UNIVERSITY OF ALBERTA

CONTROL AND FILTERING OF RANDOM PROCESSES

BY

HAILIANG YANG

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

EDMONTON, ALBERTA

SPRING 1993

ISBN 0-315-82242-2

University of Alberta Statistics and Applied Probability Inter-departmental Correspondence

Statistical Services Centre Phone: 492-5129 FAX: 492-2927

To: Faculty of Graduate Studies and Research *Date:* 24 March 1993

From: Dr. R. Elliott

Subject: Ph.D. Thesis of Hailiang Yang

Hailiang Yang has my permission to include our co-authored papers as a part of his Ph.D thesis.

Professor Robert Elliott

GROUPE HEC
HEC SCHOOL OF MANAGEMENT

HAUTES ÉTUDES COMMERCIALES
INSTITUT SUPÉRIEUR DES AFFAIRES
HEC MANAGEMENT - DOCTORAT
78351 JOUY-EN-JOSAS CEDEX (FRANCE)
TÉLÉPHONE : (1) 39 67 72 39. TÉLEX : 697 942 F.
TÉLÉCOPIE : (1) 39 67 74 40.

DOCTORAT HEC

PAGE DE GARDE TELECOPIE
(TELECOPY FRONT PAGE)

Date : April 15th, 1993

de : (from) Marc CHESNEY
Doctorat HEC

Téléphone : 33 (1) 39 67 72 39
Fax : 33 (1) 39 67 74 40

à : (to) Robert Elliot
University of Alberta Edmonton

Téléphone :
Fax : 19 1 403 492 29 27

Nombre de pages : Number of pages (including this one) : 2

Observations :

GROUPE HEC
SCHOOL OF MANAGEMENT

HAUTES ÉTUDES COMMERCIALES
INSTITUT SUPÉRIEUR DES AFFAIRES
HEC MANAGEMENT - DOCTORAT
78351 JOUY-EN-JOSAS CEDEX (FRANCE)
TÉLÉPHONE : (1) 39 67 72 39. TÉLEX : 697 942 F.
TÉLÉCOPIE : (1) 39 67 74 40.

Jouy-en-Josas, April 15th, 1993

To whom it may concern

Mr. YANG has my permission to include our joint manuscript entitled :

"Diffusion Coefficient Estimation and Asset Pricing when Risk Premia and Sensitivities are Time Varying", with the names of the different co-authors, in his Ph.D. dissertation.

M. Chesney

Marc CHESNEY
Associate Professor
Doctorat HEC

Dilip B. Madan
College of Business and Management
University of Maryland
College Park, Md. 20742

January 8 1993

Prof. Robert Elliott
Dept. of Statistics and Applied Probability
Central Academic Building, 434
University of Alberta
Edmonton, Alberta
CANADA T6G 1X5

Dear Prof. Elliott,

I am writing to give my consent to the inclusion of the paper entitled, "Diffusion Coefficient Estimation and Asset Pricing when Risk Premia and Sensitivities are Time Varying," co-authored by Mr. Hailang Yang as part of his Ph.D. thesis.

Yours Sincerely,

Dilip B. Madan

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Hailiang Yang

TITLE OF THESIS:

Control and Filtering of Random Processes

DEGREE FOR WHICH THESIS WAS PRESENTED: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: Spring 1993

Hailiang Yang

(Signed)

Permanent Address:

#25 8406 104 Street

Edmonton, Alberta

T6E 4G2

Date: March 29, 1993

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Control and Filtering of Random Processes** submitted by **Hailiang Yang** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy.**

R. J. Elliott (Supervisor)

D. Wiens

B. Schmuland

R. Karunamuni

A. T. Lau

J. B. Moore (External Examiner)

Date: March 23, 1993

TO MY WIFE AND DAUGHTER

ABSTRACT

The partially observed control problem and the related filtering problem are considered. The first chapter discusses a partially observed control problem when the control parameter appears in both the drift and diffusion coefficients. A stochastic minimum principle is obtained by improving techniques of Bensoussan.

In the second chapter, the partially observed control for a Markov chain is discussed and treated in terms of the associated Zakai equation. New equations for the adjoint processes are obtained.

In the chapter on discrete adaptive filters a discrete time Hidden Markov Model is considered. Finite closed form filters are obtained for the state of the process, and also processes such as the occupation time and number of jumps from one state to another. These enable the parameters of the model to be re-estimated from a conditional log-likelihood

The chapter on estimating the diffusion coefficient discusses increments of various powers of a scalar diffusion. An investigation of the minimum variance estimator gives a unique optimal power.

ACKNOWLEDGMENTS

I would like to thank all professors who have helped me during my study at the University of Alberta. In particular, I sincerely thank my supervisor Professor Robert J.Elliott. This work would have been impossible without his guidance and help. His constant encouragement, patience and care have give light to my life. I am also grateful to my parents and my family for their love and spiritual support throughout my graduate years.

I gratefully acknowledge the Department of Statistics and Applied Probability for continuing support during my study, the Natural Sciences and Engineering Research Council of Canada for a Postgraduate Scholarship and the University of Alberta for the Walter H Johns Graduate Fellowship.

TABLE OF CONTENTS

Introduction

The work in this thesis is presented in four papers which have been prepared for publication.

In the first chapter, "Control of Partially Observed Diffusions", the optimal control of a partially observed diffusion is discussed when the control parameter appears in both the drift and diffusion coefficients.

The minimum principle satisfied by an optimal control in a partially observed stochastic control problem has been discussed by many authors. See, for example, papers by Baras, Elliott and Kohlmann, Bensoussan, Elliott and Haussmann. In these articles, however, the control parameter occurs in only the drift coefficient. For a fully observed stochastic control problem Bensoussan and Elliott do consider the case when the control variable also appears in the diffusion coefficient, and Elliott gives an explicit equation for the adjoint process when the optimal control is Markov.

In the first chapter we consider a state process, which is only partially observed through a noisy observation process, and for which the control parameter is present in both the drift and diffusion coefficients. Using a differentiation result of Blagovescenskii and Freidlin, and adapting techniques of Bensoussan, an adjoint process is described and a stochastic minimum principle is obtained for an optimum control.

The second chapter, "Forward and Backward Equations for an Adjoint Process", discusses a Markov chain observed only through a noisy continuous observation process. Consider a system whose state is described by a Markov chain X_t. Without loss of generality the state space of the Markov chain can be taken to be the set of unit basis vectors in R^N. We shall suppose our X_t process is not observed directly. Rather X_t is observed through the noisy process y_t, where

$$y_t = \int_0^t h(X_s)ds + w_t.$$

Here w is a Brownian motion independent of X.

For simplicity a terminal cost is considered and the control problem is formulated in separated form by considering an unnormalized conditional distribution of X_t

By introducing a Gateaux derivative the minimum principle, satisfied by an optimal control, is derived. If the optimal control is Markov new forward and backward equations satisfied by the adjoint process are obtained. A similar problem for a controlled Markov chain for which only the jump times, but not the jump locations, can be observed has been discussed by Elliott.

The chapter "How to Count and Guess Well: Discrete Adaptive Filters" considers a discrete state and time Markov chain which is observed through a finite state function which is subject to random perturbations. Such a situation is often called a Hidden Markov Model. Discrete filtering consists of counting the occurrences of various states of a discrete observation process and then making the best estimates of quantities related to an unobserved state process. Our Hidden Markov Model consists of a homogeneous, finite state, discrete time Markov chain X_n. Without loss of generality, the state space of X can be taken to be the set of unit vector $S = \{e_1, \dots, e_N\}$, $e_i = (0, \dots, 1, 0, \dots, 0)' \in R^N$, and write $P = (p_{ji})$ the matrix of transition probabilities it can be shown that X has a semimartingale form

$$X_n = PX_{n-1} + m_n, \qquad n \in Z^+$$

where m_n is a martingale increment. The process X is not observed directly; instead we suppose there is a discrete time, finite state observation process Y. Again, the finite range of Y can be identified with the set of unit vectors $\{f_1, ..., f_k\}$, $f_j=(0, ..., 1, ..., 0)' \in R^k$, and with $D=(d_{ji})$ where $d_{ji}=P(Y_n=f_j | X_{n-1}=e_i)$, the relation between X and Y can also be expressed in semimartingale form as:

$$Y_n=DX_{n-1}+\mu_n, \qquad n\in Z^+,$$

where μ_n is a martingale increment.

In a recent paper, Elliott obtained finite dimensional filters and smoothers for a continuous time Markov chain observed in Gaussian noise. In addition to the filter for the state, finite dimensional filters and smoothers are obtained for the number of jumps from one state to another, for the occupation time of any state, and also of a process related to the observation. The methods of this chapter are an adaptation to discrete time and state of those of Elliott. The application of methods from the discrete time general theory of processes in this situation is not well known and sheds light on the problems. Early contributions can be found in the papers of Boel and Segall. Chapter three begins with martingale representation and Girsanov results related to multivariate point processes. The semimartingale representation of the Hidden Markov Model is next given and followed by a general filtering result. A general unnormalized, or Zakai, estimate is derived. Compared with the normalized filter this has a remarkably simple form. Then, specializing this result, we obtain recursive estimates and smoothers for the state of the process, the number of jumps from one state to another, the occupation time of a state and of a process related to the observations.

Following Elliott's paper, a particular trick used is to exploit the idempotent property of X; instead of estimating H, which would involve HX, we estimate HX. This introduces $HX \otimes X$ but this can be expressed in terms of HX itself and so, unlike H, HX has a recursive estimate. Taking the inner product with $\mathbf{1}=(1, 1, ..., 1)$ then gives an estimate for H. From these estimates new optimal values for the parameters p_{ji} and d_{ji} in the matrices P and D can be obtained. Using the new parameters, and perhaps new observations, a sequence of increasingly better models can be obtained.

We believe our model is of wide applicability and generality. Our model, by discretizing time and state, can be made an approximation to many continuous models, including non-linear diffusion models in continuous time. In particular, by discretizing and approximating the noise in the observations, the case of a Markov chain observed in Gaussian noise can be approximated by a Hidden Markov Model. Discretization is necessary for numerical implementation. Furthermore, our model re-tunes its parameters in an increasingly optimal way.

Given a diffusion process it is often easier to estimate the drift coefficient than the diffusion coefficient. Log-normal diffusion processes are frequently used to model asset prices in finance and , in a recent paper, Chesney and Elliott have used the Mihlstein approximation to estimate the diffusion coefficient, (known in finance as the volatility).

In chapter four, "Estimation of the Diffusion Coefficient", a general (scalar) diffusion x_t is considered. By introducing the process $y_t = \exp x_t$ properties of the exponential can be exploited. In the paper of Chesney and Elliott a point estimate for the diffusion coefficient is obtained by comparing expressions derived from y_t and y_t^{-1}; in this chapter estimates for the diffusion coefficient of x_t are obtained by using the Ito calculus and Mihlstein approximations, and comparing expressions for y_t and y_t^{α}, (α real). The minimum variance estimate gives a unique optimal value of α. A table illustrating optimal α values is also presented.

The scalar diffusion estimation strategy is then extended to also allow estimation of the instantaneous variation in the predictable quadratic covariation of two diffusion processes. Such a point estimate may be used to accommodate time varying risk sensitivities in asset pricing models that simultaneously permit time variation in risk premia as well. Applications to the Capital Asset Pricing Model illustrate the procedure.

Chapter One

Control of Partially Observed Diffusions

1. Introduction

The adjoint process, and related minimum principles, for partially observed stochastic control problems have been investigated in several recent papers. See, for example, the works of Bensoussan (Ref. 1), Haussmann (Ref. 2), Baras, Elliott and Kohlmann (Ref. 3) and Elliott (Ref. 4). In these papers, however, the control variable occurs in only the drift coefficient. For a fully observed stochastic control problem Bensoussan (Ref. 5) does consider the case when the control also appears in the diffusion coefficient. This case is also discussed in (Ref. 6), and, when the optimal control is Markov, an explicit equation for the adjoint process is derived.

In this paper we consider a state process, which is only partially observed through a noisy observation process, and for which the control variable is present in both the drift and diffusion coefficients. By adapting the techniques of Bensoussan (Ref. 5) an adjoint process is described and a minimum principle obtained for an optimum control. To the best of our knowledge, this is the first paper that discusses this problem for the partially observed case when the control appears in both the drift and diffusion terms.

A version of this chapter has been published. Robert J. Elliott and Hailiang Yang. *Journal of Optimization Theory and Applications.* Vol. 71, No. 3., December 1991, 485–501.

2. Dynamics

Suppose that the state of the system is described by a stochastic differential equation,

$$dx_t = f(t, x_t, u)dt + g(t, x_t, u)dw_t, \quad x_t \in R^d, \ x_0 = x_0, \ 0 \le t \le T. \tag{1}$$

The control parameter u will take values in a compact, convex subset U of some Euclidean space R^k.

We shall assume the following:

(A1) $x_0 \in R^d$ is given.

(A2) $f : [0,T] \times R^d \times U \to R^d$ is continuous, and continuously differentiable with respect to x, u.

(A3) $g : [0,T] \times R^d \times U \to R^d \otimes R^n$ is a continuous, matrix valued function, which is continuously differentiable with respect to x, u. The columns of g will be denoted by $g^{(k)}$ for $k = 1, 2, \ldots, n$.

(A4) There is a constant K such that

$$(1 + |x|)^{-1}|f(t, x, u)| + |f_x(t, x, u)| + |f_u(t, x, u)| \le K$$

$$|g(t, x, u)| + |g_x(t, x, u)| + |g_u(t, x, u)| \le K.$$

Suppose the observation process is given by

$$dy_t = h(x_t)dt + dv_t, \quad y_t \in R^m, \ y_0 = 0, \ 0 \le t \le T. \tag{2}$$

In the above equations $w = (w^1, \ldots, w^n)$ and $v = (v^1, \ldots, v^m)$ are independent Brownian motions. We also assume:

(A5) $h : R^d \to R^m$ is Borel measurable, continuously differentiable in x, and for some constant K_1,

$$|h(x)| + |h_x(x)| \leq K_1.$$

Let $\widehat{P}$ denote Wiener measure on $C([0,T], R^n)$ and μ denote Wiener measure on $C([0,T], R^m)$. Consider the space $\Omega = C([0,T], R^n) \times C([0,T], R^m)$ with coordinate functions (w_t, y_t) and define Wiener measure P on Ω by

$$P(dw, dy) = \widehat{P}(dw)\mu(dy).$$

DEFINITION 2.1. Write $\{F_t\}$ for the right continuous, complete filtration on $C([0,T], R^n)$ generated by $F_t^0 = \sigma\{w_s,\ s \leq t\}$. Write $Y = \{Y_t\}$ for the right continuous complete filtration on $C([0,T], R^m)$ generated by $Y_t^0 = \sigma\{y_s - y_r,\ 0 \leq r \leq s \leq t\}$. The set of admissible control functions $\underline{U}$ will be the Y-predictable functions on $[0,T] \times C([0,T],\ R^m)$ with values in U. Then

$$\begin{aligned} \underline{U} &\subset L^2_Y[0,T] \\ &= \{v(t,w') : v(t,w') \in L^2([0,T] \times (C([0,T], R^m)),\ dt \times d\mu;\ R^k), \\ &\qquad \text{for a.e. } t,\ v(t,\cdot) \in L^2(C([0,T], R^m), Y_t, d\mu, R^k)\}. \end{aligned}$$

For $u \in \underline{U}$, write $X^u_{s,t}(x)$ for the unique strong solution of (1) corresponding to control u, and with $X^u_{s,s}(x) = x$.

Write

$$Z^u_{s,t}(x) = \exp\left(\int_s^t h(X^u_{s,r}(x))' dy_r - \frac{1}{2}\int_s^t |h(X^u_{s,r}(x))|^2 dr\right) \tag{3}$$

and define a new probability measure P^u on Ω by

$$\frac{dP^u}{dP} = Z^u_{0,T}(x_0).$$

Then under P^u, $(X^u_{0,t}(x_0),\ y_t)$ is a solution of (1) and (2).

Cost. We shall suppose the cost is

$$C(X^u_{0,T}(x_0)) + \int_0^T \ell(r, X^u_{0,r}(x_0), u_r)dr.$$

We suppose

(A6) $|C(x)| + |C_x(x)| + |C_{xx}(x)| \leq K(1 + |x|^q)$, for some $q < \infty$.

(A7) $\ell : [0,T] \times R^d \times U \to R$ is Borel measurable and continuously differentiable in (x, u). Furthermore ℓ and its derivatives in x and u satisfy linear growth conditions in x.

The expected cost if a control $u \in \underline{U}$ is used is, therefore,

$$J(u) = E^u\Big[C(X^u_{0,T}(x_0)) + \int_0^T \ell(r, X^u_{0,r}(x_0), u_r)dr\Big].$$

In terms of P, this is

$$J(u) = E\Big[Z^u_{0,T}(x_0)\Big(C(X^u_{0,T}(x_0)) + \int_0^T \ell(r, X^u_{0,r}(x_0), u_r)dr\Big)\Big].$$

Consider the $d+1$ dimensional system given by

$$\begin{aligned} X^u_{s,t} &= x + \int_s^t f(r, X^u_{s,r}(x), u_r)dr + \int_s^t g(r, X^u_{s,r}(x), u_r)dW_r \\ Z^u_{s,t} &= z + \int_s^t Z^u_{s,r} h(X^u_{s,r}(x))dy_r. \end{aligned} \tag{4}$$

Write

$$\widetilde{X}^u_{s,t} = \begin{pmatrix} X^u_{s,t} \\ Z^u_{s,t} \end{pmatrix} \qquad \tilde{f}(r) = \begin{pmatrix} f(r, X^u_{s,r}(x), u_r) \\ 0 \end{pmatrix}$$

$$\tilde{g}(r) = \begin{pmatrix} g(r, X^u_{s,r}(x), u_r) & 0 \\ 0 & Z^u_{s,r}(x,z)h(X^u_{s,r}(x)) \end{pmatrix}$$

$$\widetilde{W}_r = \begin{pmatrix} w_r \\ y_r \end{pmatrix} \qquad \tilde{X} = \begin{pmatrix} x \\ z \end{pmatrix}.$$

Then we can write (4) as

$$\widetilde{X}^u_{s,t}(\tilde{x}) = \tilde{x} + \int_s^t \tilde{f}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r)dr + \int_s^t \tilde{g}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r)d\widetilde{W}_r. \tag{5}$$

As in (Ref. 3) we can assume the Jacobian $\dfrac{\partial \widetilde{X}^u_{s,t}(\tilde{x})}{\partial \tilde{x}} = \widetilde{D}^u_{s,t}$ exists for all s, t, $\tilde{x}$ and all w not in a set of measure zero, and is the solution of

$$\widetilde{D}^u_{s,t} = I + \int_s^t \tilde{f}_{\tilde{x}}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r)\widetilde{D}^u_{s,r}dr + \sum_{i=1}^{n+m} \int_s^t \tilde{g}^{(i)}_{\tilde{x}}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r)\widetilde{D}^u_{s,r}d\widetilde{W}^i_r. \tag{6}$$

Here I is $(d+1)\times(d+1)$ identity matrix. In fact, if the coefficients $\tilde{f}$ and $\tilde{g}$ are C^k the map $\tilde{x} \to \widetilde{X}^u_{s,t}(\tilde{x})$ is C^{k-1}.

Similarly to (Ref. 3) the matrix process $\widetilde{H}$ defined by

$$\widetilde{H}^u_{s,t} = I - \int_s^t \widetilde{H}^u_{s,r}\Big(\tilde{f}_{\tilde{x}}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r) - \sum_{k=1}^{n+m} \tilde{g}^{(k)}_{\tilde{x}}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r)^2\Big)dr$$

$$- \sum_{k=1}^{m+n} \int_s^t \widetilde{H}^u_{s,r}\tilde{g}^{(k)}_{\tilde{x}}(r, \widetilde{X}^u_{s,r}(\tilde{x}), u_r)d\widetilde{W}^k_r. \tag{7}$$

exists and $\widetilde{H}^u_{s,t} = (\widetilde{D}^u_{s,t})^{-1}$.

REMARK 2.1. Write $\|\widetilde{X}^u(\tilde{x}_0)\|_t = \sup_{0\le s\le t} |\widetilde{X}^u_{0,s}(\tilde{x}_0)|$, $\|\widetilde{D}^u\|_t = \sup_{0\le s\le t} |\widetilde{D}^u_{0,s}|$, $\|\widetilde{H}^u\|_t = \sup_{0\le s\le t} |\widetilde{H}^u_{0,s}|$. Then $\|\widetilde{X}^u(\tilde{x}_0)\|_T$, $\|\widetilde{D}^u\|_T$, $\|\widetilde{H}^u\|_T$ are in L^P, $1 \le P < \infty$.

We shall suppose there is an optimal control $u^* \in \underline{U}$, so that $J(u^*) \le J(u)$ for all other $u \in \underline{U}$.

NOTATION 2.1. We shall write $\widetilde{X}^*$ for $\widetilde{X}^{u^*}$ and $\widetilde{D}^*_{0,t}$ for $\widetilde{D}^{u^*}_{0,t}$, etc.

3. Differentiability

Suppose $u^* \in \underline{U}$ is an optimal control. Consider any other control $v \in \underline{U}$. Then for $\theta \in [0,1]$

$$u_\theta(t) = u^*(t) + \theta(v(t) - u^*(t)) \in \underline{U}$$

and

$$J(u_\theta) \geq J(u^*). \tag{8}$$

If the Gâteaux derivative $J'(u^*)$ of J, as a functional on the Hilbert space $L^2_Y[0,T]$, is well defined, differentiating (8) in θ implies

$$\langle J'(u^*),\ v(t) - u^*(t)\rangle \geq 0$$

for all $v \in \underline{U}$.

LEMMA 3.1. *Suppose $v \in \underline{U}$ is such that $u^*_\theta = u^* + \theta v \in \underline{U}$ for $\theta \in [0,\alpha]$. Write $\widetilde{X}^\theta_{0,t}(\tilde{x}_0)$ for the trajectory associated with u^*_θ. Then $M_t = \frac{\partial \widetilde{X}^\theta_{0,t}(\tilde{x}_0)}{\partial\theta}\Big|_{\theta=0}$ exists a.s. and is the unique solution of the equation*

$$\begin{aligned} M_t = &\int_0^t (\tilde{f}_{\tilde{x}}(r, \widetilde{X}^*_{0,r}(\tilde{x}_0), u^*_r)M_r + \tilde{f}_u(r, \widetilde{X}^*_{0,r}(\tilde{x}_0), u^*_r)v_r)dr \\ &+ \sum_{i=1}^{n+m} \int_0^t \tilde{g}^{(i)}_{\tilde{x}}(r, \widetilde{X}^*_{0,r}(\tilde{x}_0), u^*_r)M_r d\widetilde{W}^i_r \\ &+ \sum_{i=1}^{n} \int_0^t \tilde{g}^{(i)}_u(r, \widetilde{X}^*_{0,r}(\tilde{x}_0), u^*_r)v_r d\widetilde{W}^i_r, \end{aligned} \tag{9}$$

because for $n+1 \leq i \leq n+m$, $\tilde{g}^{(i)}_u = 0$.

Proof. The result follows from the theorem of Blagovescenskii and Freidlin (Refs. 7–8) on the differentiability of solutions of stochastic differential equations which depend on a parameter. In effect the result of (Refs. 7–8) states that, if the coefficients are differentiable, the equation for the derivative is obtained by differentiation. Considering the initial condition as a parameter this result gives, in particular, the equation for the differential or Jacobian as in (6). □

LEMMA 3.2. *Write*

$$\tilde{\eta}_{0,t} = \int_0^t (\widetilde{D}^*_{0,r})^{-1} \tilde{f}_u(r) v_r dr + \sum_{i=1}^n \int_0^t (\widetilde{D}^*_{0,r})^{-1} \tilde{g}^{(i)}_u(r) v_r d\widetilde{W}^i_r$$

$$- \sum_{i=1}^n \int_0^t (\widetilde{D}^*_{0,r})^{-1} \tilde{g}^{(i)}_{\tilde{x}}(r) \tilde{g}^{(i)}_u(r) v_r dr \tag{10}$$

where $\tilde{f}_u$, $\tilde{g}_{\tilde{x}}$, $\tilde{g}_u$ *are as in equation (9). Then* $M_t = \widetilde{D}^*_{0,t}\tilde{\eta}_{0,t}$.

