Parameter Estimation in Hidden Markov Models

by

Michael Ryan Roberts

B.A. (University of California, San Diego) 1992

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Master of Arts

in

Statistics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Steven Evans, Chair
Professor David Freedman
Professor Terry Marsh

Fall 2001

The dissertation of Michael Ryan Roberts is approved:

Chair Date

Date

Date

University of California, Berkeley

Fall 2001

Parameter Estimation in Hidden Markov Models

by

Michael Ryan Roberts

Abstract

Parameter Estimation in Hidden Markov Models

by

Michael Ryan Roberts

Master of Arts in Statistics

University of California, Berkeley

Professor Steven Evans, Chair

This thesis discusses parameter estimation in a nonlinear state-space framework with possibly non-Gaussian random components. Parameters are treated from a Bayesian perspective as unobserved random variables. A technique known as the "reference probability method" is developed in which a change of measure decouples the unobserved state process from the measured observation process. Under this "reference" measure, estimates of both the state process and parameters may be computed recursively using Zakai's Equation. The contributions of this thesis include illuminating the application of the technique to economics and finance, a unified treatment of the reference probability method as it pertains to parameter estimation, and derivations of the appropriate measure change and Zakai's equation for a general

hidden Markov model.

Professor Steven Evans
Dissertation Committee Chair

To my family: Lois, Leonard, Patricia and Matthew

Contents

Acknowledgements

Special thanks go to Steve Evans and David Freedman. Both spent numerous hours sharing their wisdom and teaching me how to think about probability and statistics. This work would have suffered greatly were it not for their patience and support. Helpful comments from Joel Hanson, Terry Marsh and Lea Popovic are also noted.

Chapter 1

Introduction

During the past twenty years, there has been a proliferation of nonlinear, non-Gaussian dynamic models in economics and finance. With these more complicated models have come new challenges in parameter estimation. Many techniques rely on approximations to the model obtained by linearizing the specification, imposing normality (quasi-maximum likelihood), or using an auxiliary model (simulated method of moments). The accuracy of such methods depends on the degree of nonlinearity, departure from normality, or similarity of the auxiliary model to the true model.

This thesis discusses, in a possibly nonlinear and non-Gaussian state-space framework, a technique for parameter estimation that avoids such approximations. The technique, originally introduced by DiMasi and Runggaldier (1982), begins by treating the parameters as unobserved random variables. Then, a procedure called the "reference probability method" introduces a change of measure that transforms the original estimation problem so that the observable random variables are independent and, possibly, identically distributed. Estimation is conceptually easier given the simplified structure of the problem, and may be performed recursively using a result called Zakai's equation. Results are interpreted under the original probability measure using a version of Bayes' rule.[1]

Previous work on nonlinear and non-Gaussian state-space models focused on different approximations to either the model or the distributional assumptions. Kitigawa (1987) proposed a piecewise linear approximation to the filtering density of non-Gaussian state-space models. Meinhold and Singpurwalla (1989) proposed approximations involving poly-t-distributions and a recursion for implementing a multivariate t-distribution based on the Kalman filter recursions. Other Bayesian approaches include Alspach and Sorenson (1972), Harrison and Stevens (1976), Smith and West (1983), West et al. (1985) and Gordon and Smith (1990). Carlin et al. (1992) present a Markov Chain Monte Carlo approach to estimating non-Gaussian and nonlinear state space models. More recent work in this area has focused on recursive estimation methods. Some examples include Pitt and Shephard (1997), Doucet (1998), Doucet et al. (2000), and Liu and West (2000).

Elliott et al. (1995) provide a textbook treatment of the reference probability method with emphasis on state estimation in special cases of the general model considered here. The primary contributions of this thesis are to illuminate the application

[1] This version of Bayes' rule is commonly referred to as the Kallianpur-Streibel formula in the filtering literature.

of the technique in economics and finance, provide a unified treatment of the reference probability method as it pertains to parameter estimation, and derive the appropriate measure change and Zakai's equation for a general discrete-time state-space framework. As DiMasi and Runggaldier note, the approach does not provide conceptually new results when compared to the traditional Bayesian approach. Instead, the main advantage lies in the simplification of the basic structure of the problem and in the use of Zakai's equation.

Chapters 1 and 2 present background material in filtering theory and parameter estimation in a state-space framework. Linear filtering theory is briefly reviewed, with a discussion of the Kalman filter. A nonlinear model is then examined. Chapter 2 discusses the problem of parameter estimation. First, the special case of estimation in a linear Gaussian model is presented, followed by a discussion of estimation in a nonlinear model. Chapter 3 presents the general theory, followed by an example. Chapter 4 concludes with a brief summary.

Chapter 2

The Filtering Problem

Filtering refers to the process of estimating the state of an unobserved dynamical system from a set of measurements, which may be contaminated by random noise. This chapter begins with a presentation of a linear Gaussian state-space model, which can be estimated using a recursion called the Kalman filter. The goal of this discussion is to provide the reader with an intuitive understanding of a basic filtering problem and its solution. The difficulties encountered in nonlinear specifications are then discussed.

2.1 The Kalman Filter

Consider a univariate random process evolving over time according to the following specification:

$$X_t = \alpha X_{t-1} + \sigma V_t \tag{2.1}$$

Here $\{V_t\}$ is a sequence of i.i.d. standard normal random variables, also called a Gaussian white noise process. Equation (2.1) is defined for time periods $t = 1, ..., T$. The initial state X_0 is assumed to be a standard normal random variable independent of $\{V_t\}$. The parameters α and σ are, for the time being, assumed to be known constants.

Now assume that X_t is not observed. Instead, measurements are made on Y_t:

$$Y_t = X_t + W_t \tag{2.2}$$

where $\{W_t\}$ is a Gaussian white noise process independent of $\{V_t\}$ and defined for $t = 1, ..., T$. Equation (2.1) is called the "state equation" and specifies the dynamic evolution of the latent process $\{X_t\}$.[1] Equation (2.2) is called the "observation equation" and relates the observable process $\{Y_t\}$ to the latent process.[2] Together, these two equations define a state-space, or filtering, framework. This example is also an instance of a hidden Markov model.

One goal of filtering is to recover the unobserved state process from the observations. For a number of loss functions, the optimal estimate of the state process is

[1] The state equation is also referred to as the "signal equation" or "transition equation".
[2] The observation equation is also referred to as the "measurement equation".

its conditional expectation given the observations. The linear Gaussian structure in equations (2.1) and (2.2) allows this conditional expectation to be computed recursively, by a procedure called the Kalman filter. The approach used here to derive this result follows that in Harvey (1989) and avoids the need for Hilbert space theory or optimization methods.[3]

Begin by considering the conditional expectation of the state process at time t given the observations up to time $t-1$. Let $Z^s = (Z_0, ..., Z_s)$ for any process $\{Z_t\}$ and let $z^s = (z_0, ..., z_s)$ denote a realization. The conditional mean of X_t given $Y^{t-1} = (Y_0, ..., Y_{t-1})$ is given by

$$\mathbb{E}\{X_t | Y^{t-1}\} = \alpha \mathbb{E}\{X_{t-1} | Y^{t-1}\} \tag{2.3}$$

The next step is to derive an expression for the contemporaneous conditional mean in (2.3). Begin by considering the joint distribution of (X_t, Y_t). At $t = 1$, the state and observation equations are

$$X_1 = \alpha X_0 + \sigma V_1 \tag{2.4}$$

$$Y_1 = X_1 + W_1 \tag{2.5}$$

From (2.4) and (2.5) it is clear that X_1 and Y_1 are jointly normally distributed with

[3]In his original derivation, Kalman (1960) used the concept of orthogonal projection. Other derivations have utilized recursive least squares (e.g. Swerling 1959, Gainer 1963, Fagin 1964, Mowery 1965, and Bryson and Ho 1969), maximum likelihood (e.g. Ho 1962, Schmidt 1966, Rauch et al. 1965, and Smith et al. 1962) and the linear minimum variance estimator (e.g. Battin 1962, Battin 1964, Schmidt 1966, and Sorenson 1966).

mean vector

$$(\alpha\mathbb{E}\{X_0\}\,,\alpha\mathbb{E}\{X_0\})' \tag{2.6}$$

and covariance matrix

$$\begin{bmatrix} \alpha^2\mathrm{Var}\{X_0\}+\sigma^2 & \alpha^2\mathrm{Var}\{X_0\}+\sigma^2 \\ \alpha^2\mathrm{Var}\{X_0\}+\sigma^2 & \alpha^2\mathrm{Var}\{X_0\}+\sigma^2+1 \end{bmatrix} \tag{2.7}$$

