-.07461 -.09325
-.00668 -.01355	.04583 .01299	.01666 .03814	.00444 .01643
.08330 .14337	.00992 .02088	.04301 .05193	.04196 .03708
-.04143 -.01790	.02356 .05147	.04825 .07033	-.00604 .00298
.03328 .03782	-.02801 -.03594	.00770 .00690	.03098 .05574
-.03890 -.00343	-.03130 -.04138	.00681 -.00421	.08971 .11645
.02336 .05951	.01181 .01845	-.02166 -.02687	.01654 .02810
.04174 .05859	.01047 .01620	.05739 .07289	-.03696 -.01562
-.00722 -.00923	-.05026 -.07071	.03117 .01889	.02127 .01059
.00304 -.00257	-.06235 -.09557	-.06301 -.07997	.05017 .05010
-.02231 -.01632	.05504 .07636	-.03151 -.04740	-.01766 -.04501
-.07634 -.05607	.05956 .05238	-.00269 -.00903	-.09977 -.12998
-.06159 -.08786	-.05059 -.07009	.07460 .07432	.04980 .06828
.04258 .08838	-.01576 -.04624	.05260 .03279	.06173 .08773
.04965 .09963	.01475 .02573	.04401 .05361	.03396 .03403
-.03662 -.04224	.00428 -.00845	-.04064 -.04601	.04228 .05205
-.00568 -.01265	-.03962 -.05032	.00018 -.02168	.09056 .11232
.02391 .06023	.03049 .03000	.00922 .00171	.00628 .00536
.01724 -.00908	-.02185 -.03300	-.00188 -.02258	.03791 .02915

-.00655 -.02543		.01038 -.00174		.04951 .06556		.01105 -.01313
-.02545 -.05051		-.04021 -.06447		-.00534 -.02362		-.04635 -.06336
-.01876 -.06291		-.00852 -.03135		.05210 .10438		-.03026 -.03454
.05259 .10404		-.00139 -.00911		-.11611 -.16980		.01522 -.01002
-.00057 .10283		.00318 .01038		-.02419 -.01766		-.04331 -.05963
-.03503 -.07316		-.01896 -.03387		-.07276 -.04621		-.08537 -.09296
-.11028 -.07791		.16800 .11779		-.04017 -.04206		-.02350 -.06631
.13483 .30024		.06037 .03453		.02901 .07982		.04685 .03106
.05499 .06829		.05150 .07621		-.06358 -.04080		-.02055 -.04364
-.03606 -.04109		.06086 .03704		.03128 .03112		-.01018 .00048
.12524 .19089		.00098 .07222		.02971 .01084		-.01099 -.01497
-.00907 -.01864		.04749 .04847		-.00737 -.00001		.00053 -.01354
.02572 .02195		-.02128 -.02419		.00515 .03402		.05806 .09372
-.03965 -.00069		-.01677 -.01623		-.01079 .00368		.00384 .01592
-.01238 -.00540		.05119 .06241		-.01548 -.00838		-.01412 -.01456
.00040 .00639		-.03944 -.02840		.04234 .07877		.00553 .00341
-.05739 -.03637		-.01217 .00650		.03182 .06494		.08347 .07615
.01896 .04454		-.01323 -.00604		.05683 .05966		.03755 .06751
-.00660 -.00768		-.10221 -.16731		.03170 .04482		.01652 .01170
.04721 .08699		-.02897 -.02915		.06199 .08643		.00680 .01737
-.01488 -.00773		.04467 .05410		.01532 .02813		.06299 .07851
-.00037 -.00908		-.06925 -.10039		.06059 .07540		.02280 .04380
.06193 .06412		-.00344 -.03120		-.10759 -.13698		.04880 .06376
.05840 .07789		.03345 .03966		.06848 .09627		.02003 .04153
.02906 .02628		.01965 .01800		.10769 .05668		-.03369 -.02104
-.04353 -.00337		.01833 .01956		.04297 .08058		-.01642 .01191
.00814 .01993		-.00789 -.00412		.00075 -.01852		-.05598 -.06503
-.05570 -.06441		.05729 .07096		.04585 .04131		-.02781 -.02020
-.02207 -.02273		-.04924 -.03775		-.00833 -.00122		.04188 .04788
-.02911 -.03329		-.01988 -.02074		-.02112 -.01444		.12520 .11348
.01264 .02535		.11569 .13465		.04713 .07346		.01615 .01450
.03695 .05305		.02794 .04766		.03343 .04331		.07215 .06774
.00373 .05105		.03830 .03808		-.03039 -.02021		.01239 -.00868
.01749 .02427		-.01824 -.03523		.02563 .04803		-.00821 -.01127
-.00888 -.00610		-.03688 -.04973		.01669 .01977		.00526 -.00599
-.05129 -.04961		.02325 .02578		-.01565 -.03727		.11144 .11793
.00205 .00682		.00331 -.00533		-.00938 -.01677		.02517 .01825
.07950 .10633		.01661 .02128		-.00037 -.00828		-.00277 -.01114
.05872 .04541		.01719 .01406		-.00351 .01465		-.00463 -.00294
-.03667 -.04905		.04462 .03495		.06884 .06549		.04554 .04245
.00737 .01326		.07374 .07056		.05560 .05967		-.01322 -.00631
.05146 .03973		.01509 .00034		-.05480 -.07165		.07312 .06207
-.07957 -.05350		.05402 .04518		.01857 .00726		-.02677 -.01967