Proof. By differentiating, we see the product $\widetilde{D}^*_{0,t}\tilde{\eta}_{0,t}$ satisfies equation (9). □

LEMMA 3.3.

$$\left.\frac{dJ(u^*_\theta)}{d\theta}\right|_{\theta=0} = E\Big[\widetilde{C}_{\tilde{x}}(\widetilde{X}^*_{0,T}(\tilde{x}_0))\widetilde{D}^*_{0,T}\tilde{\eta}_{0,T}$$

$$+ \int_0^T (\tilde{\ell}_{\tilde{x}}(r, \widetilde{X}^*_{0,r}(\tilde{x}_0), u^*_r)\widetilde{D}^*_{0,r}\tilde{\eta}_{0,r} + \tilde{\ell}_u(r, \widetilde{X}^*_{0,r}(x_0), u^*_r) v_r) dr\Big]$$

where

$$\widetilde{C}(\widetilde{X}^*_{0,T}(\tilde{x}_0)) = Z^*_{0,T}(x_0) C(X^*_{0,T}(x_0))$$

$$\tilde{\ell}(r, \widetilde{X}^*_{0,r}, u^*_r) = Z^*_{0,r}(x_0)\ell(r, X^*_{0,r}(x_0), u^*_r).$$

Proof.

$$J(u^*_\theta) = E\Big[\widetilde{C}(\widetilde{X}^\theta_{0,T}(\tilde{x}_0)) + Z^\theta_{0,T}(x_0)\int_0^T \ell(r, X^\theta_{0,r}(x_0), u^*_\theta(r))dr)\Big]$$

$$= E\Big[\widetilde{C}(\widetilde{X}^\theta_{0,T}(\tilde{x}_0)) + \int_0^T Z^\theta_{0,r}\ell(r, X^\theta_{0,r}(x_0), u^*_\theta(r))dr + \int_0^T \Big(\int_0^r \ell(s)ds\Big) dZ^\theta_{0,r}\Big]$$

$$= E\Big[\widetilde{C}(\widetilde{X}^\theta_{0,T}(\tilde{x}_0)) + \int_0^T \tilde{\ell}(r, \widetilde{X}^\theta_{0,r}(\tilde{x}_0), u^*_\theta(r))dr\Big].$$

So

$$\left.\frac{dJ(u_\theta^*)}{d\theta}\right|_{\theta=0} = E\Big[\widetilde{C}_{\tilde{x}}(\widetilde{X}_{0,T}^*(\tilde{x}_0))M_T + \int_0^T (\tilde{\ell}_{\tilde{x}}(r,\widetilde{X}_{0,r}^*(\tilde{x}_0),u_r^*)M_r$$
$$+ \tilde{\ell}_u(r,\widetilde{X}_{0,r}^*(\tilde{x}_0),u_r^*)v_r)dr\Big]$$

substituting $M_t = \widetilde{D}_{0,t}^*\tilde{\eta}_{0,t}$ the result follows. □

Consider the right continuous version of the square integrable martingale

$$N_t = E\Big[\widetilde{C}_{\tilde{x}}(\widetilde{X}_{0,T}^*(\tilde{x}_0))\widetilde{D}_{0,T}^* + \int_0^T \tilde{\ell}_{\tilde{x}}(r,\widetilde{X}_{0,r}^*(\tilde{x}_0),u_r^*)\widetilde{D}_{0,r}^*dr \mid \mathcal{G}_t\Big]$$

where $\mathcal{G}_t$ is the right continuous complete σ-field on Ω, generated by $\mathcal{G}_t^0 = F_t^0 \otimes Y_t^0$.

From (Ref. 9) N_t has a martingale representation

$$N_t = E\Big[\widetilde{C}_{\tilde{x}}(\widetilde{X}_{0,T}^*(\tilde{x}_0))\widetilde{D}_{0,T}^* + \int_0^T \tilde{\ell}_{\tilde{x}}(r,\widetilde{X}_{0,r}^*(\tilde{x}_0),u_r^*)\widetilde{D}_{0,r}^*dr\Big] + \sum_{i=1}^{n+m}\int_0^t \tilde{\gamma}_r^i d\widetilde{W}_r^i$$

where the $\tilde{\gamma}_r^i$ are $\mathcal{G}_r$ predictable processes such that

$$E\Big[\int_0^T (\tilde{\gamma}_r^i)^2 dr\Big] < \infty.$$

Write

$$\xi_t = N_t - \int_0^t \tilde{\ell}_{\tilde{x}}(r,\widetilde{X}_{0,r}^*(\tilde{x}_0),u_r^*)\widetilde{D}_{0,r}^*dr$$

$$\tilde{p}_t = \xi_t(\widetilde{D}_{0,t}^*)^{-1}$$
$$= E\Big[\widetilde{C}_{\tilde{x}}(\widetilde{X}_{0,T}^*(\tilde{x}_0))\widetilde{D}_{t,T}^* + \int_t^T \tilde{\ell}_{\tilde{x}}(r,\widetilde{X}_{0,r}^*(\tilde{x}_0),u_r^*)\widetilde{D}_{t,r}^*dr \mid \mathcal{G}_t\Big].$$

THEOREM 3.1.

$$\left.\frac{dJ(u_\theta^*)}{d\theta}\right|_{\theta=0} = E\Big[\int_0^T \tilde{p}_s\tilde{f}_u(s)v_s ds - \sum_{i=1}^n \int_0^T \tilde{p}_s\tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)v_s ds + \int_0^T \tilde{\ell}_u(s)v_s ds$$
$$+ \sum_{i=1}^n \int_0^T \tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}\tilde{g}_u^{(i)}(s)v_s ds\Big]. \qquad (11)$$

Proof. The product rule gives

$$\xi_T \cdot \tilde{\eta}_{0,T} = \int_0^T \xi_r (\widetilde{D}^*_{0,s})^{-1} \tilde{f}_u(s) v_s ds$$
$$+ \sum_{i=1}^{n} \int_0^T \xi_s (\widetilde{D}^*_{0,s})^{-1} \tilde{g}^{(i)}_u(s) v_s d\widetilde{W}^i_s - \sum_{i=1}^{n} \int_0^T \xi_s (\widetilde{D}^*_{0,s})^{-1} g^{(i)}_x(s) g^{(i)}_u(s) v_s ds$$
$$+ \sum_{i=1}^{n+m} \int_0^T \tilde{\gamma}^{(i)}_s \tilde{\eta}_{0,s} d\widetilde{W}^{(i)}_s - \int_0^T \tilde{\ell}_{\tilde{x}}(s)(\widetilde{D}^*_{0,s})^{-1} \tilde{\eta}_{0,s} ds$$
$$+ \sum_{i=1}^{n} \int_0^T (\widetilde{D}^*_{0,s})^{-1} g^{(i)}_u(s) v_s \tilde{\gamma}^{(i)}_s ds. \tag{12}$$

However, from Lemma 3.3

$$\frac{dJ(u^*_\theta)}{d\theta}\Big|_{\theta=0} = E\Big[\xi_T \cdot \tilde{\eta}_{0,T} + \int_0^T (\tilde{\ell}_{\tilde{x}}(s) \widetilde{D}^*_{0,s} \tilde{\eta}_{0,s} + \tilde{\ell}_u(s) v_s) ds\Big]. \tag{13}$$

Substituting (12) in (13) and using the definition of $\tilde{p}$, the result follows. □

REMARK 3.1. Write $\widetilde{X}^*_{t,T}(\tilde{x}) = (X^*_{t,T}(x), Z^*_{t,T}(x,z))'$ for the solution of (4) using control u^*. Then, by uniqueness,

$$Z^*_{t,T}(x,z) = z Z^*_{t,T}(x,1) \tag{14}$$

and $Z^*_{t,T}(x,1)$ is the density given by (3).

LEMMA 3.4.

$$\frac{\partial Z^*_{t,T}(x,z)}{\partial z} = Z^*_{t,T}(x,1) \tag{15}$$

$$= Z^{*-1}_{0,t}(x_0,1) Z^*_{0,T}(x,1) \tag{16}$$

and

$$\frac{\partial Z^*_{t,T}(x,1)}{\partial x} = Z^*_{t,T}(x,1)\Big(\int_0^T \frac{\partial h(x^*_{t,r})}{\partial x} \cdot D^*_{t,r} dv_r\Big) \tag{17}$$

where $D^*_{t,r} = \dfrac{\partial X^*_{t,r}}{\partial x}$.

Proof. (15) is immediate from (14). Now

$$Z^*_{t,T}(x,1) = 1 + \int_t^T Z^*_{t,r}(x,1)h(X^*_{t,r}(x))dy_r.$$

Applying the differentiation result of Blagovescenskii and Freidlin (Ref. 7–8) we have

$$\frac{\partial Z^*_{t,T}(x,1)}{\partial x} = \int_t^T \frac{\partial Z^*_{t,r}(x,1)}{\partial x} h(X^*_{t,r}(x))dy_r + \int_t^T Z^*_{t,r}(x,1)\,\frac{\partial h}{\partial x}(X^*_{t,r}(x))D^*_{t,r}dy_r.$$

This equation can be solved by variation of constants to give

$$\frac{\partial Z^*_{t,T}(x,1)}{\partial x} = Z^*_{t,T}(x,1)\Big(\int_t^T \frac{\partial h}{\partial x}(X^*_{t,r}(x))D^*_{t,r}dy_r - \int_t^T \frac{\partial h}{\partial x}(X^*_{t,r}(x))D^*_{t,r}\cdot h(X^*_{t,r}(x))dr\Big)$$

and the result follows from (2). □

NOTATION 3.1. Write $Z^*_{0,t}$ for $Z^*_{0,t}(x_0,1)$, $Z^*_{t,T}$ for $Z^*_{t,T}(x,1)$,

$$\phi(t) = \Big(C_x(X^*_{0,T}(x_0))D^*_{t,T} + C(X^*_{0,T}(x_0))\Big(\int_t^T \frac{\partial h}{\partial x}\cdot D^*_{t,r}dv_r\Big), Z^{*-1}_{0,t}C(X^*_{0,T}(x_0))\Big)$$

and

$$\psi(r) = \Big(\ell_x(r)D^*_{t,r} + \ell(r)\Big(\int_t^r \frac{\partial h}{\partial x}\; D^*_{t,\eta}\cdot dv_\eta\Big), Z^{*-1}_{0,t}\ell(r)\Big).$$

Note that the linear growth conditions of ℓ and ℓ_x, the integrability properties of D^* and the boundedness of h and h_x imply that

$$\int_t^s \Big(\int_t^r \psi(\eta)d\eta\Big)dZ^*_{t,r}$$

is a square integrable martingale.

LEMMA 3.6.

$$\tilde{p}_t = E^*\Big[Z^*_{0,t}\Big(\phi(t) + \int_t^T \psi(r)dr\Big) \mid \mathcal{G}_t\Big]. \tag{18}$$

Proof.

$$\tilde{p}_t = E\Big[\widetilde{C}_{\tilde{x}}(\widetilde{X}^*_{0,T}(\tilde{x}_0))\widetilde{D}^*_{t,T} + \int_t^T \tilde{\ell}_{\tilde{x}}(r, \widetilde{X}^*_{0,r}(\tilde{x}_0), u^*_r)\widetilde{D}^*_{t,r}dr \mid \mathcal{G}_t\Big]$$

$$= E\Big[\Big(Z^*_{0,T}C_x(X^*_{0,T}(x_0))D^*_{t,T} + Z^*_{0,t}\frac{\partial Z^*_{t,T}(x)}{\partial x}C(X^*_{0,T}(x_0)), Z^*_{t,T}C(X^*_{0,T}(x_0))\Big)$$

$$+ \int_t^T \Big(Z^*_{0,r}\ell_x(r)D^*_{t,r} + Z^*_{0,r}\ \frac{\partial Z^*_{t,r}(x)}{\partial x} \cdot \ell(r), Z^*_{t,r}\ell(r)\Big)dr \mid \mathcal{G}_t\Big].$$

Substituting (17) this is

$$= E\Big[Z^*_{0,T}\Big\{C_x(X^*_{0,T}(x_0))D^*_{t,T} + C(X^*_{0,T}(x_0)\Big(\int_t^T \frac{\partial h}{\partial x} \cdot D^*_{t,r}dv_r\Big), Z^{*-1}_{0,t}C(X^*_{0,T}(x_0))\Big\}$$

$$+ \int_t^T Z^*_{0,r}\Big\{\ell_x(r)D^*_{t,r} + \ell(r)\Big(\int_t^r \frac{\partial h}{\partial x} \cdot D^*_{t,\eta} \cdot dv_\eta\Big), Z^{*-1}_{0,t}\ell(r)\Big\}dr \mid \mathcal{G}_t\Big]$$

$$= E\Big[Z^*_{0,T}\phi(t) + \int_t^T Z^*_{0,r}\psi(r)dr \mid \mathcal{G}_t\Big]. \tag{19}$$

Now

$$E\Big[\int_t^T Z^*_{0,r}\psi(r)dr \mid \mathcal{G}_t\Big] = Z^*_{0,t}E\Big[\int_t^T Z^*_{t,r}\psi(r)dr \mid \mathcal{G}_t\Big]$$

$$= Z^*_{0,t}E\Big[Z^*_{t,T}\int_t^T \psi(r)dr - \int_t^T \Big(\int_t^r \psi(\eta)d\eta\Big)dZ^*_{t,r} \mid \mathcal{G}_t\Big].$$

However, the last term is a square integrable $(P, \mathcal{G}_t)$ martingale, so

$$E\Big[\int_t^T Z^*_{0,r}\psi(r)dr \mid \mathcal{G}_t\Big] = Z^*_{0,t}E\Big[Z^*_{t,T}\int_t^T \psi(r)dr \mid \mathcal{G}_t\Big]$$

$$= E\Big[Z^*_{0,T}\int_t^T \psi(r)dr \mid \mathcal{G}_t\Big].$$

Substituting in (19)

$$\tilde{p}_t = E\Big[Z^*_{0,T}\Big(\phi(t) + \int_t^T \psi(r)dr\Big) \mid \mathcal{G}_t\Big]$$

and using Bayes' formula, this is

$$= E^*\Big[\phi(t) + \int_t^T \psi(r)dr \mid \mathcal{G}_t\Big] Z^*_{0,t}$$

$$= E^*\Big[Z^*_{0,t}\Big(\phi(t) + \int_t^T \psi(r)dr\Big) \mid \mathcal{G}_t\Big].$$

□

DEFINITION 3.2. The adjoint process will be the process p defined by

$$p_s = E[\tilde{p}_s \mid Y_s \vee \{x\}] \, \frac{E[Z^*_{0,s} \mid Y_s]}{E[Z^*_{0,s} \mid Y_s \vee \{x\}]}$$

$$= E\Big[Z^*_{0,T}\Big(\phi(s) + \int_s^T \psi(r)dr\Big) \Big| Y_s \vee \{x\}\Big] \, \frac{E[Z^*_{0,s} \mid Y_s]}{E[Z^*_{0,s} \mid Y_s \vee \{x\}]}$$

$$= E^*\Big[\phi(s) + \int_s^T \psi(r)dr \Big| Y_s \vee \{x\}\Big] E[Z^*_{0,s} \mid Y_s]$$

$$= E^*\Big[\Big(\phi(s) + \int_s^T \psi(r)dr\Big) E[Z^*_{0,s} \mid Y_s] \Big| Y_s \vee \{x\}\Big].$$

As in Bensoussan (Ref. 1), the adjoint process depends on x, which represents the state of the process at time s. However, x is just a parameter which is integrated out in the minimum principle of Theorem 3.2.

Returning to the perturbation

$$u_\theta(t) = u^*(t) + \theta(v(t) - u^*(t))$$

of the optimal control, we have

$$\frac{dJ(u_\theta)}{d\theta}\bigg|_{\theta=0} \geq 0.$$

That is

$$E\Big[\int_0^T \tilde{p}_s \tilde{f}_u(s)(v(s)-u^*(s))ds - \sum_{i=1}^n \int_0^T \tilde{p}_s \tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))ds$$

$$+\int_0^T \tilde{\ell}_u(s)(v(s)-u^*(s))ds + \sum_{i=1}^n \int_0^T \tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))ds\Big] \geq 0$$

for all $v \in \underline{U}$. Now

$$E\Big[\int_0^T \tilde{p}_s \tilde{f}_u(s)(v(s)-u^*(s))ds - \sum_{i=1}^n \int_0^T \tilde{p}_s \tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))ds$$

$$+\int_0^T \tilde{\ell}_u(s)(v(s)-u^*(s))ds + \sum_{i=1}^n \int_0^T \tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))ds\Big]$$

$$= E\Big[\int_0^T E\Big[\tilde{p}_s \tilde{f}_u(s) - \sum_{i=1}^n \tilde{p}_s \tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s) + \tilde{\ell}_u(s) + \sum_{i=1}^n \tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}\tilde{g}_u^{(i)}(s) \mid Y_s\Big]$$

$$\cdot (v(s)-u^*(s))ds\Big] \geq 0. \tag{20}$$

Therefore, because (20) is true for all $v \in \underline{U}$, we have for a.e. t and a.s. w.

$$E\Big[\tilde{p}_s \tilde{f}_u(s)(v(s)-u^*(s)) - \sum_{i=1}^n \tilde{p}_s \tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))$$

$$+\tilde{\ell}_u(s)(v(s)-u^*(s)) + \sum_{i=1}^n \tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}\tilde{g}_u^{(i)}(s)(v(s)-u^*(s)) \,\Big|\, Y_s\Big] \geq 0 \tag{21}$$

for all $v \in \underline{U}$.

From (21)

$$E\Big[\tilde{p}_s \tilde{f}_u(s)(v(s)-u^*(s)) - \sum_{i=1}^n \tilde{p}_s \tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))$$

$$+ \tilde{\ell}_u(s)(v(s) - u^*(s)) + \sum_{i=1}^{n} \tilde{\gamma}_s^i (\widetilde{D}_{0,s}^*)^{-1} \tilde{g}_u^{(i)}(s)(v(s) - u^*(s)) \Bigm| Y_s\Big]$$

$$= E\Big[E\Big[Z_{0,T}^*(\phi(s) + \int_s^T \psi(r)dr) \Bigm| \mathcal{G}_s\Big] \tilde{f}_u(s)(v(s) - u^*(s))$$

$$- \sum_{i=1}^{n} E\Big[Z_{0,T}^*(\phi(s) + \int_s^T \psi(r)dr) \Bigm| \mathcal{G}_s\Big] \tilde{g}_{\tilde{x}}^{(i)}(s) \tilde{g}_u^{(i)}(s)(v(s) - u^*(s))$$

$$+ Z_{0,T}^*(Z_{0,T}^{*-1} \tilde{\ell}_u(s))(v(s) - u^*(s))$$

$$+ Z_{0,T}^* \sum_{i=1}^{n} Z_{0,T}^{*-1} \tilde{\gamma}_s^i (\widetilde{D}_{0,s}^*)^{-1} \tilde{g}_u^{(i)}(s)(v(s) - u^*(s)) \Bigm| Y_s\Big]$$

$$= E^*\Big[\Big(\phi(s) + \int_s^T \psi(r)dr\Big) \tilde{f}_u(s)(v(s) - u^*(s))$$

$$- \sum_{i=1}^{n} \Big(\phi(s) + \int_s^T \psi(r)dr\Big) \tilde{g}_{\tilde{x}}^{(i)}(s) \tilde{g}_u^{(i)}(s)(v(s) - u^*(s))$$

$$+ Z_{0,T}^{*-1} \tilde{\ell}_u(s)(v(s) - u^*(s))$$

$$+ \sum_{i=1}^{n} Z_{0,T}^{*-1} \tilde{\gamma}_s^i (\widetilde{D}_{0,s}^*)^{-1} \tilde{g}_u^{(i)}(s)(v(s) - u^*(s)) \Bigm| Y_s\Big] \cdot E\Big[Z_{0,s}^* \Bigm| Y_s\Big]$$

$$= E^*\Big[\Big(\phi(s) + \int_s^T \psi(r)dr\Big) E[Z_{0,s}^* \mid Y_s] \tilde{f}_u(s)(v(s) - u^*(s))$$

$$- \sum_{i=1}^{n} \Big(\phi(s) + \int_s^T \psi(r)dr\Big) E[Z_{0,s}^* \mid Y_s] \tilde{g}_{\tilde{x}}^{(i)}(s) \tilde{g}_u^{(i)}(s)(v(s) - u^*(s))$$

$$+ Z_{0,T}^{*-1} \tilde{\ell}_u(s) E[Z_{0,s}^* \mid Y_s](v(s) - u^*(s))$$

$$+ \sum_{i=1}^{n} Z_{0,T}^{*-1} \tilde{\gamma}_s^i (\widetilde{D}_{0,s}^*)^{-1} E[Z_{0,s}^* \mid Y_s] \tilde{g}_u^{(i)}(s)(v(s) - u^*(s)) \mid Y_s\Big] \geq 0. \tag{22}$$

Write

$$\tilde{\tilde{\ell}}(s) = E[\tilde{\ell}(s) \mid Y_s \vee \{x\}] \ \frac{E[Z_{0,s}^* \mid Y_s]}{E[Z_{0,s}^* \mid Y_s \vee \{x\}]} = E^*[\ell(s) \mid Y_s \vee \{x\}] E[Z_{0,s}^* \mid Y_s]$$

$$\gamma_s^i = E[\tilde{\gamma}_s^i (D_{0,s}^*)^{-1} \mid Y_s \vee \{x\}] \ \frac{E[Z_{0,s}^* \mid Y_s]}{E[Z_{0,s}^* \mid Y_s \vee \{x\}]}.$$

Define the Hamiltonian by

$$H(\tilde{x},v,t,p(t)) = p_t\tilde{f}_u(t,\tilde{x},v) - \sum_{i=1}^{n} p_t\tilde{g}_{\tilde{x}}^{(i)}(t,\tilde{x},u_t^*)\tilde{g}(t,\tilde{x},v) + \tilde{\tilde{\ell}}(t,\tilde{x},v) + \sum_{i=1}^{n} \gamma_t^i\tilde{g}^{(i)}(t,\tilde{x},v).$$

THEOREM 3.2. *If u^* is the optimal control, then a.e. s*

$$E^*\Big[\frac{\partial H}{\partial v}\,(\tilde{x},u^*,s,p(s))(v(s)-u^*(s))\,\Big|\,Y_s\Big] \geq 0 \qquad \text{a.s.}$$

Proof. From (14), $\tilde{f}_u(s)$ and $\tilde{g}^{(i)}(s)$ $(i \leq n)$ are $Y_s \vee \{x\}$ measurable. Therefore,

$$\begin{aligned}
0 \leq E^*\Big[&\Big(\phi(s)+\int_s^T \psi(r)dr\Big)E[Z_{0,s}^* \mid Y_s]\tilde{f}_u(s)(v(s)-u^*(s)) \\
&-\sum_{i=1}^{n}\Big(\phi(s)+\int_s^T \psi(r)dr\Big)E[Z_{0,s}^* \mid Y_s]\tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s)) \\
&+Z_{0,T}^{*-1}\tilde{\ell}_u(s)E[Z_{0,s}^* \mid Y_s](v(s)-u^*(s)) \\
&+\sum_{i=1}^{n} Z_{0,T}^{*-1}\tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}E[Z_{0,s}^* \mid Y_s]\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))\,\Big|\,Y_s\Big] \\
= E^*\Big\{E^*\Big[&\Big(\phi(s)+\int_s^T \psi(r)dr\Big)E[Z_{0,s}^* \mid Y_s]\tilde{f}_u(s)(v(s)-u^*(s)) \\
&-\sum_{i=1}^{n}\Big(\phi(s)+\int_s^T \psi(r)dr\Big)E[Z_{0,s}^* \mid Y_s]\tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s)) \\
&+Z_{0,T}^{*-1}\tilde{\ell}_u(s)E[Z_{0,s}^* \mid Y_s](v(s)-u^*(s)) \\
&+\sum_{i=1}^{n} Z_{0,T}^{*-1}\tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1}E[Z_{0,s}^* \mid Y_s]\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))\,\Big|\,Y_s \vee \{x\}\Big]\,\Big|\,Y_s\Big\} \\
= E^*\Big[&p_s\tilde{f}_u(s)(v(s)-u^*(s)) - \sum_{i=1}^{n} p_s\tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s)) \\
&+E^*[Z_{0,T}^{*-1}\tilde{\ell}_u(s) \mid Y_s \vee \{x\}]\cdot E[Z_{0,s}^* \mid Y_s](v(s)-u^*(s)) \\
&+\sum_{i=1}^{n} E^*[Z_{0,T}^{*-1}\tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1} \mid Y_s \vee \{x\}]E[Z_{0,s}^* \mid Y_s]\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))\,\Big|\,Y_s\Big]
\end{aligned}$$

$$= E^*\Big[p_s\tilde{f}_u(s)(v(s)-u^*(s)) - \sum_{i=1}^n p_s\tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))$$

$$+ E[\tilde{\ell}_u(s) \mid Y_s \vee \{x\}]\ \frac{E[Z_{0,s}^* \mid Y_s]}{E[Z_{0,s}^* \mid Y_s \vee \{x\}]}\cdot(v(s)-u^*(s))$$

$$+ \sum_{i=1}^n E[\tilde{\gamma}_s^i(\widetilde{D}_{0,s}^*)^{-1} \mid Y_s \vee \{x\}]\ \frac{E[Z_{0,s}^* \mid Y_s]}{E[Z_{0,s}^* \mid Y_s \vee \{x\}]}\cdot \tilde{g}_u^{(i)}(s)(v(s)-u^*(s))\ \Big|\ Y_s\Big]$$

$$= E^*\Big[p_s\tilde{f}_u(s)(v(s)-u^*(s)) - \sum_{i=1}^n p_s\tilde{g}_{\tilde{x}}^{(i)}(s)\tilde{g}_u^{(i)}(s)(v(s)-u^*(s))$$

$$+ \tilde{\tilde{\ell}}_u(s)(v(s)-u^*(s)) + \sum_{i=1}^n \gamma_s^i\tilde{g}_u^{(i)}(s)(v(s)-u^*(s)) \mid Y_s\Big]$$

$$= E^*\Big[\frac{\partial H}{\partial v}(\tilde{x},u^*,s,p(s))(v(s)-u^*(s))\ \Big|\ Y_s\Big].$$

So the result follows. □

Referencesplus1pt

[1] BENSOUSSAN, A., *Maximum Principle and Dynamic Programming Approaches of the Optimal Control of Partially Observed Diffusions*, Stochastics, Vol. 9, pp. 169–222, 1983.