As a consequence of joint normality, X_1 is conditionally Gaussian given Y_1. The conditional mean is [4]

$$\begin{aligned} \mathbb{E}\{X_1\,|\,Y^1\} &= \mathbb{E}\{X_1\}+\frac{\mathrm{Cov}\{X_1,Y_1\}}{\mathrm{Var}\{Y_1\}}\,(Y_1-\mathbb{E}\{Y_1\}) & (2.8)\\ &= \alpha\mathbb{E}\{X_0\}+\frac{\alpha^2\mathrm{Var}\{X_0\}+\sigma^2}{\alpha^2\mathrm{Var}\{X_0\}+\sigma^2+1}\,(Y_1-\alpha\mathbb{E}\{X_0\}) & (2.9)\end{aligned}$$

Noting that $\mathbb{E}\{X_1\,|\,Y^0\}=\mathbb{E}\{X_1\}$ enables (2.9) to be written as

$$\mathbb{E}\{X_1\,|\,Y^1\}=\mathbb{E}\{X_1\,|\,Y^0\}+\frac{\alpha^2\mathrm{Var}\{X_0\}+\sigma^2}{\alpha^2\mathrm{Var}\{X_0\}+\sigma^2+1}\left(Y_1-\mathbb{E}\{X_1\,|\,Y^0\}\right) \tag{2.10}$$

Equation (2.10) expresses the contemporaneous conditional expectation $\mathbb{E}\{X_1\,|\,Y^1\}$ as a function of the previous period's one-step prediction of the state, $\mathbb{E}\{X_1\,|\,Y^0\}$. Continuing this process for $t=2,3,..,T$ will establish a recursion for the conditional mean. However, this expression would quickly become messy because of the second moments in the expression for the conditional mean. The next step, therefore, is to show that the second order moments in (2.8) can be written as a set of recursions as well.

[4]See Harvey (1989) for a discussion of the relevant properties of normal random variables.

Following the approach used to derive the recursion for the conditional mean,

$$\begin{aligned}\mathrm{Var}\{X_t\,|\,Y^{t-1}\} &\equiv \mathbb{E}\Big\{\big(X_t - \mathbb{E}\{X_t\,|\,Y^{t-1}\}\big)^2\Big|\,Y^{t-1}\Big\}\\ &= \mathbb{E}\Big\{\big(\alpha\left[X_{t-1} - \mathbb{E}\{X_{t-1}\,|\,Y^{t-1}\}\right] + \sigma V_t\big)^2\Big|\,Y^{t-1}\Big\}\\ &= \alpha^2\mathrm{Var}\{X_{t-1}\,|\,Y^{t-1}\} + \sigma^2 \qquad (2.11)\end{aligned}$$

Note that, by construction, the estimation error is orthogonal to the space spanned by Y^{t-1}. In a Gaussian framework, orthogonality is equivalent to independence so that the conditional variance above is equal to the unconditional variance.

Returning to equation (2.10), the contemporaneous mean at time $t = 1$ may now be written in terms of $\mathrm{Var}\{X_1\,|\,Y^0\}$ using the result in (2.11).

$$\mathbb{E}\{X_1\,|\,Y^1\} = \mathbb{E}\{X_1\,|\,Y^0\} + \frac{\mathrm{Var}\{X_1\,|\,Y^0\}}{\mathrm{Var}\{X_1\,|\,Y^0\} + 1}\big(Y_1 - \mathbb{E}\{X_1\,|\,Y^0\}\big)$$

Continuing this recursion over time yields the general update formula for the conditional mean,

$$\mathbb{E}\{X_t\,|\,Y^t\} = \mathbb{E}\{X_t\,|\,Y^{t-1}\} + \frac{\mathrm{Var}\{X_t\,|\,Y^{t-1}\}}{\mathrm{Var}\{X_t\,|\,Y^{t-1}\} + 1}\big(Y_t - \mathbb{E}\{X_t\,|\,Y^{t-1}\}\big) \qquad (2.12)$$

The remaining issue is computation of the contemporaneous error variance in equation (2.11), that is, $\mathrm{Var}\{X_{t-1}\,|\,Y^{t-1}\}$. Beginning at time $t = 1$, it was shown that (X_1, Y_1) is jointly normal (see equations 2.4 and 2.5) with mean vector and covariance matrix given in equations (2.6) and (2.7). Because of the joint normality of X_1 and

Y_1, the conditional variance of X_1 given Y_1 is

$$\begin{aligned}\text{Var}\{X_1|Y^1\} &= \text{Var}\{X_1\} - \frac{\text{Cov}\{X_1, Y_1\}^2}{\text{Var}\{Y_1\}} \\ &= \alpha^2\text{Var}\{X_0\} + \sigma^2 - \frac{(\alpha^2\text{Var}\{X_0\} + \sigma^2)^2}{\alpha^2\text{Var}\{X_0\} + \sigma^2 + 1} \qquad (2.13)\end{aligned}$$

Recognizing that $\text{Var}\{X_1|Y^0\} = \alpha^2\text{Var}\{X_0\} + \sigma^2$ allows (2.13) to be written as:

$$\text{Var}\{X_1|Y^1\} = \text{Var}\{X_1|Y^0\} - \frac{\text{Var}\{X_1|Y^0\}^2}{\text{Var}\{X_1|Y^0\} + 1}$$

Continuing for time $t = 2, 3, ..., T$ yields the general update formula,

$$\text{Var}\{X_t|Y^t\} = \text{Var}\{X_t|Y^{t-1}\} - \frac{\text{Var}\{X_t|Y^{t-1}\}^2}{\text{Var}\{X_t|Y^{t-1}\} + 1} \qquad (2.14)$$

Equations (2.3), (2.11), (2.12), and (2.14) form the Kalman filter. The first two equations (2.3 and 2.11) are called the "prediction" equations and the latter two (2.12 and 2.14) are called the "update" equations. The necessary conditions for this result are:

1. additive Gaussian errors in both the state and observation equations, and

2. a linear specification in the state process.

Under these assumptions, the conditional density function is Gaussian and completely characterized by its mean and variance.

Sometimes a function of the state $g(X_t)$, for example, must be computed. In this case, the conditional density of the state given the observations is needed and not just an estimate of the state itself. The popularity of the Kalman filter is due to the fact that it presents a closed form solution for the update or filtering density, $f_{X_t|Y^t}(x_t|y^t)$. This was implicitly shown above by noting that the conditional distribution of the state given the observations is Gaussian with mean and variance given by (2.12 and 2.14). For almost all other specifications, though a recursive result is still possible, the update density can rarely be found in closed form.

Consider the prediction density, which will be denoted by $f_{X_t|Y^{t-1}}(x_t|y^{t-1})$. The Chapman-Kolmogorov equation says.[5]

$$f_{X_t|Y^{t-1}}\left(x_t|y^{t-1}\right) = \int f_{X_t|X_{t-1}}(x_t|x_{t-1})\, f_{X_{t-1}|Y^{t-1}}\left(x_{t-1}|y^{t-1}\right) dx_{t-1} \tag{2.15}$$

For the model defined in equations (2.1) and (2.2), at time $t = 1$ the prediction density follows from (2.15).[6]

$$f_{X_1|Y^0}\left(x_1|y^0\right) = (2\pi)^{-1/2}\left(\alpha^2+\sigma^2\right)^{-1/2}\exp\left\{-1/2\left(\alpha^2+\sigma^2\right)^{-1}x_1^2\right\} \tag{2.16}$$

This result shows that X_1 is normal with zero mean and variance $(\alpha^2+\sigma^2)$ and it agrees with the prediction equations in (2.3) and (2.11). Continuing for each t produces the general prediction density, $f_{X_t|Y^{t-1}}(x_t|y^{t-1})$, which may be written as:

$$(2\pi)^{-1/2}\left(\alpha^2 \mathrm{Var}\left\{X_{t-1}\,|\,Y^{t-1}\right\}+\sigma^2\right)^{-1/2}\times$$
$$\exp\left\{-1/2\left(\alpha^2 \mathrm{Var}\left\{X_{t-1}\,|\,Y^{t-1}\right\}+\sigma^2\right)^{-1}\left(x_t-\alpha\mathbb{E}\left\{X_{t-1}\,|\,Y^{t-1}\right\}\right)^2\right\} \tag{2.17}$$

The update density may be found recursively using the prediction density in con-

[5]The Chapman-Kolmogorov equation and its derivation may be found in Jazwinski (1970). For ease of reference, the derivation is reproduced in the appendix.

