University of Alberta

APPLICATIONS OF HIDDEN MARKOV CHAINS TO CREDIT RISK MODELLING

by

Malgorzata Wiktoria Korolkiewicz ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in

Finance

School of Business

Edmonton, Alberta
Spring 2004

Your file Votre référence
ISBN: 0-612-96288-1
Our file Notre référence
ISBN: 0-612-96288-1

Abstract

We propose that the credit rating evolution can be described by a Markov chain but that we do not observe this Markov chain directly. Rather, it is hidden in "noisy" observations represented by the posted credit ratings. We consider the discrete time model with a Markov Chain observed in martingale noise (Hidden Markov Model). By introducing a new probability measure we are able to obtain unnormalized, recursive estimates for the state of the Markov chain governing the credit rating evolution. We use the so-called EM (Expectation Maximization) algorithm to estimate the parameters of the model, namely probabilities of migration between "true" credit quality states and probabilities of observing a particular rating given the "true" credit worthiness of the issuer. The model is then applied to a data set of credit ratings obtained from the Standard and Poor's COMPUSTAT database. We also consider a Kalman filtering model for estimating the dynamics of credit quality aimed to overcome some of the challenges posed by the nature of available credit rating data.

Finally, we introduce an intensity-based credit migration model of default risk. We take default to be an unpredictable event governed by a hazard process defined in terms of intensity. The value of a zero-recovery defaultable zero-coupon bond is then its value if it were risk-free, adjusted by the probability of no default before

maturity. This probability is calculated explicitly in terms of intensity and the issuer's credit quality. We suppose that the latter is governed by a Markov chain and distinguish two cases. First we take the issuer's credit rating to represent the "true" credit quality and then extend the model to value zero-recovery defaultable bonds when "true" credit quality is not observed directly but only through noisy observations given by posted ratings. We also consider valuation of defaultable bonds when a fraction of face value is paid at the time of default.

Acknowledgement

I would like to thank my supervisors, Robert J. Elliott and Felipe Aguerrevere, committee members Vikas Mehrotra and Abel Cadenillas, the external examiner Dilip B. Madan, Paul Malcolm, PhD Office, Department of Finance and Management Science, fellow PhD students and my friend Maggie Charlton for all the help and support along the way.

Table of Contents

Introduction

The market for credit derivatives has experienced spectacular growth in recent years, creating for investors many new opportunities for higher returns and diversification. Credit derivatives are financial instruments whose payoffs depend on the credit characteristics of a reference asset's value. Swaps and options on corporate debt are but two examples of credit derivatives. Given the sensitivity of credit derivatives to the credit quality of underlying assets, pricing models must make good use of credit risk information.

Credit ratings published in a timely manner by rating agencies are an invaluable source of credit risk information: investors can use the ratings to assess firms' abilities to meet their debt obligations and to estimate the payoffs from the corresponding credit derivatives. For many reasons, however, credit ratings change from time to time and so reflect firms' unpredictable credit risk.

The question of information value of credit ratings is well represented in the literature but some of the earlier works reported mixed results. Two recent studies, Kliger and Sarig [21], and Dichev and Piotroski [8] provide evidence that credit rating changes have impact on returns, and so they contain information that investors cannot obtain from other sources. Kliger and Sarig find that although rating information does not affect firm value, announcements of rating changes have impact on debt and equity values. Dichev and Piotroski report evidence that rating downgrades are followed by negative abnormal returns, which the authors attribute to underreaction to the announcement of downgrades.

The role of credit rating agencies and the transparency of their rating policies has recently come under scrutiny, especially after the collapse of Enron Corporation. The leading credit rating agencies have long been regarded as the most methodical and independent financial research firms, but their cautious approach to rating debt obligations, once the agencies' biggest asset, has been under attack as of late as their clients voiced concerns over the timeliness of the rating reports. The leading rating agencies, such as Moody's Investor Service and Standard & Poor's, have been accused of reacting too slowly to the disaster at Enron: Moody's and Standard & Poor's both continued to rate Enron's debt as investment grade until just days before the company went bankrupt. In response, rating agencies started looking at ways to react more rapidly to changes in the debt market and possibly to trade some of the comprehensiveness of their analysis for timeliness. Just after the Enron collapse, Moody's in particular seems to have shifted to "faster and tougher" coverage of corporate debt by issuing a number of rating cuts. For example, it swiftly downgraded Kmart Corporation's debt to junk status, a call that was vindicated a month later when the company filed for bankruptcy. However, these rating cuts in turn produced accusations of overreaction to Enron's collapse and fears of increased volatility in the bond market. Consequently, the reliability of ratings as indicators of creditworthiness is of even greater importance today and there is an even greater need for gaining a better understanding of the credit rating dynamics.

Existing models of default risk fall into two broad categories: the *structural models* and the *reduced-form* models. *Structural models* are concerned with modelling and pricing default risk specific to a particular corporate borrower. Default is triggered by movements of the firm's value relative to some barrier. A major issue within this framework is the evolution of the firm's value and of the firm's capital structure. In the *reduced-form* approach, the firm value and its capital structure are not modelled at all, and default is specified exogeneously. Within the class of *reduced-form models*, we find the so-called *intensity-based models* and *credit migration models.* The latter take the evolution of credit rating over time to be governed by a Markov chain.

The key assumption behind the Markov chain representation of credit rating evolution is the Markov property, which implies that the rating process should have no memory of its past behaviour. In other words, predictions of future rating evolution based on the past rating history and a current rating are no better than predictions based on the current rating only. Markov Chain Models of credit migration also assume that the dynamics are stationary so that the probability of a transition from one rating category to the next does not change over time. The pioneering work in the direction of using Markov Chain models, not only to describe the dynamics of a firm's credit rating but also to value credit derivatives, was done by Jarrow and Turnbull [18], hereafter JT, and Jarrow, Lando and Turnbull [19], hereafter JLT. JT proposed a model for pricing and hedging derivative securities that takes as given a stochastic term structure of default-free interest rates and the firm's bankruptcy process. As a result, financial instruments subject to credit risk can be priced in an arbitrage-free manner using the equivalent martingale measure technology. The JLT model follows the JT model in spirit and explicitly incorporates credit rating information into the valuation process.

Two empirical studies, Carty and Fons [5], and Carty and Lieberman [6] suggest that the credit rating process may have memory. It is reported that prior rating changes can have predictive power for the direction of future rating changes. These studies find in particular that a firm upgraded (downgraded) is more likely to be subsequently upgraded (downgraded). More recently, Lando and Skodeberg [24] tested the data set provided by Standard and Poor's for the presence of "momentum" or "rating drift." Using the theory of Markov chain modeling, they concluded that there seem to be strong non-Markov effects for downgrades in the entire population of rated firms.

The credit rating process has been subject to many studies aimed at selecting an appropriate statistical technique for estimating a function to explain and predict bond ratings. Although credit rating agencies generally insist that they consider factors that cannot be quantified, studies have shown that approximately two-thirds of ratings can be predicted on the basis of a fairly small number of financial variables. One common approach is to perform an ordered probit analysis and the rationale is as follows (cf. Kaplan and Urwitz [20], Ederington [10]). It is supposed that the bond rater tries to measure the risk of default of bond issues. Unfortunately, the rater can only make an ordinal ranking of bond issues, that is say for example that AAA bonds are less risky than AA, AA bonds are less risky than A bonds and so on. The default risk, measured on interval scale, is then the dependent variable of interest which if observed would satisfy a linear model with financial variables as explanatory variables. Instead, only an ordinal version of the variable of interest is observed which does not satisfy the linear model. The probit model then aims to estimate intervals corresponding to each rating category and bond issues are classified based on estimated probability that an estimate of default risk as a function of the chosen set of financial variables falls within a particular interval. This classification technique could be a source of "memory" in the credit rating process as one can imagine two bond issues with close scores on the explanatory variables being assigned to two different rating categories by virtue of their scores straddling an interval end point.

Markov-type models take the probabilities computed from credit rating data as elements of the correct transition matrix for the credit rating evolution process represented as a Markov chain. The empirical findings, however, suggest that this may not be the case. In other words, the observed rating process is corrupted by what we may call "noise." This is precisely the premise behind *Hidden Markov Models* (HMM). In the HMM framework, we assume that the credit rating evolution can be described by a Markov chain but that we do not observe this Markov chain directly. Rather, it is hidden in "noisy" observations represented by the posted credit ratings. The HMM approach allows us to filter out the "noise" from the observations by purely quantitative means, without investigating the rating assignment process or explicitly looking for factors that cause the observed rating process to have memory. It thus provides a framework for quantitatively assessing the credibility of internal and external rating systems used by financial institutions. The outcome of the HMM applied to the evolution of credit ratings is a probability distribution for a rating at some time k given the information (observed ratings) up to time k. The technique also allows for reestimation of parameters, namely the elements of the transition matrix and the probabilities of observing a particular rating given the "true" credit rating. The latter property is especially valuable since credit transition matrices are at the centre of credit risk management. The reports on rating migrations published by Moody's and Standard and Poor's are studied by credit risk managers everywhere and several of the most prominent risk management tools, such as JP Morgan's CreditMetrics, are built around estimates of rating migration probabilities. The HMM framework provides a tool for evaluating the reliability of these estimates. It can also be incorporated into pricing models for risky debt and credit derivatives to account for "non-Markovian" effects in the behaviour of rating over time. In other words, prices of bonds and credit derivatives would then be derived conditional on the information about the issuer's "true" rating implied by the observed rating history.