[2] HAUSSMANN, U.G., *The Maximum Principle for Optimal Control of Diffusions with Partial Information*, SIAM Journal of Control and Optimization, Vol. 25, pp. 341–361, 1987.

[3] BARAS, J., ELLIOTT, R.J., and KOHLMANN, M., *The Partially Observed Stochastic Minimum Principle*, SIAM Journal of Control and Optimization (to appear).

[4] ELLIOTT, R.J., *The Optimal Control of a Stochastic System*, SIAM Journal of Control and Optimization, Vol. 15, pp. 756–778, 1977.

[5] BENSOUSSAN, A., *Lectures on Stochastic Control*, Springer–Verlag, Berlin, Germany, 1982.

[6] ELLIOTT, R.J., *The Optimal Control of Diffusions*, University of Alberta, Technical Report 89.12, 1989.

[7] BLAGOVESCENSKII, J.N., and FREIDLIN, M.I., *Some Properties of Diffusion Process Depending on a Parameter*, Doklady Akademia Nauk SSSR pg. 138, 1961.

[8] BLAGOVESCENSKII, J.N., and FREIDLIN, M.I., *Some Properties of Diffusion Process Depending on a Parameter*, Soviet Mathematics, Vol. 2, pp. 633–636, 1961.

[9] ELLIOTT, R.J., *Stochastic Calculus and Applications*, Applications of Mathematics, Vol. 18, Springer-Verlag, Berlin, Germany, 1982.

Chapter Two

Forward and Backward Equations for an Adjoint Process

1. Introduction

Without loss of generality the state space of a finite state Markov chain can be taken to be the set of unit basis vectors in R^N. We suppose such a chain X_t, $0 \leq t \leq T$, is observed through the noisy process y, where

$$y_t = \int_0^t h(X_s)ds + w_t. \tag{1.1}$$

Here w is a Brownian motion independent of X.

For simplicity a terminal cost of the form $\langle \ell, X_T \rangle$ is considered and, following Davis, [5], the control problem is formulated in separated form by considering an unnormalized conditional distribution of X_t.

An adjoint process is introduced and shown to satisfy forward and backward equations. Early works of Bismut [3], [4] have discussed the adjoint process using different methods. Some of our techniques are related to those of Bensoussan [1]. A similar problem for a controlled Markov chain for which only the jump times, but not the jump locations, can be observed is discussed in [7].

A version of this chapter has been published. Robert J. Elliott and Hailiang Yang. Stochastic Processes. A Festschrift in honour of Gopinath Kallianpur. Springer-Verlag, 1993, 61–70.

2. The System

The formulation in this section is well known. Let $\{X_t\}$, $t \in [0, T]$ be a Markov chain, defined on a probability space (Ω, F, P), whose state space is the set

$$S = \{e_1, \ldots, e_N\}$$

where $e_i = (0, \ldots, 1, 0, \ldots, 0)'$, $i = 1, 2, \ldots, N$, is a unit vector of R^N.

Write $p_t^i = P(X_t = e_i)$, $0 \leq i \leq W$. We shall suppose that for some family of matrices which depend on the control parameter $A_t(u)$, $p_t = (p_t^1, \ldots, p_t^N)'$ satisfies the forward Kolmogorov equation

$$\frac{dp_t}{dt} = A_t(u) p_t. \tag{2.1}$$

$A = (a_{ij}(t, u))$, $t \geq 0$, is the family of Q matrices of the process. We shall suppose the $a_{ij}(t, u)$ are differentiable in u.

Suppose y_t is a Brownian motion process on (Ω, F, P) independent of X_t and write Y_t for the right continuous, complete filtration generated by y. The set $\underline{U}$ of admissible controls will be the set of Y-predictable functions with values in a compact, convex set $U \subset R^k$. Suppose h is a real valued function on S, (so h is just given by a vector $h = (h(e_1), \ldots, h(e_N))$. For $u \in \underline{U}$ write $\Lambda_{s,t}^u = \exp\left\{\int_s^t h(X_r^u) dy_r - \frac{1}{2}\int_s^t |h(X_r^u)|^2 dr\right\}$ and define a new probability measure P^u by

$$\frac{dP^u}{dP} = \Lambda_{0,T}^u. \tag{2.2}$$

Then according to Girsanov's Theorem, P^u is a probability measure, and under P^u the process W_t is a Brownian motion, where W_t is defined by

$$y_t = \int_0^t h(X_s^u) ds + W_t. \tag{2.3}$$

Also $\{X_t\}$ and $\{W_t\}$ are independent, and $\{X_t\}$ has the same distribution as under measure P. Note that under P^u the process y represents a noisy observation of $\int_0^t h(X_s^u)ds$ as in (1.1).

From Davis [5] or Elliott [6] we know that if $\hat{p}_t^i$ is the Y_t optional projection of $I_{\{X_t=e_i\}}$ under P^u, then $\hat{p}_t^i = E^u[I_{\{X_t=e_i\}} \mid Y_t] = P^u[X_t = e_i \mid Y_t]$ a.s. and $\hat{p}_t = (\hat{p}_t^1, \ldots, \hat{p}_t^w)'$ satisfies

$$d\hat{p}_t = A(u_t)\hat{p}_t dt + (H - \tilde{h}'\hat{p}_t I)\hat{p}_t dv_t, \tag{2.4}$$

where $\tilde{h}' = (h(e_1), \ldots, h(e_N))$, H is the $N \times N$ diagonal matrix with elements $H_{ii} = h(e_i)$ and I is the $N \times N$ identity matrix. v_t is the innovations process, given by

$$dv_t = dy_t - \tilde{h}'\hat{p}_t dt. \tag{2.5}$$

Now $\Lambda_{0,T} = \Lambda_{0,t} \cdot \Lambda_{t,T}$ so

$$\begin{aligned}\hat{p}_t^i &= E^u[I_{\{X_t=e_i\}} \mid Y_t] \\ &= \frac{E[\Lambda_{0,t}^u I_{\{X_t=e_i\}} \mid Y_t]}{E[\Lambda_{0,t}^u \mid Y_t]}.\end{aligned}$$

Let q_t^i be the Y_t optimal projection of $\Lambda_{0,t}^u I_{\{X_t=e_i\}}$. Then

$$q_t^i = E[\Lambda_{0,t}^u I_{\{X_t=e_i\}} \mid Y_t]$$

a.s. and $q_t = (q_t^1, \ldots, q_t^N)'$, the unnormalized density of X_t conditional on Y_t, satisfies

$$q_t = \hat{p}_t E[\Lambda_{0,t}^u \mid Y_t] = \hat{p}_t \overline{\Lambda}_{0,t}^u. \qquad 6)$$

Here $\overline{\Lambda}^u$ is the Y-optional projection of Λ^u under measure P, so that $\overline{\Lambda}^u_t = E[\Lambda^u_{0,t} \mid Y_t]$ a.s. Now $\overline{\Lambda}^u_{0,t}$, satisfies

$$d\overline{\Lambda}^u_{0,t} = \overline{\Lambda}^u_{0,t}\hat{h}(X_t)dy_t. \tag{2.7}$$

Here

$$\hat{h}(X_t) = E[\Lambda^u_{0,t}h(X_t) \mid Y_t]/E[\Lambda^u_{0,t} \mid Y_t]$$

$$= E^u[h(X_t) \mid Y_t].$$

Therefore, by Itô's rule from (2.6), (2.4) and (2.7)

$$dq_t = \hat{p}_t\overline{\Lambda}_{0,t}\hat{h}(X_t)dy_t + A\hat{p}_t\overline{\Lambda}_{0,t}dt$$

$$+ (H - \tilde{h}'\hat{p}_tI)\hat{p}_t\overline{\Lambda}_{0,t}dv_t + (H - \tilde{h}'\hat{p}_tI)\hat{p}_t\overline{\Lambda}_{0,t}\hat{h}(X_t)dt.$$

Since

$$\tilde{h}'\hat{p}_t = \sum_{i=1}^{N} h(e_i)E^u[I_{\{X_t=e_i\}} \mid Y_t]$$

$$= E^u\Big[\sum_{i=1}^{N} h(e_i)I_{\{X_t=e_i\}} \mid Y_t\Big] = E^u[h(X_t) \mid Y_t].$$

we see

$$dq_t = q_t\hat{h}(X_t)dy_t + Aq_tdt + Hq_tdy_t$$

$$- \hat{h}(X_t)q_tdy_t - (H - \hat{h}(X_t)I)q_t\hat{h}(X_t)dt$$

$$+ (H - \hat{h}(X_t)I)q_t\hat{h}(X_t)dt.$$

That is, q_t satisfies the Zakai equation

$$dq_t = A_t(u)q_t dt + Hq_t dy_t. \tag{2.8}$$

Cost: The cost function will be

$$J(u) = E^u[\langle \ell, X_T^u\rangle] = E[\Lambda_{0,T}^u \langle \ell, X_T^u\rangle]$$

$$= E[\langle \ell, \Lambda_T^u X_T^u\rangle] = E[\langle \ell, E[\Lambda_T^u X_T^u \mid Y_t]\rangle]$$

$$= E[\langle \ell, q_T^u\rangle].$$

The control problem has, therefore, been formulated in separated form: find $u \in \underline{U}$ which minimizes

$$J(u) = E[\langle \ell, q_T^u\rangle] \tag{2.9}$$

where q is given by (2.8).

3. Differentiation

For $u \in \underline{U}$ write $\Phi^u(t,s)$ for the fundamental matrix solution of

$$d\Phi^u(t,s) = A_t(u)\Phi^u(t,s)dt + H\Phi^u(t,s)dy_t \tag{3.1}$$

with initial condition $\Phi^u(s,s) = I$, the $N \times N$ identity matrix.

LEMMA 3.1. *For $u \in \underline{U}$, consider the matrix Ψ^u defined by the equation*

$$\Psi^u(t,s) = I - \int_s^t \Psi^u(r,s)A_r(u)dr$$

$$- \int_s^t \Psi^u H dy_r + \int_s^t \Psi^u H^2 dr. \tag{3.2}$$

Then $\Psi^u \Phi^u = I$ *for* $t \geq s$.

Proof: Using the Itô rule we see $d(\Psi\Phi) = 0$, $\Psi(s,s)\Phi(s,s) = I$.

We shall suppose there is an optimal control $u^* \in \underline{U}$. Write q^* for q^{u^*}, Φ^* for Φ^{u^*} etc. Consider any other control $v \in \underline{U}$. Then for $\theta \in [0,1]$, $u_\theta(t) = u^*(t) + \theta(v(t) - u^*(t)) \in \underline{U}$.

Now

$$J(u_\theta) \geq J(u^*). \tag{3.3}$$

Therefore, if the Gâteaux derivative $J'(u^*)$ of J, as a functional on the Hilbert space $H = L^2[\Omega \times [0,T], R^k]$, is well defined, differentiating (3.3) in θ and letting $\theta = 0$, we have

$$\langle J'(u^*),\ v(t) - u^*(t)\rangle \geq 0 \tag{3.4}$$

for all $v \in \underline{U}$.

LEMMA 3.2. *Suppose* $v \in \underline{U}$ *is such that* $u_\theta^* = u^* + \theta v \in \underline{U}$ *for* $\theta \in [0,\alpha]$. *Write* $q_t(\theta)$ *for the solution* $q_t(u_\theta^*)$ *of (2.8). Then* $z_t = \left.\frac{\partial q_t(\theta)}{\partial \theta}\right|_{\theta=0}$ *exists and is the unique solution of the equation*

$$z_t = \int_0^t \Big(\frac{\partial A}{\partial u}(u^*)\Big) v_r q_r^* dr + \int_0^t A(u_r^*) z_r dr + \int_0^t H z_r dy_r. \tag{3.5}$$

Proof: Differentiating under the integrals gives the result. This is justified by the result of [4].

LEMMA 3.3. *Write*

$$\eta_{0,t} = \int_0^t \Psi^*(r,0)\Big(\frac{\partial A}{\partial u}(u^*)\Big) v_r q_r^* dr. \tag{3.6}$$

Then $z_t = \Phi^*(t,0)\eta_{0,t}$.

Proof: By Itô's rule we see $\Phi^*(t,0)\eta_{0,t}$ satisfies the equation (3.5).

COROLLARY 3.4. *Because* $J(u_\theta^*) = E[\langle \ell, q_T^{u_\theta^*} \rangle]$ *we see*

$$\frac{\partial J}{\partial \theta}(u_\theta^*)\Big|_{\theta=0} = E[\langle \ell, \Phi^*(T,0)\eta_{0,T} \rangle]. \tag{3.7}$$

Write $M_t = E[\Phi^*(T,0)'\ell \mid Y_t]$. Then M_t is a square integrable martingale on the Y-filtration; hence, (see [6]), M_t has representation

$$M_t = E[\Phi^*(T,0)'\ell] + \int_0^t \gamma_r dy_r \tag{3.8}$$

where γ is a $\{Y_t\}$ predictable process, such that

$$\int_0^T E|\gamma_r^2| dr < \infty.$$

DEFINITION 3.5 The adjoint process is

$$p_t := \Psi^*(t,0)' M_t$$

where the prime $'$ denotes the transpose of the matrix.

THEOREM 3.6.

$$\frac{\partial J(u_\theta^*)}{\partial \theta}\Big|_{\theta=0} = \int_0^T E\Big[\Big\langle p_r, \frac{\partial A}{\partial u}(u^*) v_r q_r^* \Big\rangle\Big] dr. \tag{3.9}$$

Proof. Using (3.6) and (3.8)

$$\langle M_T, \eta_{0,T} \rangle = \int_0^T \Big\langle M_r, \Psi^*(r,0) \frac{\partial A}{\partial u}(u^*) v_r q_r^* \Big\rangle dr + \int_0^T \langle \gamma_r, \eta_{0,r} \rangle dy_r.$$

From (3.7)

$$\frac{\partial J(u_\theta^*)}{\partial \theta}\Big|_{\theta=0} = E[\langle M_T, \eta_{0,T} \rangle],$$

so the result follows. □

Under integrability conditions J' is in H, and so a Gâteaux derivative.

Now consider perturbations of u^* of the form $u_\theta(t) = u^*(t) + \theta(v(t) - u^*(t))$ for $\theta \in [0,1]$, and any $v \in \underline{U}$. Then

$$\frac{\partial J(u_\theta)}{\partial \theta}\Big|_{\theta=0} = \langle J'(u^*),\ v - u^* \rangle \geq 0.$$

for all $v \in \underline{U}$. So we have the following

THEOREM 3.7. *Suppose $u^* \in \underline{U}$ is an optimal control. Then a.s. in w and a.e. in t*

$$\left\langle p_t, \frac{\partial A}{\partial u}(u^*)(v_t - u_t^*)q_t^* \right\rangle \geq 0. \tag{3.10}$$

4. Equations for the Adjoint Process

Suppose the optimal control u^* is a Markov, feedback control in the state variable q.

We have the following expression for the integrand in (3.8).

LEMMA 4.1.

$$\gamma_r = \Phi^*(r,0)' \frac{\partial p_r}{\partial q} H q_r + \Phi^*(r,0)' H p_r. \tag{4.1}$$

Proof: $\Phi^*(t,0)'p_t = M_t = E[\Phi^*(T,0)'\ell] + \int_0^t \gamma_r dy_r$. If u^* is Markov, q^* is also Markov. Write $q = q_t^*$, $\Phi = \Phi^*(t,0)$, then by the Markov property

$$\begin{aligned} E[\Phi^*(T,0)'\ell \mid Y_t] &= E[\Phi'\Phi^*(T,t)'\ell \mid q, \Phi] \\ &= \Phi' E[\Phi^*(T,t)'\ell \mid q]. \end{aligned}$$

So $p_t = E[\Phi^*(T,t)'\ell \mid q]$ is a function of q only. Therefore,

$$\begin{aligned} p_t = p_0 &+ \int_0^t \frac{\partial p}{\partial q}(Aq_r dr + Hq_r dy_r) \\ &+ \int_0^t \frac{\partial p_r}{\partial r}\, dr + \frac{1}{2}\sum_{i,j=1}^N \int_0^t \frac{\partial^2 p_r}{\partial q^i \partial q^j}\, h(e_i)h(e_j)q_r^i q_r^j dr \\ = p_0 &+ \int_0^t \left[\frac{\partial p_r}{\partial q}\, Aq_r + \frac{1}{2}\sum_{i,j=1}^N \frac{\partial^2 p_r}{\partial q^i \partial q^j} h(e_i)h(e_j)q_r^i q_r^j + \frac{\partial p_r}{\partial r}\right] dr \\ &+ \int_0^t \frac{\partial p_r}{\partial q}\, Hq_r dy_r. \end{aligned}$$

Then

$$M_t = \Phi^*(t,0)'p_t$$

$$= p_0 + \int_0^t \Phi^*(r,0)' \Big[\frac{\partial p_r}{\partial q} A q_r + \frac{1}{2}\sum_{ij=1}^{N} \frac{\partial^2 p_r}{\partial q_r^i \partial q_r^j} h(e_i)h(e_j) q_r^i q_r^j + \frac{\partial p_r}{\partial r}\Big] dr$$

$$+ \int_0^t \Phi^*(r,0)' \frac{\partial p_r}{\partial q} \; H q_r dy_r + \int_0^t \Phi^*(r,0)' A(u)' p_r dr$$

$$+ \int_0^t \Phi^*(r,0)' H p_r dy_r + \int_0^t \Phi^{*\prime} H \frac{\partial p_r}{\partial q} \; H q_r dr. \tag{4.2}$$

Since M_t is a Martingale the sum of the dr integrals in (4.2) must be 0, and, therefore,

$$\gamma_r = \Phi^*(r,0)' \; \frac{\partial p_r}{\partial q} \; H q_r + \Phi^*(r,0)' H p_r.$$

□

THEOREM 4.2.

$$p_t = E[\Phi^*(T,0)'\ell] + \int_0^t \frac{\partial p_r}{\partial q} \; H q_r dy_r - \int_0^t (A' p_r + H \; \frac{\partial p_r}{\partial q} \; H q_r) dr. \tag{4.3}$$

Proof:

$$p_t = \Psi^*(t,0)' M_t = E[\Phi^*(T,0)'\ell]$$

$$+ \int_0^t \Psi^{*\prime} \Big(\Phi^{*\prime} \frac{\partial p_r}{\partial q} \; H q_r + \Phi^{*\prime} H p_r \Big) dy_r$$

$$- \int_0^t A' \Psi^{*\prime} M_r dr - \int_0^t H \Psi^{*\prime} M_r dy_r$$

$$+ \int_0^t H^2 \Psi^{*\prime} M_r dr - \int_0^t H \Psi^{*\prime} \Big(\Phi^{*\prime} \frac{\partial p_r}{\partial q} \; H q_r + \Phi^{*\prime} H p_r \Big) dr$$

$$= E[\Phi^*(T,0)'\ell] + \int_0^t \Big(\frac{\partial p_r}{\partial q}\ Hq_r + Hp_r\Big)dy_r$$
$$- \int_0^t A'p_r dr - \int_0^t Hp_r dy_r + \int_0^t H^2 p_r dr$$
$$- \int_0^t \Big(H\ \frac{\partial p_r}{\partial q}\ Hq_r + H^2 p_r\Big)dr.$$

So

$$p_t = E[\Phi^*(T,0)'\ell] + \int_0^t \frac{\partial p_r}{\partial q}\ Hq_r dy_r$$
$$- \int_0^t \Big(A'p_r + H\ \frac{\partial p_r}{\partial q}\ Hq_r\Big)dr.$$

□

From (4.2), equating the dr integrals to zero we also obtain the following result.

THEOREM 4.3. *p_t satisfies the backward parabolic system*

$$\frac{\partial p_t}{\partial t} + \frac{\partial p_t}{\partial q}\ Aq_t + H\ \frac{\partial p_t}{\partial q}\ Hq_t + \frac{1}{2}\sum_{ij=1}^{N} \frac{\partial^2 p_t}{\partial q^i \partial q^j} h(e_i)h(e_j)q_t^i q_t^j + A(u^*)'p_t = 0. \quad (4.4)$$

with terminal condition

$$p_T = \ell.$$

REMARKS 4.4. In [3] Bismut considers a forward equation, with a terminal condition, for the adjoint process.

References

[1] A. Bensoussan, Lectures on stochastic control. In *Lecture Notes in Mathematics*, Vol. 972. Springer-Verlag, Berlin, Heidelberg, New York, 1982.

[2] J.M. Bismut, Linear quadratic optimal stochastic control with random coefficients. *S.I.A.M. J. Control and Optimization*, Vol. 14, No. 3 (1976), 419–444.

[3] J.M. Bismut, An introductory approach to duality in optimal stochastic control. *S.I.A.M. Review* 20 (1978), 62–78.

[4] J.N. Blagovescenskii and M.I. Freidlin, Some properties of diffusion process depending on a parameter. *Dokl. Akad. Nauk.* 138 (1961); *Soviet Math.*, 2 (1961), 633–636.

[5] M.H.A. Davis, Nonlinear semigroups in the control of partially observable stochastic systems. In *Lecture Notes in Mathematics*, Vol. 695, Springer-Verlag, Berlin, Heidelberg, New York, 1977, 37–49.

[6] R.J. Elliott, *Stochastic Calculus and Applications.* Springer-Verlag, Berlin, Heidelberg, New York, 1982.

[7] R.J. Elliott, A partially observed control problem for Markov chains. To appear.

Chapter Three

How to Count and Guess Well: Discrete Adaptive Filters

1. Introduction

Discrete filtering consists of counting the occurences of various states of a discrete observation process and then making the best estimates of quantities related to an unobserved state process. The situation we consider in this paper is that of a Hidden Markov Model, HMM. This consists of a homogeneous, finite state, discrete time Markov chain X_ℓ, $\ell \in Z^+$. Without loss of generality, the state space of X can be taken to be the set of unit vectors $S = \{e_1, \ldots, e_N\}$, $e_i = (0, \ldots, 1, 0, \ldots, 0)' \in R^N$, and with $P = (p_{ji})$ the matrix of transition probabilities it is shown in Section 4 that X has a semimartingale form

$$X_\ell = PX_{\ell-1} + m_\ell, \qquad \ell \in Z^+, \tag{1.1}$$

where m_ℓ is a martingale increment. The process X is not observed directly; instead we suppose there is a discrete time, finite state observation process Y. Again, the finite range of Y can be identified with the set of unit vectors $\{f_1, \ldots, f_k\}$, $f_j = (0, \ldots, 1, 0, \ldots, 0)' \in R^k$, and with $D = (d_{ji})$ where $d_{ji} = P(Y_\ell = f_j \mid X_{\ell-1} = e_i)$, the relation between X and Y can also be expressed in semimartingale form as:

$$Y_\ell = DX_{\ell-1} + \mu_\ell, \qquad \ell \in Z^+, \tag{1.2}$$

A version of this chapter has been submitted for publication. Robert J. Elliott and Hailiang Yang.

where μ_ℓ is a martingale increment.

This situation is often called a Hidden Markov Model, HMM, and such structures have been found useful in many areas of probabilistic modelling, including speech processing; see Rabiner [5]. We believe our model is of wide applicability and generality. Many state and observation processes of the form (1.1) and (1.2) arise in the literature. Our model, by discretizing time and state, can be made an approximation to many continuous models, including non-linear diffusion models in continuous time. In particular, by discretizing and approximating the noise in the observations, the case of a Markov chain observed in Gaussian noise can be approximated by a HMM. Discretization is necessary for numerical implementation. Furthermore, our model re-tunes its parameters in an increasingly optimal way. In addition, certain time series models can be approximated by an HMM.

The methods of this paper are an adaptation to discrete time and state of those of Elliott [4]. The application of methods from the discrete time general theory of processes in this situation is not well known and sheds light on the problems. Early contributions can be found in the papers of Boel [3] and Segall [6]. The latter is in the spirit of this paper, but discusses only a single counting observation process. Boel has considered multi-dimensional point processes, but has not introduced Zakai equations or the change of measure. Our paper begins with martingale representation and Girsanov results related to multivariate point processes. The semimartingale representation of the HMM is next given and followed by a general filtering result. A general unnormalized, or Zakai, estimate is derived in Section 6.

Compared with the normalized filter this has a remarkably simple form. Specializing this result in Section 7 gives recursive estimates and smoothers for the state of the process, the number of jumps from one state to another, the occupation time of a state and of a process related to the observations.

Following Elliott [4], a particular trick used is to exploit the idempotent property of X; instead of estimating H, which would involve HX, we estimate HX. This introduces $HX \otimes X$ but this can be expressed in terms of HX itself and so, unlike H, HX has a recursive estimate. Taking the inner product with $\underline{1} = (1, 1, \dots, 1)$ then gives an estimate for H. From these estimates new optimal values for the parameters p_{ji} and d_{ji} in the matrices P and D can be obtained. Our model is, therefore, adaptive or 'self tuning' to the observations. Using the new parameters and perhaps new observations a sequence of increasingly better models can be obtained.

2. Discrete Time Martingale Representation

Processes and random variables considered in the sequel are supposed defined on a complete probability space (Ω, F, P). The discrete time parameter will take values in $Z^+ = \{1, 2, \dots\}$. A filtration $\{F_\ell\}$, $\ell \in Z^+$, is given; that is the F_ℓ are an increasing family $F_1 \subset F_2 \subset \dots$ of sub-σ fields of F, and F_1 is complete. F_0 is the trivial σ-field (Ω, ϕ).