Bibliography

[1] Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control*, **13** , 162-173.

[2] Anderson, B.D.O and Moore, J.B. (1979). *Optimal Filtering.* Englewood Cliffs, New Jersey : Prentice Hall.

[3] Ansley, C.F. and Kohn, R. (1984). On the estimation of ARIMA models with missing values. In *Time Series Analysis of Irregularly Observed Data*, Parzen E. (ed), New York : Springer-Verlag.

[4] Ansley, C.F. and Kohn, R. (1985a). A structured state space approach to computing the likelihood of an ARIMA process and its derivatives. *Journal of Statistical Computation and Simulation*, **21** , 135-169.

[5] Ansley, C.F. and Kohn, R. (1985b). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *Annals of Statistics*, **13** , 1286-1316.

[6] Ansley, C.F. and Kohn, R. (1987). Efficient generalised cross-validation for state space models. *Biometrika*, **74** , 139-148.

[7] Ansley, C.F. and Kohn, R. (1990). Filtering and smoothing in state space models with partially diffuse initial conditions. *Journal of Time Series Analysis*, **11** , 275-293.

[8] Bell, W. and Hillmer, S. (1991). Initializing the Kalman Filter for nonstationary time series models. *Journal of Time Series Analysis*, **12** , 283-300.

[9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedascity. *Journal of Econometrics*, **37** , 307-327.

[10] Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis : Forecasting and Control.* San Francisco : Holden Day.

[11] Boyles, R.A. (1983). On the convergence of the EM algorithm. *Journal of the Royal Statistical Society*, Series B, **45** , 47-50.

[12] Brown, C.B. (1988). *A Second Course In Linear Algebra.* New York : Wiley.

[13] Brown, R.L., Durbin, J. and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, Series B, **37** , 141-192.

[14] 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.

[15] Burridge, P. and Wallis, K.F. (1985). Calculating the variance of seasonally adjusted series. *Journal of the American Statistical Association*, **80** , 541-552.

[16] Chen, N.F., Roll, R. and Ross, S. (1986). Economic forces and the stock market. *Journal of Business*, **59** , 383-403.

[17] Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* New York : Chapman and Hall.

[18] Cooper, D.M. and Thompson, R. (1977) A note on the estimation of parameters of the autoregressive-moving average process. *Biometrika*, **64** , 625-628.

[19] Crafts, N.F.R. Leybourne S.J. and Mills T.C. (1989). Trends and cycles in British industrial production 1700-1913. *Journal of the Royal Statistical Society*, Series A, **152** , 43-60.

[20] De Jong, Piet (1988a). The likelihood for a state space model. *Biometrika*, **75** , 165-169.

[21] De Jong, Piet (1988b). A cross-validation filter for time series models. *Biometrika*, **75** , 594-600.

[22] De Jong, Piet (1989). Smoothing and interpolation with the state space model. *Journal of the American Statistical Association*, **84** , 1085-1088.

[23] De Jong, Piet (1991a). Stable algorithms for the state space model. *Journal of Time Series Analysis*, **12** 143-157.

[24] De Jong, Piet (1991b). The diffuse Kalman filter. *Annals of Statistics*, **19** , 1073-1083.

[25] De Jong, Piet (1991c). *Linear Time Series for Statistical Analysis and Prediction.* Unpublished manuscript.

[26] Dempster, A.P. Laird, N.M. and Rubin D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B, **39** , 1-38.

[27] Den Butter, F.A.G. and Mourik, T.J. (1990). Seasonal adjustment using structural time series models : an application and a comparison with the Census X-11 method. *Journal of Business and Economic Statistics*, **8** , 385-394.