Hidden Markov models, when Markov chains are observed in Gaussian noise, have been subject to extensive studies. See for example the book by Elliott, Aggoun and Moore [14] and references therein. Here we consider the discrete time model with a Markov chain observed in martingale noise. By introducing a new probability measure we are able to obtain unnormalized, recursive estimates for the state of the Markov chain governing the evolution of the credit rating process. We use the so-called EM (Expectation Maximization) algorithm to estimate the parameters of the model. The method allows for the parameters to be revised as new information is obtained. The resulting filters are, therefore, adaptive and "self-tuning."

The thesis is organized as follows. Chapter 1 describes the dynamics of the Markov chain and observations. The reference probability and the forward filter for the "true" credit rating process are also introduced. The chapter concludes with recursive formulae for updating the parameters of the model for both the filtering and smoothing case. The filtering results of Chapter 1 are then applied to a data set of issuer credit ratings obtained from the Standard and Poor's COMPUSTAT database. The ratings data and simulation results are discussed in Chapter 2. In Chapter 3 we present a Kalman filtering model for estimating the dynamics of credit quality aimed to overcome some of the challenges posed by the nature of available credit rating data. Finally, in Chapter 4 we describe a *reduced-form* model of default risk that provides formulae for calculating the value of a defaultable bond conditional on the issuer's credit quality.

Chapter 1

Hidden Markov Model

1. INTRODUCTION

Our goal is to estimate, from the published credit ratings, the state of the Markov chain that represents the evolution of a "true" credit rating process. We shall use a procedure known as *filtering* in the stochastic processes literature to obtain an algorithm to be later used to assess "true" credit quality. In general, filtering concerns optimal recursive estimation of a noisy signal given a sequence of observations. We shall suppose that the signal process is a *Markov Chain* which we do not observe directly. We also assume that the observation process has zero delay to the signal process so that the current observation contains information about the current signal value. In our case, a firm's "true" credit quality is a signal and posted credit labels are the noisy observations. Given the zero-delay assumption, we shall then use the rating history up to and including time k to estimate the state of the "true" credit rating process at time k. We also derive formulae for *smoothed* estimates, where we use all available observations to extract information about the signal at time k. In both cases we use the so-called EM (Expectation Maximization) algorithm to provide recursive formulae for estimating the parameters of the model.

2. DYNAMICS OF THE MARKOV CHAIN AND OBSERVATIONS

Formally, a discrete-time, finite-state, time homogeneous Markov chain is a stochastic process $\{X_k\}$ with the state space $S = \{1, 2, \dots, N\}$ and a transition matrix $A = (a_{ji})_{1 \leq i,j \leq N}$. Without loss of generality, we can assume that the elements of S are identified with the standard unit vectors $\{e_1, e_2, \dots, e_N\}$, $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{R}^N$. That is, without loss of generality we can assume that $S = \{e_1, e_2, \dots, e_N\}$. At each time $k \in \{0, 1, 2, \dots\}$ X_k is then one of the unit vectors e_i, $1 \leq i \leq N$. Write $\mathcal{F}_k = \sigma\{X_0, X_1, \dots, X_k\}$ for the *σ-field* containing all the information about the process X up to and including time k. Then, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \mathcal{F}_k$, so that we learn more and more about the process X as time passes. The family of σ-fields $\{\mathcal{F}_k\}$ is a *filtration* that models all possible histories of X. Note that the *Markov property* implies that $P(X_{k+1} = e_j | \mathcal{F}_k) = P(X_{k+1} = e_j | X_k)$. In other words, knowing the current state of the process X is sufficient to make inferences about its future behaviour.

Now, an (i, j)-th entry of A is defined as $a_{ji} := P(X_{k+1} = e_j | X_k = e_i)$, the probability of the process X moving from state i to state j within one unit of time. The relationship between the state process at time k and the state of the process at time $k + 1$ is then as follows:

Lemma 1.1. $E[X_{k+1}|X_k] = AX_k$.

Proof. See Appendix I.

Define $V_{k+1} = X_{k+1} - AX_k \in \mathbb{R}^N$. Then, the *semimartingale representation* of the chain X is

$$X_{k+1} = AX_k + V_{k+1}, \quad k = 0, 1, \dots,$$

where V_{k+1} is a martingale increment with $E[V_{k+1}|\mathcal{F}_k] = 0 \in \mathbb{R}^N$.

Let $p_k = (p_1, \dots, p_N)' = E[X_k]$. Then, $p_{k+1} = Ap_k = A^{k+1}p_0 \in \mathbb{R}^N$. We have:

Lemma 1.2. $Var V_k = E[V_k V_k'] = \text{diag}\,(Ap_{k-1}) - A\,\text{diag}\,(p_{k-1})A'$.

Proof. See Appendix I.

Suppose we do not observe X directly. Rather, we observe a process Y such that

$$Y_k = c(X_k, \omega_k), \quad k = 0, 1, \dots,$$

where c is a function with values in a finite set and $\{\omega_k\}$ is a sequence of i.i.d. random variables independent of X. Random variables $\{\omega_k\}$ represent the noise present in the system. Suppose the range of c consists of M points which are identified with unit vectors $\{f_1, f_2, \dots, f_M\}$, $f_j = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^M$.

Recall $\mathcal{F}_k = \sigma\{X_0, X_1, \dots, X_k\}$. Write

$$\mathcal{Y}_k = \sigma\{Y_0, Y_1, \dots, Y_k\}$$

$$\text{and} \quad \mathcal{G}_k = \sigma\{X_0, \dots, X_k, Y_0, \dots, Y_k\}.$$

These increasing families of σ-fields are then filtrations representing possible histories of the state process X, the observation process Y and both processes (X, Y).

Write $c_{ji} = P(Y_k = f_j | X_k = e_i)$, $1 \le i \le N$, $1 \le j \le M$, for the probability of observing a state f_j when the signal process is in fact in state e_i. Then,

Lemma 1.3. *$E[Y_k | X_k] = CX_k$, where $C = (c_{ji})_{1 \le i,j \le M}$ is a matrix with $c_{ji} \ge 0$ and $\sum_{j=1}^{M} c_{ji} = 1$.*

Proof. In Appendix I.

Define $W_k = Y_k - CX_k$. Then, the semimartingale representation of the process Y is

$$Y_k = CX_k + W_k, \quad k = 0, 1, \dots,$$

where W is a martingale increment with $E[W_k | \mathcal{G}_{k-1} \vee \{X_k\}] = 0 \in \mathbb{R}^M$. Note that we are assuming zero delay between X_k and its observation Y_k. We have the following result:

Lemma 1.4.

$$Var W_k = E[W_k W_k'] = E[(Y_k - CX_k)(Y_k - CX_k)'] = \operatorname{diag}(Cp_k) - C \operatorname{diag}(p_k)\, C'.$$

Proof. In Appendix I.

In summary, our model for the Markov Chain X hidden in martingale noise is as follows:

Hidden Markov Model (HMM)

Under a probability measure P,

$$X_{k+1} = AX_k + V_{k+1} \quad \textit{(signal equation, "true" credit quality)}$$

$$Y_k = CX_k + W_k \quad \textit{(observation equation, posted rating)}$$

A and C are matrices of transition probabilities whose entries satisfy $\sum_{j=1}^{N} a_{ji} = 1$, $a_{ji} \geq 0$, $\sum_{j=1}^{M} c_{ji} = 1$, $c_{ji} \geq 0$.

V_k and W_k are martingale increments satisfying

$$E[V_{k+1}|\mathcal{F}_k] = 0, \quad VarV_k = \text{diag}\,(Ap_{k-1}) - A\,\text{diag}\,(p_{k-1})A',$$

$$E[W_{k+1}|\mathcal{G}_k \vee \{X_{k+1}\}] = 0, \quad VarW_k = \text{diag}\,(Cp_k) - C\,\text{diag}\,(p_k)\,C'.$$

3. REFERENCE PROBABILITY

Suppose that under some probability measure $\bar{P}$ on $(\Omega, \mathcal{F})$, $\{Y_k\}$ is a sequence of i.i.d. uniform variables, i.e. $\bar{P}(Y_{k+1} = f_j|\mathcal{G}_k) = \bar{P}(Y_{k+1} = f_j) = \frac{1}{M}$. Further, under $\bar{P}$, X is Markov chain independent of Y, with state space $S = \{e_1, \ldots, e_N\}$ and transition matrix $A = (a_{ji})$. That is, $X_{k+1} = AX_k + V_{k+1}$, where $\bar{E}[V_{k+1}|\mathcal{G}_k] = \bar{E}[V_{k+1}|\mathcal{F}_k] = 0 \in \mathbb{R}^N$. Suppose $C = (c_{ji})$, $1 \leq i \leq N$, $1 \leq j \leq M$, is a matrix with $c_{ji} \geq 0$, and $\sum_{j=1}^{M} c_{ji} = 1$. We have the following result:

Lemma 1.5. *Define $\bar{\lambda}_l = M \sum_{j=1}^{M} \langle CX_l, f_j \rangle \langle Y_l, f_j \rangle$*[1] *and $\bar{\Lambda}_k = \prod_{l=1}^{k} \bar{\lambda}_l$. Define a new probability measure P by putting $\frac{dP}{d\bar{P}}|_{\mathcal{G}_k} = \bar{\Lambda}_k$. Then, under P, X remains a Markov chain with transition matrix A and $P(Y_k = f_j|X_k = e_i) = c_{ji}$. That is, under P, $X_{k+1} = AX_k + V_{k+1}$ and $Y_k = CX_k + W_k$.*

Proof. Appendix I.