[6]A derivation of (2.16) may be found in the appendix.

junction with[7]

$$f_{X_t|Y^t}(x_t|y^t) = \frac{f_{Y_t|X_t}(y_t|x_t)\, f_{X_t|Y^{t-1}}(x_t|y^{t-1})}{f_{Y_t|Y^{t-1}}(Y_t|y^{t-1})} \tag{2.18}$$

Postponing the derivation to the appendix, the update density at time $t = 1$ is

$$f_{X_1|Y^1}(x_1|y^1) = (2\pi)^{-1/2}\left(\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1}\right)^{-1} \exp\left\{-\frac{1}{2}\left(\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1}\right)^{-1} \times \left[x_1 - \frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1}y_1\right]^2\right\} \tag{2.19}$$

This density corresponds to a normal random variable with mean

$$\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1}y_1$$

and variance

$$\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1}$$

which agree with the update equations in (2.12) and (2.14). Continuing for each t, the general update density, $f_{X_t|Y^t}(x_t|y^t)$, may be written as

$$(2\pi)^{-1/2}\left[\mathrm{Var}\{X_t|Y^{t-1}\} - \frac{(\mathrm{Var}\{X_t|Y^{t-1}\})^2}{(\mathrm{Var}\{X_t|Y^{t-1}\}+1)}\right]^{-1/2} \exp\left\{-\frac{1}{2}\left[\mathrm{Var}\{X_t|Y^{t-1}\} - \frac{(\mathrm{Var}\{X_t|Y^{t-1}\})^2}{(\mathrm{Var}\{X_t|Y^{t-1}\}+1)}\right]^{-1} \times \left[x_t - \left(\mathbb{E}\{X_t|Y^{t-1}\} + \frac{\mathrm{Var}\{X_t|Y^{t-1}\}}{\mathrm{Var}\{X_t|Y^{t-1}\}+1}\left(y_t - \mathbb{E}\{X_t|Y^{t-1}\}\right)\right)\right]^2\right\} \tag{2.20}$$

[7] A derivation of equation (2.18) may be found in the appendix.

2.2 A Nonlinear Filter

Now consider a modification to the observation process in (2.2)

$$Y_t = \exp(X_t)\, W_t \tag{2.21}$$

Equations (2.1) and (2.21) correspond to a popular specification in finance for asset prices called a "stochastic volatility model". In this framework, the observed process Y_t corresponds to the change in an asset price or the return on an asset. The unobserved state X_t is the volatility of the asset. The popularity of this model is due to its generalization of the Black-Scholes option pricing model by allowing for volatility clustering.[8]

The problem that the nonlinearity in (2.21) introduces is most clearly seen by considering the update density, equation (2.18). The denominator of this expression is $f_{Y_t|Y^{t-1}}(y_t|y^{t-1})$. At time $t = 1$, this density may be computed using the Chapman-Kolmogorov equation as

$$f_{Y_1|Y^0}(y_1|y^0) = \int f_{Y_1|X_1}(y_1|x_1)\, f_{X_1|Y_0}(x_1|y_0)\, dx_1 \tag{2.22}$$

The first density in (2.22) follows immediately upon inspection of (2.21). Conditional on X_1, Y_1 is Gaussian with mean 0 and variance $\exp(2X_1)$. The second density is the prediction density of X_1 and is shown above in equation (2.16). Substituting explicit

[8]There are numerous references on stochastic volatility models. Several treatments include Hull and White (1987), Harvey et al. (1994) and Jacquier et al. (1993). For a review of the literature, see Ghysels et al. 1996.

expressions for these densities into (2.22) yields

$$
\begin{aligned}
f_{Y_1|Y^0}\left(y_1|y^0\right) \quad = \quad & \int (2\pi)^{1/2} \exp\left\{-\frac{1}{2}\frac{y_1}{\exp(2x_1)}\right\} \\
& \times (2\pi)^{1/2} \left(\alpha^2+\sigma^2\right)^{-1/2} \exp\left(-\frac{1}{2}\frac{x_1^2}{(\alpha^2+\sigma^2)}\right) dx_1 \qquad (2.23)
\end{aligned}
$$

The integral in equation (2.23) cannot be evaluated in closed form. Since the update density feeds into the prediction density, future prediction densities ($t = 2, ..., T$) will be intractable as well. Thus, in this model and almost all other nonlinear models, the filtering and prediction densities must rely on numerical integration each period.

Chapter 3

Parameter Estimation

In most Economic and Financial applications of state-space models, the parameters of the system are unknown and must be estimated from the data along with the state process. This chapter begins with a presentation of the parameter estimation problem and one common solution in the linear Gaussian case. This is followed by a discussion of the problem in a nonlinear setting and the associated difficulties.

3.1 Linear Gaussian Models

Consider the linear Gaussian model in equations (2.1) and (2.2). Assuming the parameters α and σ are now unknown constants, one estimation technique is maximum likelihood. Let $\theta = (\alpha, \sigma)'$ denote the parameter vector. The likelihood function will be denoted by $l(Y_1, ..., Y_T|\theta)$ where

$$l(Y_1, ..., Y_T|\theta) = f_{Y_1}(y_1) \prod_{s=2}^{T} f_{Y_s|Y^{s-1}}\big(y_s|y^{s-1}, \theta\big) \tag{3.1}$$

Here, f_{Y_1} is the marginal density of the initial observation and $f_{Y_s|Y^{s-1}}$ is the conditional density of Y_s given the path of the process up to the previous time period, Y^{s-1}. Since Y_t is Gaussian, the conditional densities are Gaussian and fully specified by the conditional mean and variance. The conditional mean is

$$\begin{aligned} \mathbb{E}\big\{Y_t\,|\,Y^{t-1}\big\} &= \mathbb{E}\big\{X_t + W_t\,|\,Y^{t-1}\big\} \\ &= \mathbb{E}\big\{X_t\,|\,Y^{t-1}\big\} \end{aligned} \tag{3.2}$$

and the conditional variance is

$$\begin{aligned} \mathrm{Var}\big\{Y_t\,|\,Y^{t-1}\big\} &= \mathbb{E}\Big\{\big(Y_t - \mathbb{E}\big\{Y_t\,|\,Y^{t-1}\big\}\big)^2\Big|\,Y^{t-1}\Big\} \\ &= \mathbb{E}\Big\{\big(X_t - \mathbb{E}\big\{X_t\,|\,Y^{t-1}\big\} + W_t\big)^2\Big|\,Y^{t-1}\Big\} \\ &= \mathrm{Var}\big\{X_t\,|\,Y^{t-1}\big\} + 1 \end{aligned} \tag{3.3}$$

The likelihood function may be written explicitly as

$$l(Y_0, ..., Y_T|\theta) \;=\; (2\pi)^{-T/2} \prod_{s=1}^{T} \exp\left\{-\frac{1}{2}\frac{\big(y_s - \mathbb{E}\{X_s\,|\,Y^{s-1}\}\big)^2}{\mathrm{Var}\{X_s\,|\,Y^{s-1}\} + 1}\right\} \tag{3.4}$$

This is not a closed form expression for the likelihood function since it involves the Kalman filter recursions, which are nonlinear functions of previous estimates. Maximization of the likelihood function may be performed numerically with the Kalman filter supplying estimates of $\mathbb{E}\{X_t\,|\,Y^{t-1}\}$ and $\mathrm{Var}\{X_t\,|\,Y^{t-1}\}$ (and consequently $\mathbb{E}\{X_t\,|\,Y^t\}$ and $\mathrm{Var}\{X_t\,|\,Y^t\}$).

3.2 Nonlinear and Non-Gaussian Models

Now consider performing maximum likelihood estimation in the nonlinear setting of equations (2.1) and (2.21) . The likelihood may still be written as in (3.1), however, evaluation of each term in the product will require numerical integration as indicated by equation (2.23). The resulting likelihood function is a high dimensional integral that must be maximized, a daunting task beyond a few observations.

Though Markov Chain Monte Carlo has been used for the purpose of evaluating high-dimensional integrals, it's accuracy and efficiency is closely tied to the dimension of the problem. In addition, incorporating new observations as they become available is non-trivial since the simulation must be re-run on an even higher- dimensional integral.

Taking a Bayesian approach and treating the parameters as random variables does not avoid the problem of an intractable likelihood function. For example, one might use the posterior expectation of a parameter as a possible point estimate. Unfortunately, computation of this expectation requires the likelihood function, as Bayes' rule makes all too clear.

What DiMasi and Runggaldier (1982) proposed was to treat the parameters of the system as elements of the state process. This enables the use of the filtering recursions so that computation of parameter estimates may be performed sequentially through time. This does not solve the problem of intractable densities, but it does eliminate the need for performing high-dimensional integration. The approach effectively translates a high dimensional integration problem into a series of low dimensional integrals. The remainder of this thesis will focus on this approach.

Chapter 4

The Reference Probability Method

This chapter presents a parameter estimation technique using the reference probability method. It begins by discussing the general theory, followed by an example to illustrate the application.

The basic approach of the technique may be summarized as follows. The parameters of the system are treated as unobserved elements of the state process. A new probability measure is constructed such that the observation process is a sequence of independent random variables. This is accomplished using a discrete time version of Girsanov's theorem to decouple the observation and state equations. Under the new measure, a recursion is derived for an unnormalized conditional density of the state process given the observation path. Parameter estimates (e.g. condtional mean, mode, etc.) are then computed sequentially using the recursion and translated back under the original measure using a version of Bayes' theorem.