Suppose for each $\ell \in Z^+$, $y_\ell^1, \dots, y_\ell^k$ are random variables each of which takes either the value 0 or 1, and for each time ℓ one and only one of the $y_\ell^i = 1$. That is $y_\ell^i y_\ell^j = 0$ if $i \neq j$ and

$$\sum_{i=1}^{k} y_\ell^i = 1. \tag{2.1}$$

Write $y_\ell = (y_\ell^1, \ldots, y_\ell^k)' \in R^k$. (As we wish to consider column vectors the prime $'$ denotes the transpose.) With $1 \le j \le k$ write $f_j = (0, \ldots, 1, 0, \ldots, 0)'$ for the set of standard unit vectors of R^k; then we see $y_\ell = f_j$ for some j.

Write

$$Y_n^1 = \sum_{\ell=1}^n y_\ell^1, \ldots, Y_n^k = \sum_{\ell=1}^n y_\ell^k$$

$$Y_n = (Y_n^1, \ldots, Y_n^k)' = \sum_{\ell=1}^n y_\ell.$$

We take $Y_0 = 0 \in R^k$.

DEFINITION 2.1. $a_\ell^i = E[y_\ell^i | F_{\ell-1}], \quad \ell \ge 1.$

Write

$$a_\ell = (a_\ell^1, \ldots, a_\ell^k)'.$$

From (2.1) we see that

$$\sum_{i=1}^k a_\ell^i = \sum_{i=1}^k E[y_\ell^i \mid F_{\ell-1}] = 1. \tag{2.2}$$

If $\mu_\ell^i := y_\ell^i - a_\ell^i$ then $E[\mu_\ell^i \mid F_{\ell-1}] = 0$ so μ_ℓ^i is a martingale difference.

Write $\mu_\ell = (\mu_\ell^1, \ldots, \mu_\ell^k)'$. From (2.1) and (2.2)

$$\sum_{i=1}^k \mu_\ell^i = 0 \tag{2.3}$$

so the dimension of the space spanned by the martingale differences μ_ℓ^i, $1 \le i \le k$, is at most $k-1$. For $n \in Z^+$ write $\mathcal{Y}_n = \sigma\{Y_\ell : \ell \le n\} = \sigma\{y_\ell : \ell \le n\}$ so that $\mathcal{Y}_n \subset F_n$.

NOTATION 2.2. For any vector $\alpha = (\alpha_1, \ldots, \alpha_k)' \in R^k$ write $\bar{\alpha}$ for the $(k-1)$ vector $(\alpha_1, \ldots, \alpha_{k-1})' \in R^{k-1}$.

REMARK 2.3. Note in the sequel that the particular role played by the kth component could be taken by any other component.

From (2.1) we see that

$$\mathcal{Y}_n = \sigma\{y_\ell : \ell \leq n\} = \sigma\{\bar{y}_\ell : \ell \leq n\}.$$

Write

$$\hat{a}_\ell^i = E[y_\ell^i \mid \mathcal{Y}_{\ell-1}], \qquad \ell \geq 1,$$

$$\hat{a}_\ell = (\hat{a}_\ell^1, \ldots, \hat{a}_\ell^k)'$$

Also, denote

$$\nu_\ell^i = y_\ell^i - \hat{a}_\ell^i, \qquad 1 \leq i \leq k,$$

$$\text{and} \quad \nu_\ell = (\nu_\ell^1, \ldots, \nu_\ell^k)'.$$

The following martingale representation results are similar to those of Boel [3].

THEOREM 2.4. *Suppose M_n, $n \in Z^+$, is a d-dimensional vector $(P, \mathcal{Y}_n)$ martingale with $M_0 = 0$. Then there is a $\mathcal{Y}$-predictable $d \times k$ matrix process H_ℓ such that*

$$M_n = \sum_{\ell=1}^{n} H_\ell (y_\ell - \hat{a}_\ell) = \sum_{\ell=1}^{n} H_\ell \cdot \nu_\ell.$$

($\mathcal{Y}$-predictable means that for each $\ell \in Z^+$ H_ℓ is $\mathcal{Y}_{\ell-1}$ measurable.)

Proof. Because M_n is $\mathcal{Y}_n$ measurable there is a d-vector function f such that

$$M_n = f(y_1, y_2, \ldots, y_{n-1}, y_n).$$

Write H_n for the $d \times k$, $\mathcal{Y}_{n-1}$ measurable matrix with columns $f(y_1, \ldots, y_{n-1}, f_i)$, $1 \leq i \leq k$.

REMARK 2.5. From (2.1) and (2.2) we see

$$\sum_{i=1}^{k} \nu_\ell^i = \sum_{i=1}^{k} (y_\ell^i - \hat{a}_\ell^i) = 0$$

so there is redundancy in the representation (2.4). For any component, say component k,

$$\nu_\ell^k = -\sum_{i=1}^{k-1} \nu_\ell^i,$$

and if we define h_ℓ to be the $d \times (k-1)$ $\mathcal{Y}$-predictable matrix process with columns $f(f_i) - f(f_k)$, $1 \leq i \leq k-1$, for $\ell = 1$, and $f(y_1, \ldots, y_{\ell-1}, f_i) - f(y_1, \ldots, y_{\ell-1}, f_k)$, $1 \leq i \leq k-1$, for $\ell > 1$, we have the following result:

COROLLARY 2.6.

$$M_n = \sum_{\ell=1}^{n} h_\ell \cdot \begin{pmatrix} y_\ell^1 - \hat{a}_\ell^1 \\ \vdots \\ y_\ell^{k-1} - \hat{a}_\ell^{k-1} \end{pmatrix} = \sum_{\ell=1}^{n} h_\ell \cdot \bar{\nu}_\ell.$$

3. A Girsanov Theorem.

Suppose y_ℓ^i, $1 \leq i \leq k$, $\ell \in Z^+$, and a_ℓ^i are as defined in Section 2. In this section we further require that

$$a_\ell^i > 0, \quad 1 \leq i \leq k, \quad \ell \in Z^+.$$

Suppose $b_\ell = (b_\ell^1, \ldots, b_\ell^k)'$ is a second sequence of random vectors which are F-predictable and satisfy $b_\ell^i > 0$, $\sum_{i=1}^k b_\ell^i = 1$. In the sequel $\langle\ ,\ \rangle$ will denote the inner product of vectors in Euclidean space. Write

$$\gamma_\ell = \prod_{i=1}^{k} \left(\frac{b_\ell^i}{a_\ell^i}\right)^{y_\ell^i}.$$

Then

$$E[\gamma_\ell \mid F_{\ell-1}] = \sum_{i=1}^{k} \left(\frac{b_\ell^i}{a_\ell^i}\right) P(y_\ell^i = 1 \mid F_{\ell-1}) = 1.$$

Consequently, if $\Lambda_n = \prod_{\ell=1}^{n} \gamma_\ell$ then $\{\Lambda_n\}$ is a positive $\{F_n\}$ martingale. Define a new probability measure $\overline{P}_n$ on F_n by putting

$$\frac{d\overline{P}_n}{dP} = \Lambda_n.$$

For $m > n$ we have $F_n \subset F_m$ and for $C \in F_n$

$$\begin{aligned}\overline{P}_m(C) &= \int_C \Lambda_m dP = \int_C E[\Lambda_m \mid F_n] dP \\ &= \int_C \Lambda_n dP = \overline{P}_n(C),\end{aligned}$$

so the restriction of $\overline{P}_m$ to F_n coincides with $\overline{P}_n$. Consequently, by Kolmogorov's existence theorem there is a probability measure $\overline{P}$ on $\left(\Omega, \bigcup_\ell F_\ell\right)$ which restricted to F_n coincides with $\overline{P}_n$. Write $\overline{E}$ for expectation under $\overline{P}$.

LEMMA 3.1. $\overline{E}[y_\ell \mid F_{\ell-1}] = b_\ell$.

Proof.

$$\overline{E}[y_\ell \mid F_{\ell-1}] = E[\Lambda_\ell y_\ell \mid F_{\ell-1}] / E[\Lambda_\ell \mid F_{\ell-1}]$$

$$= E[\gamma_\ell y_\ell \mid F_{\ell-1}] = b_\ell.$$

4. Hidden Markov Models

Consider a system whose state is described by a discrete time, homogeneous, finite state space Markov chain X_ℓ, $\ell \in Z^+$, defined on (Ω, F, P). We suppose that X_0 is given, or its distribution known. Without loss of generality the state space of X_ℓ can be taken to be the set $S = \{e_1, \ldots, e_N\}$ for a suitable $N \in Z^+$, where the $e_i = (0, \ldots, 1, \ldots, 0)'$, $1 \le i \le N$, are the standard unit vectors of R^N.

Suppose $\{F_n\}$, $n \in Z^+$, is the complete filtration generated by X. That is, for each $n \in Z^+$, F_n is the completion of the σ-field generated by X_ℓ, $\ell \le n$. From the Markov property $P(X_n = e_j \mid F_{n-1}) = P(X_n = e_j \mid X_{n-1})$. Write $p_{ji} = P(X_n = e_j \mid X_{n-1} = e_i)$. P will denote the $N \times N$ matrix (p_{ji}). Note $\sum_{j=1}^N p_{ji} = 1$.

LEMMA 4.1. *For $n \in Z^+$,*

$$X_n = PX_{n-1} + m_n \tag{4.1}$$

where m_n is a martingale increment.

Proof.

$$E[m_n \mid F_{n-1}] = E[X_n - PX_{n-1} \mid X_{n-1}]$$

$$= E[X_n \mid X_{n-1}] - PX_{n-1} = 0.$$

COROLLARY 4.2. Write $Q = P - I$. The semimartingale representation of X is:

$$X_n = X_0 + \sum_{\ell=1}^{n} QX_{\ell-1} + M_n. \tag{4.2}$$

Here $M_n = \sum_{\ell=1}^n m_\ell$ is a (P, F_n) martingale.

We now wish to calculate the predictable quadratic variation $\langle m_\ell \rangle = E[m_n \otimes m_n \mid F_{n-1}]$ of m_ℓ. Recall X_ℓ is one of the vectors $e_i \in S$. $\otimes$ will denote the tensor product.

LEMMA 4.3.

$$m_n \otimes m_n = \operatorname{diag}(PX_{n-1}) + \operatorname{diag} m_n - PX_{n-1} \otimes PX_{n-1} - PX_{n-1} \otimes m_n - m_n \otimes PX_{n-1} \tag{4.3}$$

and

$$\begin{aligned}\langle m_n \rangle := E[m_n \otimes m_n \mid F_{n-1}] &= E[m_n \otimes m_n \mid X_{n-1}] \\ &= \operatorname{diag}(PX_{n-1}) - PX_{n-1} \otimes PX_{n-1}.\end{aligned} \tag{4.4}$$

Proof. From (4.1) $X_n \otimes X_n = PX_{n-1} \otimes PX_{n-1} + PX_{n-1} \otimes m_n + m_n \otimes PX_{n-1} + m_n \otimes m_n$. However, $X_n \otimes X_n = \operatorname{diag} X_n = \operatorname{diag} PX_{n-1} + \operatorname{diag} m_n$. Equation (4.3) follows. The terms on the right side of (4.3) involving m_n are martingale increments; conditioning on X_{n-1} we see $\langle m_n \rangle = E[m_n \otimes m_n \mid X_{n-1}] = \operatorname{diag}(PX_{n-1}) - PX_{n-1} \otimes PX_{n-1}$.

NOTATION 4.4. Recall $\{F_n\}$, $n \in Z^+$, is the filtration generated by X, and, as in Section 2, $\{\mathcal{Y}_n\}$ is the filtration generated by y. We shall take $y_0 = 0 \in R^k$. Write $\{G_n\}$ for the filtration generated by X and y. A process $\{Z_\ell\}$, $\ell \in Z^+$, is adapted to the filtration $\{G_\ell\}$ if Z_ℓ is G_ℓ measurable for each ℓ. A process $\{Z_\ell\}$ is a (P, G_ℓ) martingale if Z_ℓ is integrable and for $m \geq \ell$, $E[Z_m \mid G_\ell] = Z_\ell$.

Following Rabiner [5] we shall make the following definition:

DEFINITION 4.5. A Hidden Markov Model, HMM, consists of a Markov chain $\{X_n\}$, $n \in Z^+$, and a random, finite state space function f of X_n whose values are observed.

We shall suppose our X process is not observed directly. Rather, as in Åström [1] or Rabiner [5], there is a function f with a finite range and we observe the values

$y_\ell = f(X_{\ell-1}, \kappa_\ell)$, $\ell \in Z^+$, where κ_ℓ is a sequence of independent random variables. We have assumed f is independent of the time parameter ℓ, that is, the y observation process is time homogeneous. If the range of f consists of k points we can, without loss of generality, identify the range of f with the unit vectors $f_1, \dots, f_k$ of R^k, where $f_i = (0, \dots, 1, \dots, 0)'$. Write $d_{ji} = P(y_\ell = f_j \mid X_{\ell-1} = e_i)$. Note again

$$\sum_{j=1}^{k} d_{ji} = 1 \tag{4.5}$$

and $d_{ji} \geq 0$, $1 \leq j \leq k$, $1 \leq i \leq N$. D will denote the $k \times N$ matrix (d_{ji}).

Similarly to Lemma 4.1 we have the following representation.

LEMMA 4.6. *For $n \in Z^+$*

$$y_n = DX_{n-1} + K_n \tag{4.6}$$

where K_n is a (P, G_n) martingale increment.

Proof.

$$E[K_n \mid G_{n-1}] = E[y_n - DX_{n-1} \mid X_{n-1}]$$

$$= DX_{n-1} - DX_{n-1} = 0.$$

NOTATION 4.7. For each $\ell \in Z^+$, y_ℓ is one of the unit vectors $f_1, \dots, f_k$. Writing $y_\ell^i = \langle y_\ell, f_i \rangle$, $1 \leq i \leq k$, we see $y_\ell = (y_\ell^1, \dots, y_\ell^k)'$, $\ell \in Z^+$. Therefore, exactly one component of y_ℓ is equal to 1 for each $\ell \in Z^+$; the remainder are 0. Consequently, the process y_ℓ is of the form discussed in Section 2. As in (2.1) $\sum_{j=1}^{k} y_\ell^j = 1$ and

$$\mathcal{Y}_n = \sigma\{y_\ell : \ell \leq n\} = \sigma\{\bar{y}_\ell : \ell \leq n\}.$$

NOTATION 4.8. For any process ϕ_ℓ, $\ell \in Z^+$, write $\check{\phi}_\ell = E[\phi_\ell \mid \mathcal{Y}_\ell]$ for its $\mathcal{Y}$-optional projection, and, consistently with Section 2, write $\dot{\phi}_\ell = E[\phi_\ell \mid \mathcal{Y}_{\ell-1}]$ for its $\mathcal{Y}$-predictable projection.

LEMMA 4.9.

$$y_n = D\check{X}_{n-1} + \nu_n \tag{4.7}$$

where the innovations $\nu_n = D(X_{n-1} - \check{X}_{n-1}) + K_n$ is a $(P, \mathcal{Y})$-martingale increment.

Proof. From Lemma 4.6 $y_n = DX_{n-1} + K_n = D\check{X}_{n-1} + \nu_n$. Now

$$E[\nu_n \mid \mathcal{Y}_{n-1}] = E[D(X_{n-1} - \check{X}_{n-1}) \mid \mathcal{Y}_{n-1}] + E[K_n \mid \mathcal{Y}_{n-1}]$$

$$= 0 + E[E[K_n \mid G_{n-1}]\mathcal{Y}_{n-1}] = 0.$$

NOTATION 4.10. In the sequel we shall write:

$$a_\ell^i = E[y_\ell^i \mid G_{\ell-1}] = \langle DX_{\ell-1}, f_i \rangle, \qquad 1 \leq i \leq k$$

$$\hat{a}_\ell^i = E[y_\ell^i \mid \mathcal{Y}_{\ell-1}] = \langle D\check{X}_{\ell-1}, f_i \rangle, \qquad 1 \leq i \leq k.$$

Note

$$\sum_{i=1}^{k} a_\ell^i = 1 \tag{4.8}$$

and

$$\sum_{i=1}^{k} \hat{a}_\ell^i = 1. \tag{4.9}$$

Because of the redundancy noted in equation (2.3) we shall use only $(k-1)$ components of the observation vector y_ℓ, $\ell \in Z^+$; without loss of generality we shall use the first $(k-1)$.

Note if $a_\ell^i = 0$ for some i and ℓ then $y_\ell^i = 1$ with zero probability. The range space of the observation process at that time ℓ can, therefore, be reduced by one dimension for each a_ℓ^i which is 0. We shall suppose $a_\ell^i > 0$, $1 \leq i \leq k$, $\ell \in Z^+$.

Write $\overline{D}$ for the $(k-1) \times N$ matrix (d_{ji}), $1 \le j \le (k-1)$, $1 \le i \le N$. The analogs of (4.6) and (4.7) are

$$\bar{y}_n = \overline{D} X_{n-1} + \overline{K}_n \tag{4.10}$$

$$\bar{y}_n = \overline{D} \check{X}_{n-1} + \bar{\nu}_n. \tag{4.11}$$

Consequently $\bar{a}_n = \overline{D} X_{n-1}$ and $\hat{\bar{a}}_n = \overline{D} \check{X}_{n-1}$. At most one component of $\bar{y}_n$ is 1; otherwise they are 0.

LEMMA 4.11.

$$E[\bar{\nu}_n \otimes \bar{\nu}_n \mid \mathcal{Y}_{n-1}] = \widehat{\overline{\Phi}}_n$$

$$= \operatorname{diag}(\overline{D}\check{X}_{n-1}) - (\overline{D}\check{X}_{n-1}) \otimes (\overline{D}\check{X}_{n-1})$$

$$= \operatorname{diag} \hat{\bar{a}}_n - \hat{\bar{a}}_n \otimes \hat{\bar{a}}_n$$

$$= \begin{pmatrix} \hat{a}_n^1(1-\hat{a}_n^1) & -\hat{a}_n^1\hat{a}_n^2 & \dots & -\hat{a}_n^1\hat{a}_n^{k-1} \\ -\hat{a}_n^2\hat{a}_n^1 & \hat{a}_n^2(1-\hat{a}_n^2) & \dots & -\hat{a}_n^2\hat{a}_n^{k-1} \\ \vdots & \vdots & & \vdots \\ -\hat{a}_n^{k-1}\hat{a}_n^1 & -\hat{a}_n^{k-1}\hat{a}_n^2 & \dots & \hat{a}_n^{k-1}(1-\hat{a}_n^{k-1}) \end{pmatrix} \tag{4.12}$$

Proof.

$$\bar{\nu}_n \otimes \bar{\nu}_n = (\bar{y}_n - \overline{D}\check{X}_{n-1}) \otimes (\bar{y}_n - \overline{D}\check{X}_{n-1})$$

$$= \operatorname{diag} \bar{y}_n - \bar{y}_n \otimes \overline{D}\check{X}_{n-1} - \overline{D}\check{X}_{n-1} \otimes \bar{y}_n + (\overline{D}\check{X}_{n-1}) \otimes (\overline{D}\check{X}_{n-1})$$

$$= \operatorname{diag} \bar{y}_n - \bar{y}_n \otimes \hat{\bar{a}}_n - \hat{\bar{a}}_n \otimes \bar{y}_n + \hat{\bar{a}}_n \otimes \hat{\bar{a}}_n.$$

Conditioning on $\mathcal{Y}_{n-1}$ we see

$$E[\bar{\nu}_n \otimes \bar{\nu}_n \mid \mathcal{Y}_{n-1}] = \operatorname{diag} \hat{\bar{a}}_n - \hat{\bar{a}}_n \otimes \hat{\bar{a}}_n.$$

REMARKS 4.12. A simple calculation gives

$$\det \widehat{\Phi}_n = \prod_{i=1}^{k} \hat{a}_n^i$$

and

$$\widehat{\Phi}_n^{-1} = \frac{1}{\hat{a}_n^k} \begin{pmatrix} \frac{\hat{a}_n^k}{\hat{a}_n^1} + 1 & 1 & \dots & 1 \\ 1 & \frac{\hat{a}_n^k}{\hat{a}_n^2} + 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & \frac{\hat{a}_n^k}{\hat{a}_n^{k-1}} + 1 \end{pmatrix}. \tag{4.13}$$

It is clear from (4.13) why we require $\hat{a}_n^i > 0$, $1 \le i \le k$. Note a similar calculation shows the $k \times k$ matrix $E[\nu_n \otimes \nu_n \mid \mathcal{Y}_{n-1}]$ is singular; this is why we use only $(k-1)$ components of the observation process.

Finally, we note the following product formula.

LEMMA 4.13. *Suppose $W_n = W_0 + \sum_{\ell=1}^{n} \alpha_\ell$ and $V_n = V_0 + \sum_{\ell=1}^{n} \beta_\ell$ are N and k dimensional processes, respectively. Then*

$$W_n \otimes V_n = W_0 \otimes V_0 + \sum_{\ell=1}^{n} W_{\ell-1} \otimes \beta_\ell + \sum_{\ell=1}^{n} \alpha_\ell \otimes V_{\ell-1} + \sum_{\ell=1}^{n} \alpha_\ell \otimes \beta_\ell.$$

Proof. By induction.

5. A General Filter

Suppose H_n, $n \in Z^+$, is a d-dimensional (P, G_n) process of the form

$$H_n = H_0 + \sum_{\ell=1}^{n} \alpha_\ell + \sum_{\ell=1}^{n} \beta_\ell m_\ell + \sum_{\ell=1}^{n} \delta_\ell \bar{y}_\ell + \sum_{\ell=1}^{n} \lambda_\ell m_\ell \otimes m_\ell + \sum_{\ell=1}^{n} \sigma_\ell \bar{y}_\ell \otimes m_\ell. \tag{5.1}$$

Here α, β, δ, λ, σ are (P, G_n) predictable processes of appropriate dimensions, that is, α is d-dimensional, β is $d \times N$ matrix valued, and so on.

THEOREM 5.1.

$$\check{H}_n = E[H_n \mid \mathcal{Y}_n] = \check{H}_0 + \sum_{\ell=1}^{n} (\hat{\alpha}_\ell + \widehat{\delta_\ell \bar{a}_\ell} + E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}]) + \sum_{\ell=1}^{n} h_\ell \bar{\nu}_\ell. \quad (5.2)$$

where

$$h_\ell = \Big\{ E[H_{\ell-1} \otimes \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - \check{H}_{\ell-1} \otimes \hat{\bar{a}}_\ell + E[\alpha_\ell \otimes \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - \hat{\alpha}_\ell \otimes \hat{\bar{a}}_\ell + E[\delta_\ell \operatorname{diag} \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}]$$

$$- \widehat{\delta_\ell \bar{a}_\ell} \otimes \hat{\bar{a}}_\ell + E[\lambda_\ell \langle m_\ell \rangle \otimes \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \otimes \hat{\bar{a}}_\ell \Big\} \widehat{\Phi}_\ell^{-1}.$$

Proof. First note that because the random variables κ_ℓ in the observations $y_\ell = f(X_{\ell-1}, \kappa_\ell)$ are independent the martingale increments in the signal X and the observation y are independent, so the terms $\sigma_\ell \bar{y}_\ell \otimes m_\ell$ are martingale increments, as is the difference $m_\ell \otimes m_\ell - \langle m_\ell \rangle$. Consequently, from (5.1)

$$\begin{aligned}
\check{H}_n = E[H_n \mid \mathcal{Y}_n] &= \check{H}_0 + (E[H_0 \mid \mathcal{Y}_n] - \check{H}_0) + \sum_{\ell=1}^{n} \hat{\alpha}_\ell \\
&+ \sum_{\ell=1}^{n} (E[\alpha_\ell \mid \mathcal{Y}_n] - \hat{\alpha}_\ell) + E\Big[\sum_{\ell=1}^{n} \beta_\ell m_\ell \mid \mathcal{Y}_n\Big] \\
&+ E\Big[\sum_{\ell=1}^{n} \delta_\ell (\bar{y}_\ell - \bar{a}_\ell) \mid \mathcal{Y}_n\Big] + \sum_{\ell=1}^{n} (E[\delta_\ell \bar{a}_\ell \mid \mathcal{Y}_n] - \widehat{\delta_\ell \bar{a}_\ell}) + \sum_{\ell=1}^{n} \widehat{\delta_\ell \bar{a}_\ell} \\
&+ \Big(E\Big[\sum_{\ell=1}^{n} \lambda_\ell m_\ell \otimes m_\ell \mid \mathcal{Y}_n\Big] - \sum_{\ell=1}^{n} E[\lambda_\ell m_\ell \otimes m_\ell \mid \mathcal{Y}_{\ell-1}] \Big) \\
&+ \Big(\sum_{\ell=1}^{n} E[\lambda_\ell m_\ell \otimes m_\ell \mid \mathcal{Y}_{\ell-1}] - \sum_{\ell=1}^{n} E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \Big] \\
&+ \sum_{\ell=1}^{n} E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] + E\Big[\sum_{\ell=1}^{n} \sigma_\ell \bar{y}_\ell \otimes m_\ell \mid \mathcal{Y}_n\Big].
\end{aligned}$$

That is,

$$\check{H}_n = \check{H}_0 + \sum_{\ell=1}^{n} (\hat{\alpha}_\ell + \widehat{\delta_\ell \bar{a}_\ell} + E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}])$$
$$+ \text{ a term which is a } (P, \mathcal{Y}_n) \text{ martingale}.$$

From Corollary 2.6 we can, therefore, write

$$\check{H}_n = \check{H}_0 + \sum_{\ell=1}^{n} \big(\hat{\alpha}_\ell + \widehat{\delta_\ell \bar{a}_\ell} + E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}]\big) + \sum_{\ell=1}^{n} h_\ell \cdot \bar{\nu}_\ell. \tag{5.3}$$

Here h_ℓ is a $d \times (k-1)$ dimensional $\mathcal{Y}$-predictable process which we now proceed to find by calculating $\overline{Y}_n \otimes \check{H}_n$ in two ways.