Note. P represents the "real world" probability measure. However, measure $\bar{P}$ is easier to work with since under $\bar{P}$, $\{Y_k\}$ is i.i.d. uniform and independent of X.

---

[1] For any vectors a and b, $\langle a, b \rangle = a'b$.

Lemma 1.5 provides a useful link between our "real world" probability measure P and a "reference" probability measure $\bar{P}$ which preserves all the properties of the process X.

4. RECURSIVE FILTER

Suppose we observe $Y_0, \ldots, Y_k$, and we wish to estimate $X_0, \ldots, X_k$. The best (mean-square) estimate of X_k given $\mathcal{Y}_k = \sigma\{Y_0, \ldots, Y_k\}$ is $E[X_k|\mathcal{Y}_k] \in \mathbb{R}^N$. However, $\bar{P}$ is a much easier measure under which to work. Using Bayes' Theorem, we have

$$E[X_k|\mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k]}.$$

Write $q_k := \bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_k] \in \mathbb{R}^N$. q_k is then an unnormalized conditional expectation of X_k given the observations $\mathcal{Y}_k$.

Lemma 1.6. $\bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k] = \langle q_k, \mathbf{1}\rangle$, *where* $\mathbf{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^N$.

Proof. Appendix I.

It follows that $E[X_k|\mathcal{Y}_k] = \frac{q_k}{\langle q_k, \mathbf{1}\rangle}$. Hence, to estimate $E[X_k|\mathcal{Y}_k]$ we need to know the dynamics of q. The following theorem shows how the unnormalized filter is updated with arrival of each new observation.

Theorem 1.1. *Write* $B(Y_{k+1})$ *for the diagonal matrix with entries*

$$M(\sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle).$$

Then, $q_{k+1} = B(Y_{k+1})Aq_k$.

Proof. Appendix I.

To summarize, given the parameters of the model, namely matrices A and C, the distribution of X_k given information in $\mathcal{Y}_k$ is $E[X_k|\mathcal{Y}_k] = \frac{q_k}{\langle q_k, \mathbf{1}\rangle}$, where $q_{k+1} = B(Y_{k+1})Aq_k$.

5. PARAMETER ESTIMATES

To estimate parameters of the model, matrices A and C, we need estimates of the following processes:

$$J_k^{ij} = \sum_{n=1}^{k} \langle X_{n-1}, e_i\rangle\langle X_n, e_j\rangle, \quad 1 \leq i, j \leq N,$$
$$O_k^{i} = \sum_{n=1}^{k} \langle X_{n-1}, e_i\rangle, \quad 1 \leq i \leq N,$$
$$T_k^{ij} = \sum_{n=0}^{k} \langle X_n, e_i\rangle\langle Y_n, f_j\rangle, \quad 1 \leq i \leq N, \quad 1 \leq j \leq M.$$

The above processes are interpreted as follows:

J_k^{ij} - the number of jumps of X from state e_i to state e_j up to time k.

O_k^{i} - the amount of time the chain has spent in state e_i up to time $k-1$.

T_k^{ij} - the amount of time process X has spent in state e_i when process Y was in state f_j up to time k.

Remark 1.1. Note that $\sum_{j=1}^{N} J_k^{ij} = O_k^i$ and $\sum_{j=1}^{M} T_k^{ij} = O_{k+1}^i$.

Consider first the jump process $\{J_k^{ij}\}$. We wish to estimate J_k^{ij} given the observations $Y_0, \ldots, Y_k$. Using Bayes' Theorem, the best (mean-square) estimate is

$$E[J_k^{ij}|\mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k J_k^{ij}|\mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k]} := \frac{\sigma(J^{ij})_k}{\langle q_k, \mathbf{1}\rangle}.$$

We wish to know how $\sigma(J^{ij})_k$ is updated as time passes and new information arrives. However, there does not exist a recursion formula for $\sigma(J^{ij})_k$. Instead, we consider a vector process $\sigma(J^{ij}X)_k := \bar{E}[\bar{\Lambda}_k J_k^{ij} X_k | \mathcal{Y}_k]$ for which recursive formulae can be derived. We then readily obtain the quantity of interest, namely $\sigma(J^{ij})_k$, since $\sigma(J^{ij})_k = \langle \sigma(J^{ij}X)_k, \mathbf{1}\rangle$. We have the following result:

Theorem 1.2.

$$\sigma(J^{ij}X)_{k+1} = B(Y_{k+1})A\sigma(J^{ij}X)_k + (M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle a_{ji} e_j.$$

Proof. Appendix I.

Similarly, we consider the best (mean square) estimates of O_k^i and T_k^{ij} given $\mathcal{Y}_k$:

$$E[O_k^i | \mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k O_k^i | \mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k | \mathcal{Y}_k]} := \frac{\sigma(O^i)_k}{\langle q_k, \mathbf{1}\rangle},$$

$$E[T_k^{ij} | \mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k T_k^{ij} | \mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k | \mathcal{Y}_k]} := \frac{\sigma(T^{ij})_k}{\langle q_k, \mathbf{1}\rangle}.$$

Recursive formulae for processes $\sigma(O^i X)_k := \bar{E}[\bar{\Lambda}_k O_k^i X_k | \mathcal{Y}_k]$ and $\sigma(T^{ij}X)_k := \bar{E}[\bar{\Lambda}_k T_k^{ij} X_k | \mathcal{Y}_k]$ are as follows:

Theorem 1.3.

$$\sigma(O^i X)_{k+1} = B(Y_{k+1})A\sigma(O^i X)_k + \langle q_k, e_i\rangle B(Y_{k+1})Ae_i,$$

$$\sigma(T^{ij} X)_{k+1} = B(Y_{k+1})A\sigma(T^{ij} X)_k + Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle e_i.$$

Proof. Appendix I.

Note that $\sigma(O^i)_k = \langle \sigma(O^i X)_k, \mathbf{1}\rangle$ and $\sigma(T^{ij})_k = \langle \sigma(T^{ij} X)_k, \mathbf{1}\rangle$.

Remark 1.2. Define $O1_k^i := \sum_{j=1}^{M} T_k^{ij} = O_{k+1}^i$. Then,

$$\sigma(O1^i X)_{k+1} = \sigma(O^i X)_{k+1} + (M \sum_{s=1}^{M} c_{si} \langle Y_k, f_s \rangle) \langle Aq_{k-1}, e_i \rangle e_i$$

$$= B(Y_{k+1}) A \sigma(O^i X)_k + \langle Aq_k, e_i \rangle B(Y_{k+1}) Ae_i + (M \sum_{s=1}^{M} c_{si} \langle Y_k, f_s \rangle) \langle Aq_{k-1}, e_i \rangle e_i$$

and

$$\sigma(O1^i)_k = \sigma(O^i)_k + (M \sum_{s=1}^{M} c_{si} \langle Y_{k+1}, f_s \rangle) \langle Aq_k, e_i \rangle.$$

Proof. Appendix I.

Our model is determined by parameters $\theta = \{a_{ji}, 1 \leq i, j \leq N; c_{ji}, 1 \leq i \leq N, 1 \leq j \leq M\}$, $a_{ji} \geq 0$, $\sum_{j=1}^{N} a_{ji} = 1$, $c_{ji} \geq 0$, $\sum_{j=1}^{M} c_{ji} = 1\}$. We want to determine a new set of parameters $\hat{\theta} = \{\hat{a}_{ji}, 1 \leq i, j \leq N; \hat{c}_{ji}, 1 \leq i \leq N, 1 \leq j \leq M\}$ given the arrival of new information embedded in the values of the observation process Y. This requires maximum likelihood estimation. We proceed by using the so-called EM (Expectation Maximization) algorithm.

The EM algorithm is a broadly applicable method that provides an iterative procedure for computing Maximum Likelihood Estimators (MLEs) in situations where maximum likelihood estimation would be straightforward if more data were available. The algorithm starts with an initial estimate of the unknown parameter θ and iteratively replaces this estimate with its conditional expectation given the data actually observed. The EM algorithm is typically easily implemented as it required only complete-data computations. It is also numerically stable: each iteration monotonically increases the log-likelihood and, under fairly general conditions, starting from an arbitrary point θ_0 in the parameter space, convergence to a local maximizer is nearly always guaranteed, with the exception of very bad luck in the choice of θ_0 or some pathology in the log-likehood function. The method can also be used in problems that are not of incomplete-data type, but where the EM algorithm reduces the complexity of the maximum likelihood computation, which is the case here. The application of the algorithm to estimating parameters of our Hidden Markov Model is described next.

Suppose $\{P_\theta, \theta \in \Theta\}$ is a family of probability measures on a measurable space $(\Omega, \mathcal{F})$. Suppose also that there is another σ-field $\mathcal{Y} \subset \mathcal{F}$. The likelihood function for computing an estimate of θ based on information given in $\mathcal{Y}$ is

$$L(\theta) = E_0[log \frac{dP_\theta}{dP_0} | \mathcal{Y}].$$

The maximum likelihood estimate (MLE) of θ is then

$$\hat{\theta} \in arg \max_{\theta \in \Theta} L(\theta).$$

However, MLE is hard to compute. The expectation maximization (EM) algorithm provides an alternative approximate method:

Step 1: Set $p = 0$ and choose $\hat{\theta}_0$.

Step 2: (E-step) Set $\theta^* = \hat{\theta}_p$ and compute

$$Q(\theta, \theta^*) = E_{\theta^*}[log \frac{dP_\theta}{dP_{\theta^*}} | \mathcal{Y}].$$

Step 3: (M-step) Find

$$\hat{\theta}_{p+1} \in arg \max_{\theta \in \Theta} Q(\theta, \theta^*).$$

Step 4: Replace p by $p+1$ and repeat from Step 2 until a stopping criterion is satisfied.