4.1 General Setup

The general setup begins with the state equation:

$$X_t = a_t(X_{t-1}, \theta) + B_t(X_{t-1}, \theta)\, V_t \tag{4.1}$$

Equation (4.1) defines an $\mathbb{R}^n$-valued stochastic process $\{X_t\}$ for $t = 1, ...T$. The driving process $\{V_t\}$ is independent over time, $\mathbb{R}^q$-valued, and has corresponding densities $\{\pi_t\}$. These densities need not be Gaussian and may vary over time. The drift function $a_t \in \mathbb{R}^n$ and volatility matrix $B_t \in \mathbb{R}^{n\times q}$ are Borel-measurable for all t. At time $t = 0$, the process starts at $X_0 = V_0$ implying that the density of X_0 is π_0.

The observation equation is given by:

$$Y_t = c_t\big(X_t, Y^{t-1}, \theta\big) + D_t\big(X_t, Y^{t-1}, \theta\big)\, W_t \tag{4.2}$$

Equation (4.2) defines an $\mathbb{R}^m$-valued stochastic process $\{Y_t\}$ for $t = 1, ..., T$. The driving process $\{W_t\}$ is independent over time, $\mathbb{R}^m$-valued, and has corresponding densities $\{\varphi_t\}$. These densities are not necessarily Gaussian and may vary over time as well. The driving processes, $\{V_t\}$ and $\{W_t\}$, are assumed independent of one another. The functions $c_t \in \mathbb{R}^m$ and $D_t \in \mathbb{R}^{m\times m}$ are Borel-measurable for all t. In addition, D_t is assumed to be nonsingular for all t. The parameter vector $\theta \in \mathbb{R}^k$ has density ρ_θ and is assumed independent of both $\{V_t\}$ and $\{W_t\}$. All random variables are defined on a probability triple, $(\Omega, \mathcal{F}, \mathbb{P})$.

4.1.1 Measure Change

This subsection constructs a new probability measure $\mathbb{Q} \ll \mathbb{P}$ such that the observation process is independent of the state process and parameters. Inspection of (4.2) reveals that both the drift term c_t and the volatility term D_t of the observation process must be changed in order to remove the dependence of Y_t on X_t, Y^{t-1}, and θ. In discrete time, changing measures is equivalent to performing a change of variables. Thus, by an appropriate change of variable, both the conditional mean and variance of the process may be altered.

Define $\mathbb{Q}$ as

$$
\begin{aligned}
d\mathbb{Q} &= L_T^{-1} d\mathbb{P}, && (4.3)\\
&= \prod_{s=1}^{T} \psi_s^{-1}(X_s, Y^s, \theta)\, d\mathbb{P} && (4.4)
\end{aligned}
$$

where

$$
\psi_s^{-1}(X_s, Y^s, \theta) = \left|\det\big(D_s\big(X_s, Y^{s-1}, \theta\big)\big)\right| \frac{\varphi_s(Y_s)}{\varphi_s(W_s)} \qquad (4.5)
$$

To show that $\mathbb{Q}$ is indeed a probability measure, first note that $\mathbb{Q}$ is nonnegative by construction. Letting $\mathcal{F}_t^+ = \sigma\left\{X^{t+1}, Y^t, \theta\right\}$

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left\{L_t^{-1} \,\middle|\, \mathcal{F}_{t-1}^+\right\} &= L_{t-1}^{-1}\mathbb{E}_{\mathbb{P}}\left\{\psi_t^{-1}\big(X_t, Y^t, \theta\big) \,\middle|\, \mathcal{F}_{t-1}^+\right\}\\
&= L_t^{-1}\left|\det\big(D_t\big(X_t, Y^{t-1}, \theta\big)\big)\right| \mathbb{E}_{\mathbb{P}}\left\{\frac{\varphi_t(Y_t)}{\varphi_t(W_t)} \,\middle|\, \mathcal{F}_{t-1}^+\right\}\\
&= L_{t-1}^{-1}\left|\det\big(D_t\big(X_t, Y^{t-1}, \theta\big)\big)\right| \int_{\mathbb{R}^m} \frac{\varphi_t(X_t + w_t)}{\varphi_t(w_t)} \varphi_t(w_t)\, dw_t \qquad (4.6)
\end{aligned}
$$

The change of variable given by the observation equation (4.2) implies that the integral in (4.6) may be written as

$$\int_{\mathbb{R}^m} \varphi_t(X_t + w_t)\, dw_t = \int_{\mathbb{R}^m} \varphi_t(y_t)\, \frac{dy_t}{|\det(D_t(X_t, Y^{t-1}, \theta))|} \tag{4.7}$$

Substituting (4.7) into (4.6) implies

$$\mathbb{E}_{\mathbb{P}}\left\{ L_t^{-1} \,\middle|\, \mathcal{F}_{t-1}^{+} \right\} \;=\; L_{t-1}^{-1}$$

showing that L_t^{-1} is a $\mathbb{P}$-martingale with respect to $\mathcal{F}_t^{+}$.[1] The unconditional expectation of a martingale is just the expectation of the initial value so that, repeating a previous argument

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left\{ L_1^{-1} \right\} &= \mathbb{E}_{\mathbb{P}}\left\{ L_1^{-1} \,\middle|\, \mathcal{F}_0^{+} \right\} \\
&= \mathbb{E}_{\mathbb{P}}\left\{ \psi\big(X_1, Y^0, \theta\big) \,\middle|\, \mathcal{F}_0^{+} \right\} \\
&= \left|\det\big(D_1\big(X_1, Y^0, \theta\big)\big)\right| \mathbb{E}_{\mathbb{P}}\left\{ \frac{\varphi_1(Y_1)}{\varphi_1(W_1)} \,\middle|\, \mathcal{F}_0^{+} \right\} \\
&= \left|\det\big(D_1\big(X_1, Y^0, \theta\big)\big)\right| \int_{\mathbb{R}^m} \frac{\varphi_1(X_1 + w_1)}{\varphi_1(w_1)} \varphi_1(w_1)\, dw_1 \\
&= \left|\det\big(D_1\big(X_1, Y^0, \theta\big)\big)\right| \int_{\mathbb{R}^m} \varphi_1(X_1 + w_1)\, dw_1
\end{aligned}$$

The change of variable illustrated in (4.7) shows that this last expression is equal to one. Thus, $\mathbb{Q}$ is a nonnegative measure that integrates to one, and is therefore a probability measure.

[1] Using this result and the conditional version of Bayes theorem, L_t can be shown to be a $\mathbb{Q}$-martingale.

$$\mathbb{E}_{\mathbb{Q}}\left\{ L_t \,|\, \mathcal{F}_{t-1}^{+} \right\} = \frac{\mathbb{E}_{\mathbb{P}}\left\{ L_t L_t^{-1} \,|\, \mathcal{F}_{t-1}^{+} \right\}}{\mathbb{E}_{\mathbb{P}}\left\{ L_t^{-1} \,|\, \mathcal{F}_{t-1}^{+} \right\}} = \frac{1}{L_{t-1}^{-1}} = L_{t-1}$$

Before deriving the distributional properties of Y_t under $\mathbb{Q}$, the following lemma is presented as it will be used throughout the remainder of the thesis.[2]

Lemma 1 *Consider a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and let* $\mathcal{G}$ *be a sub-sigma field of* $\mathcal{F}$. *Let* $\mathbb{Q}$ *be a probability measure absolutely continuous with respect to* $\mathbb{P}$ *and denote the Radon-Nikodym derivative by* $L = d\mathbb{P}/d\mathbb{Q}$. *Then, for any set* $B \in \mathcal{F}$

$$\mathbb{E}_{\mathbb{P}}\{1_B \,|\, \mathcal{G}\}\, \mathbb{E}_{\mathbb{Q}}\{L \,|\, \mathcal{G}\} = \mathbb{E}_{\mathbb{Q}}\{1_B \,|\, \mathcal{G}\} \tag{4.8}$$

where equality is $\mathbb{P} - a.s.$

It is now conjectured that under the measure $\mathbb{Q}$ defined in (4.3):

1. the observation process is independent of the state process and parameter vector, and

2. the observation process is independent over time with densities $\{\varphi_t\}$

Let f be a test function and define $\mathcal{F}_t^{-1} = \sigma\{X^t, Y^{t-1}, \theta\}$. Then, by Lemma 1,

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}\left\{f(Y_t) \,|\, \mathcal{F}_t^{-1}\right\} &= \frac{\mathbb{E}_{\mathbb{P}}\left\{f(Y_t)\, L_t^{-1} \,\middle|\, \mathcal{F}_t^{-}\right\}}{\mathbb{E}_{\mathbb{P}}\left\{L_t^{-1} \,\middle|\, \mathcal{F}_t^{-}\right\}} \\ &= \frac{L_{t-1}^{-1}\mathbb{E}_{\mathbb{P}}\left\{f(Y_t)\, \psi_t^{-1}(X_t, Y^t, \theta) \,\middle|\, \mathcal{F}_t^{-}\right\}}{L_{t-1}^{-1}\mathbb{E}_{\mathbb{P}}\left\{\psi_t^{-1}(X_t, Y^t, \theta) \,\middle|\, \mathcal{F}_t^{-}\right\}} \\ &= \frac{\mathbb{E}_{\mathbb{P}}\left\{f(Y_t)\, |\det(D_t(X_t, Y^{t-1}, \theta))|\, \varphi_t(Y_t)\, / \varphi_t(W_t) \,|\, \mathcal{F}_t^{-}\right\}}{\mathbb{E}_{\mathbb{P}}\left\{|\det(D_t(X_t, Y^{t-1}, \theta))|\, \varphi_t(Y_t)\, / \varphi_t(W_t) \,|\, \mathcal{F}_t^{-}\right\}} \end{aligned} \tag{4.9}$$

[2]This lemma and its proof are given by Karatzas and Shreve (1991), page 193. A proof of the lemma as it is stated here may be found in the appendix for completeness, but does not differ in any meaningful way from that found in Karatzas and Shreve.