Recall

$$\overline{Y}_n = \sum_{\ell=1}^{n} \bar{y}_\ell \tag{5.4}$$

and $\overline{Y}_0 = 0$. From (5.1) and (5.4), using Lemma 4.13,

$$\begin{aligned}
\overline{Y}_n \otimes H_n = \sum_{\ell=1}^{n} \Big\{ & \overline{Y}_{\ell-1} \otimes \alpha_\ell + \overline{Y}_{\ell-1} \otimes \beta_\ell m_\ell + \overline{Y}_{\ell-1} \otimes \delta_\ell \bar{y}_\ell \\
& + \overline{Y}_{\ell-1} \otimes (\lambda_\ell m_\ell \otimes m_\ell) + \overline{Y}_{\ell-1} \otimes (\sigma_\ell \bar{y}_\ell \otimes m_\ell) \\
& + (\bar{y}_\ell - \bar{a}_\ell) \otimes H_{\ell-1} + \bar{a}_\ell \otimes H_{\ell-1} + \bar{y}_\ell \otimes \alpha_\ell \\
& + \bar{y}_\ell \otimes \beta_\ell m_\ell + \bar{y}_\ell \otimes \delta_\ell \bar{y}_\ell + \bar{y}_\ell \otimes (\lambda_\ell m_\ell \otimes m_\ell) \\
& + \bar{y}_\ell \otimes (\sigma_\ell \bar{y}_\ell \otimes m_\ell) \Big\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E[\overline{Y}_n \otimes H_n \mid \mathcal{Y}_n] &= \overline{Y}_n \otimes \check{H}_n \\
&= \sum_{\ell=1}^{n} \Big\{ \overline{Y}_{\ell-1} \otimes \hat{\alpha}_\ell + \overline{Y}_{\ell-1} \otimes \widehat{\delta_\ell \bar{a}_\ell} + \overline{Y}_{\ell-1} \otimes E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \\
&\quad + E[\bar{a}_\ell \otimes H_{\ell-1} \mid \mathcal{Y}_{\ell-1}] + E[\bar{a}_\ell \otimes \alpha_\ell \mid \mathcal{Y}_{\ell-1}] \\
&\quad + E[\operatorname{diag} \bar{a}_\ell \cdot \delta'_\ell \mid \mathcal{Y}_{\ell-1}] + E[\bar{a}_\ell \otimes \lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \Big\} \\
&\quad + \text{a term which is a } (P, \mathcal{Y}_n) \text{ martingale.}
\end{aligned} \tag{5.5}$$

However, from (5.3) and (5.4), using Lemma 4.13,

$$\begin{aligned}
\overline{Y}_n \otimes \check{H}_n = \sum_{\ell=1}^{n} \Big\{ & \overline{Y}_{\ell-1} \otimes \hat{\alpha}_\ell + \overline{Y}_{\ell-1} \otimes \widehat{\delta_\ell \bar{a}_\ell} + \overline{Y}_{\ell-1} \otimes E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \\
& + \bar{y}_\ell \otimes \check{H}_{\ell-1} + \bar{y}_\ell \otimes \hat{\alpha}_\ell + \bar{y}_\ell \otimes \widehat{\delta_\ell \bar{a}_\ell} + \bar{y}_\ell \otimes E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \\
& + \bar{y}_\ell \otimes h_\ell \bar{\nu}_\ell + \overline{Y}_{\ell-1} \otimes h_\ell \bar{\nu}_\ell \Big\} \\
= \sum_{\ell=1}^{n} \Big\{ & \overline{Y}_{\ell-1} \otimes \hat{\alpha}_\ell + \overline{Y}_{\ell-1} \otimes \widehat{\delta_\ell \bar{a}_\ell} + \overline{Y}_{\ell-1} \otimes E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \\
& + \hat{\bar{a}}_\ell \otimes \check{H}_{\ell-1} + \hat{\bar{a}}_\ell \otimes \hat{\alpha}_\ell + \hat{\bar{a}}_\ell \otimes \widehat{\delta_\ell \bar{a}_\ell} + \hat{\bar{a}}_\ell \otimes E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \\
& + E[\bar{\nu}_\ell \otimes \bar{\nu}_\ell \mid \mathcal{Y}_{\ell-1}] h_\ell' \Big\} \\
& + \text{a term which is a } (P, \mathcal{Y}_n) \text{ martingale.}
\end{aligned} \tag{5.6}$$

Equating the increments in (5.5) and (5.6) and conditioning on $\mathcal{Y}_{\ell-1}$ we have

$$\begin{aligned}
\widehat{\Phi}_\ell \cdot h_\ell' = {} & E[\bar{a}_\ell \otimes H_{\ell-1} \mid \mathcal{Y}_{\ell-1}] - \hat{\bar{a}}_\ell \otimes \check{H}_{\ell-1} + E[\bar{a}_\ell \otimes \alpha_\ell \mid \mathcal{Y}_{\ell-1}] - \hat{\bar{a}}_\ell \otimes \hat{\alpha}_\ell \\
& + \operatorname{diag} E[\bar{a}_\ell \cdot \delta_\ell' \mid \mathcal{Y}_{\ell-1}] - \hat{\bar{a}}_\ell \otimes \widehat{\delta_\ell \bar{a}_\ell} \\
& + E[\bar{a}_\ell \otimes \lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] - \hat{\bar{a}}_\ell \otimes E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}].
\end{aligned}$$

Therefore,

$$\begin{aligned}
h_\ell = \Big\{ & E[H_{\ell-1} \otimes \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - \check{H}_{\ell-1} \otimes \hat{\bar{a}}_\ell \\
& + E[\alpha_\ell \otimes \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - \hat{\alpha}_\ell \otimes \hat{\bar{a}}_\ell + E[\delta_\ell \cdot \operatorname{diag} \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - \widehat{\delta_\ell \bar{a}_\ell} \otimes \hat{\bar{a}}_\ell \\
& + E[\lambda_\ell \langle m_\ell \rangle \otimes \bar{a}_\ell \mid \mathcal{Y}_{\ell-1}] - E[\lambda_\ell \langle m_\ell \rangle \mid \mathcal{Y}_{\ell-1}] \otimes \hat{\bar{a}}_\ell \Big\} \widehat{\Phi}_\ell^{-1}.
\end{aligned}$$

REMARKS 5.2. Suppose H_n is a scalar process of the form

$$H_n = H_0 + \sum_{\ell=1}^{n} (\alpha_\ell + \beta_\ell m_\ell + \delta_\ell \bar{y}_\ell).$$

Then

$$\begin{aligned}
H_n X_n &= H_0 X_0 + \sum_{\ell=1}^{n} (\alpha_\ell X_{\ell-1} + \beta_\ell m_\ell X_{\ell-1} + \delta_\ell \bar{y}_\ell X_{\ell-1}) \\
&\quad + \sum_{\ell=1}^{n} (H_{\ell-1} Q X_{\ell-1} + H_{\ell-1} m_\ell) + \sum_{\ell=1}^{n} (\alpha_\ell + \beta_\ell m_\ell + \delta_\ell \bar{y}_\ell)(Q X_{\ell-1} + m_\ell) \\
&= H_0 X_0 + \sum_{\ell=1}^{n} (\alpha_\ell X_{\ell-1} + H_{\ell-1} Q X_{\ell-1} + \alpha_\ell Q X_{\ell-1}) \\
&\quad + \sum_{\ell=1}^{n} (\beta_\ell m_\ell X_{\ell-1} + H_{\ell-1} m_\ell + \alpha_\ell m_\ell + \beta_\ell m_\ell Q X_{\ell-1}) \\
&\quad + \sum_{\ell=1}^{n} (\delta_\ell \bar{y}_\ell X_{\ell-1} + \delta_\ell \bar{y}_\ell Q X_{\ell-1}) + \sum_{\ell=1}^{n} m_\ell \otimes m_\ell \cdot \beta'_\ell + \sum_{\ell=1}^{n} \delta_\ell y_\ell \cdot m_\ell.
\end{aligned}$$

Recalling $I + Q = P$ we can, therefore, apply Theorem 5.1 with α_ℓ replaced by $\alpha_\ell P X_{\ell-1} + H_{\ell-1} Q X_{\ell-1}$, δ_ℓ replaced by $(PX)_{\ell-1} \otimes \delta_\ell$ and λ_ℓ replaced by β'_ℓ to obtain:

THEOREM 5.3.

$$\begin{aligned}
E[H_n X_n \mid \mathcal{Y}_n] = E[H_0 X_0] &+ \sum_{\ell=1}^{n} E\Big[(\alpha_\ell P X_{\ell-1} + H_{\ell-1} Q X_{\ell-1} \\
&+ P X_{\ell-1} \langle \delta_\ell, \overline{D} X_{\ell-1} \rangle + \langle m_\ell \rangle \beta'_\ell) \mid \mathcal{Y}_{\ell-1} \Big] + \sum_{\ell=1}^{n} h^*_\ell \bar{\nu}_\ell, \quad (5.7)
\end{aligned}$$

where

$$h_\ell^* = \Big\{E\big[H_{\ell-1}X_{\ell-1}\otimes\overline{D}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}\big] - E\big[H_{\ell-1}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}\big]\otimes\overline{D}\check{X}_{\ell-1}$$

$$+ E[(\alpha_\ell PX_{\ell-1}+H_{\ell-1}QX_{\ell-1})\otimes\overline{D}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}]$$

$$- E[\alpha_\ell PX_{\ell-1}\mid\mathcal{Y}_{\ell-1}]\otimes\overline{D}\check{X}_{\ell-1} - E[H_{\ell-1}QX_{\ell-1}\mid\mathcal{Y}_{\ell-1}]\otimes\overline{D}\check{X}_{\ell-1}$$

$$+ E[(PX_{\ell-1})\otimes\delta_\ell\,\mathrm{diag}\,\overline{D}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}] - E[(PX_{\ell-1})\otimes\delta_\ell\mid\mathcal{Y}_{\ell-1}]\otimes\overline{D}\check{X}_{\ell-1}$$

$$+ E[\langle m_\ell\rangle\beta'_\ell\otimes\overline{D}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}] - E[\langle m_\ell\rangle\beta'_\ell\mid\mathcal{Y}_{\ell-1}]\otimes\overline{D}\widehat{X}_{\ell-1}\Big\}\widehat{\Phi}_\ell^{-1}. \tag{5.8}$$

Note h_ℓ^* is $\mathcal{Y}_{\ell-1}$ measurable.

REMARK 5.4. Taking $H_n = H_0 = 1$, $\alpha_\ell = 0$, $\beta_\ell = 0$, $\delta_\ell = 0$, we obtain the following filter for X_n:

$$\check{X}_n = \check{X}_0 + \sum_{\ell=1}^{n} Q\check{X}_{\ell-1} + \sum_{\ell=1}^{n}\{E[PX_{\ell-1}\otimes\overline{D}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}] - P\check{X}_{\ell-1}\otimes\overline{D}\check{X}_{\ell-1}\}\widehat{\Phi}_\ell^{-1}\bar{\nu}_\ell. \tag{5.9}$$

The product term $PX_{\ell-1}\otimes\overline{D}X_{\ell-1}$ can be written $\sum_{i=1}^N \langle X_{\ell-1}, e_i\rangle Pe_i\otimes\overline{D}e_i$. Therefore, $E[PX_{\ell-1}\otimes\overline{D}X_{\ell-1}\mid\mathcal{Y}_{\ell-1}] = \sum_{i=1}^N \langle\check{X}_{\ell-1}, e_i\rangle Pe_i\otimes\overline{D}e_i$ and so, recalling the forms of $\bar{\nu}_\ell = (\bar{y}_\ell - \overline{D}\check{X}_{\ell-1})$ and $\widehat{\Phi}_\ell^{-1}$ from (4.13), equation (5.9) gives a closed, recursive, finite dimensional filter for $\check{X}_n$. However, we shall see in the next section that the unnormalized Zakai filters have much simpler forms, so we shall not pursue normalized filters any further.

6. The Zakai Equation

As in Section 4 $a_n = DX_{n-1}$ and $\hat{a}_n = D\check{X}_{n-1}$, so that $\bar{a}_n = \overline{D}X_{n-1}$ and $\hat{\bar{a}}_n = \overline{D}\check{X}_{n-1}$. $\sum_{i=1}^k a_n^i = 1$ and $\sum_{i=1}^k \hat{a}_n^i = 1$ for $n \in Z^+$. We shall suppose $a_n^i > 0$, $1 \le i \le k$, $n \in Z^+$. Again $b_n = (b_n^1, \dots, b_n^k)'$ will be a second sequence of F-predictable random vectors such that $b_n^i > 0$ and $\sum_{i=1}^k b_n^i = 1$. Following Section 3 take

$$\gamma_\ell = \prod_{i=1}^k \left(\frac{b_\ell^i}{a_\ell^i}\right)^{y_\ell^i}, \quad \text{and}$$
$$\Lambda_n = \prod_{\ell=1}^n \gamma_\ell.$$

Define $\overline{P}_n$ by putting the Radon-Nikodym derivative $\left.\frac{d\overline{P}_n}{dP}\right|_{G_n} = \Lambda_n$; the existence of a measure P such that $\overline{P}|_{G_n} = \overline{P}_n$ follows from Kolmogorov's theorem as in Section 3.

NOTATION 6.1. Suppose from now on that

$$b_\ell = (\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k})' = \frac{1}{k} \cdot \underline{1}, \quad \ell \in Z^+,$$

where $\underline{1} = (1, 1, \dots, 1)' \in R^k$.

If $\overline{P}$ is constructed for this b_ℓ sequence then, with $\overline{E}$ denoting expectation under $\overline{P}$,

$$\overline{E}[y_\ell^i \mid G_{\ell-1}] = \overline{E}[y_\ell^i \mid \mathcal{Y}_{\ell-1}] = \frac{1}{k}, \quad 1 \le i \le k, \quad \ell \in Z^+.$$

We now wish to take $\overline{P}$ as the "reference probability" and, starting with $\overline{P}$, construct the measure P on $(\Omega, \bigcup_{n=1}^\infty G_n)$ such that under P: $E[y_\ell^i \mid G_{\ell-1}] = a_\ell^i$.

The roles of a_ℓ and b_ℓ are, therefore, interchanged, so we define

$$\bar{\gamma}_\ell = \prod_{i=1}^{k} \left(\frac{a_\ell^i}{b_\ell^i}\right)^{y_\ell^i} \quad \text{and}$$

$$\Lambda_n^* = \prod_{\ell=1}^{n} \bar{\gamma}_\ell.$$

P_n is then defined on G_n by putting its Radon-Nikodym derivative with respect to P equal to Λ_n^*; the existence of P again follows from Kolmogorov's theorem. For this b_ℓ the analog of the $(k-1) \times (k-1)$ matrix $\widehat{\Phi}_n$, (equation (4.12)) is

$$\Psi_\ell = \Psi = \frac{1}{k^2} \begin{pmatrix} (k-1) & -1 & \dots & -1 \\ -1 & (k-1) & \dots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & & (k-1) \end{pmatrix} \tag{6.1}$$

so that

$$\Psi^{-1} = k \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & & 2 \end{pmatrix}. \tag{6.2}$$

NOTATION 6.2. If ϕ_ℓ, $\ell \in Z^+$, is a G-adapted integrable process a version of Bayes' theorem states that

$$\begin{aligned} \check{\phi}_\ell = E[\phi_\ell \mid \mathcal{Y}_\ell] &= \frac{\overline{E}[\Lambda_\ell^* \phi_\ell \mid \mathcal{Y}_\ell]}{\overline{E}[\Lambda_\ell^* \mid \mathcal{Y}_\ell]} \\ &= \sigma_\ell(\phi_\ell)/\sigma_\ell(1), \quad \text{say}, \end{aligned} \tag{6.3}$$

where $\sigma_\ell(\phi_\ell) = \overline{E}[\Lambda_\ell^* \phi_\ell \mid \mathcal{Y}_\ell]$ is an unnormalized conditional expectation of ϕ_ℓ under $\overline{P}$.

For $m \leq n$ we shall [illegible]

$$\sigma_n(\phi_m) = \overline{E}[\Lambda_n^* \phi_m \mid \mathcal{Y}_n].$$

LEMMA 6.3.

$$\sigma_n(1) = 1 + \sum_{\ell=1}^{n} \sigma_{\ell-1}(1)\langle \hat{\bar{\gamma}}_\ell, (\bar{y}_\ell - \bar{b}_\ell)\rangle. \tag{6.4}$$

Proof. By definition

$$\Lambda_n^* = 1 + \sum_{\ell=1}^{n} \Lambda_{\ell-1}^* \langle \bar{\gamma}_\ell, \bar{y}_\ell - \bar{b}_\ell \rangle$$

and under $\overline{P}$ $\bar{y}_\ell - \bar{b}_\ell$ is a martingale increment. We wish to calculate the optional projection under $\overline{P}$:

$$\sigma_n(1) = \overline{E}[\Lambda_n^* \mid \mathcal{Y}_n].$$

This is, therefore, a filtering problem under measure $\overline{P}$ and the result is a special case of Theorem 5.1 with $H_n = \Lambda_n^*$, $\alpha_\ell = 0$, $\beta_\ell = 0$, $\lambda_\ell = 0$, $\sigma_\ell = 0$ and $\delta_\ell = \Lambda_{\ell-1}^* \bar{\gamma}_\ell$. Consequently, recalling $\bar{a}_\ell$ is replaced by $\bar{b}_\ell$ and P by $\overline{P}$,

$$\begin{aligned} h_\ell &= \left\{ \widehat{\delta}_\ell \operatorname{diag} \bar{b}_\ell - \widehat{\delta}_\ell \bar{b}_\ell \otimes \bar{b}_\ell \right\} \Psi^{-1} \\ &= \widehat{\delta}_\ell \cdot \Psi \cdot \Psi^{-1} = \widehat{\delta}_\ell = \overline{E}[\Lambda_{\ell-1}^* \bar{\gamma}_\ell \mid \mathcal{Y}_{\ell-1}] \end{aligned}$$

and by the Bayes' rule, (6.3), this is

$$= \overline{E}[\Lambda_{\ell-1}^* \mid \mathcal{Y}_{\ell-1}] E[\bar{\gamma}_\ell \mid \mathcal{Y}_{\ell-1}].$$

That is

$$h_\ell = \sigma_{\ell-1}(1) \hat{\bar{\gamma}}_\ell,$$

and the result follows from Theorem 5.1.

We shall also need the following identity which is verified by calculation:

Lemma 6.4. *For $\ell \in Z^+$,*

$$\widehat{\Phi}_\ell^{-1}(\bar{y}_\ell - \hat{a}_\ell)\langle\hat{\bar{\gamma}}_\ell, \bar{y}_\ell - \bar{b}_\ell\rangle = \Psi^{-1}(\bar{y}_\ell - \bar{b}_\ell) - \widehat{\Phi}_\ell^{-1}(\bar{y}_\ell - \hat{a}_\ell). \tag{6.5}$$

Notation 6.5. As in Remarks 5.2 suppose H_n is a scalar process of the form

$$H_n = \sum_{\ell=1}^{n}(\alpha_\ell + \beta_\ell \cdot m_\ell + \delta_\ell \bar{y}_\ell).$$

Here α_ℓ, β_ℓ, δ_ℓ are G-predictable processes of appropriate dimensions, that is α_ℓ, β_ℓ, δ_ℓ are $G_{\ell-1}$ measurable, α_ℓ is scalar, β_ℓ is N dimensional and δ_ℓ is $(k-1)$ dimensional. Write

$$Z_\ell = \alpha_\ell P X_{\ell-1} + H_{\ell-1} Q X_{\ell-1} + P X_{\ell-1}\langle\delta_\ell, \overline{D} X_{\ell-1}\rangle + \langle m_\ell\rangle\beta_\ell'.$$

When compared with the filtering result, Theorem 5.3, the following recursive, Zakai equation for the unnormalized estimate

$$\sigma_n(H_n X_n) = \overline{E}[\Lambda_n^* H_n X_n \mid \mathcal{Y}_n]$$

is remarkable for its simplicity. Of course, $\sigma_n(H_n X_n)$ could be computed by evaluating the product $\Lambda_n^* H_n X_n$ and then taking its $\mathcal{Y}$-optional projection. The martingale representation result under $\overline{P}$ would be used. However, we have preferred to develop the normalized filter first.

Theorem 6.6. *For $1 \le j \le k$ write $d_j = De_j = (d_{1j}, \ldots, d_{kj})'$ for the jth column of $D = (d_{ij})$. Then*

$$\begin{aligned}\sigma_n(H_n X_n) &= \overline{E}[\Lambda_n^* H_n X_n \mid \mathcal{Y}_n] \\ &= k \sum_{j=1}^{N} \sigma_{n-1}\{(H_{n-1} X_{n-1} + Z_n)\langle X_{n-1}, e_j\rangle\}\langle d_j, y_n\rangle.\end{aligned} \tag{6.6}$$

REMARKS 6.7. Note equation (6.6) involves all components of the observation process y_n. With $\underline{1} = (1, 1, \ldots, 1)' \in R^N$ we have $\langle X_n, \underline{1}\rangle = 1$ for all $n \in Z^+$. Formula (6.6) provides a recursive way of determining $\sigma_n(H_n X_n)$; taking the inner product with $\underline{1}$ we, therefore, obtain an expression for $\sigma_n(H_n) = \sigma_n(H_n\langle X_n, \underline{1}\rangle)$. The expression for $\sigma_n(H_n)$ itself is not recursive because, in particular, it involves a term of the form $\sigma_{n-1}(QH_{n-1}X_{n-1})$. Roughly speaking, conditioning H involves a product with X. By starting with HX the conditioning involves a term of the form $HX \otimes X$. However, this can be written $\sum_{i=1}^N \langle HX, e_i\rangle e_i \otimes e_i$ and so no 'higher powers' of X are introduced.

Finally, we note that $\sigma_n(\langle X_n, \underline{1}\rangle) = \sigma_n(1) = \overline{E}[\Lambda_n^* \mid \mathcal{Y}_n]$ and we can obtain the normalized estimate of H_n from Bayes' rule, (6.3), as

$$\check{H}_n = E[H_n \mid \mathcal{Y}_n] = \sigma_n(H_n)/\sigma_n(1). \tag{6.7}$$

We now proceed with the proof of Theorem 6.6.