Our model is determined by the parameters

$$\theta = \{a_{ji}, 1 \leq i, j \leq N; c_{ji}, 1 \leq i \leq N, 1 \leq j \leq M\}.$$

Suppose our model is given by such a set of parameters and we wish to derive a new set

$$\hat{\theta} = \{\hat{a}_{ji}, 1 \leq i, j \leq N; \hat{c}_{ji}, 1 \leq i \leq N, 1 \leq j \leq M\}$$

which maximizes the analogs of the Q functions.

Consider first the parameter a_{ji}. Suppose that under measure P_θ, X is a Markov chain with transition matrix $A = (a_{ji})$. We define a new probability measure $P_{\hat{\theta}}$ such that under $P_{\hat{\theta}}$, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{ji})$, i.e.

$$P_{\hat{\theta}}(X_{k+1} = e_j | X_k = e_i) = \hat{a}_{ji},$$

$\hat{a}_{ji} \geq 0$, $\sum_{j=1}^{N} \hat{a}_{ji} = 1$. Define

$$\begin{aligned}\Lambda_0 &= 1 \\ \Lambda_k &= \prod_{l=1}^{k} \Big(\sum_{r,s=1}^{N} \big(\frac{\hat{a}_{sr}}{a_{sr}}\big) \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle \Big)\end{aligned}$$

In case $a_{ji} = 0$ take $\hat{a}_{ji} = 0$ and $\frac{\hat{a}_{sr}}{a_{sr}} = 1$. Define $P_{\hat{\theta}}$ by setting $\frac{dP_{\hat{\theta}}}{dP_\theta}|\mathcal{F}_k = \Lambda_k$.

Lemma 1.7. *Under $P_{\hat{\theta}}$, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{ji})$.*

Proof. Appendix I.

Lemma 1.8. *Given the observations up to time k, $\{Y_0, Y_1, \dots, Y_k\}$, and given the parameter set $\theta = \{a_{ji}, 1 \le i, j \le N; c_{ji}, 1 \le i \le N, 1 \le j \le M\}$, the EM estimates $\hat{a}_{ji}$ are given by*

$$\hat{a}_{ji} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}.$$

Consider now the parameter c_{ji}. Suppose that under measure P_θ, $Y_k = CX_k + W_k$, where $C = (c_{ji})$. We define a new probability measure $P_{\hat{\theta}}$ as follows. Put

$$\Lambda_0 = 1$$
$$\Lambda_k = \prod_{l=1}^{k} \Big(\sum_{r,s=1}^{N} \Big(\frac{\hat{c}_{sr}}{c_{sr}}\Big) \langle X_l, e_r \rangle \langle Y_l, f_s \rangle \Big).$$

In case $c_{ji} = 0$ take $\hat{c}_{ji} = 0$ and $\frac{\hat{c}_{sr}}{c_{sr}} = 1$. Define $P_{\hat{\theta}}$ by setting $\frac{dP_{\hat{\theta}}}{dP_\theta}|\mathcal{G}_k = \Lambda_k$.

Lemma 1.9. *Under $P_{\hat{\theta}}$, $Y_k = \hat{C}X_k + \hat{W}_k$, i.e. $P_{\hat{\theta}}(Y_k = f_s | X_k = e_r) = \hat{c}_{sr}$.*

Proof. Appendix I.

Lemma 1.10. *Given the observations up to time k, $\{Y_0, Y_1, \dots, Y_k\}$, and given the parameter set $\theta = \{a_{ji}, 1 \le i, j \le N; c_{ji}, 1 \le i \le N, 1 \le j \le M\}$, the EM estimates $\hat{c}_{ji}$ are given by*

$$\hat{c}_{ji} = \frac{\sigma(T^{ij})_k}{\sigma(O^i)_k + (M \sum_{s=1}^{M} c_{si} \langle Y_k, f_s \rangle) \langle A q_{k-1}, e_i \rangle}.$$

Proof. Appendix I.

6. SMOOTHED ESTIMATES

Suppose $0 \leq k \leq T$ and we are given the information $\mathcal{Y}_{0,T} = \sigma\{Y_0, Y_1, \dots, Y_T\}$. We wish to estimate X_k given $\mathcal{Y}_{0,T}$. From Bayes' Theorem,

$$E[X_k|\mathcal{Y}_{0,T}] = \frac{\bar{E}[\bar{\Lambda}_{0,T}X_k|\mathcal{Y}_{0,T}]}{\bar{E}[\bar{\Lambda}_{0,T}|\mathcal{Y}_{0,T}]},$$

where $\bar{\Lambda}_{0,T} = \prod_{k=0}^{T} \bar{\lambda}_k$, $\bar{\lambda}_k = M\sum_{j=1}^{M}\langle CX_k, f_j\rangle\langle Y_k, f_j\rangle$. From Lemma 1.6, the denominator is

$$\bar{E}[\bar{\Lambda}_{0,T}|\mathcal{Y}_{0,T}] = \langle q_T, \mathbf{1}\rangle,$$

where $q_T = \bar{E}[\bar{\Lambda}_{0,T}X_T|\mathcal{Y}_{0,T}]$ and $\mathbf{1} = (1, 1, \dots, 1)' \in \mathbb{R}^N$. Now,

$$\bar{E}[\bar{\Lambda}_{0,T}X_k|\mathcal{Y}_{0,T}] = \bar{E}[\bar{\Lambda}_{0,k}\bar{\Lambda}_{k+1,T}X_k|\mathcal{Y}_{0,T}] = \bar{E}[\bar{\Lambda}_{0,k}X_k\bar{E}[\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T} \vee \mathcal{F}_k]|\mathcal{Y}_{0,T}],$$

where $\bar{\Lambda}_{k+1,T} = \prod_{l=k+1}^{T} \bar{\lambda}_l$. Consider $\bar{E}[\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T} \vee \mathcal{F}_k] = \bar{E}[\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T} \vee X_k]$ using the Markov property. Write $v_k = (v_k^1, \dots, v_k^N)'$, where $v_k^i := \bar{E}[\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]$.

Lemma 1.11. *v satisfies the backwards dynamics, dual to q, of the form*

$$v_k = A'B(Y_{k+1})v_{k+1}.$$

Proof. Appendix I.

Lemma 1.12. $v_T = (1, \dots, 1)' \in \mathbb{R}^N$.

Proof. Appendix I.

Remark 1.3. Since $v_T = \mathbf{1}$, we have $v_k = A'B(Y_{k+1})A'B(Y_{k+2})\cdots A'B(Y_T)\mathbf{1}$.

Theorem 1.4. *The unnormalized smoothed estimate is*

$$\bar{E}[\bar{\Lambda}_{0,T} X_k | \mathcal{Y}_{0,T}] = \text{diag}\,(q_k \cdot v_k').$$

Proof. Appendix I.

It follows that $E[X_k|\mathcal{Y}_{0,T}] = \frac{\text{diag}\,(q_k \cdot v_k')}{\langle q_T, 1 \rangle}$. Hence, to estimate $E[X_k|\mathcal{Y}_{0,T}]$ we need only know the dynamics of q and v, which are, respectively:

$$q_k = B(Y_k)AB(Y_{k-1})A \cdots B(Y_0)Aq_0,$$

where q_0 is the initial distribution for X_0, and

$$v_k = A'B(Y_{k+1})A'B(Y_{k+2}) \cdots A'B(Y_T) \cdot \mathbf{1}.$$

As described in detail in Section 4, the EM algorithm re-estimates the parameters of the model as

$$\hat{a}_{ji} = \frac{\sigma(J^{ij})_T}{\sigma(O^i)_T}$$

$$\hat{c}_{ji} = \frac{\sigma(T^{ij})_T}{\sigma(O1^i)_T} = \frac{\sigma(T^{ij})_T}{\sigma(O^i)_T + (M\sum_{s=1}^{M} c_{si}\langle Y_T, f_s \rangle)\langle Aq_{T-1}, e_i \rangle}.$$

Given observations $\mathcal{Y}_{0,T} = \sigma\{Y_0, Y_1, \ldots, Y_T\}$, we are interested in smoothed estimates of the number of jumps, the occupation time and the time spent. These processes were described in Section 4.