The denominator of (4.9) is equal to

$$\begin{aligned}
& \int_{\mathbb{R}^m} \left|\det\big(D_t\big(X_t, Y^{t-1}, \theta\big)\big)\right| \frac{\varphi_t(X_t + w_t)}{\varphi_t(w_t)} dw_t \\
= & \int_{\mathbb{R}^m} \left|\det\big(D_t\big(X_t, Y^{t-1}, \theta\big)\big)\right| \varphi_t(X_t + w_t)\, dw_t \\
= & \int_{\mathbb{R}^m} \varphi_t(y_t)\, dy_t \\
= & \ 1
\end{aligned} \tag{4.10}$$

which follows from the change of variable argument used above (equation 4.7). A similar argument shows that the numerator of (4.9) is equal to

$$\begin{aligned}
& \int_{\mathbb{R}^m} f(Y) \left|\det\big(D_t\big(X_t, Y^{t-1}, \theta\big)\big)\right| \varphi_t(X_t + w_t)\, dw_t \\
= & \int_{\mathbb{R}^m} f(Y)\, \varphi_t(y_t)\, dy_t
\end{aligned}$$

Defining $f(y_t)$ to be the indicator function $I(y_{1t} \leq k_1, ..., y_{mt} \leq k_m)$ shows that the denisty of Y_t under the measure $\mathbb{Q}$ is given by φ_t. Note that this expression is a function of only the scalar constants $(k_1, ..., k_m)$ and is therefore independent of the state process, the parameter vector and all previous observations. Since the time period is arbitrary, this independence holds over all time periods.

4.1.2 Zakai's Equation

The next step is to derive Zakai's equation, which will enable recursive computation of an unnormalized conditional density of the state process and parameter vector given the observations. From (4.3) and (4.4), $L_T = d\mathbb{P}/d\mathbb{Q}$ has the general form:

$$L_T = \prod_{s=1}^{T} \psi_s(Z_s, Y^s) \tag{4.11}$$

where, under the measure $\mathbb{Q}$, $\{Z_t\}$ is a vector-valued Markov process independent of $\{Y_t\}$ with density $p_t(z_t)$ and transition density $p(z_t|z_{t-1})$. Define

$$U_t \equiv u_t\big(Z_t, Y^t\big) = \mathbb{E}_{\mathbb{Q}}\{L_t\,|\,\mathcal{F}_t^y \vee \sigma\{Z_t\}\} \tag{4.12}$$

and, for measurable f, define

$$V_t^f \equiv v_t^f\big(Y^t\big) = \mathbb{E}_{\mathbb{Q}}\{f(Z_t)\,L_t\,|\,\mathcal{F}_t^y\} \tag{4.13}$$

From the definition of U_t and V_t^f,

$$\begin{aligned} V_t^f &= \mathbb{E}_{\mathbb{Q}}\{f(Z_t)\,L_t\,|\,\mathcal{F}_t^y\} \\ &= \mathbb{E}_{\mathbb{Q}}\{f(Z_t)\,\mathbb{E}_{\mathbb{Q}}\{L_t\,|\,\mathcal{F}_t^y \vee \sigma\{Z_t\}\}\,|\,\mathcal{F}_t^y\} \\ &= \int f(z)\,u_t\big(z, Y^t\big)\,p_t(z)\,dz \end{aligned} \tag{4.14}$$

The second line above follows from the law of iterated expectations. The third line above follows from the independence of $\{Z_t\}$ and $\{Y_t\}$ under $\mathbb{Q}$ so that the expectation is with respect to the unconditional density of Z_t.

From Lemma 1, it follows that

$$\mathbb{E}_{\mathbb{P}}\{Z_t\,|\,\mathcal{F}_t^y\} = \frac{\mathbb{E}_{\mathbb{Q}}\{Z_t L_t\,|\,\mathcal{F}_t^y\}}{\mathbb{E}_{\mathbb{Q}}\{L_t\,|\,\mathcal{F}_t^y\}} \tag{4.15}$$

Defining

$$q_t\big(Z_t|Y^t\big) \equiv u_t\big(Z_t, Y^t\big)\,p_t(Z_t) \tag{4.16}$$

enables (4.15) to be written as

$$\mathbb{E}_{\mathbb{P}}\{f(Z_t)\,|\,\mathcal{F}_t^y\} = \int f(z_t)\, q_t\big(z_t|Y^t\big)\, dz_t \Big/ \int q_t\big(z_t|Y^t\big)\, dz_t \tag{4.17}$$

The function $q_t(Z_t|Y^t)$ is referred to as an "unnormalized" condtional density of Z_t for reasons made clear by equation (4.17). The next theorem shows how $q_t(Z_t|Y^t)$ may be computed recursively.

Theorem 1 (Zakai's Equation) *Let $\{Z_t\}$ be a Markov process independent of $\{Y_t\}$ with transition density $p(z_t|z_{t-1})$ and unconditional density $p_t(z_t)$ under the probability measure $\mathbb{Q}$. Suppose further that the Radon-Nikodym derivative is given by $L_T = d\mathbb{P}/d\mathbb{Q}$ and may be expressed as in (4.11). Then the unnormalized conditional density $q_t(Z_t|Y^t)$ defined by (4.16) may be computed recursively as follows:*

$$q_t\big(Z_t|Y^t\big) = \psi_t\big(Z_t, Y^t\big) \int p(Z_t|z_{t-1})\, q_{t-1}\big(z_{t-1}|Y^{t-1}\big)\, dz_{t-1} \tag{4.18}$$

$$q_0(Z_0) = p_0(Z_0) \tag{4.19}$$

Proof: From the definitions in (4.12) and (4.16),

$$\begin{aligned} q_t\big(Z_t|Y^t\big) &= \mathbb{E}_{\mathbb{Q}}\left\{L_t\,|\,Y^t, Z_t\right\} p_t(z_t) \\ &= \mathbb{E}_{\mathbb{Q}}\left\{\prod_{s=0}^{t} \psi_s(Z_s, Y^s)\,\middle|\, Y^t, Z_t\right\} p_t(Z_t) \\ &= \psi_t\big(Z_t, Y^t\big)\, \mathbb{E}_{\mathbb{Q}}\left\{\prod_{s=0}^{t-1} \psi_s(Z_s, Y^s)\,\middle|\, Y^t, Z_t\right\} p_t(Z_t) \\ &= \psi_t\big(Z_t, Y^t\big)\, \mathbb{E}_{\mathbb{Q}}\left\{L_{t-1}\,|\,Y^t, Z_t\right\} p_t(Z_t) \end{aligned} \tag{4.20}$$

Consider the conditional expectation of the product function,

$$\mathbb{E}_{\mathbb{Q}}\left\{g\big(Y^{t-1}\big)\, h\big(Z^{t-1}\big)\,\big|\,Y^t, Z_t\right\}$$

This expectation equals

$$
\begin{aligned}
& g\big(Y^{t-1}\big)\, \mathbb{E}_{\mathbb{Q}}\left\{h\big(Z^{t-1}\big)\,\big|\,Y^t, Z_t\right\} \\
= \;& g\big(Y^{t-1}\big)\, \mathbb{E}_{\mathbb{Q}}\left\{h\big(Z^{t-1}\big)\,\big|\,Z_t\right\} \\
= \;& g\big(Y^{t-1}\big) \int h\big(z^{t-1}\big)\, \frac{p_{Z^t}(z_0, \ldots, z_{t-1}, z_t)}{p_t(Z_t)} dz_0 \cdots dz_{t-1} \\
= \;& \int g\big(Y^{t-1}\big)\, h\big(z^{t-1}\big)\, \frac{p_{Z^t}(z_0, \ldots, z_{t-1}, z_t)}{p_t(Z_t)} dz_0 \cdots dz_{t-1} \\
= \;& \int g\big(Y^{t-1}\big)\, h\big(z^{t-1}\big)\, \frac{p_{Z^{t-1}}(z_0, \ldots, z_{t-1})\, p(z_t|z_0, \ldots, z_{t-1})}{p_t(Z_t)} dz_0 \cdots dz_{t-1} \\
= \;& \int g\big(Y^{t-1}\big)\, h\big(z^{t-1}\big)\, \frac{p_{Z^{t-1}}(z_0, \ldots, z_{t-1})\, p(z_t|z_{t-1})}{p_t(Z_t)} dz_0 \cdots dz_{t-1} \\
= \;& \frac{\mathbb{E}_{\mathbb{Q}}\{g(Y^{t-1})\, h(Z^{t-1})\, p(Z_t|Z_{t-1})\}}{p_t(Z_t)}
\end{aligned} \tag{4.21}
$$