Proof. From Bayes' rule, (6.3),

$$\sigma_n(H_n X_n) = E[H_n X_n \mid \mathcal{Y}_n] \cdot \sigma_n(1).$$

Using equations (5.7), (6.4) and Lemma 4.13

$$\begin{aligned}\sigma(H_n X_n) = E[H_0 X_0] &+ \sum_{\ell=1}^n \widehat{Z}_\ell \sigma_{\ell-1}(1) + \sum_{\ell=1}^n h_\ell^* \sigma_{\ell-1}(1)(\bar{y}_\ell - \hat{a}_\ell) \\ &+ \sum_{\ell=1}^n \sigma_{\ell-1}(1) E[H_{\ell-1} X_{\ell-1} \mid \mathcal{Y}_{\ell-1}]\langle \hat{\bar{\gamma}}_\ell, (\bar{y}_\ell - \bar{b}_\ell)\rangle \\ &+ \sum_{\ell=1}^n \sigma_{\ell-1}(1)\langle \hat{\bar{\gamma}}_\ell, \bar{y}_\ell - \bar{b}_\ell\rangle \{\widehat{Z}_\ell + h_\ell^*(\bar{y}_\ell - \hat{a}_\ell)\}.\end{aligned} \tag{6.8}$$

From (5.8) we see h_ℓ^* involves the factor $\widehat{\Phi}_\ell^{-1}$ and from Lemma 6.4

$$\widehat{\Phi}_\ell^{-1}(\bar{y}_\ell - \hat{a}_\ell)\langle\hat{\bar{\gamma}}_\ell, \bar{y}_\ell - \bar{b}_\ell\rangle = \Psi^{-1}(\bar{y}_\ell - \bar{b}_\ell) - \widehat{\Phi}_\ell^{-1}(\bar{y}_\ell - \hat{a}_\ell).$$

Substituting in the last term of (6.8) and using Bayes' rule (6.3):

$$\begin{aligned}\sigma_n(H_nX_n) = \sigma_0(H_0X_0) &+ \sum_{\ell=1}^{n}\sigma_{\ell-1}(Z_\ell) + \sum_{\ell=1}^{n}\sigma_{\ell-1}(H_{\ell-1}X_{\ell-1})\langle\Psi^{-1}(\hat{a}_\ell - \bar{b}_\ell), \bar{y}_\ell - \bar{b}_\ell\rangle \\ &+ \sum_{\ell=1}^{n}\sigma_{\ell-1}(Z_\ell)\langle\Psi^{-1}(\hat{a}_\ell - \bar{b}_\ell), \bar{y}_\ell - \bar{b}_\ell\rangle \\ &+ \sum_{\ell=1}^{n}\sigma_{\ell-1}(1)h_\ell^*\widehat{\Phi}_\ell \cdot \Psi^{-1}(\bar{y}_\ell - \bar{b}_\ell). \qquad (6.9)\end{aligned}$$

From (5.8) e final summation in (6.9) is

$$\begin{aligned}&\sum_{\ell=1}^{n}\Big\{\sigma_{\ell-1}(H_{\ell-1}X_{\ell-1} \otimes \overline{D}X_{\ell-1}) - \sigma_{\ell-1}(H_{\ell-1}X_{\ell-1}) \otimes \hat{a}_\ell \\ &+ \sigma_{\ell-1}((\alpha_\ell PX_{\ell-1} + H_{\ell-1}QX_{\ell-1}) \otimes \overline{D}X_{\ell-1}) - \sigma_{\ell-1}(\alpha_\ell PX_{\ell-1} + H_{\ell-1}QX_{\ell-1}) \otimes \hat{a}_\ell \\ &\qquad + \sigma_{\ell-1}((PX_{\ell-1}) \otimes \delta_\ell \operatorname{diag}\overline{D}X_{\ell-1}) - \sigma_{\ell-1}((PX_{\ell-1}) \otimes \delta_\ell) \otimes \hat{a}_\ell \\ &\qquad\qquad + \sigma_{\ell-1}(\langle m_\ell\rangle\beta_\ell' \otimes \overline{D}X_{\ell-1}) - \sigma_{\ell-1}(\langle m_\ell\rangle\beta_\ell') \otimes \hat{a}_\ell\Big\}\Psi^{-1}(\bar{y}_\ell - \bar{b}_\ell).\end{aligned}$$

The terms involving $\hat{a}$ cancel in (6.9) and we have

$$\begin{aligned}\sigma_n(H_nX_n) &= \sigma_0(H_0X_0) + \sum_{\ell=1}^{n}\sigma_{\ell-1}(Z_\ell) \\ &- \sum_{\ell=1}^{n}\sigma_{\ell-1}(H_{\ell-1}X_{\ell-1})\langle\Psi^{-1}\bar{b}_\ell, \bar{y}_\ell - \bar{b}_\ell\rangle - \sum_{\ell=1}^{n}\sigma_{\ell-1}(Z_\ell)\langle\Psi^{-1}\bar{b}_\ell, \bar{y}_\ell - \bar{b}_\ell\rangle \\ &+ \sum_{\ell=1}^{n}\sigma_{\ell-1}\{H_{\ell-1}X_{\ell-1} \otimes \overline{D}X_{\ell-1} + (\alpha_\ell PX_{\ell-1} + H_{\ell-1}QX_{\ell-1}) \otimes \overline{D}X_{\ell-1} \\ &+ (PX_{\ell-1}) \otimes \delta_\ell \operatorname{diag}\overline{D}X_{\ell-1} + \langle m_\ell\rangle\beta_\ell' \otimes \overline{D}X_{\ell-1}\}\Psi^{-1}(\bar{y}_\ell - \bar{b}_\ell). \qquad (6.10)\end{aligned}$$

Note the term in braces is $(H_{\ell-1}X_{\ell-1} + Z_\ell) \otimes \overline{D}X_{\ell-1}$. Writing (6.10) recursively we have

$$\sigma_n(H_nX_n) = \sigma_{n-1}(H_{n-1}X_{n-1}) + \sigma_{n-1}(Z_n)$$

$$- \sigma_{n-1}(H_{n-1}X_{n-1})\langle \bar{b}_n, \Psi^{-1}(\bar{y}_n - \bar{b}_n)\rangle - \sigma_{n-1}(Z_n)\langle \bar{b}_n, \Psi^{-1}(\bar{y}_n - \bar{b}_n)\rangle$$

$$+ \sigma_{n-1}\{(H_{n-1}X_{n-1} + Z_n) \otimes \overline{D}X_{n-1}\}\Psi^{-1}(\bar{y}_n - \bar{b}_n). \tag{6.11}$$

Recall Ψ^{-1} is given in (6.2). If $\bar{y}_n \neq 0$, so $y_n^j = 1$ for some j, $1 \leq j \leq k-1$ then $\Psi^{-1}(\bar{y}_n - \bar{b}_n) = k\bar{y}_n$ and $\langle \bar{b}_n, \Psi^{-1}(\bar{y}_n - \bar{b}_n)\rangle = 1$. In this case, substituting in (6.11)

$$\sigma_n(H_nX_n) = k\sigma_{n-1}\{(H_{n-1}X_{n-1} + Z_n)\langle \overline{D}X_{n-1}, \bar{y}_n\rangle\}$$

$$= k\sum_{j=1}^{N} \sigma_{n-1}\{(H_{n-1}X_{n-1} + Z_n)\langle X_{n-1}, e_j\rangle\}\langle \bar{d}_j, \bar{y}_n\rangle$$

$$= k\sum_{j=1}^{N} \sigma_{n-1}\{(H_{n-1}X_{n-1} + Z_n)\langle X_{n-1}, e_j\rangle\}\langle d_j, y_n\rangle. \tag{6.12}$$

If $\bar{y}_n = 0$, so that $y_n^k = 1$, then

$$\Psi^{-1}(\bar{y}_n - \bar{b}_n) = -\Psi^{-1}\bar{b}_n = -k\underline{1} \in R^{k-1}$$

and $\langle \bar{b}_n, \Psi^{-1}(\bar{y}_n - \bar{b}_n)\rangle = 1 - k$. Substituting in (6.11) in this case we have

$$\sigma_n(H_nX_n) = \sigma_{n-1}(H_{n-1}X_{n-1}) + \sigma_{n-1}(Z_n) + (k-1)\sigma_{n-1}(H_{n-1}X_{n-1})$$

$$+ (k-1)\sigma_{n-1}(Z_n) - k\sigma_{n-1}\{(H_{n-1}X_{n-1} + Z_n)\langle \overline{D}X_{n-1}, \underline{1}\rangle\}.$$

Now

$$\langle \overline{D}X_{n-1}, 1\rangle = \sum_{i=1}^{k-1} a_n^i = 1 - a_n^k = 1 - \sum_{j=1}^{N} \langle X_{n-1}, e_j\rangle\langle d_j, y_n\rangle.$$

Therefore, if $y_n^k = 1$,

$$\sigma_n(H_nX_n) = k\sum_{j=1}^{N} \sigma_{n-1}\{(H_{n-1}X_{n-1} + Z_n)\langle X_{n-1}, e_j\rangle\}\langle d_j, y_n\rangle$$

and the formula is proved in all cases.

7. Special Cases

We now obtain particular forms of the Zakai equation

7.1. Estimates and Smoothers for the State

Take $H_n = H_0 = 1$, $\alpha_\ell = 0$, $\beta_\ell = 0$ and $\delta_\ell = 0$. Applying Theorem 6.6 we have the following recursive estimate for $\sigma_n(X_n) = \overline{E}[\Lambda_n^* X_n \mid \mathcal{Y}_n]$.

$$\sigma_n(X_n) = k \sum_{j=1}^{N} \sigma_{n-1}\{(PX_{n-1})\langle X_{n-1}, e_j\rangle\}\langle d_j, y_n\rangle.$$

Writing $q_n = \sigma_n(X_n)$ and $p_j = Pe_j = (p_{1j}, \ldots, p_{Nj})'$ for the jth column of $P = I + Q$, this is

$$q_n = k \sum_{j=1}^{N} \langle q_{n-1}, e_j\rangle\langle d_j, y_n\rangle p_j. \tag{7.1}$$

q_n is a vector in R^N with non-negative components q_n^i, $1 \le i \le N$. We have noted in Remarks 6.7 that

$$\sigma_n(\langle X_n, 1\rangle) = \langle q_n, 1\rangle = \sum_{i=1}^{N} q_n^i = \sigma_n(1) = \overline{E}[\Lambda_n^* \mid \mathcal{Y}_n]. \tag{7.2}$$

Consequently, the normalized estimate is

$$\hat{X}_n = q_n \langle q_n, 1\rangle^{-1}. \tag{7.3}$$

This form is similar to that given by Åström [1] and Stratonovich [7].

We can also obtain a recursive form for the unnormalized conditional expectation of $\langle X_m, e_i\rangle$ given $\mathcal{Y}_n$, $m < n$. For this we take $H_n = H_m = \langle X_m, e_i\rangle$, $1 \le i \le N$, $\alpha_\ell = 0$, $\beta_\ell = 0$ and $\delta_\ell = 0$. Applying Theorem 6.6 we have

$$\begin{aligned}\sigma_n(\langle X_m, e_i\rangle X_n) &= \overline{E}[\Lambda_n \langle X_m, e_i\rangle X_n \mid \mathcal{Y}_n] \\ &= k \sum_{j=1}^{N} \sigma_{n-1}(\langle X_m, e_i\rangle P X_{n-1}\langle X_{n-1}, e_j\rangle)\langle d_j, y_n\rangle \\ &= k \sum_{j=1}^{N} \sigma_{n-1}(\langle X_m, e_i\rangle\langle X_{n-1}, e_j\rangle)\langle d_j, y_n\rangle p_j. \end{aligned} \tag{7.4}$$

Writing $\langle X_m, e_i\rangle X_n = \phi_n$ we see the right hand side of (7.4) involves ϕ_{n-1}; this is why we consider $H_n X_n$ to obtain a recursive equation. Having obtained a recursive estimate for

$$\overline{E}[\Lambda_n^* \langle X_m, e_i\rangle X_n \mid \mathcal{Y}_n]$$

taking the inner product with $\underline{1}$ gives the smoothed, unnormalized estimate

$$\overline{E}[\Lambda_n^* \langle X_m, e_i\rangle \mid \mathcal{Y}_n].$$

Dividing by $\sigma_n(1)$ from (7.2) gives the smoothed normalized estimate

$$\overline{E}[\Lambda_n^* \langle X_m, e_i\rangle \mid \mathcal{Y}_n]\sigma_n(1)^{-1} = E[\langle X_m, e_i\rangle \mid \mathcal{Y}_n].$$

7.2. Estimates and Smoothers for the Number of Jumps

Recall from Lemma 4.1

$$X_\ell = PX_{\ell-1} + m_\ell. \tag{7.5}$$

Now the Markov chain X jumps from state e_r at time $\ell - 1$ to state e_s at time ℓ, $1 \le r, s \le N$, if $\langle X_{\ell-1}, e_r\rangle\langle X_\ell, e_s\rangle = 1$. Note we can have $e_r = e_s$. The number of jumps from e_r to e_s in time n is, therefore,

$$N_n^{rs} = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle\langle X_\ell, e_s\rangle$$

and from (7.5) this is

$$= \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle\langle PX_{\ell-1}, e_s\rangle + \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle\langle m_\ell, e_s\rangle$$

$$= \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle p_{sr} + \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle\langle m_\ell, e_s\rangle.$$

Applying Theorem 6.6 with $H_n = N_n^{rs}$, $H_0 = 0$, $\alpha_\ell = \langle X_{\ell-1}, e_r\rangle p_{sr}$, $\beta_\ell = \langle X_{\ell-1}, e_r\rangle e_s$, $\delta_\ell = 0$ we have

$$\begin{aligned}
\sigma_n(N_n^{rs}X_n) &= k\sum_{j=1}^{N}\sigma_{n-1}\Big[\Big\{(N_{n-1}^{rs}PX_{n-1} + \langle X_{n-1}, e_r\rangle p_{sr}PX_{n-1} \\
&\quad + \langle X_{n-1}, e_r\rangle(\operatorname{diag} PX_{n-1} - (PX_{n-1})\otimes(PX_{n-1}))e_s\Big\}\langle X_{n-1}, e_j\rangle\langle d_j, y_n\rangle \\
&= k\sum_{j=1}^{N}\sigma_{n-1}(\langle N_{n-1}^{rs}X_{n-1}, e_j\rangle)\langle d_j, y_n\rangle p_j \\
&\quad + k\langle\sigma_{n-1}(X_{n-1}), e_r\rangle\langle d_r, y_n\rangle\{p_{sr}p_r + p_{sr}(e_s - p_r)\} \\
&= k\sum_{j=1}^{N}\sigma_{n-1}(\langle N_{n-1}^{rs}X_{n-1}, e_j\rangle)\langle d_j, y_n\rangle p_j \\
&\quad + k\langle\sigma_{n-1}(X_{n-1}), e_r\rangle\langle d_r, y_n\rangle p_{sr}e_s. \qquad (7.6)
\end{aligned}$$

Together with the recursive equation (7.1) for $q_n = \sigma_n(X_n)$ we have in (7.6) a recursive estimator for $\sigma_n(N_n^{rs}X_n)$. Taking its inner product with $\underline{1}$, that is summing its components, we obtain $\sigma_n(N_n^{rs})$. Finally dividing by $\sigma_n(1)$ we obtain

$$E[N_n^{rs} \mid \mathcal{Y}_n] = \sigma_n(N_n^{rs})\sigma_n(1)^{-1}.$$

Taking $H_n = H_m = N_m^{rs}$, $\alpha_\ell = 0$, $\beta_\ell = 0$, $\delta_\ell = 0$, and applying Theorem 6.6 we obtain for $n > m$

$$\begin{aligned}
\sigma_n(N_m^{rs}X_n) &= \overline{E}[\Lambda_n^* N_m^{rs}X_n \mid \mathcal{Y}_n] \\
&= k\sum_{j=1}^{N}\sigma_{n-1}\{(N_m^{rs}PX_{n-1})\langle X_{n-1}, e_j\rangle\}\langle d_j, y_n\rangle \\
&= k\sum_{j=1}^{N}\sigma_{n-1}\{\langle N_m^{rs}X_{n-1}, e_j\rangle\}\langle d_j, y_n\rangle p_j. \qquad (7.7)
\end{aligned}$$

Again, by considering the product $N_m^{rs} X_n$ a recursive form is obtained. Taking the inner product with $\underline{1}$ gives the smoothed unnormalized estimate

$$\overline{E}[\Lambda_n^* N_m^{rs} \mid \mathcal{Y}_n]$$

and dividing by $\sigma_n(1)$ gives the smoothed estimate

$$E[N_m^{rs} \mid \mathcal{Y}_n].$$

7.3. Estimates and Smoothers for the Occupation Time

The number of occasions up to time n for which the Markov chain X has been in state e_r, $1 \leq r \leq N$, is

$$J_n^r = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle.$$

Taking $H_n = J_n^r$, $H_0 = 0$, $\alpha_\ell = \langle X_{\ell-1}, e_r \rangle$, $\beta_\ell = 0$, $\delta_\ell = 0$ and applying Theorem 6.6 we have:

$$\begin{aligned}\sigma_n(J_n^r X_n) &= k \sum_{j=1}^{N} \sigma_{n-1}\{(J_{n-1}^r P X_{n-1} + \langle X_{n-1}, e_r \rangle P X_{n-1}) \langle X_{n-1}, e_j \rangle\} \langle d_j, y_n \rangle \\ &= k \sum_{j=1}^{N} \langle \sigma_{n-1}(J_{n-1}^r X_{n-1}), e_j \rangle \langle d_j, y_n \rangle p_j + k \langle \sigma_{n-1}(X_{n-1}), e_r \rangle \langle d_r, y_n \rangle p_r. \end{aligned} \tag{7.8}$$

Together with (7.1) for $\sigma_n(X_{n-1})$ this equation gives a recursive expression for $\sigma_n(J_n^r X_n)$. Taking the inner product with $\underline{1}$ gives $\sigma_n(J_n^r)$ and dividing by $\sigma_n(1)$ gives $E[J_n^r \mid \mathcal{Y}_n] = \sigma_n(J_n^r)\sigma_n(1)^{-1}$. For the related smoother take $n > m$, $H_n = H_m = J_m^r$, $\alpha_\ell = 0$, $\beta_\ell = 0$, $\delta_\ell = 0$ and apply Theorem 6.6 to obtain:

$$\begin{aligned}\sigma_n(J_m^r X_n) &= k \sum_{j=1}^{N} \sigma_{n-1}\{(J_m^r P X_{n-1}) \langle X_{n-1}, e_j \rangle\} \langle d_j, y_n \rangle \\ &= k \sum_{j=1}^{N} \langle \sigma_{n-1}(J_m^r X_{n-1}), e_j \rangle \langle d_j, y_n \rangle p_j. \end{aligned} \tag{7.9}$$

Again this is recursive in $\sigma_n(J_m^r X_n)$. The normalized smoother is again obtained as

$$E[J_m^r \mid \mathcal{Y}_n] = \langle \sigma_n(J_m^r X_n), \underline{1} \rangle \sigma_n(1)^{-1}.$$

7.4. Estimates and Smoothers Related to the Observations

In estimating the parameters of our model in the next section we shall require estimates and smoothers of the process

$$G_n^{rs} = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \langle \bar{f}_s, \bar{y}_\ell \rangle$$

which counts the number of times up to time n that the observation process is in state f_s given the Markov chain at the preceding time is in state e_r, $1 \leq r \leq N$, $1 \leq s \leq k-1$.

Taking $H_n = G_n^{rs}$, $H_0 = 0$, $\alpha_\ell = 0$, $\beta_\ell = 0$, $\delta_\ell = \langle X_{\ell-1}, e_r \rangle \bar{f}_s$ and applying Theorem 6.6 we obtain

$$\begin{aligned}
\sigma_n(G_n^{rs} X_n) &= k \sum_{j=1}^{N} \sigma_{n-1}\big(G_{n-1}^{rs} P X_{n-1}\big) \langle d_j, y_n \rangle \langle X_{n-1}, e_j \rangle \\
&\quad + \sigma_{n-1}\big\{ \langle X_{n-1}, e_r \rangle \langle f_s, y_n \rangle P X_{n-1} d_{sr} \big\} \\
&= k \sum_{j=1}^{N} \langle \sigma_{n-1}(G_{n-1}^{rs} X_{n-1}), e_j \rangle \langle d_j, y_n \rangle p_j \\
&\quad + k \langle \sigma_{n-1}(X_{n-1}), e_r \rangle d_{sr} \langle f_s, y_n \rangle p_r. \qquad (7.10)
\end{aligned}$$

Together with equation (7.1) for $\sigma_n(X_n)$ we have a recursive expression for $\sigma_n(G_n^{rs} X_n)$. Normalizing as before we have

$$E[G_n^{rs} \mid \mathcal{Y}_n] = \langle \sigma_n(G_n^{rs} X_n), \underline{1} \rangle \sigma_n(1)^{-1}.$$

To obtain the related smoother take $n > m$, $H_n = H_m = G_m^{rs}$, $\alpha_\ell = 0$, $\beta_\ell = 0$, $\delta_\ell = 0$ and apply Theorem 6.6 to obtain

$$\sigma_n(G_m^{rs} X_n) = k \sum_{j=1}^{N} \langle \sigma_{n-1}(G_m^{rs} X_{n-1}), e_j \rangle \langle d_j, y_n \rangle p_j. \tag{7.11}$$

This is recursive in $\sigma_n(G_m^{rs} X_n)$. The normalized smoother is

$$E[G_m^{rs} \mid \mathcal{Y}_n] = \langle \sigma_n(G_m^{rs} X_n), \underline{1} \rangle \sigma_n(1)^{-1}.$$

8. Parameter Estimation of a Hidden Markov Model

We now show how our normalized and unnormalized filters and smoothers can be used to update the parameters of the model. Suppose as in Section 4 X_n, $n \in Z^+$, is a finite state Markov chain. Then from Lemma 4.1

$$X_n = P X_{n-1} + m_n \tag{8.1}$$

where $P = (p_{ji})$, $1 \le i, j \le N$, is a probability transition matrix with entries satisfying $p_{ij} \ge 0$.

$$\sum_{j=1}^{N} p_{ji} = 1. \tag{8.2}$$

The observation process in the Hidden Markov Model is described in Lemma 4.6 by the equation

$$Y_n = D X_{n-1} + K_n. \tag{8.3}$$

Here $d_{ji} = P(y_n = f_j \mid X_{n-1} = e_i) \ge 0$ and

$$\sum_{j=1}^{k} d_{ji} = 1. \tag{8.4}$$

Our model is, therefore, determined by the set of parameters

$$\theta := (p_{ji},\ 1 \le i, j \le N,\ d_{ji},\ 1 \le j \le k,\ 1 \le i \le N)$$

which are also subject to the constraints (8.2) and (8.4).

Suppose our model is determined by such a set θ and we wish to determine a new set

$$\hat{\theta} = (\pi_{ji},\ 1 \le i, j \le N,\ \delta_{ji},\ 1 \le j \le k,\ 1 \le i \le N)$$

which maximizes the log-likelihoods defined below. Consider first the parameters p_{ji}. From Section 7.2 the number of jumps of X from e_r to e_s in time n is

$$\begin{aligned} N_n^{rs} &= \sum_{\ell=1}^{n} (\langle X_{\ell-1}, e_r \rangle p_{sr} + \langle X_{\ell-1}, e_r \rangle \langle m_\ell, e_s \rangle) \\ &= N_{n-1}^{rs} + \langle X_{n-1}, e_r \rangle p_{sr} + \langle X_{n-1}, e_r \rangle \langle m_n, e_s \rangle. \end{aligned}$$

(Note e_r can be the same as e_s.) Therefore,

$$\Delta N_{n-1}^{rs} = N_n^{rs} - N_{n-1}^{rs} = \langle X_{n-1}, e_r \rangle (p_{sr} + \langle m_n, e_s \rangle).$$

Now $\{\Delta N_n^{rs}\}$, $1 \le r, s \le N$, is an N^2-dimensional process just one component of which is 1 for any $n \in Z^+$, the remaining components being 0. Further

$$E[\Delta N_n^{rs} \mid F_{n-1}] = \langle X_{n-1}, e_r \rangle p_{sr} \ge 0.$$

We are, therefore, in the situation of Section 3 with the process y_n replaced by $\{\Delta N_n^{rs}\}$ and

$$a_n^{rs} = \langle X_{n-1}, e_r \rangle p_{sr}. \tag{8.5}$$

Note

$$\sum_{s=1}^{N} \langle X_{n-1}, e_r \rangle p_{sr} = \langle X_{n-1}, e_r \rangle$$

and

$$\sum_{r,s=1}^{N} \langle X_{n-1}, e_r \rangle p_{sr} = 1.$$

The dynamic form of the constraint (8.2) is then

$$\sum_{\ell=1}^{n} \sum_{r,s=1}^{N} \langle X_{\ell-1}, e_r \rangle p_{sr} = n. \tag{8.6}$$

If we wish to replace the parameters p_{sr} by π_{sr}, $1 \leq s, r \leq N$, we are replacing the a_n^{rs} in (8.5) by

$$b_n^{rs} = \langle X_{n-1}, e_r \rangle \pi_{sr}.$$

That is, the vector b_n of Section 3 is the N^2 dimensional vector with components b_n^{rs}.

As in Section 3 the restriction of the required Radon–Nikodym derivative to F_n is

$$\Lambda_n = \prod_{\ell=1}^{n} \prod_{r,s=1}^{N} \left(\frac{\pi_{sr}}{p_{sr}} \right)^{\Delta N_\ell^{rs}}.$$

Consequently,

$$\begin{aligned} \log \Lambda_n &= \sum_{r,s=1}^{N} \sum_{\ell=1}^{n} \Delta N_\ell^{rs} (\log \pi_{sr} - \log p_{sr}) \\ &= \sum_{r,s=1}^{N} N_n^{rs} \log \pi_{sr} + R(p) \end{aligned}$$

where $R(p)$ is independent of π. Therefore,

$$E[\log \Lambda_n \mid \mathcal{Y}_n] = \sum_{r,s=1}^{N} \check{N}_n^{rs} \log \pi_{sr} + \check{R}(p). \tag{8.7}$$

Now the π_{sr} must also satisfy the analog of (8.6), that is

$$\sum_{\ell=1}^{n} \sum_{r,s=1}^{N} \langle X_{\ell-1}, e_r \rangle \pi_{sr} = n. \tag{8.8}$$

Conditioning on $\mathcal{Y}_n$, the constraint (8.8) states that

$$\sum_{r,s=1}^{N} \check{J}_n^r \pi_{sr} - n = 0. \tag{8.9}$$

We wish, therefore, to choose the π_{sr} to maximize the conditional log-likelihood (8.7) subject to the constraint (8.9). Write λ for the Lagrange multiplier and put

$$L(\pi, \lambda) = \sum_{r,s=1}^{N} \check{N}_n^{rs} \log \pi_{sr} + \check{R}(p) + \lambda \Big(\sum_{r,s=1}^{N} \check{J}_n^r \pi_{sr} - n \Big).$$

Differentiating in λ and π_{sr}, and equating the derivatives to 0, we have the optimum choice of π_{sr} is given by the equations

$$\frac{1}{\pi_{sr}} \check{N}_n^{rs} + \lambda \check{J}_n^r = 0 \tag{8.10}$$

$$\sum_{r,s=1}^{N} \check{J}_n^r \pi_{sr} - n = 0. \tag{8.11}$$

From (8.11) we see that $\lambda = -1$ so the optimum choice of π_{sr}, $1 \leq s, r \leq N$, is

$$\pi_{sr} = \frac{\check{N}_n^{rs}}{\check{J}_n^r} = \frac{\sigma_n(N_n^{rs})}{\sigma_n(J_n^r)}. \tag{8.12}$$

Consider now the parameters d_{ji} in the matrix D. From Section 7.4 the number of times up to time n the observation process jumps to state f_s, $1 \leq s \leq k-1$, given the Markov chain X is in state e_r, $1 \leq r \leq N$, at the preceding time, is

$$G_n^{rs} = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \langle \bar{f}_s, \bar{y}_\ell \rangle \tag{8.13}$$

$$= \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \langle \bar{f}_s, \overline{D} X_{\ell-1} \rangle + \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \langle \bar{f}_s, \bar{y}_\ell - \overline{D} X_{\ell-1} \rangle$$

$$= \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle d_{sr} + \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \langle \bar{f}_s, \bar{\nu}_\ell \rangle. \tag{8.14}$$

Here $1 \leq s \leq k-1$, $1 \leq r \leq N$. The reason we did not include $s = k$ in (8.13) was so that the inner product could be written in terms of $\bar{y}_\ell$ and our general Zakai equation of Theorem 6.6 used. However, we can write (8.14) in terms of $\nu_\ell = y_\ell - a_\ell = y_\ell - DX_{\ell-1}$ as:

$$G_n^{rs} = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle d_{sr} + \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r \rangle \langle f_s, \nu_\ell \rangle. \tag{8.15}$$

This equation is valid for $1 \leq s \leq k$, $1 \leq r \leq N$, and

$$\Delta G_n^{rs} = G_n^{rs} - G_{n-1}^{rs} = \langle X_{n-1}, e_r \rangle (d_{sr} + \langle f_s, \nu_n \rangle).$$

Now $\{\Delta G_n^{rs}\}$ is an Nk dimensional process, just one component of which is 1 for any $n \in Z^+$, the remaining components being 0. Further,

$$E[\Delta G_n^{rs} \mid G_{n-1}] = \langle X_{n-1}, e_r \rangle d_{sr} \geq 0.$$

The procedure now is the same as that above, when we wished to replace in an optimal way the p_{sr} by π_{sr}.