Consider first the smoothed estimate $E[J_k^{ij} X_k|\mathcal{Y}_{0,T}]$. Using Bayes' theorem,

$$E[J_k^{ij} X_k|\mathcal{Y}_{0,T}] = \frac{\bar{E}[\bar{\Lambda}_{0,T} J_k^{ij} X_k|\mathcal{Y}_{0,T}]}{\bar{E}[\bar{\Lambda}_{0,T}|\mathcal{Y}_{0,T}]}.$$

The numerator is $\bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}X_k\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T}]$. Consider the l-th component:

$$\begin{aligned}
&\bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}X_k\bar{\Lambda}_{k+1,T}\langle X_k, e_l\rangle|\mathcal{Y}_{0,T}] \\
&= \bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}X_k\bar{E}[\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T} \vee \{X_k = e_l\}]\langle X_k, e_l\rangle|\mathcal{Y}_{0,T}] \\
&= \bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}X_k v_k^l\langle X_k, e_l\rangle|\mathcal{Y}_{0,T}] \\
&= \bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}X_k\langle X_k, e_l\rangle|\mathcal{Y}_{0,T}]v_k^l.
\end{aligned}$$

Then,

$$\begin{aligned}
\bar{E}[\bar{\Lambda}_{0,T}J_k^{ij}X_k|\mathcal{Y}_{0,T}] &= \sum_{l=1}^{N} \bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}\langle X_k, e_l\rangle e_l|\mathcal{Y}_{0,T}]v_k^l \\
&= \sum_{l=1}^{N} \bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}\langle X_k, e_l\rangle|\mathcal{Y}_{0,T}]v_k^l e_l \\
&= \sum_{l=1}^{N}\langle\bar{E}[\bar{\Lambda}_{0,k}J_k^{ij}X_k|\mathcal{Y}_{0,T}], e_l\rangle v_k^l e_l.
\end{aligned}$$

Recall $\sigma(J^{ij}X)_k = \bar{E}[\bar{\Lambda}_k J_k^{ij}X_k|\mathcal{Y}_k]$. We then have that

$$\begin{aligned}
\bar{E}[\bar{\Lambda}_{0,T}J_k^{ij}X_k|\mathcal{Y}_{0,T}] &= \sum_{l=1}^{N}\langle\sigma(J^{ij}X)_k, e_l\rangle v_k^l e_l \\
&= \sum_{l=1}^{N}\sigma(J^{ij}X)_k^l v_k^l e_l \\
&= \operatorname{diag}(\sigma(J^{ij}X)_k \cdot v_k^l).
\end{aligned}$$

Therefore, $\mathbf{1}'\operatorname{diag}(\sigma(J^{ij}X)_k \cdot v_k^l) = \langle\sigma(J^{ij}X)_k, v_k\rangle = \bar{E}[\bar{\Lambda}_{0,T}J_k^{ij}|\mathcal{Y}_{0,T}]$ is the unnormalized, smoothed estimate of J_k^{ij} given $\mathcal{Y}_{0,T}$.

Given observations $\mathcal{Y}_{0,T} = \sigma\{Y_0, Y_1, \ldots, Y_T\}$, we are interested in $\sigma(J^{ij})_T$.

Theorem 1.5.

$$\sigma(J^{ij})_T = a_{ji}\sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle\langle v_k, e_j\rangle(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle).$$

Proof. Appendix I.

Corollary 1.1.

$$\sigma(O^i)_T = \sum_{k=1}^{T} \langle q_{k-1}, e_i \rangle \langle v_{k-1}, e_i \rangle$$

Proof. Appendix I.

Remark 1.4. Again by Bayes' Theorem,

$$E[T_k^{ij} X_k | \mathcal{Y}_{0,T}] = \frac{\bar{E}[\bar{\Lambda}_{0,T} T_k^{ij} X_k | \mathcal{Y}_{0,T}]}{\bar{E}[\bar{\Lambda}_{0,T} | \mathcal{Y}_{0,T}]}.$$

As before, $\mathbf{1}' \operatorname{diag}\left((T^{ij}X)_k \cdot v_k'\right) = \bar{E}[\bar{\Lambda}_{0,T} T_k^{ij} | \mathcal{Y}_{0,T}] = \langle \sigma(T^{ij}X)_k, v_k \rangle$.

Theorem 1.6.

$$\sigma(T^{ij})_T = \sum_{k=1}^{T} M c_{ji} \langle Y_k, f_j \rangle \langle A q_{k-1}, e_i \rangle \langle v_k, e_i \rangle.$$

Proof. Appendix I.

Corollary 1.2.

$$\sigma(O1^i)_T = \sum_{k=1}^{T} (M \sum_{k=1}^{M} c_{si} \langle Y_k, f_s \rangle) \langle v_k, e_i \rangle \langle A q_{k-1}, e_i \rangle.$$

Proof. Appendix I.

7. UPDATING SMOOTHED ESTIMATES

Write $V_{k+1,T} = A'B(Y_{k+1}) \cdots A'B(Y_T)$ so that

$$v_k = v_{k,T}, \text{ where}$$

$$v_{k,T} = V_{k+1,T} \cdot \mathbf{1}.$$

The methods to update smoothed estimates from the previous section have required recalculation of all backward estimates v. Below we note results that provide for more efficient computations.

Lemma 1.13. $v_{k,T+1} = V_{k+1,T+1}\mathbf{1}$, *where* $V_{k+1,T+1} = V_{k+1,T}A'B(Y_{T+1})$ □.

From Theorem 1.5,

$$\begin{aligned}
\sigma(J^{ij})_T &= a_{ji}\sum_{k=1}^{T}(M\sum_{s=1}^{M}c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle\langle v_k, e_j\rangle \\
&= a_{ji}\sum_{k=1}^{T}(M\sum_{s=1}^{M}c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle e_j' v_k \\
&= a_{ji}\sum_{k=1}^{T}(M\sum_{s=1}^{M}c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle e_j' A'B(Y_{k+1})\cdots A'B(Y_T)\mathbf{1} \\
&= \Gamma_T'\mathbf{1},
\end{aligned}$$

where $\Gamma_T' = a_{ji}\sum_{k=1}^{T}(M\sum_{s=1}^{M}c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle e_j' A'B(Y_{k+1})\cdots A'B(Y_T)$.

Lemma 1.14.

$$\Gamma_{T+1}' = \Gamma_T' A'B(Y_{T+1}) + a_{ji}(M\sum_{s=1}^{M}c_{sj}\langle Y_{T+1}, f_s\rangle)\langle q_T, e_i\rangle e_j'.$$

Proof. Appendix I.

Corollary 1.3. $\sigma(O^i)_T = K_T'\mathbf{1}$, *where*

$$K_T' = \sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle e_i' A'B(Y_k)\cdots A'B(Y_T).$$

Then,

$$K_{T+1}' = K_T' A'B(Y_{T+1}) + \langle q_T, e_i\rangle e_i'. \quad \square$$

From Theorem 1.6 we have

$$\begin{aligned}\sigma(T^{ij})_T &= \sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle\langle v_k, e_i\rangle \\ &= \sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle e_i' v_k \\ &= \sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle e_i' A'B(Y_{k+1})\cdots A'B(Y_T)\mathbf{1} \\ &= H_T'\mathbf{1},\end{aligned}$$

where $H_T' = \sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle e_i' A'B(Y_{k+1})\cdots A'B(Y_T)$.

Lemma 1.15.

$$H_{T+1}' = H_T' A'B(Y_{T+1})\mathbf{1} + Mc_{ji}\langle Y_{T+1}, f_j\rangle\langle Aq_T, e_i\rangle e_i'.$$

Proof. Appendix I.

Corollary 1.4. *In particular,* $\sigma(O1^i)_T = \Delta_T'\mathbf{1}$, *where*

$$\Delta_T' = \sum_{k=1}^{T}\langle Aq_{k-1}, e_i\rangle(M\sum_{s=1}^{M} c_{si}\langle Y_k, f_s\rangle)e_i' A'B(Y_{k+1})\cdots A'B(Y_T).$$

Then, $\Delta_{T+1}' = \Delta_T' A'B(Y_{T+1}) + \langle Aq_T, e_i\rangle(M\sum_{s=1}^{M}\langle Y_T, f_s\rangle)e_i'$. □

Appendix I
Proofs of Results in Chapter 1

Proof of **Lemma 1.1.**

$$
\begin{aligned}
E[X_{k+1}|X_k] &= \sum_{i=1}^{N} E[X_{k+1}|X_k = e_i]\langle X_k, e_i\rangle \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} E[\langle X_{k+1}, e_j\rangle | X_k = e_i]\langle X_k, e_i\rangle e_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} P(X_{k+1} = e_j | X_k = e_i)\langle X_k, e_i\rangle e_j \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ji}\langle X_k, e_i\rangle e_j = AX_k. \quad \square
\end{aligned}
$$

Proof of **Lemma 1.2.**

$$
\begin{aligned}
Var V_k = E[V_k V_k'] &= E[(X_k - AX_{k-1})(X_k - AX_{k-1})'] \\
&= E[X_k X_k' - AX_{k-1}X_k' - X_k X_{k-1} A' + AX_{k-1}X_{k-1}' A'] \\
&= E[E[X_k X_k' - AX_{k-1}X_k' - X_k X_{k-1} A' + AX_{k-1}X_{k-1}' A' | \mathcal{F}_{k-1}]] \\
&= E[E[\operatorname{diag}(X_k) - A \operatorname{diag}(X_{k-1}) A' | \mathcal{F}_{k-1}]] \\
&= E[\operatorname{diag}(AX_{k-1}) - A \operatorname{diag}(X_{k-1}) A'] \\
&= \operatorname{diag}(Ap_{k-1}) - A \operatorname{diag}(p_{k-1}) A'. \quad \square
\end{aligned}
$$

Proof of **Lemma 1.3.**

$$\begin{aligned} E[Y_k|X_k] &= \sum_{i=1}^{N} E[Y_k|X_k = e_i]\langle X_k, e_i\rangle \\ &= \sum_{i=1}^{N} \Big(\sum_{j=1}^{N} P(Y_k = f_j|X_k = e_i) f_j\Big)\langle X_k, e_i\rangle \\ &= \sum_{i=1}^{N} \Big(\sum_{j=1}^{N} c_{ji} f_j\Big)\langle X_k, e_i\rangle = CX_k. \quad \square \end{aligned}$$