The monotone class theorem allows one to extend this argument to L_{t-1}, which is a function of (Z^{t-1}, Y^{t-1}). Thus,

$$
\mathbb{E}_{\mathbb{Q}}\left\{L_{t-1}\,|\,Y^t, Z_t\right\} = \frac{\mathbb{E}_{\mathbb{Q}}\{L_{t-1} p(Z_t|Z_{t-1})\}}{p_t(z_t)} \tag{4.22}
$$

Substituting this result into equation (4.20) implies

$$
\begin{aligned}
q_t\big(Z_t|Y^t\big) &= \psi_t\big(Z_t, Y^t\big)\, \frac{\mathbb{E}_{\mathbb{Q}}\{L_{t-1} p(Z_t|Z_{t-1})\}}{p_t(z_t)} p_t(Z_t) \\
&= \psi_t\big(Z_t, Y^t\big)\, \mathbb{E}_{\mathbb{Q}}\{L_{t-1} p(Z_t|Z_{t-1})\} \\
&= \psi_t\big(Z_t, Y^t\big)\, \mathbb{E}_{\mathbb{Q}}\left\{\mathbb{E}_{\mathbb{Q}}\left\{L_{t-1} p(Z_t|Z_{t-1})\,|\,Y^{t-1}, Z_{t-1}\right\}\right\} \\
&= \psi_t\big(Z_t, Y^t\big)\, \mathbb{E}_{\mathbb{Q}}\left\{p(Z_t|Z_{t-1})\, \mathbb{E}_{\mathbb{Q}}\left\{L_{t-1}\,|\,Y^{t-1}, Z_{t-1}\right\}\right\} \\
&= \psi_t\big(Z_t, Y^t\big) \int p(Z_t|z_{t-1})\, \mathbb{E}_{\mathbb{Q}}\left\{L_{t-1}\,|\,Y^{t-1}, Z_{t-1}\right\} p_{t-1}(z_{t-1})\, dz_{t-1} \\
&= \psi_t\big(Z_t, Y^t\big) \int p(Z_t|z_{t-1})\, q_{t-1}\big(z_{t-1}|Y^{t-1}\big)\, dz_{t-1}
\end{aligned} \tag{4.23}
$$

completing the proof. ■

For the model defined in equations (4.1) and (4.2), Zakai's equation is

$$\begin{aligned} q_t\big(X_t, \theta | Y^t\big) &= \left|\det\left(D_t\big(X_t, Y^{t-1}, \theta\big)\right)\right|^{-1} \times \\ & \frac{\varphi_t\big(D_t^{-1}(X_t, Y^{t-1}, \theta)\,[Y_t - c_t(X_t, Y^{t-1}, \theta)]\big)}{\varphi_t(Y_t)} \times \\ & \int p(X_t|x_{t-1})\, q_{t-1}(x_{t-1}, \theta | Y_{t-1})\, dx_{t-1} \qquad (4.24) \\ q_0(X_0, \theta) &= \pi_0(X_0)\, \rho_\theta(\theta) \qquad (4.25) \end{aligned}$$

4.2 An Example

This section applies the technique discussed above to the nonlinear example from chapter 2. For ease of reference, the state and observation equations are repeated here.

$$X_t = \alpha X_{t-1} + \sigma V_t \qquad (4.26)$$

$$Y_t = \exp(X_t)\, W_t \qquad (4.27)$$

where $\{V_t\}$ and $\{W_t\}$ are independent Gaussian white noise processes. The initial state X_0 is assumed to be standard normal and independent of the parameter vector (α, σ), which is assumed to have a bivariate standard normal distribution.[3]

[3] While the model parameterization and prior assumptions on the initial state and parameter vector would be suspect in most applications, the point here is merely to illustrate the technique with as little complication as possible.

4.2.1 Measure Change

The first step is to construct a new probability measure $\mathbb{Q}$, absolutely continuous with respect to $\mathbb{P}$, such that $\{Y_t\}$ is an i.i.d. sequence of standard normal random varables. Referring to equation (4.27), this entails "removing" the $\exp(X_t)$ term so that, in effect, $Y_t = W_t$. This is accomplished by defining $\mathbb{Q}$ as:

$$d\mathbb{Q} = \prod_{s=1}^{T} \frac{\phi(Y_s)}{\phi(W_s)} d\mathbb{P} \tag{4.28}$$

$$= L^{-1} d\mathbb{P} \tag{4.29}$$

where $\phi(\cdot)$ is the standard normal density. Under $\mathbb{Q}$ the observation process is Gaussian white noise, independent of the state process and parameters. This is easily confirmed using the same arguments as in the previous subsection.

In order to explicitly derive Zakai's equation for the model in (4.26) and (4.27), two components are needed. The first component is the distribution of the initial state variable and parameter vector (X_0, α, σ). By assumption, this distribution is simply the product of three standard normal distributions. Thus,

$$q_0(X_0, \alpha, \sigma) = \phi(X_0)\phi(\alpha)\phi(\sigma) \tag{4.30}$$

Next, the transition density of the state process is needed. Examination of (4.26) reveals that

$$p(X_t | X_{t-1}) = \phi\left(\frac{X_t - \alpha X_{t-1}}{\sigma}\right) |\sigma|^{-1} \tag{4.31}$$

This implies Zakai's equation for $t = 2, ..., T$ is

$$q_t\big(X_t, \alpha, \sigma | Y^t\big) = \frac{\phi(Y_t/\exp(X_t))}{\phi(Y_t)} \times \int \phi\left(\frac{X_t - \alpha x_{t-1}}{\sigma}\right) |\sigma|^{-1} q_t\big(x_{t-1}, \alpha, \sigma | Y^{t-1}\big)\, dx_{t-1} \quad (4.32)$$

One possible point estimate of α in equation (4.26) is the posterior conditional mean, $\mathbb{E}_{\mathbb{P}}\left\{\alpha \,|\, Y^T\right\}$. According to equation (4.17), this is equal to

$$\frac{\int\int\int \alpha q_T(x_T, \alpha, \sigma | Y^t)\, dx_T d\alpha d\sigma}{\int\int\int q_T(x_T, \alpha, \sigma | Y^t)\, dx_T d\alpha d\sigma} \quad (4.33)$$

with $q_T(X_T, \alpha, \sigma | Y^t)$ defined in (4.32).

Evaluation of the integrals in (4.33) may be carried out by sequential Monte Carlo methods.[4]

[4]See Doucet (1998), Doucet et al. (2000), and Pitt and Shephard (1997) for a discussion of such methods.

Chapter 5

Conclusion

This thesis has discussed a technique for parameter estimation in state-space models that are possibly nonlinear and non-Gaussian. The approach allows for on-line computation of estimates by relying on recursive computation of an unnormalized conditional density of the parameters. Thus, as new data becomes available, parameter estimates can be updated far more efficiently than if a non-recursive approach had been used.

One potential drawback to this approach is the accumulation of approximation error. At each time step, numerical integration must be performed. As one iterates through time, the errors at each time step are propagated into the next time step. Therefore, evaluation beyond a certain number of time steps should be performed with caution and perhaps calibrated against a full estimation of the model by Markov

Chain Monte Carlo methods, for example.[1]

[1]For a discussion of this issue see Liu and West (2000).

Bibliography

[1] Alspach, D. L. and W. H. Sorenson, 1972, "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, 17, 439-447.

[2] Carlin, B. P., Polson, N. G. and D. S. Stoffer, 1992, "A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling," *Journal of the American Statistical Association*, 87, 493-500.

[3] DiMasi, G. B. and W. J. Runggaldier, 1982, "On Measure Transformations for Combined Filtering and Parameter Estimation in Discrete Time," *Systems & Control Letters*, Vol. 2, No. 1.

[4] Doucet, A., 1998, "On Sequential Simulation-Based Methods for Bayesian Filtering," Technical Report CUED/F-INFENG/TR.310, University of Cambridge, Department of Engineering.

[5] Doucet, A., Godsill, S. and C. Andrieu, 2000, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," Cambridge University, Department of

Engineering.

[6] Durrett, R., 1996, *Probability Theory and Examples*, Wadsworth Publishing Co.

[7] Elliott, R. J., Aggoun, L. and J. B. Moore, 1995, *Hidden Markov Models: Estimation and Control*, Springer-Verlag.