Again

$$\sum_{s=1}^{k} \langle X_{n-1}, e_r \rangle d_{sr} = \langle X_{n-1}, e_r \rangle$$

so

$$\sum_{r=1}^{N} \sum_{s=1}^{k} \langle X_{n-1}, e_r \rangle d_{sr} = 1.$$

The dynamic form of the constraint (8.4) is

$$\sum_{\ell=1}^{n} \sum_{r=1}^{N} \sum_{s=1}^{k} \langle X_{n-1}, e_r \rangle d_{sr} = n. \tag{8.16}$$

To replace the parameters d_{sr} by δ_{sr} we must now consider the Radon–Nikodym derivative

$$\widetilde{\Lambda}_n = \prod_{\ell=1}^{n} \prod_{r=1}^{N} \prod_{s=1}^{k} \left(\frac{\delta_{sr}}{d_{sr}}\right)^{\Delta G_\ell^{rs}}.$$

Then

$$E[\log \widetilde{\Lambda}_n \mid \mathcal{Y}_n] = \sum_{r=1}^{N} \sum_{s=1}^{k} \breve{G}_n^{rs} \log \delta_{sr} + \widetilde{R}(p) \tag{8.17}$$

where $\widetilde{R}(p)$ is independent of δ. Now the δ_{sr} must also satisfy the conditioned version of (8.16), that is

$$\sum_{r=1}^{N} \sum_{s=1}^{k} \breve{J}_n^r \delta_{sr} - n = 0. \tag{8.18}$$

We wish, therefore, to choose the δ_{sr} to maximize the conditional log-likelihood (8.17) subject to the constraint (8.18).

Writing λ for the Lagrange multiplier we have that the optimum choice of δ_{sr} is given by the equations

$$\frac{1}{\delta_{sr}} \breve{G}_n^{rs} + \lambda \breve{J}_n^r = 0$$

$$\sum_{r=1}^{N} \sum_{s=1}^{k} \breve{J}_n^r \delta_{sr} - n = 0.$$

Again we must have $\lambda = -1$ and

$$\delta_{sr} = \frac{\breve{G}_n^{rs}}{\breve{J}_n^r} = \frac{\sigma_n(G_n^{rs})}{\sigma_n(J_n^r)}. \tag{8.19}$$

We have obtained in Section 7.4 the Zakai estimates for $\sigma_n(G_n^{rs})$ only when $1 \leq s \leq k-1$. Together with the estimates for $\sigma_n(J_n^r)$ in Section 7.3 we can determine the optimal choice for δ_{sr}, $1 \leq s \leq k-1$, $1 \leq r \leq N$. However, $\sum_{s=1}^{k} \delta_{sr} = 1$ for each r, so the remaining δ_{kr} can also be found.

REMARKS 8.1. As is known from the work of Baum and Petrie [2] and Rabiner [5] the revised parameters π_{sr}, δ_{sr} determined by (8.12) and (8.19) give new probability measures for the model. The contribution of this paper is to give new, recursive, discrete time filters and smoothers for the quantities $\sigma_n(N_n^{rs})$, $\sigma_n(G_n^r)$, $\sigma_n(J_n^r)$ which are used to determine these parameters. The sequence of densities Λ_n and $\widetilde{\Lambda}_n$ is increasing by construction, and so the model is then adaptive, or 'self tuning' to the observations.

References

[1] K.J. Åström. Optimal control of Markov processes with incomplete state information, *Jour. Math. Anal. and App.* 10 (1965), 174–205.

[2] L.E. Baum and T. Petrie, Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.* 37 (1966), 1554–1563.

[3] R. Boel. Discrete time martingales in filtering and control. Report: Division of Engineering Cybernetics. NTH. Trondheim, Norway 1976.

[4] R.J. Elliott. New finite dimensional filters and smoothers related to the Wonham filter. Preprint June 1991.

[5] L.R. Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. I.E.E.E.* 77 (1989), 257–286.

[6] A. Segall. Recursive estimation from discrete time point processes, *I.E.E.E. Trans. on Info. Theory* IT-22 (1976), 422–431.

[7] R.L. Stratonovich, Conditional Markov processes, *Theory Prob. App.* 5 (1960), 156–178.

Chapter Four

Estimation of the Diffusion Coefficient

1. Introduction

Given a diffusion process it is often easier to estimate the drift coefficient than the diffusion coefficient. Log-normal diffusion processes are frequently used to model asset prices in finance and, in a recent paper, [1], Chesney and Elliott have used the Mihlstein approximation, [2], to estimate the diffusion coefficient, (known in finance as the volatility). The application they had in mind was that of an exchange rate between two currencies: if S_t represents the U.S. dollar to French Franc rate, then $1/S_t$ represents the French Franc to U.S. dollar rate. The result of [1] follows by using the Itô calculus and properties of a log-normal diffusion.

In this paper a general (scalar) diffusion x_t is considered. By introducing the process $y_t = \exp x_t$ properties of the exponential can be exploited. In the paper of Chesney and Elliott a point estimate for the diffusion coefficient is obtained by comparing expressions derived from S_t and S_t^{-1}; in the present paper estimates for the diffusion coefficient of x_t are obtained by using the Itô calculus and Mihlstein approximations, and comparing expressions for y_t and y_t^α, (α real). The minimum variance estimate gives a unique optimal value of α. A table illustrating optimal α values is also presented.

A version of this chapter has been accepted for publication as part of the paper: Diffusion Coefficient Estimation and Asset Pricing when Risk Premia and Sensitivities are Time Varying. Marc Chesney, Robert J. Elliott, Dilip Madan and Hailiang Yang. *Mathematical Finance*.

The scalar diffusion estimation strategy is then extended to also allow estimation of the instantaneous variation in the predictable quadratic covariation of two diffusion processes. Such a point estimate may be used to accommodate time varying risk sensitivities in asset pricing models that simultaneously permit time variation in risk premia as well. Applications to the Capital Asset Pricing Model illustrate the procedure.

Section 2 presents approximations for the increment in stochastic differential equations for the scalar case. These approximations are used to develop a class of diffusion coefficient estimates in section 3. Section 4 identifies a minimum variance optimal estimate in this class. The results are extended to the covariation in section 5. Section 6 applies the point estimates of section 5 to develop a procedure for estimating asset pricing models that permit simultaneous time variation in risk sensitivities and premia. Results of an application to the capital asset pricing model are presented in section 7. Section 8 concludes.

2. Approximation for S.D.E. Increments

Suppose x_t, $t \geq 0$, is a (real) process defined as the solution of a stochastic differential equation

$$dx_t = f(t, x_t)dt + g(t, x_t)dw_t. \tag{2.1}$$

Here w_t, $t \geq 0$. is a real Brownian motion defined on a probability space (Ω, F, P).

If x_t is known $x_{t+\Delta t}$ is given by

$$x_{t+\Delta t} = x_t + \int_t^{t+\Delta t} f(r, x_r)dr + \int_t^{t+\Delta t} g(r, x_r)dw_r \tag{2.2}$$

and a first approximation is to write

$$x_{t+\Delta t} \approx x_t + f(t, x_t)\Delta t + g(t, x_t)\Delta w_t \tag{2.3}$$

where $\Delta w_t = w_{t+\Delta t} - w_t$.

From Mihlstein, [2], a better approximation for $x_{t+\Delta t}$ is given by

$$x_{t+\Delta t} \approx x_t + f(t, x_t)\Delta t + g(t, x_t)\Delta w_t + \frac{1}{2}\frac{\partial g(t, x_t)}{\partial x} g(t, x_t)\Big((\Delta w_t)^2 - \Delta t\Big). \tag{2.4}$$

3. Dynamics and Estimation of Diffusion Coefficient

Suppose the state of the system is described by a stochastic differential equation:

$$dx_t = \mu(t, x_t)dt + \sigma(t, x_t)dw_t, \qquad x_t \in R, \quad t \geq 0, \quad x_0 \text{ given.} \tag{3.1}$$

Here w_t is a real Brownian motion.

Consider the process

$$y_t = \exp x_t = \exp x_0 \cdot \exp\Big\{\int_0^t \mu(s, x_s)ds + \int_0^t \sigma(s, x_s)dw_s\Big\}. \tag{3.2}$$

By Itô's formula,

$$dy_t = y_t\Big(\mu(t, x_t) + \frac{1}{2}\sigma(t, x_t)^2\Big)dt + y_t\sigma(t, x_t)dw_t. \tag{3.3}$$

Let f be a twice differentiable function and $f'' \neq 0$. Then

$$df(y_t) = \Big[f'(y_t)y_t\Big(\mu(t, x_t) + \frac{1}{2}\sigma(t, x_t)^2\Big) + \frac{1}{2}f''(y_t)y_t^2\sigma(t, x_t)^2\Big]dt + f'(y_t)y_t\sigma(t, x_t)dw_t. \tag{3.4}$$

Using (2.4) we have

$$y_{t+\Delta t} \approx y_t + y_t\Big(\mu(t,x_t) + \frac{1}{2}\sigma(t,x_t)^2\Big)\Delta t + y_t\sigma(t,x_t)\Delta w_t$$
$$+ \frac{1}{2}\Big\{y_t\sigma(t,x_t)^2 + y_t^2\frac{\partial\sigma(t,x_t)}{\partial y}\sigma(t,x_t)\Big\}\big((\Delta w_t)^2 - \Delta t\big) \quad (3.5)$$

and

$$f(y_{t+\Delta t}) \approx f(y_t) + \Big[f'(y_t)y_t\Big(\mu(t,x_t) + \frac{1}{2}\sigma(t,x_t)^2\Big) + \frac{1}{2}f''(y_t)y_t^2\sigma(t,x_t)^2\Big]\Delta t$$
$$+ f'(y_t)y_t\sigma(t,x_t)\Delta w_t$$
$$+ \frac{1}{2}\Big\{\frac{\partial f'(y_t)}{\partial f(y_t)}f'(y_t)y_t^2\sigma(t,x_t)^2 + f'(y_t)^2\frac{\partial y_t}{\partial f(y_t)}\sigma(t,x_t)^2 y_t$$
$$+ f'(y_t)^2y_t^2\frac{\partial\sigma(t,x_t)}{\partial f(y_t)}\sigma(t,x_t)\Big\}\big((\Delta w_t)^2 - \Delta t\big)$$
$$= f(y_t) + \Big[f'(y_t)y_t\Big(\mu(t,x_t) + \frac{1}{2}\sigma(t,x_t)^2\Big) + \frac{1}{2}f''(y_t)y_t^2\sigma(t,x_t)^2\Big]\Delta t$$
$$+ f'(y_t)y_t\sigma(t,x_t)\Delta w_t$$
$$+ \frac{1}{2}\Big\{f''(y_t)y_t^2\sigma(t,x_t)^2 + f'(y_t)\sigma(t,x_t)^2y_t$$
$$+ f'(y_t)y_t^2\frac{\partial\sigma(t,x_t)}{\partial y_t}\sigma(t,x_t)\Big\}\big((\Delta w)^2 - \Delta t\big). \quad (3.6)$$

From (3.5) we have

$$\frac{y_{t+\Delta t} - y_t}{y_t} \approx \Big(\mu(t,x_t) + \frac{1}{2}\sigma(t,x_t)^2\Big)\Delta t + \sigma(t,x_t)\Delta w_t$$
$$+ \frac{1}{2}\Big(\sigma(t,x_t)^2 + y_t\frac{\partial\sigma(t,x_t)}{\partial y}\sigma(t,x_t)\Big)\big((\Delta w_t)^2 - \Delta t\big). \quad (3.7)$$

From (3.6) we have

$$\frac{f(y_{t+\Delta t}) - f(y_t)}{f'(y_t)y_t} \approx \Big(\mu(t,x_t)+\frac{1}{2}\sigma(t,x_t)^2\Big)\Delta t+\sigma(t,x_t)\Delta w_t+\frac{1}{2}\frac{f''(y_t)}{f'(y_t)}y_t\sigma(t,x_t)^2\Delta t$$
$$+ \frac{1}{2}\Big(\frac{f''(y_t)}{f'(y_t)}y_t\sigma(t,x_t)^2 + \sigma(t,x_t)^2 + y_t\frac{\partial\sigma(t,x_t)}{\partial y}\sigma(t,x_t)\Big)\big((\Delta w_t)^2 - \Delta t\big). \quad (3.8)$$

So

$$\frac{f(y_{t+\Delta t}) - f(y_t)}{f'(y_t)y_t} - \frac{y_{t+\Delta t} - y_t}{y_t} \approx \frac{1}{2}\frac{f''(y_t)y_t}{f'(y_t)}\sigma(t, x_t)^2(\Delta w_t)^2.$$

Since $E(\Delta w_t)^2 = \Delta t$, an estimate of $\sigma(t, x_t)^2$ is given by

$$V = \frac{2f'(y_t)}{f''(y_t)y_t}\Big[\frac{f(y_{t+\Delta t}) - f(y_t)}{f'(y_t)y_t} - \frac{y_{t+\Delta t} - y_t}{y_t}\Big]\frac{1}{\Delta t}. \tag{3.9}$$

Consider a power function $f(y) = y^{1+\alpha}$. We then have the following estimate of $\sigma(t, x_t)^2$:

$$V = \frac{2}{\alpha}\Big[\frac{y_{t+\Delta t}^{1+\alpha} - y_t^{1+\alpha}}{(1+\alpha)y_t^{1+\alpha}} - \frac{y_{t+\Delta t} - y_t}{y_t}\Big]\frac{1}{\Delta t}. \tag{3.10}$$

4. Optimal Power α

We will consider the estimation of $\sigma(t, x_t)^2$ given by (3.10) and determine a necessary condition for α to be optimal. In the following we write $\mu(t, x_t) = \mu$, $\sigma(t, x_t) = \sigma$.

Since

$$y_{t+\Delta t} = y_t \exp\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big),$$

$$\begin{aligned}
E[V \mid x_t] &= \frac{2}{\alpha}\Big[\frac{1}{1+\alpha}\Big(E\Big[\exp\Big((1+\alpha)\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big)\Big) \mid x_t\Big] - 1\Big) \\
&\quad - E\Big[\exp\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big) \mid x_t\Big] + 1\Big]\frac{1}{\Delta t} \\
&= \frac{2}{\alpha\Delta t}\Big[\frac{1}{1+\alpha}\Big(\exp\left((1+\alpha)\mu\Delta t + \sigma^2(1+\alpha)\alpha\Delta t/2\right) - 1\Big) - \exp(\mu\Delta t) + 1\Big] \\
&\approx \sigma^2 + \Big(\mu^2 + \mu\sigma^2 + \alpha\mu\sigma^2 + \frac{\alpha\sigma^4}{4} + \frac{\alpha^2\sigma^4}{4}\Big)\Delta t + o(\Delta t).
\end{aligned} \tag{4.1}$$

Clearly this converges to σ^2 as $\Delta t \to 0$.

We wish to find α, such that the conditional variance of V given x_t is minimized.

$$\mathrm{Var}[V \mid x_t] = E[V^2 \mid x_t] - (E[V \mid x_t])^2,$$

$$\begin{aligned}
E[V^2 \mid x_t] &= \frac{4}{\alpha^2(\Delta t)^2} E\Big\{ \Big[\frac{1}{1+\alpha} \exp\Big((1+\alpha)\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big)\Big) \\
&\quad - \frac{1}{1+\alpha} - \exp\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big) + 1\Big]^2 \mid x_t\Big\} \\
&= \frac{4}{\alpha^2(\Delta t)^2} \Big[\frac{1}{(1+\alpha)^2} E\Big\{ \exp\Big(2(1+\alpha)\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big)\Big) \mid x_t\Big\} \\
&\quad + \frac{\alpha^2}{(1+\alpha)^2} + E\Big\{ \exp\Big(2\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big)\Big) \mid x_t\Big\} \\
&\quad + \frac{2\alpha}{(1+\alpha)^2} E\Big\{ \exp\Big((1+\alpha)\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big)\Big) \mid x_t\Big\} \\
&\quad - \frac{2}{1+\alpha} E\Big\{ \exp\Big((2+\alpha)\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big)\Big) \mid x_t\Big\} \\
&\quad - \frac{2\alpha}{1+\alpha} E\Big\{ \exp\Big((\mu - \frac{\sigma^2}{2})\Delta t + \sigma\Delta w\Big) \mid x_t\Big\}\Big] \\
&= \frac{4}{\alpha^2(\Delta t)^2} \Big[\frac{1}{(1+\alpha)^2} \exp\Big(2(1+\alpha)\Big(\mu + (\frac{1}{2} + \alpha)\sigma^2\Big)\Delta t\Big) \\
&\quad + \frac{\alpha^2}{(1+\alpha)^2} + \exp\Big(2\Big(\mu + \frac{\sigma^2}{2}\Big)\Delta t\Big) \\
&\quad + \frac{2\alpha}{(1+\alpha)^2} \exp\Big((1+\alpha)\Big(\mu + \frac{\alpha}{2}\sigma^2\Big)\Delta t\Big) \\
&\quad - \frac{2}{1+\alpha} \exp\Big((2+\alpha)\Big(\mu + \frac{\sigma^2}{2} + \frac{\alpha\sigma^2}{2}\Big)\Delta t\Big) - \frac{2\alpha}{1+\alpha} \exp(\mu\Delta t)\Big].
\end{aligned}$$

After some calculation, we have

$$\begin{aligned}
E[V^2 \mid x_t] &= 3\sigma^4 + \frac{4\sigma^2}{1+\alpha} \Big[\frac{89}{24}\alpha\sigma^4 + \frac{3}{2}\mu^2 + \frac{7}{2}\mu\sigma^2 + 6\alpha\mu\sigma^2 \\
&\quad + \frac{17}{12}\sigma^4 + \frac{10}{3}\alpha^2\sigma^4 + \frac{3}{2}\alpha\mu^2 + \frac{5}{2}\alpha^2\mu\sigma^2 + \frac{25}{24}\alpha^3\sigma^4\Big]\Delta t + o(\Delta t) \\
&= 3\sigma^4 + 4\sigma^2 \Big[\frac{25}{24}\sigma^4\alpha^2 + \frac{55}{24}\sigma^4\alpha + \frac{17}{12}\sigma^4 \\
&\quad + \frac{3}{2}\mu^2 + \frac{7\mu\sigma^2}{2} + \frac{5\mu\sigma^2}{2}\alpha\Big]\Delta t + o(\Delta t). \qquad (4.2)
\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{Var}[V \mid x_t] &= E[V^2 \mid x_t] - (E[V \mid x_t])^2 \\ &= 2\sigma^4 + 2\sigma^2\Big[2\mu^2 + 6\mu\sigma^2 + 4\mu\sigma^2\alpha \\ &\quad + \frac{11}{6}\sigma^4\alpha^2 + \frac{13}{3}\sigma^4\alpha + \frac{17}{6}\sigma^4\Big]\Delta t + o(\Delta t) \\ &= g(\alpha) + o(\Delta t). \end{aligned} \tag{4.3}$$

A minimum value of this conditional variance will occur when $g'(\alpha) = 0$; that is, when

$$11\sigma^2\alpha + 13\sigma^2 + 12\mu = 0 \tag{4.4}$$

i.e., when $\alpha = -\frac{13}{11} - \frac{12}{11}\frac{\mu}{\sigma^2}$.

The table below gives these values of α for a variety of values of μ, σ, and also reports the values of EV.

σ	μ	α	EV
0.5	0.2	-2.054544	0.2500463
0.5	0.4	-2.927272	0.2501683
0.5	0.5	-3.363636	0.2502055
0.5	1.0	-5.545454	0.2506804
1.0	0.0	-1.181818	1.001331
1.0	0.5	-1.727272	1.00068
1.0	0.9	-2.163635	1.000911
2.0	0.0	-1.181818	4.001203
2.0	0.5	-1.318181	4.003946
2.0	1.0	-1.454545	4.004741
3.0	0.0	-1.181818	9.012572
3.0	0.5	-1.242424	9.015
3.0	1.0	-1.303030	9.017202

5. Estimates for the Covariation of Diffusion Processes

Suppose now that v_t, $t \geq 0$, is a (vector) process defined as the solution of a stochastic differential equation

$$dv_t = \alpha(t, v_t)dt + \theta(t, v_t)dW_t \tag{5.1}$$

where $v_t \in \mathbb{R}^n$, $\alpha(t, v_t) \in \mathbb{R}^n$, $t \geq 0$, v_0 is given, $\theta(t, v_t)$ is an $n \times m$ matrix with $\theta\theta'$ nonsingular for all t and W_t is a standard m-dimensional Brownian motion.

The Mihlstein approximation of section 2 may also be extended to the vector case for an arbitrary vector diffusion process X_t defined as the solution of a stochastic differential equation

$$dX_t = F(t, X_t)dt + G(t, X_t)dW_t \tag{5.2}$$

where X_t, $F(t, X_t)$ are vector valued, $G(t, X_t)$ is matrix valued with GG' nonsingular and W is as above. The approximation is given by

$$X_{t+\Delta t} \approx X_t + F(t, X_t)\Delta t + G(t, X_t)\Delta W_t + \sum_{kj=1}^{m} c_{kj} z_{kj} \tag{5.3}$$

where

$$z_{kj} = \int_t^{t+\Delta t} W_k(s)dW_j(s)$$

is a zero mean scalar random variable for all pairs k, j (note that $z_{kk} = (\Delta W_k)^2 - \Delta t)$) and c_{kj} is a $\mathcal{F}_t$ measurable vector defined by

$$c_{kj}(t, X_t) = \Sigma_i G_{ik}(t, X_t) \frac{\partial G_j(t, X_t)}{\partial X_{it}}$$

where G_j is the jth column of the matrix G and has the dimension of X.

Following the construction of section 4, consider the processes

$$u_{it} = \exp v_{it} = \exp v_{i0} \cdot \exp\left\{ \int_0^t \alpha_i(s, v_s)ds + \int_0^t \Sigma_k \theta_{ik}(s, v_s)dW_k(s) \right\} \tag{5.4}$$

for $i = 1, \dots, n$. By Ito's formula we have that

$$du_{it} = u_{it}(\alpha_i(t, v_t) + \frac{1}{2}\zeta_i(t, v_t))dt + u_{it}\Sigma_k \theta_{ik}(t, v_t)dW_k(t) \tag{5.5}$$

where $\zeta_i(t, v_t) = (\theta(t, v_t)\theta'(t, v_t))_{ii}$. Also for f a twice differentiable function of a scalar argument with $f'' \neq 0$ we have that

$$df(u_{it}) = \left[f'(u_{it})u_{it}\Big(\alpha_i(t, v_t) + \frac{1}{2}\zeta_i(t, v_t)\Big) + \frac{1}{2}f''(u_{it})u_{it}^2\zeta_i(t, v_t)\right]dt$$
$$+ f'(u_{it})u_{it}\Sigma_k \theta_{ik}(t, v_t)dW_k(t). \tag{5.6}$$

Applying Mihlstein's approximation to the stochastic differential equations (5.5) and (5.6) respectively we obtain that

$$u_{i,t+\Delta t} \approx u_{it} + u_{it}\Big(\alpha_i(t, v_t) + \frac{1}{2}\zeta_i(t, v_t)\Big)\Delta t + \text{ terms in } \Delta W_k \text{ and } z_{kj}. \tag{5.7}$$

and that

$$f(u_{i,t+\Delta t}) \approx f(u_{it}) + \left[f'(u_{it})u_{it}\Big(\alpha_i(t, v_t) + \frac{1}{2}\zeta_i(t, v_t)\Big) + \frac{1}{2}f''(u_{it})u_{it}^2\zeta_i(t, v_t)\right]\Delta t$$
$$+ \text{ terms in } \Delta W_k \text{ and } z_{kj}. \tag{5.8}$$

It follows on constructing percentage changes, subtracting the approximation (5.7) from (5.8), and taking expectations conditional on $\mathcal{F}_t$ that

$$E_t\left[\frac{f(u_{i,t+\Delta t}) - f(u_{it})}{f'(u_{it})u_{it}} - \frac{u_{i,t+\Delta t} - u_{it}}{u_{it}}\right] \approx \frac{1}{2}\frac{f''(u_{it})u_{it}}{f'(u_{it})}\zeta_i(t, v_t)\Delta t. \tag{5.9}$$

Therefore defining

$$\Gamma_t = \frac{2f'(u_{it})}{f''(u_{it})u_{it}}\left[\frac{f(u_{i,t+\Delta t}) - f(u_{it})}{f'(u_{it})u_{it}} - \frac{u_{i,t+\Delta t} - u_{it}}{u_{it}}\right]\frac{1}{\Delta t} \tag{5.10}$$

we have that

$$E_t[\Gamma_t] \approx \zeta_i(t, v_t). \tag{5.11}$$

Choosing for f the power function $f(y) = y^{1+\alpha}$ yields the explicit estimate

$$\Gamma_i = \frac{2}{\alpha}\left[\frac{u_{i,t+\Delta t}^{1+\alpha} - u_{it}^{1+\alpha}}{(1+\alpha)u_{it}^{1+\alpha}} - \frac{u_{i,t+\Delta t} - u_{it}}{u_{it}}\right]\frac{1}{\Delta t} \tag{5.12}$$

for $\zeta_i(t, v_t)$.