Proof of **Lemma 1.4.**

$$\begin{aligned} VarW_k &= E[W_k W_k'] = E[(Y_k - CX_k)(Y_k - CX_k)'] \\ &= E[Y_k Y_k' - Y_k X_k' C' - C Y_k X_k' + C X_k X_k' C'] \\ &= E[E[Y_k Y_k' - Y_k X_k' C' - C Y_k X_k' + C X_k X_k' C' | X_k]] \\ &= E[E[\operatorname{diag}(Y_k) - C \operatorname{diag}(X_k) C' | X_k]] \\ &= E[\operatorname{diag}(CX_k) - C \operatorname{diag}(X_k) C'] \\ &= \operatorname{diag}(Cp_k) - C \operatorname{diag}(p_k) C'. \quad \square \end{aligned}$$

Proof of **Lemma 1.5.** First we show that $\bar{E}[\bar{\lambda}_k|\mathcal{G}_{k-1} \vee \{X_k\}] = 1$. We will make use of this fact in proofs that follow as well.

$$\begin{aligned} \bar{E}[\bar{\lambda}_k|\mathcal{G}_{k-1} \vee \{X_k\}] &= \bar{E}[M \sum_{j=1}^{M} \langle CX_k, f_j\rangle\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}] \\ &= M \sum_{j=1}^{M} \bar{E}[\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}]\langle CX_k, f_j\rangle \\ &= M \sum_{j=1}^{M} \bar{P}(Y_k = f_j|\mathcal{G}_{k-1} \vee \{X_k\})\langle CX_k, f_j\rangle \end{aligned}$$

$$= M \sum_{j=1}^{M} \bar{P}(Y_k = f_j)\langle CX_k, f_j\rangle$$
$$= \sum_{j=1}^{M} \sum_{i=1}^{N} c_{ji}\langle X_k, f_j\rangle = \sum_{i=1}^{N}\langle X_k, f_j\rangle \sum_{j=1}^{M} c_{ji} = 1.$$

Then,

$$P(X_{k+1} = e_i|\mathcal{G}_k) = E[\langle X_{k+1}, e_i\rangle|\mathcal{G}_k]$$
$$= \frac{\bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_i\rangle|\mathcal{G}_k]}{\bar{E}[\bar{\Lambda}_{k+1}|\mathcal{G}_k]}$$
$$= \frac{\bar{E}[\bar{\lambda}_{k+1}\langle X_{k+1}, e_i\rangle|\mathcal{G}_k]}{\bar{E}[\bar{\lambda}_{k+1}|\mathcal{G}_k]}$$
$$= \bar{E}[\bar{\lambda}_{k+1}\langle X_{k+1}, e_i\rangle|\mathcal{G}_k] \text{ using the claim proved above}$$
$$= \bar{E}[\big(M \sum_{j=1}^{M}\langle CX_{k+1}, f_j\rangle\langle Y_{k+1}, f_j\rangle\big)\langle X_{k+1}, e_i\rangle|\mathcal{G}_k]$$
$$= \bar{E}[\big(M \sum_{i=1}^{N}\big(\sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle\big)\langle X_{k+1}, e_i\rangle\big)\langle X_{k+1}, e_i\rangle|\mathcal{G}_k]$$
$$= \bar{E}[M\big(\sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle\big)\langle AX_k + V_{k+1}, e_i\rangle|\mathcal{G}_k]$$
$$= \bar{E}[M\big(\sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle\big)\langle AX_k, e_i\rangle|\mathcal{G}_k]$$

(since V_{k+1} is a martingale increment independent of $\mathcal{Y}_{k+1}$)

$$= \langle AX_k, e_i\rangle M \sum_{j=1}^{M} c_{ji}\bar{E}[\langle Y_{k+1}, f_j\rangle|\mathcal{G}_k]$$
$$= \langle AX_k, e_i\rangle M \sum_{j=1}^{M} c_{ji}\bar{P}(Y_{k+1} = f_j|\mathcal{G}_k)$$
$$= \langle AX_k, e_i\rangle \sum_{j=1}^{M} c_{ji} = \langle AX_k, e_i\rangle.$$

Hence, $P(X_{k+1} = e_i|\mathcal{G}_k) = \langle AX_k, e_i\rangle$ depends only on X_k. Therefore,

$$P(X_{k+1} = e_i|\mathcal{G}_k) = P(X_{k+1} = e_i|X_k).$$

Suppose now that $X_k = e_j$. Then, $P(X_{k+1} = e_i | X_k) = \langle Ae_j, e_i \rangle = a_{ji}$. It follows that $E[X_{k+1}|X_k] = AX_k$. Define $V_{k+1} = X_{k+1} - AX_k$. Then,

$$E[V_{k+1}|\mathcal{G}_k] = E[X_{k+1} - AX_k|\mathcal{G}_k] = E[X_{k+1}|X_k] - AX_k = 0 \in \mathbb{R}^N.$$

Now, by the tower property,

$$E[V_{k+1}|\mathcal{F}_k] = E\big[E[V_{k+1}|\mathcal{G}_k]|\mathcal{F}_k\big] = 0 \in \mathbb{R}^N.$$

It follows that, under P, X remains a Markov chain with semimartingale representation

$$X_{k+1} = AX_k + V_{k+1}, \quad E[V_{k+1}|\mathcal{F}_k] = 0 \in \mathbb{R}^N.$$

Now,

$$\begin{aligned}
P(Y_k = f_j|\mathcal{G}_{k-1} \vee \{X_k\}) &= E[\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}] \\
&= \frac{\bar{E}[\bar{\Lambda}_{k+1}\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}]}{\bar{E}[\bar{\Lambda}_{k+1}|\mathcal{G}_{k-1} \vee \{X_k\}]} \\
&= \bar{E}[\bar{\lambda}_k\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}] \\
&= \bar{E}[(M\sum_{j=1}^{M}\langle CX_k, f_j\rangle\langle Y_k, f_j\rangle)\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}] \\
&= \bar{E}[M\langle CX_k, f_j\rangle\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}] \\
&= M\langle CX_k, f_j\rangle\bar{E}[\langle Y_k, f_j\rangle|\mathcal{G}_{k-1} \vee \{X_k\}] \\
&= M\langle CX_k, f_j\rangle\bar{P}(Y_k = f_j|\mathcal{G}_{k-1} \vee \{X_k\}) \\
&= \langle CX_k, f_j\rangle.
\end{aligned}$$

Hence, $P(Y_k = f_j|\mathcal{G}_{k-1} \vee \{X_k\}) = \langle CX_k, f_j\rangle$. Since this depends only on X_k, we have

$$P(Y_k = f_j|\mathcal{G}_{k-1} \vee \{X_k\}) = P(Y_k = f_j|X_k).$$

Suppose that $X_k = e_i$. Then,

$$P(Y_k = f_j | X_k = e_i) = \langle Ce_i, f_j \rangle = c_{ji}.$$

It follows that $E[Y_k|X_k] = CX_k$. Define $W_k = Y_k - CX_k$. Then, $E[W_k|\mathcal{G}_{k-1} \vee \{X_k\}] = E[Y_k - CX_k|\mathcal{G}_{k-1} \vee \{X_k\}] = E[Y_k|\mathcal{G}_{k-1} \vee \{X_k\}] - CX_k = 0 \in \mathbb{R}^M$. Consequently,

$$E[W_k|\mathcal{Y}_{k-1}] = E\big[E[W_k|\mathcal{G}_{k-1} \vee \{X_k\}]|\mathcal{Y}_{k-1}\big] = 0 \in \mathbb{R}^M.$$

It follows that Y_k has the required semimartingale representation under P. □

Proof of **Lemma 1.6.**

$$\begin{aligned}\langle q_k, 1\rangle &= \langle \bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_k], 1\rangle = \bar{E}[\langle \bar{\Lambda}_k X_k, 1\rangle|\mathcal{Y}_k] = \bar{E}[\bar{\Lambda}_k \langle X_k, 1\rangle|\mathcal{Y}_k] \\ &= \bar{E}[\bar{\Lambda}_k \sum_{i=1}^{N} \langle X_k, e_i\rangle|\mathcal{Y}_k] = \bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k]. \quad \square\end{aligned}$$

Proof of **Theorem 1.1.**

$$\begin{aligned}q_{k+1} &= \bar{E}[\bar{\Lambda}_{k+1} X_{k+1}|\mathcal{Y}_{k+1}] \\ &= \bar{E}[\bar{\Lambda}_k (M \sum_{j=1}^{M} \langle CX_{k+1}, f_j\rangle \langle Y_{k+1}, f_j\rangle) X_{k+1}|\mathcal{Y}_{k+1}] \\ &= \sum_{i=1}^{N} \bar{E}[\bar{\Lambda}_k (M \sum_{j=1}^{M} c_{ji} \langle Y_{k+1}, f_j\rangle) \langle X_{k+1}, e_i\rangle e_i|\mathcal{Y}_{k+1}]\end{aligned}$$

$$= \sum_{i=1}^{N} \bar{E}[\bar{\Lambda}_k \langle X_{k+1}, e_i\rangle | \mathcal{Y}_{k+1}](M \sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle) e_i$$

$$= \sum_{i=1}^{N} \langle \bar{E}[\bar{\Lambda}_k X_{k+1} | \mathcal{Y}_{k+1}], e_i\rangle (M \sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle) e_i$$

$$= \sum_{i=1}^{N} \langle \bar{E}[\bar{\Lambda}_k (AX_k + V_{k+1}) | \mathcal{Y}_{k+1}], e_i\rangle (M \sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle) e_i$$

$$= \sum_{i=1}^{N} \langle A\bar{E}[\bar{\Lambda}_k X_k | \mathcal{Y}_k], e_i\rangle (M \sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle) e_i$$