[28] Downing, D.J. Pike, D.H. and Morrison G.W. (1980). Application of the Kalman Filter to inventory control. *Technometrics*, **22** , 17-22.

[29] Duncan, D.B. and Horn, S.D. (1972). Linear dynamic recursive estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association*, **67** , 815-821.

[30] Engle, R.F. (1982). Autoregressive conditional heteroskedascity with estimates the variance of of U.K. inflation. *Econometrica*, **50** , 987-1007.

[31] Gardner, G. Harvey, A.C. and Phillips, G.D.A. (1980). Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Applied Statistics*, **29** , 311-317.

[32] Granger, C.W.T. and Andersen, A. (1978). An introduction to bilinear time series models. Gottingen : Vandenhoeck and Ruprecht.

[33] Haggan, V. and Ozaki T. (1981). Modelling non-linear random vibrations using an amplitude- dependent autoregressive time series model. *Biometrika*, **68** , 189-196.

[34] Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society*, Series B, **38** , 205-247.

[35] Harvey, A.C. (1981). *Time series models.* New York : Wiley.

[36] Harvey, A.C. (1984). A unified view of statistical forecasting (with discussion). *Journal of Forecasting*, **3** , 245-283.

[37] Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter.* Cambridge : Cambridge University Press.

[38] Harvey, A.C. and Durbin J. (1986). The effects of seat belt legislation on British road casualties : a case study in structural time series modelling. *Journal of the Royal Statistical Society*, Series A, **149** , 187-227.

[39] Harvey, A.C. and Fernandes, C. (1989). Time series models for count or qualitative observations (with discussion). *Journal of Business and Economic Statistics*, **7** , 407-422.

[40] Harvey, A.C. and Koopman, S.J. (1991). Diagnostic checking of unobserved components time series models. Research memorandum T1-1191/66. Tinbergen Instituut.

[41] Harvey, A.C. and Peters, S. (1990) Estimation procedures for structural time series models. *Journal of Forecasting*, **9** , 89-108.

[42] Harvey, A.C. and Pierse, R.G. (1984) Estimating missing observations in economic time series. *Journal of the American Statistical Association*, **79** , 125-131.

[43] Harvey, A.C. and Phillips, G.D.A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, **66** , 49-58.

[44] 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.

[45] Helzer, G. (1983). *Applied Linear Algebra with APL.* Toronto : Little, Brown and Company.

[46] Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory.* New York : Academic Press.

[47] Kalbfleish, J.D. and Sprott, D.A. (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society*, Series B, **32** , 175-194.

[48] Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82** , 34-45.

[49] Kalman, R.E. and Bucy, R.S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, **83** , 95-108.

[50] Kitagawa, G. (1981). A nonstationary time series model and its fitting by a recursive filter. *Journal of Time Series Analysis*, **2** , 103-116.

[51] Kitagawa, G. (1987). Non-Gaussian state space modelling of nonstastionary time series. (with discussion) *Journal of the American Statistical Association*, **82** , 1032-1044.

[52] Kitagawa, G. (1989). Non-Gaussian state space modelling of time series. *1989 Proceedings of the American Statistical Association, Business and Economic Statistics Section.*

[53] Kohn, R. and Ansley, C.F. (1986). Estimation, prediction and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, **81** , 751-761.

[54] Kohn, R. and Ansley, C.F. (1987a). A new algorithm for spline smoothing based on smoothing a stochastic process. *SIAM Journal on Scientific and Statistical Computation*, **8** , 33-48.

[55] Kohn, R. and Ansley, C.F. (1987b). Signal extraction for finite nonstationary time series. *Biometrika*, **74** , 411-421.

[56] 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.

[57] Koopman, S.J. (1991). Efficient smoothing algorithms for time series models. Discussion paper. Department of Statistics, London School of Economics.

[58] Laird, N. , Lange, N. and Stram, D. (1987). Maximum likelihood computations with repeated measures : Application of the EM algorithm. *Journal of the American Statistical Association*, **82** , 97-105.

[59] Lancaster, P. and Tismenetsky, M. (1985). *The Theory of Matrices*. New York : Academic Press.

[60] Lauritzen, S.L. (1981). Time series analysis in 1880 : A discussion of contributions made by T.N. Thiele. *International Statistical Review*, **49** , 319-331.

[61] Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*, Series B, , **44** , 226-233.

[62] Mehra, R.K. (1979). Kalman Filters and their applications to forecasting. *Tims Studies in the Management Sciences*, **12** , 75-94.