The instantaneous variation in the quadratic covariation between v_i and v_j is given by

$$C_{ij}(t, v_t) = \Sigma_k \theta_{ik}(t, v_t)\theta_{jk}(t, v_t) = [\theta\theta']_{ij} \tag{5.13}$$

and this may be obtained as one fourth of the difference between the instantaneous variations in the quadratic variations of $v_i + v_j$ and $v_i - v_j$. Specifically define the variables

$$v_{ipj} = v_i + v_j$$

and

$$v_{imj} = v_i - v_j$$

and construct Γ_{ipj} and Γ_{imj}, with respect to v_{ipj} and v_{imj} respectively, as described above. It follows on analysis that

$$E_t\left[\frac{\Gamma_{ipj} - \Gamma_{imj}}{4}\right] \approx C_{ij}.$$

Hence an estimate for C_{ij} is given by

$$\Phi_{ij} = \frac{\Gamma_{ipj} - \Gamma_{imj}}{4}. \tag{5.14}$$

6. Covariation Estimates and the Merton Intemporal Capital Asset Pricing Model

The Merton intertemporal capital asset pricing model (ICAPM) asserts that the conditional drift in asset prices is linear across assets in the instantaneous beta's of the asset with respect to the market portfolio. Specifically suppose that for each asset $i = 1, \ldots, n$, the asset price process S_{it} is a diffusion given by the solution to the stochastic differential equation

$$dS_{it} = S_{it}\mu_{it}dt + S_{it}\sigma_{it}dW \tag{6.1}$$

where μ_{it} is the instantaneous scalar drift in the asset price or the instantaneous rate of return, σ_{it} is a row vector of diffusion coefficients and W is a standard m-dimensional Brownian motion. Define by S_{Mt} the process of the market value of the market portfolio consisting of the total value of all assets in the economy. Suppose that S_{Mt} satisfies the stochastic differential equation

$$dS_{Mt} = S_{Mt}\mu_{Mt}dt + S_{Mt}\sigma_{Mt}dW. \tag{6.2}$$

Define the instantaneous asset beta's by

$$\beta_{it} = \frac{\sigma_{it}\sigma'_{Mt}}{\sigma_{Mt}\sigma'_{Mt}} \tag{6.3}$$

then the Merton ICAPM asserts that there exist a process γ_{0t} such that

$$\mu_{it} = \gamma_{0t} + \beta_{it}(\mu_{Mt} - \gamma_{0t})$$

for all i and t. Define the process γ_{1t} by

$$\gamma_{1t} = \frac{\mu_{Mt} - \gamma_{0t}}{\sigma_{Mt}\sigma'_{Mt}} \tag{6.4}$$

and rewrite the content of the Merton ICAPM as follows,

$$\mu_{it} = \gamma_{0t} + \sigma_{it}\sigma'_{Mt}\gamma_{1t} \tag{6.5}$$

for all i and t. Define

$$v_{it} = \ln S_{it}$$

for $i = 1, \ldots, n$ and let

$$v_{Mt} = \ln S_{Mt}$$

and observe that the diffusion coefficient of v_{it} is precisely the vector σ_{it} for all i, while the diffusion coefficient for v_{Mt} is σ_{Mt}. Hence in terms of the notation of section 5, the ICAPM asserts that

$$\mu_{it} = \gamma_{0t} + C_{iMt}\gamma_{1t} \tag{6.6}$$

for all i and t.

In the absence of point estimates of C_{iMt}, previous empirical research has focused on supposing constancy through time of the covariations that define systematic risk sensitivities as well as constancy of the coefficients γ_{0t} and γ_{1t} that are interpreted as the appropriate market risk premia. Recently Ferson (1991) has relaxed the assumption of constant risk premia while Harvey (1991) has relaxed the assumption of constant covariations or beta's. Using the point estimates of section 5 we allow for simultaneous time variation in both risk sensitivities and risk premia.

Let R_{it} be the instantaneous rate of return in the asset price,

$$R_{it} = \frac{dS_{it}}{S_{it}} \approx \frac{\Delta S_{it}}{S_{it}} \tag{6.7}$$

let Φ_{iMt} be as constructed in section 5 from $v_{it} = \ln S_{it}$ and $v_{Mt} = \ln S_{Mt}$ and define

$$\varepsilon_{it} = R_{it} - \gamma_{0t} - \Phi_{iMt}\gamma_{1t}$$

then the ICAPM asserts the existence of processes γ_{0t}, γ_{1t} such that

$$E_t[\varepsilon_{it}] = 0$$

for all i and t. In particular for any variable Z_{jt} in the information set at time t we have the orthogonality condition

$$E_t[\varepsilon_{it} Z_{jt}] = 0 \tag{6.8}$$

for all i, j and t.

Given the cross sectional variation in i and j these conditions provided a potentially powerful basis with which to investigate time variation in the risk premia using Generalized Moment Methods (GMM) for estimation of the coefficients of the risk premia processes. Risk sensitivity variations are explicitly allowed for through variations in Φ_{iMt}, constructed as described in section 5.

7. Results for the ICAPM using Covariation Estimates

Stock return series are taken from the CRSP (Center for Research in Securities Prices) tapes. Twelve stock portfolios are formed following the procedures outlined in Breeden, Gibbons and Litzenberger (1989). All the portfolios are value weighted to closely simulate the return on a "buy and hold" strategy. Every return, except firms with a SIC number 39 (i.e., miscellaneous manufacturing), on the tape from

1959 to 1986 is included. Returns on value weighted and equally weighted portfolios of all stocks are also obtained. Let R_{it} denote the return on the twelve portfolios indexed by $i = 1, \ldots, 12$. Let R_{vt} and R_{et} be the series of returns on the value weighted and equally weighted portfolios respectively. We use R_{vt} and R_{et} as proxies for the market portfolio.

Let v_{it}, $i = 1, \ldots, 12$, v_{Mvt} and V_{Met} be the series of log portfolio prices for the twelve stock portfolios and the value and equally weighted portfolios respectively. These series are constructed by cummulating the natural logs of one plus the associated return. We then construct point estimates of covariations between the v_{it}'s and the value weighted and equally weighted log portfolio prices, v_{Mvt} and v_{Met} respectively. The corresponding series are Φ_{iMvt} and Φ_{iMet} constructed using the power function with power -1.1818 that is optimal for large volatilities. Figure 1 below presents a sample (for Petroleum, Finance and Real Estate and Consumer Durables) of covariation estimates covering the period June 1959 to December 1986. As may be observed, there is considerable variation in the covariations and the assumption of constant covariations employed in previous studies could be problematic. This is particularly true for the years 1974, 1980, 1982 and 1984.

In order to assess the content of Merton's ICAPM we first proceeded on the assumption of constant γ coefficients in (6.6) and regressed returns defined in (6.7) on our covariation estimates Φ_{iMt}. The first set of regressions does not impose constancy of the γ coefficients across the twelve portfolios and the results are reported in Table 1. The second set imposes this restriction and uses the SUR (seemingly unrelated regression) procedure. The SUR results are reported in Table 2. These

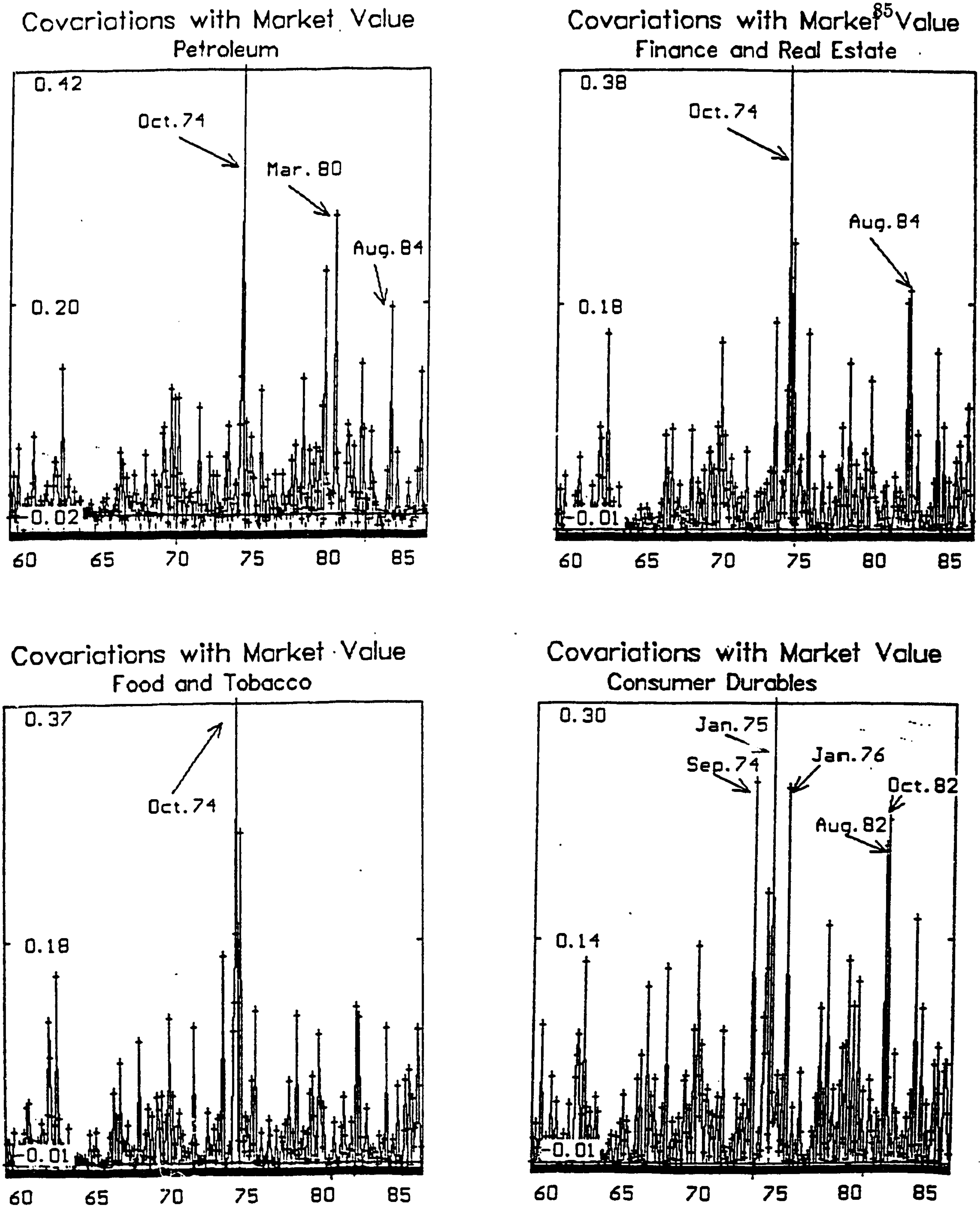
Covariations with Market Value
Petroleum
0.42
0.20
-0.02
Oct.74
Mar.80
Aug.84
60
65
70
75
80
85
Covariations with Market Value
Finance and Real Estate
0.38
0.18
-0.01
Oct.74
Aug.84
60
65
70
75
80
85
Covariations with Market Value
Food and Tobacco
0.37
0.18
-0.01
Oct.74
60
65
70
75
80
85
Covariations with Market Value
Consumer Durables
0.30
0.14
-0.01
Jan.75
Sep.74
Jan.76
Oct.82
Aug.82
60
65
70
75
80
85

FIG. 1

results though instructive, could be biased through possible correlation between the regression residuals and the explanatory variables Φ_{iMt} employed. This problem is avoided later when we employ GMM to investigate time variation in the risk premia.

It may be observed from Table 1 that the estimated market risk premium is significant in all cases exce portfolio 12 for the Leisure industry. Excluding the case of Leisure, the variation of risk premia across portfolios is not that marked, ranging from .24 to .41 with an average value of .30 in the case of the value weighted index as the market portfolio proxy. This range, .20 to .27, is much smaller for the equally weighted index as the market portfolio proxy, with an average value of .22.

The theoretical restriction of the ICAPM, of equality of γ coefficients across portfolios, is imposed in Table 2. We observe from Table 2 that the risk of the market portfolio is highly significant in explaining the variation in portfolio returns both cross-sectionally and over time. This is because our results come from a pooled cross-section and time series estimation conducted simultaneously in both dimensions. The subperiod results are however indicative of considerable variation in risk premia over time with a range of .1215 to 1.033 for the value weighted index as the market portfolio proxy, and −.1130 to .8178 for the equally weighted index as the proxy.

Given point estimates for the covariations it is possible to estimate time varying risk premia using the GMM procedure applied to the pooled cross-section and time series data as indicated in section 6. However, there are 331 months in the data set and with two parameters for each month we have a total of 662 parameters to be estimated. This is a very large estimation problem. We restrict the time variation in the parameters to a smaller dimensional entity by requiring the parameters to lie on a natural cubic spline with knot points at 2.5 year intervals. We then have a total of 24 parameters, 12 each for γ_0 and γ_1. The parameter values for each month are a predetermined linear function of the values at these knot points. The instruments used for the GMM procedure included the constant term, all portfolio returns lagged for one and two months and their squares. The results are presented in Table 3. Figure 2 presents graphs of the estimated time series of the risk premia for both the value weighted and equally weighted proxies. The general pattern is comparable for both market proxies and indicative of considerable variation in these coefficients. There is a sharp drop from late 76 to late 79 early 80 indicated in both graphs. A finer analysis with more knot points is probably necessary before we comment further on the sources of these variations.

8. Conclusion

A point estimate for the diffusion coefficient, of a stochastic differential equation on a Brownian filtration, is developed in terms of the difference in percentage changes in powers of the exponential function applied to solutions of the s.d.e. Results for the choice of an optimal power are presented. The method is then extended to include point estimates of the instantaneous variation in the predictable quadratic covariation. This leads to the possibility of investigating the Merton intertemporal capital asset pricing model, allowing for simultaneous variation in risk sensitivities across assets and time and in risk premia across time. Results on twelve industry

Table 1

Results for Individual Portfolio Return Regressions on Covariations over the Period June 1959 to December 1986

$$R_{it} = \gamma_{0i} + \Phi_{iMt}\gamma_{1i} + \varepsilon_{it}$$

	Value Weighted Proxy			Equally Weighted Proxy		
Portfolio	$\hat{\gamma}_{0i}$ (t-value)	$\hat{\gamma}_{1i}$ (t-value)	R^2	$\hat{\gamma}_{0i}$ (t-value)	$\hat{\gamma}_{1i}$ (t-value)	R^2
Petroleum	.0034 (1.100)	.3341 (4.939)	.069	.0056 (1.790)	.2201 (3.352)	.033
Finance and Real Estate	.0022 (0.755)	.3287 (5.085)	.073	.0032 (1.095)	.2385 (4.699)	.063
Consumer Durables	.0016 (0.480)	.2883 (4.163)	.050	.0021 (.666)	.2192 (4.519)	.058
Basic Industries	.0010 (.356)	.2907 (4.671)	.062	.0011 (.378)	.2537 (4.646)	.062
Food and Tobacco	.0060 (2.302)	.2398 (3.929)	.045	.0060 (2.356)	.2028 (4.294)	.053
Construction	-.0001 (-0.16)	.2811 (4.334)	.054	-.0001 (-.030)	.2224 (5.006)	.071
Capital Goods	.0029 (.933)	.2457 (3.704)	.040	.0026 (.837)	.2266 (4.239)	.052
Transportation	.0004 (.113)	.3163 (4.533)	.059	.0014 (.386)	.2170 (4.139)	.050
Utilities	.0026 (1.189)	.4050 (6.159)	.103	.0039 (1.858)	.2694 (5.637)	.088
Textiles and Trade	.0041 (1.233)	.2559 (3.724)	.040	.0038 (1.207)	.2151 (4.832)	.066
Services	.0025 (.638)	.2805 (4.257)	.052	.0031 (.817)	.1987 (4.428)	.056
Leisure	.0108 (2.705)	.0303 (.445)	.001	.0087 (2.290)	.0784 (1.757)	.009

specific portfolios for the period 1959 to 1986 illustrate the methods and indicate considerable variation in both dimensions. A more detailed investigation of the sources of these variations is the subject of further research.

Table 2

Results for SUR Imposing Parameter Constancy Across Portfolios for the Period June 1959 to December 1986 and Subperiods

$$R_{it} = \gamma_0 + \Phi_{iMt}\gamma_1 + \varepsilon_{it}$$

	Value Weighted Proxy		Equally Weighted Proxy	
Period	$\hat{\gamma}_0$ (*t*-value)	$\hat{\gamma}_1$ (*t*-value)	$\hat{\gamma}_0$ (*t*-value)	$\hat{\gamma}_1$ (*t*-value)
June 59 - Dec. 86	.0026 (1.521)	.5092 (18.861)	.0034 (1.986)	.4204 (18.954)
June 59 - Dec. 64	.0059 (2.203)	1.0330 (12.540)	.0063 (2.211)	.8103 (10.246)
Jan. 65 - Dec. 69	-.0037 (-1.231)	.9066 (10.106)	-.0036 (-1.215)	.8178 (11.966)
Jan. 70 - Dec. 74	.0031 (.6465)	.1215 (2.119)	.0038 (.7714)	-.1130 (-2.198)
Jan. 75 - Dec. 79	-.0011 (-.356)	.9815 (19.971)	.0020 (.6534)	.5744 (18.192)
Jan. 80 - Dec. 86	.0050 (1.646)	.8209 (15.294)	.0059 (1.964)	.7371 (15.248)

Table 3

Results for GMM Allowing Variation of Parameters

Across Time but Imposing Constancy Across Portfolios

$R_{it} = \gamma_{0t} + \Phi_{iMt}\gamma_{1t} + \varepsilon_{it}$

Time Period: June 1959 - December 1986

Parameters on Natural Cubic Spline at 2.5 year Knot Points

	Value Weighted Proxy		Equally Weighted Proxy	
Time Period for Coefficient	$\hat{\gamma}_{0t}$ (*t*-value)	$\hat{\gamma}_{1t}$ (*t*-value)	$\hat{\gamma}_{0t}$ (*t*-value)	$\hat{\gamma}_{1t}$ (*t*-value)
June 59	-.0062 (-.128)	1.9884 (2.165)	-.0206 (-.416)	2.4746 (2.380)
Dec. 61	.0289 (1.382)	.3887 (1.148)	.0294 (1.383)	.1285 (.423)
June 64	-.0282 (-1.40)	-.3261 (-.389)	-0.0254 (-1.246)	.0636 (.095)
Dec. 66	-.0073 (-.395)	1.1372 (3.106)	-.0033 (-.175)	.9411 (3.401)
June 69	.0085 (.521)	-.0758 (-.2978)	.0136 (.820)	-.0998 (-.498)
Dec. 71	.0264 (1.444)	.4307 (1.416)	.0309 (1.651)	.0659 (.300)
June 74	-.0202 (-1.579)	.1756 (2.552)	0.0210 (-1.602)	.2315 (3.520)
Dec. 76	.0044 (.258)	1.7700 (5.472)	.0176 (1.019)	1.172 (4.923)
June 79	.0039 (.230)	.1536 (.688)	-.0121 (-.696)	.2242 (1.329)
Dec. 81	.0093 (.650)	.4190 (2.143)	.0081 (.557)	.9277 (5.302)
June 84	-.0002 (-.0090)	.9867 (2.868)	-.0049 (-.222)	.5007 (1.907)
Dec. 86	.0085 (.237)	.1751 (.604)	.0170 (.463)	.4974 (1.425)

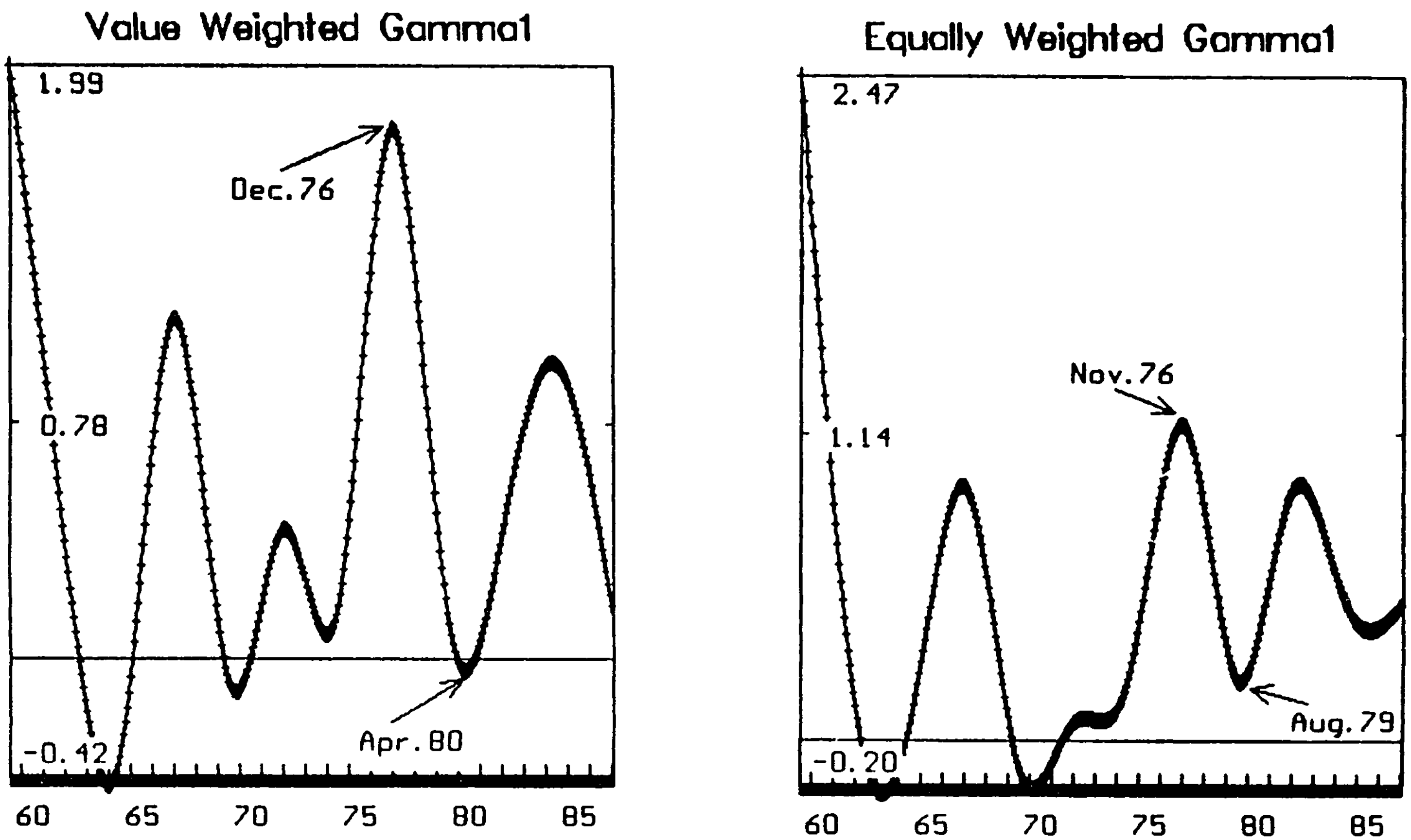
Value Weighted Gamma1
1.99
0.78
-0.42
Dec.76
Apr.80
60
65
70
75
80
85
Equally Weighted Gamma1
2.47
1.14
-0.20
Nov.76
Aug.79
60
65
70
75
80
85

FIG. 2

References

[1] Breeden, D., M.R. Gibbons and R.H. Litzenberger, 1989. Empirical tests of the consumption-oriented CAPM, *Journal of Finance*, 44, 231–262.

[2] Chesney, M. and R.J. Elliott, 1992. Estimating the volatility of an exchange rate. Preprint. Dept. of Statistics and Applied Probability, University of Alberta.

[3] Ferson, Wayne E. and Campbell R. Harvey, 1991. The variation of economic risk premiums, *Journal of Political Economy*. Forthcoming.

[4] Harvey, Campbell R., 1989. Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics*, 24, 289–317.

[5] Merton, R.C., 1990. *Continuous-Time Finance*, Chap. 15, Basil Blackwell, Cambridge, MA.

[6] Mihlstein, G.N., 1974. Approximate integration of stochastic differential equations. *Theory of Probability and Applications* 19, 557–562.