(since $V_{k+1}$is a martingale increment independent of $\mathcal{Y}_{k+1}$)

$$= \sum_{i=1}^{N} \langle Aq_k, e_i\rangle (M \sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle) e_i$$

$$= B(Y_{k+1}) A q_k,$$

where $B(Y_{k+1})$ is a diagonal matrix with entries $(M \sum_{j=1}^{M} c_{ji}\langle Y_{k+1}, f_j\rangle)$. □

Proof of **Theorem 1.2.**

$$\sigma(J^{ij}X)_{k+1} = \bar{E}[\bar{\Lambda}_{k+1} J^{ij}_{k+1} X_{k+1} | \mathcal{Y}_{k+1}]$$

$$= \bar{E}[\bar{\Lambda}_k \bar{\lambda}_{k+1} (J^{ij}_k + \langle X_k, e_i\rangle \langle X_{k+1}, e_j\rangle) X_{k+1} | \mathcal{Y}_{k+1}]$$

$$= \bar{E}[\bar{\Lambda}_k \bar{\lambda}_{k+1} J^{ij}_k X_{k+1} | \mathcal{Y}_{k+1}] + \bar{E}[\bar{\Lambda}_k \bar{\lambda}_{k+1} \langle X_k, e_i\rangle \langle X_{k+1}, e_j\rangle X_{k+1} | \mathcal{Y}_{k+1}].$$

Now,

$$\bar{E}[\bar{\Lambda}_k \bar{\lambda}_{k+1} J^{ij}_k X_{k+1} | \mathcal{Y}_{k+1}] = \bar{E}[\bar{\Lambda}_k (M \sum_{s=1}^{M} \langle CX_{k+1}, f_s\rangle \langle Y_{k+1}, f_s\rangle) J^{ij}_k X_{k+1} | \mathcal{Y}_{k+1}]$$

$$= \sum_{r=1}^{N} \bar{E}[\bar{\Lambda}_k (M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) \langle X_{k+1}, e_r\rangle J^{ij}_k e_r | \mathcal{Y}_{k+1}]$$

$$= \sum_{r=1}^{N} \bar{E}[\bar{\Lambda}_k \langle X_{k+1}, e_r\rangle J^{ij}_k | \mathcal{Y}_{k+1}](M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) e_r$$

$$= \sum_{r=1}^{N} \bar{E}[\bar{\Lambda}_k \langle AX_k, e_r\rangle J_k^{ij} | \mathcal{Y}_{k+1}](M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) e_r$$
$$+ \sum_{r=1}^{N} \bar{E}[\bar{\Lambda}_k \langle V_{k+1}, e_r\rangle J_k^{ij} | \mathcal{Y}_{k+1}](M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) e_r$$
$$= \sum_{r=1}^{N} \langle \bar{E}[\bar{\Lambda}_k J_k^{ij} AX_k | \mathcal{Y}_{k+1}], e_r\rangle (M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) e_r$$
$$= \sum_{r=1}^{N} \langle A\bar{E}[\bar{\Lambda}_k J_k^{ij} X_k | \mathcal{Y}_k], e_r\rangle (M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) e_r$$
$$= \sum_{r=1}^{N} \langle A\sigma(J^{ij}X)_k, e_r\rangle (M \sum_{s=1}^{M} c_{sr}\langle Y_{k+1}, f_s\rangle) e_r$$
$$= B(Y_{k+1}) A\sigma(J^{ij}X)_k.$$

Also,

$$\bar{E}[\bar{\Lambda}_k \bar{\lambda}_{k+1} \langle X_k, e_i\rangle \langle X_{k+1}, e_j\rangle X_{k+1} | \mathcal{Y}_{k+1}] =$$
$$= \bar{E}[\bar{\Lambda}_k (M \sum_{s=1}^{M} \langle CX_{k+1}, f_s\rangle \langle Y_{k+1}, f_s\rangle) \langle X_k, e_i\rangle \langle X_{k+1}, e_j\rangle X_{k+1} | \mathcal{Y}_{k+1}]$$
$$= \bar{E}[\bar{\Lambda}_k (M \sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle) \langle X_k, e_i\rangle \langle X_{k+1}, e_j\rangle e_j | \mathcal{Y}_{k+1}]$$
$$= \bar{E}[\bar{\Lambda}_k (M \sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle) \langle X_k, e_i\rangle \langle AX_k, e_j\rangle e_j | \mathcal{Y}_{k+1}]$$
$$+ \bar{E}[\bar{\Lambda}_k (M \sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle) \langle X_k, e_i\rangle \langle V_{k+1}, e_j\rangle e_j | \mathcal{Y}_{k+1}]$$
$$= (M \sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle) \bar{E}[\bar{\Lambda}_k \langle X_k, e_i\rangle \langle AX_k, e_j\rangle | \mathcal{Y}_{k+1}] e_j$$
$$+ (M \sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle) \bar{E}[\bar{\Lambda}_k \langle X_k, e_i\rangle \langle V_{k+1}, e_j\rangle | \mathcal{Y}_{k+1}] e_j$$
$$= (M \sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle) \bar{E}[\bar{\Lambda}_k \langle X_k, e_i\rangle a_{ji} | \mathcal{Y}_{k+1}] e_j$$

$$= (M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle \bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_{k+1}], e_i\rangle a_{ji}e_j$$
$$= (M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle \bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_{k}], e_i\rangle a_{ji}e_j$$
$$= (M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle a_{ji}e_j.$$

Therefore,

$$\sigma(J^{ij}X)_{k+1} = B(Y_{k+1})A\sigma(J^{ij}X)_k + (M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle a_{ji}e_j,$$

as required. □

Proof of **Theorem 1.3.**

$$\sigma(O^iX)_{k+1} = \bar{E}[\bar{\Lambda}_{k+1}O^i_{k+1}X_{k+1}|\mathcal{Y}_{k+1}]$$
$$= \bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}(O^i_k + \langle X_k, e_i\rangle)X_{k+1}|\mathcal{Y}_{k+1}]$$
$$= \bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}O^i_kX_{k+1}|\mathcal{Y}_{k+1}] + \bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}\langle X_k, e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}].$$

Now,

$$\bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}O^i_kX_{k+1}|\mathcal{Y}_{k+1}] = \bar{E}[\bar{\Lambda}_k(M\sum_{j=1}^{M}\langle CX_{k+1}, f_j\rangle\langle Y_{k+1}, f_j\rangle)O^i_kX_{k+1}|\mathcal{Y}_{k+1}]$$
$$= \sum_{r=1}^{N}\bar{E}[\bar{\Lambda}_k(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1}, f_j\rangle)\langle X_{k+1}, e_r\rangle O^i_k e_r|\mathcal{Y}_{k+1}]e_r$$
$$= \sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1}, f_j\rangle)\bar{E}[\bar{\Lambda}_k\langle X_{k+1}, e_r\rangle O^i_k|\mathcal{Y}_{k+1}]e_r$$
$$= \sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1}, f_j\rangle)\langle\bar{E}[\bar{\Lambda}_k AX_kO^i_k|\mathcal{Y}_{k+1}], e_r\rangle e_r$$

$$+\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\langle\bar{E}[\bar{\Lambda}_k O_k^i V_{k+1}|\mathcal{Y}_{k+1}],e_r\rangle e_r$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\langle A\bar{E}[\bar{\Lambda}_k O_k^i X_k|\mathcal{Y}_k],e_r\rangle e_r$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\langle A\sigma(O^iX)_k,e_r\rangle e_r$$
$$=B(Y_{k+1})A\sigma(O^iX)_k.$$

Also,

$$\bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}\langle X_k,e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}]$$
$$=\bar{E}[\bar{\Lambda}_k(M\sum_{j=1}^{M}\langle CX_{k+1},f_j\rangle\langle Y_{k+1},f_j\rangle)\langle X_k,e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}]$$
$$=\sum_{r=1}^{N}\bar{E}[\bar{\Lambda}_k(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\langle X_{k+1},e_r\rangle\langle X_k,e_i\rangle e_r|\mathcal{Y}_{k+1}]$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\bar{E}[\bar{\Lambda}_k\langle X_{k+1},e_r\rangle\langle X_k,e_i\rangle|\mathcal{Y}_{k+1}]e_r$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\bar{E}[\bar{\Lambda}_k\langle AX_k,e_r\rangle\langle X_k,e_i\rangle|\mathcal{Y}_{k+1}]e_r$$
$$+\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\bar{E}[\bar{\Lambda}_k\langle V_{k+1},e_r\rangle\langle X_k,e_i\rangle|\mathcal{Y}_{k+1}]e_r$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\bar{E}[\bar{\Lambda}_k a_{ri}\langle X_k,e_i\rangle|\mathcal{Y}_{k+1}]e_r$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\bar{E}[\bar{\Lambda}_k\langle X_k,e_i\rangle|\mathcal{Y}_{k+1}]a_{ri}e_r$$
$$=\sum_{r=1}^{N}(M\sum_{j=1}^{M}c_{jr}\langle Y_{k+1},f_j\rangle)\langle q_k,e_i\rangle a_{ri}e_r$$
$$=\langle q_k,e_i\rangle B(Y_{k+1})Ae_i.$$