[63] Meilijson, I. (1989). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society B*, **51** , 127-138.

[64] Meinhold, R.J. and Singpurwalla, N.D. (1983). Understanding the Kalman Filter. *The American Statistician*, **37** , 123-127.

[65] Meinhold, R.J. and Singpurwalla, N.D. (1987). A Kalman-Filter smoothing approach for extrapolations in certain dose-response, damage-assessment, and accelerated-life-testing studies. *The American Statistician*, **41** , 101-106.

[66] Nicholls, D.F. and Pagan,A.R. (1985). Varying coefficient regression. In *Handbook of Statistics*, Volume 5, Hannan, E.J., Krishnaiah, P.R. and Rao, M.M. (eds). New York : Elsevier Science Publishers B.V.

[67] Orchard, T. and Woodbury, M.A. (1972). A missing information principle : theory and applications. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics*, **1** , 697-715.

[68] Osborn, D.R. (1976). Maximum likelihood estimation of moving average processes. *Annals of Economic and Social Measurement*, **5** , 75-87.

[69] Patterson, H.D. and Thompson, R. (1975). Maximum likelihood estimation of components of variance. *Proceedings Eighth International Biometric Conference*, Corsten L.C.A. and Postelnicu, T. (eds). Bucharest : Academy of the Socialist Republic of Romania.

[70] Pena, D. and Guttman, I. (1988). Bayesian approaches to robustifying the Kalman Filter. In *Bayesian Analysis of Time Series and Dynamic Models*, Spall, J.C. (ed). New York : Marcel Dekker.

[71] Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys (with discussion). *Journal of Business and Economic Statistics*, **9** , 163-177.

[72] Phadke, M.S. (1981). Quality audit using adaptive Kalman Filtering. *ASQC Quality Congress Transactions-San Francisco*, 1045-1052.

[73] Plackett, R.L. (1950). Some theorems in least squares. *Biometrika*, **37** , 149-157.

[74] Pole, A. and West, M. (1989). Reference analysis for the DLM. *Journal of Time Series Analysis*, **10** , 131-147.

[75] Priestley, M.B. (1988). Non-linear and non-stationary time series analysis. London : Academic Press.

[76] 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.

[77] Sallas, W.M. and Harville,D.A. (1981). Best linear recursive estimation for mixed linear models. *Journal of the American Statistical Association*, **76** , 860-869.

[78] Sallas,W.M. and Harville,D.A. (1988). Noninformative priors and restricted maximum likelihood estimation in the Kalman Filter. In *Bayesian Analysis of Time Series and Dynamic Models*, Spall J.C. (ed), New York : Marcel Dekker.

[79] Schweppe, F.C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions in Information Theory*, **11** , 61-70.

[80] Sharpe, W.F. (1964). Capital asset prices : A theory of market equilibrium. *Journal of Finance*, **19** 425-442.

[81] 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.

[82] Shumway, R.H. (1988). *Applied Statistical Time Series Analysis.* Englewood Cliffs, NJ : Prentice Hall.

[83] Shumway, R.H. and Stoffer, D.S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, **3** , 253-264.

[84] Stoffer, D.S. (1982). *Estimation of Parameters in a Linear Dynamic System with Missing Observations.* PhD thesis, University of California, Davis.

[85] Subba Rao, T. (1981). On the theory of bilinear time series models. *Journal of the Royal Statistical Society*, Series B, **43** , 244-255.

[86] Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, **1** , 49-58

[87] Thiele, T.N. (1880). *Sur la compensation de quelques erreurs quasi-systematiques par la methode des moindres carrees.* Kobenhavn : Reitzel.

[88] Tong, H. and Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society*, Series B, **42** , 245-292.

[89] Tsay, R.S. and Tiao, G.C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Annals of Statistics*, **18** , 220-250.

[90] Tunnicliffe Wilson, G. (1989). On the use of marginal likelihood in time series model estimation. *Journal of the Royal Statistical Society*, Series B, **51** , 15-27.

[91] Vardi, Y., Shepp, L.A. and Kaufmann, L. (1985). A statistical model for positron emission tomography (with comments). *Journal of the American Statistical Association*, **80** , 8-37.

[92] Watson, M.W. and Engle, R. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics*, **23** , 385-400.

[93] West, M. and Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models.* New York : Springer Verlag.

[94] Wu, C.F.J. (1983). On the convergence of the EM algorithm. *Annals of Statistics*, **11** , 95-103.