[8] Girsanov, I. V., 1960, "On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measures," *Theory Probab. Appl.*, 5, 285-301.

[9] Harrison, P J. and C. F. Stevens, 1976, "Bayesian Forecasting," *Journal of the Royal Statistical Society*, Ser. B, 38, 205-247.

[10] Harvey, A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

[11] Harvey, A. C., 1993, *Time Series Models,* MIT Press.

[12] Harvey, A. C., Ruiz, E. and N. Shephard, 1994, "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247-264.

[13] Hull, J. and A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42, 281-300.

[14] Jacquier, E., Polson, N. G. and P. E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models (with discussion)," *Journal of Business and Economic Statistics*, 12, 371-417.

[15] Jazwinski, A., 1970, *Stochastic Processes and Filtering Theory*, Academic Press.

[16] Kalman, R. E., 1960, "A New Approach to Linear Filtering and Prediction Problems," *J. of Bas. Engin.*, 35-45.

[17] Kalman, R. E. and R. S. Bucy, 1961, "New Results in Linear Filtering and Prediction Theory," *J. of Bas. Engin.*, 95-108.

[18] Karatzas, I. and S. Shreve, 1991, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer-Verlag.

[19] Kitigawa, G., 1987, "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82, 1032-1063.

[20] Liu, J. and M. West, 2000, "Combined Parameter and State Estimation in Simulation-Based Filtering," Duke University.

[21] Meinhold, R. J. and N. D. Singpurwalla, 1989, "Robustification of Kalman Filter Models," *Journal of the American Statistical Association*, 84, 479-486.

[22] Pitt, M. K. and N. Shephard, 1997, "Filtering Via Simulation: Auxillary Particle Filters," Nuffield College, Oxford University.

[23] Priestley, M. B., 1981, *Spectral Analysis and Time Series*, Academic Press.

[24] Smith, A. F. M. and M. West, 1983, "Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter," *Biometrics* 39, 867-878.

[25] West, M., Harrison, P. J. and H. S. Migon, 1985, "Dynamic Generalized Linear Models and Bayesian Forecasting," *Journal of the American Statistical Association*, 80, 73-97.

[26] Zakai, M., 1969, "On the Optimal Filtering of Diffusion Processes," *Z. Wahrsch. Verw. Geb.*, 11, 230-243.

Appendix A

Derivations

A.1 Bayes' Rule

This section proves Lemma 1 (Bayes' rule) as it is presented in equation (4.8).

It must be shown that for all $A \in \mathcal{G}$ and $B \in \mathcal{F}$

$$\int_A \mathbb{E}_{\mathbb{P}}\{1_B \,|\, \mathcal{G}\}\, \mathbb{E}_{\mathbb{Q}}\{L \,|\, \mathcal{G}\}\, d\mathbb{Q} = \int_A \mathbb{E}_{\mathbb{Q}}\{1_B L \,|\, \mathcal{G}\}\, d\mathbb{Q} \tag{A.1}$$

The right hand side of (A.1) is

$$\begin{aligned}
\int_A \mathbb{E}_{\mathbb{Q}}\{1_B L \,|\, \mathcal{G}\}\, d\mathbb{Q} &= \int_A 1_B L\, d\mathbb{Q} \\
&= \int_A 1_B\, d\mathbb{P} \\
&= \mathbb{P}\{A \cap B\}
\end{aligned}$$

Letting $\mathbb{P}_{\mathcal{G}}$ and $\mathbb{Q}_{\mathcal{G}}$ denote the restriction of $\mathbb{P}$ and $\mathbb{Q}$ to $\mathcal{G}$, the left hand side of (A.1) is

$$
\begin{aligned}
\int_A \mathbb{E}_{\mathbb{P}}\{1_B|\mathcal{G}\}\,\mathbb{E}_{\mathbb{Q}}\{L|\mathcal{G}\}\,d\mathbb{Q} &= \int_A \mathbb{E}_{\mathbb{P}}\{1_B|\mathcal{G}\}\,\frac{d\mathbb{P}\mathcal{G}}{d\mathbb{Q}_{\mathcal{G}}}\,d\mathbb{Q}_{\mathcal{G}} \\
&= \int_A \mathbb{E}_{\mathbb{P}}\{1_B|\mathcal{G}\}\,d\mathbb{P}\mathcal{G} \\
&= \int_A 1_B d\mathbb{P}\mathcal{G} \\
&= \int_A \mathbb{P}\{B|\mathcal{G}\}\,d\mathbb{P} \\
&= \mathbb{P}\{A\cap B\}
\end{aligned}
$$

The equivalence in (A.1) has been proven and the result follows. ■

A.2 Chapman-Kolmogorov Equation

Let $\{X_t\}$ be a Markov process and $\{Y_t\}$ another stochastic process. Then the Chapman-Kolmogorov equation says

$$
f_{X_t|Y_{t-1}}(x_t|y_{t-1}) = \int f_{X_t|X_{t-1}}(x_t|x_{t-1})\,f_{X_{t-1}|Y_{t-1}}(x_{t-1}|y_{t-1})\,dx_{t-1} \tag{A.2}
$$

PROOF: The approach is to show that the expectation of both sides with respect to Y^{t-1} are equal. Beginning with the left hand side of (A.2)

$$
\begin{aligned}
\mathbb{E}\big\{f_{X_t|Y^{t-1}}\big(x_t|y^{t-1}\big)\big\} &= \int f_{X_t|Y^{t-1}}\big(x_t|y^{t-1}\big)\,f_{Y_{t-1}}(y_t)\,dy^{t-1} \\
&= \int f_{X_t,Y^{t-1}}\big(x_t,y^{t-1}\big)\,dy^{t-1} \\
&= f_{X_t}(x_t)
\end{aligned}
$$

The expectation of the right hand side of (A.2) with respect Y_{t-1} is

$$
\begin{aligned}
& \mathbb{E}\left\{\int f_{X_t|X_{t-1}}(x_t|x_{t-1})\, f_{X_{t-1}|Y_{t-1}}(x_{t-1}|y_{t-1})\, dx_{t-1}\right\} \\
= & \int\int f_{X_t|X_{t-1}}(x_t|x_{t-1})\, f_{X_{t-1}|Y_{t-1}}(x_{t-1}|y_{t-1})\, f_{Y_{t-1}}(y_t)\, dx_{t-1}dy^{t-1} \\
= & \int\int f_{X_t|X_{t-1}}(x_t|x_{t-1})\, f_{X_{t-1}|Y_{t-1}}(x_{t-1}|y_{t-1})\, f_{Y_{t-1}}(y_t)\, dy^{t-1}dx_{t-1} \\
= & \int f_{X_t|X_{t-1}}(x_t|x_{t-1}) \int f_{X_{t-1}|Y_{t-1}}(x_{t-1}|y_{t-1})\, f_{Y_{t-1}}(y_t)\, dy^{t-1}dx_{t-1} \\
= & \int f_{X_t|X_{t-1}}(x_t|x_{t-1}) \int f_{X_{t-1},Y_{t-1}}(x_{t-1}, y_{t-1})\, dy^{t-1}dx_{t-1} \\
= & \int f_{X_t|X_{t-1}}(x_t|x_{t-1})\, f_{X_{t-1}}(x_{t-1})\, dx_{t-1} \\
= & f_{X_t}(x_t)
\end{aligned}
$$

where the third line above follows from Fubini. The result follows. ■

A.3 Derivation of Equation (2.18)

The relation in equation (2.18) begins with Bayes' rule

$$f_{X_t|Y^t}\left(x_t|y^t\right) = \frac{f_{Y^t|X_t}(y^t|x^t)\, f_{X_t}(x_t)}{f_{Y^t}(y^t)} \tag{A.3}$$

The next step is to use the conditional independence of Y_t given X_t. This enables the factorization of $f_{Y^t|X_t}(y^t|x^t)$

$$= \frac{f_{Y_t|X_t}(y_t|x_t)\, f_{Y^{t-1}|X_t}(y^{t-1}|x_t)\, f_{X_t}(x_t)}{f_{Y^t}(y^t)} \tag{A.4}$$

The remaining steps follow from basic properties of joint and conditional densities.

$$
\begin{aligned}
&= \frac{f_{Y_t|X_t}(y_t|x_t)\, f_{Y^{t-1},X_t}(y^{t-1}, x_t)}{f_{Y^{t-1}}(y^{t-1})\, f_{Y_t|Y^{t-1}}(y_t|y^{t-1})} \\
&= \frac{f_{Y_t|X_t}(y_t|x_t)\, f_{X_t,Y_{t-1}}(x_t, y^{t-1})}{f_{Y_t|Y^{t-1}}(y_t|y^{t-1})}
\end{aligned}
$$

A.4 Derivation of the Prediction and Update Densities

This section derives the prediction and update densities found in equations (2.17) and (2.20).

A.4.1 Prediction Density

From the Chapman-Kolmogorov equation, the prediction density at time $t = 1$ is computed as follows.

$$
\begin{aligned}
f_{X_1|Y_0}(x_0|y_0) &= \int f_{X_1|X_0}(x_1|x_0)\, f_{X_1|Y_0}(x_0|y_0)\, dx_0 \\
&= \int \left(2\pi\sigma^2\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\left(x_1 - \alpha x_0\right)^2\right\} \times \qquad \text{(A.5)} \\
&\quad (2\pi)^{-1/2} \exp\left\{-1/2\left(x_0\right)^2\right\} dx_0 \\
&= (2\pi)^{-1}(\sigma)^{-1} \exp\left\{-\frac{1}{2\sigma^2}\left(x_1^2 - 2\alpha x_1 x_0 + \alpha^2 x_0^2\right) - \frac{1}{2}x_0^2\right\} \\
&= (2\pi)^{-1}(\sigma)^{-1} \exp\left\{-\frac{1}{2\sigma^2}x_1^2\right\} \times \\
&\quad \exp\left\{-\frac{\alpha^2}{2\sigma^2}x_0^2 - \frac{1}{2}x_0^2 + \frac{\alpha x_1}{\sigma^2}x_0\right\} \qquad \text{(A.6)}
\end{aligned}
$$

An intermediate step is to complete the square in the exponential.

$$\left(-\frac{\alpha^2}{2\sigma^2} - \frac{1}{2}\right) x_0^2 + \frac{\alpha x_1}{\sigma^2} x_0 \quad = \quad \left(\frac{-\alpha^2 - \sigma^2}{2\sigma^2}\right) x_0^2 + \frac{\alpha x_1}{\sigma^2} x_0 \tag{A.7}$$

Define

$$\begin{aligned} a &= \left(\frac{-\alpha^2 - \sigma^2}{2\sigma^2}\right) \\ b &= \frac{\alpha x_1}{\sigma^2} \end{aligned}$$

then

$$\begin{aligned} \left(\frac{-\alpha^2 - \sigma^2}{2\sigma^2}\right) x_0^2 + \frac{\alpha x_1}{\sigma^2} x_0 &= a x_0^2 + b x_0 \\ &= a \left(x_0^2 + \frac{b}{a} x_0 \pm \left(\frac{b}{2a}\right)^2\right) \\ &= a \left(x_0^2 + \frac{b}{a}\right) - \frac{b^2}{4a} \end{aligned}$$

Substituting for a and b gives

$$-\frac{1}{2}\left(\frac{\alpha^2 + \sigma^2}{\sigma^2}\right)\left(x_0 + \frac{2\sigma^2 \alpha x_1}{2\sigma^2\left(-\alpha - \sigma^2\right)}\right)^2 + \frac{1}{2}\left(\frac{\alpha^2 x_1^2}{\sigma^2\left(\alpha^2 + \sigma^2\right)}\right) \tag{A.8}$$

Equation (A.6) may now be written as

$$\begin{aligned} f_{X_1|Y_0}(x_0|y_0) \quad = \quad & (2\pi)^{-1/2} (\sigma)^{-1} \exp\left\{-\frac{1}{2\sigma^2} x_1^2 + \left(\frac{\alpha^2}{\sigma^2\left(\alpha^2 + \sigma^2\right)}\right) x_1^2\right\} \times \\ & \int (2\pi)^{-1/2} \exp\left(-\frac{1}{2}\left(\frac{\sigma^2}{\alpha^2 + \sigma^2}\right)^{-1}\left(x_0 + \frac{2\sigma^2 \alpha x_1}{2\sigma^2\left(-\alpha - \sigma^2\right)}\right)^2\right) dx_0 \end{aligned}$$

The integral above is a normal density kernel with incomplete constant of integration. Therefore, the integral evaluates to $\sqrt{\sigma^2/\left(\alpha^2 + \sigma^2\right)}$. After some algebra, the prediction density becomes

$$f_{X_1|Y_0}(x_0|y_0) = (2\pi)^{-1/2}\left(\alpha^2 + \sigma^2\right)^{-1/2} \exp\left(-\frac{1}{2}\left(\alpha^2 + \sigma^2\right)^{-1} x_1^2\right) \tag{A.9}$$

implying that the conditional density of X_1 given Y_0 (which is simply the unconditional density of X_1) is normal with mean 0 and variance $(\alpha^2 + \sigma^2)$.

A.4.2 Update Density

The update density follows from the relation in equation (2.18). The derivation begins at time $t = 1$ with the denominator of (2.18).

$$
\begin{aligned}
f_{Y_1|Y_0}(y_1|y_0) &= \int f_{Y_1|X_1}(y_1|x_1)\, f_{X_1|Y_0}(x_1|y_0)\, dx_1 \\
&= \int (2\pi)^{-1/2} \exp\left(-\frac{1}{2}\left(y_1 - x_1\right)^2\right) \times \\
&\quad (2\pi)^{-1/2}\left(\alpha^2+\sigma^2\right)^{-1/2} \exp\left(-\frac{1}{2}\left(\alpha^2+\sigma^2\right)^{-1} x_1^2\right) dx_1 \\
&= (2\pi)^{-1/2}\left(\alpha^2+\sigma^2\right)^{-1/2} \exp\left(-\frac{1}{2} y_1^2\right) \times \\
&\quad \int (2\pi)^{-1/2} \exp\left(-\frac{1}{2}\left(-2y_1x_1 + x_1^2 + \frac{x_1^2}{\alpha^2+\sigma^2}\right)\right) dx_1 \qquad \text{(A.10)}
\end{aligned}
$$

In order to evaluate the integral, the exponential is rewritten to resemble a normal density kernel by completing the square. Define

$$
\begin{aligned}
a &= -\frac{\alpha^2+\sigma^2+1}{2\left(\alpha^2+\sigma^2\right)} \\
b &= y_1
\end{aligned}
$$

Completing the square yields

$$
a\left(x_1 + \frac{b}{2a}\right)^2 - \frac{b^2}{4a}
$$

Substituting for a and b gives

$$
\left(-\frac{\alpha^2+\sigma^2+1}{2\left(\alpha^2+\sigma^2\right)}\right)\left(x_1 - \frac{2\left(\alpha^2+\sigma^2\right)}{\alpha^2+\sigma^2+1} y_1\right)^2 + \frac{\left(\alpha^2+\sigma^2\right)}{2\left(\alpha^2+\sigma^2+1\right)} y_1^2
$$

The integral in equation (A.10) may now be written as

$$\int (2\pi)^{-1/2} \exp\left\{ \left(-\frac{\alpha^2+\sigma^2+1}{2\left(\alpha^2+\sigma^2\right)} \right) \times \left(x_1 - \frac{2\left(\alpha^2+\sigma^2\right)}{\alpha^2+\sigma^2+1} y_1 \right)^2 + \frac{\left(\alpha^2+\sigma^2\right)}{2\left(\alpha^2+\sigma^2+1\right)} y_1^2 \right\} dx_1$$

This integrand is a normal density kernel with incomplete normalizing constant. The integral therefore evaluates to

$$\left(\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1} \right)^{1/2}$$

Substituting this result into (A.10) and some algebra reveals that $f_{Y_1|Y_0}(y_1|y_0)$ is

$$(2\pi)^{-1/2} \left(\alpha^2+\sigma^2+1\right)^{-1/2} \exp\left\{ -\frac{1}{2} \left(\alpha^2+\sigma^2+1\right)^{-1} y_1^2 \right\} \tag{A.11}$$

Thus, the conditional density of Y_1 given Y_1 is normal with mean 0 and variance $(\alpha^2+\sigma^2+1)$.

The update density may now be found using the relation in (2.18).

$$\begin{aligned} f_{X_1|Y_1}(x_1|y_1) &= \frac{f_{Y_1|X_1}(y_1|x_1)\, f_{X_1|Y_0}(x_1|y_0)}{f_{Y_1|Y_0}(y_1|y_0)} \\ &= \frac{(2\pi)^{-1/2} \exp\left\{ -\frac{(y_1-x_1)^2}{2} \right\} \times \left(2\pi\left[\alpha^2+\sigma^2\right]\right)^{-1/2} \exp\left\{ -\frac{x_1^2}{2(\alpha^2+\sigma^2)} \right\}}{\left(2\pi\left[\alpha^2+\sigma^2+1\right]\right)^{-1/2} \exp\left\{ -\frac{y_1^2}{(\alpha^2+\sigma^2+1)} \right\}} \end{aligned}$$

Algebra and completing the square of the resulting exponential reveals that the update density for the initial time period to be

$$(2\pi)^{-1/2} \left(\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1} \right)^{-1/2} \exp\left\{ -\frac{1}{2} \left(\frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1} \right)^{-1} \left[x_1 - \frac{\alpha^2+\sigma^2}{\alpha^2+\sigma^2+1} y_1 \right]^2 \right\} \tag{A.12}$$

Using the prediction and update densities in conjunction with the Chapman-Kolmogorov theorem and the relation in equation (2.18) produces the general prediction and update densities.