We follow the same procedure to obtain the recursion for the dynamics of the vector process $\sigma(T^{ij}X)_k$:

$$\begin{aligned}\sigma(T^{ij}X)_{k+1} &= \bar{E}[\bar{\Lambda}_{k+1}T^{ij}_{k+1}X_{k+1}|\mathcal{Y}_{k+1}] \\ &= \bar{E}[\bar{\Lambda}_{k+1}(T^{ij}_k + \langle X_{k+1}, e_i\rangle\langle Y_{k+1}, f_j\rangle)X_{k+1}|\mathcal{Y}_{k+1}] \\ &= \bar{E}[\bar{\Lambda}_{k+1}T^{ij}_k X_{k+1}|\mathcal{Y}_{k+1}] + \bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_i\rangle\langle Y_{k+1}, f_j\rangle X_{k+1}|\mathcal{Y}_{k+1}].\end{aligned}$$

Now,

$$\begin{aligned}\bar{E}[\bar{\Lambda}_{k+1}T^{ij}_k X_{k+1}|\mathcal{Y}_{k+1}] &= \bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}T^{ij}_k X_{k+1}|\mathcal{Y}_{k+1}] \\ &= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)T^{ij}_k X_{k+1}|\mathcal{Y}_{k+1}] \\ &= \sum_{r=1}^{N}(M\sum_{s=1}^{M}c_{sr}\langle Y_{k+1}, f_s\rangle)\bar{E}[\bar{\Lambda}_k T^{ij}_k\langle X_{k+1}, e_r\rangle|\mathcal{Y}_{k+1}]e_r \\ &= \sum_{r=1}^{N}(M\sum_{s=1}^{M}c_{sr}\langle Y_{k+1}, f_s\rangle)\bar{E}[\Lambda_k T^{ij}_k\langle AX_k, e_r\rangle|\mathcal{Y}_{k+1}]e_r \\ &+ \sum_{r=1}^{N}(M\sum_{s=1}^{M}c_{sr}\langle Y_{k+1}, f_s\rangle)\bar{E}[\bar{\Lambda}_k T^{ij}_k\langle V_{k+1}, e_r\rangle|\mathcal{Y}_{k+1}]e_r \\ &= \sum_{r=1}^{N}(M\sum_{s=1}^{M}c_{sr}\langle Y_{k+1}, f_s\rangle)\langle A\bar{E}[\bar{\Lambda}_k T^{ij}_k X_k|\mathcal{Y}_{k+1}], e_r\rangle e_r \\ &= \sum_{r=1}^{N}(M\sum_{s=1}^{M}c_{sr}\langle Y_{k+1}, f_s\rangle)\langle A\sigma(T^{ij}X)_k, e_r\rangle e_r \\ &= B(Y_{k+1})A\sigma(T^{ij}X)_k.\end{aligned}$$

Also,

$$\begin{aligned}&\bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_i\rangle\langle Y_{k+1}, f_j\rangle X_{k+1}|\mathcal{Y}_{k+1}] = \\ &= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle X_{k+1}, e_i\rangle\langle Y_{k+1}, f_j\rangle X_{k+1}|\mathcal{Y}_{k+1}]\end{aligned}$$

$$= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle Ce_i, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle X_{k+1}, e_i\rangle\langle Y_{k+1}, f_j\rangle e_i|\mathcal{Y}_{k+1}]$$

$$= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}c_{si}\langle Y_{k+1}, f_s\rangle)\langle X_{k+1}, e_i\rangle\langle Y_{k+1}, f_j\rangle|\mathcal{Y}_{k+1}]e_i$$

$$= Mc_{ji}\langle Y_{k+1}, f_j\rangle\bar{E}[\bar{\Lambda}_k\langle X_{k+1}, e_i\rangle|\mathcal{Y}_{k+1}]e_i$$

$$= Mc_{ji}\langle Y_{k+1}, f_j\rangle\bar{E}[\bar{\Lambda}_k\langle AX_k, e_i\rangle|\mathcal{Y}_{k+1}]e_i + Mc_{ji}\langle Y_{k+1}, f_j\rangle\bar{E}[\bar{\Lambda}_k\langle V_{k+1}, e_i\rangle|\mathcal{Y}_{k+1}]e_i$$

$$= Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle A\bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_{k+1}], e_i\rangle e_i$$

$$= Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle e_i.$$

The result follows. □

Proof of **Remark 1.2.**

$$\begin{aligned}
\sigma(O1^iX)_{k+1} &= \bar{E}[\bar{\Lambda}_{k+1}O1^i_{k+1}X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}[\bar{\Lambda}_{k+1}O^i_{k+2}X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}[\bar{\Lambda}_{k+1}(O^i_{k+1} + \langle X_{k+1}, e_i\rangle)X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}[\bar{\Lambda}_{k+1}O^i_{k+1}X_{k+1}|\mathcal{Y}_{k+1}] \\
&+ \bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \sigma(O^iX)_{k+1} + \bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}].
\end{aligned}$$

Now,

$$\bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}] =$$

$$= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle X_{k+1}, e_i\rangle X_{k+1}|\mathcal{Y}_{k+1}]$$

$$= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle X_{k+1}, e_i\rangle e_i|\mathcal{Y}_{k+1}]$$

$$= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle AX_k + V_{k+1}, e_i\rangle e_i|\mathcal{Y}_{k+1}]$$

$$= \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle AX_k, e_i\rangle e_i|\mathcal{Y}_{k+1}]$$

$$+ \bar{E}[\bar{\Lambda}_k(M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle V_{k+1}, e_i\rangle e_i|\mathcal{Y}_{k+1}]$$

$$= (M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\bar{E}[\bar{\Lambda}_k\langle AX_k, e_i\rangle e_i|\mathcal{Y}_{k+1}]$$

$$= (M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle A\bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_{k+1}], e_i\rangle e_i$$

$$= (M\sum_{s=1}^{M}\langle CX_{k+1}, f_s\rangle\langle Y_{k+1}, f_s\rangle)\langle Aq_k, e_i\rangle e_i.$$

The result follows. □

Proof of **Lemma 1.7.** Using Bayes' Theorem, we have

$$P_{\hat{\theta}}(X_{k+1} = e_j|\mathcal{F}_k) = E_{\hat{\theta}}[\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]$$

$$= \frac{E_{\theta}[\Lambda_{k+1}\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]}{E_{\theta}[\Lambda_{k+1}|\mathcal{F}_k]}$$

$$= \frac{E_{\theta}[\lambda_{k+1}\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]}{E_{\theta}[\lambda_{k+1}|\mathcal{F}_k]}$$

$$= \frac{E_{\theta}[\big(\sum_{i,j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle\langle X_k, e_i\rangle\big)\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]}{E_{\theta}[\sum_{i,j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle\langle X_k, e_i\rangle|\mathcal{F}_k]}.$$

Now, the denominator is

$$E_{\theta}[\sum_{i,j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle\langle X_k, e_i\rangle|\mathcal{F}_k] =$$

$$= \sum_{i=1}^{N}E_{\theta}[\sum_{j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N} E_\theta[\sum_{j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle | X_k = e_i]\langle X_k, e_i\rangle \quad \text{(by the Markov property)}$$

$$= \sum_{i=1}^{N}\big(\sum_{j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})E_\theta[\langle X_{k+1}, e_j\rangle | X_k = e_i]\big)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\big(\sum_{j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})P_\theta(\langle X_{k+1}, e_j\rangle | X_k = e_i)\big)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\big(\sum_{j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})a_{ji}\big)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\big(\sum_{j=1}^{N}\hat{a}_{ji}\big)\langle X_k, e_i\rangle = \sum_{i=1}^{N}\langle X_k, e_i\rangle = 1.$$

The numerator is

$$E_\theta[(\sum_{i,j=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle\langle X_k, e_i\rangle)\langle X_{k+1}, e_j\rangle | \mathcal{F}_k] =$$

$$= E_\theta[\sum_{i=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})\langle X_{k+1}, e_j\rangle\langle X_k, e_i\rangle | \mathcal{F}_k]$$

$$= \sum_{i=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})E_\theta[\langle X_{k+1}, e_j\rangle | \mathcal{F}_k]\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}(\frac{\hat{a}_{ji}}{a_{ji}})P_\theta(X_{k+1} = e_j | \mathcal{F}_k)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\hat{a}_{ji}\langle X_k, e_i\rangle = \langle \hat{A}X_k, e_j\rangle.$$

This means that

$$P_{\hat{\theta}}(X_{k+1} = e_j | \mathcal{F}_k) = P_{\hat{\theta}}(X_{k+1} = e_j | X_k = e_i) = \hat{a}_{ji}.$$

It follows that under $P_{\hat{\theta}}$, X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{ji})$.

□

Proof of **Lemma 1.8.**

$$\frac{dP_{\hat{\theta}}}{dP_\theta}|\mathcal{F}_k = \prod_{l=1}^{k} \Big(\sum_{r,s=1}^{N} \big(\frac{\hat{a}_{sr}}{a_{sr}}\big)\langle X_l, e_s\rangle\langle X_{l-1}, e_r\rangle\Big)$$

so

$$\begin{aligned}
\log \frac{dP_{\hat{\theta}}}{dP_\theta} &= \sum_{l=1}^{k} \log \Big(\sum_{r,s=1}^{N} \big(\frac{\hat{a}_{sr}}{a_{sr}}\big)\langle X_l, e_s\rangle\langle X_{l-1}, e_r\rangle\Big) \\
&= \sum_{l=1}^{k}\sum_{r,s=1}^{N} \langle X_l, e_s\rangle\langle X_{l-1}, e_r\rangle(\log \hat{a}_{sr} - \log a_{sr}) \\
&= \sum_{r,s=1}^{N} \Big(\sum_{l=1}^{k}\langle X_l, e_s\rangle\langle X_{l-1}, e_r\rangle\Big)(\log \hat{a}_{sr} - \log a_{sr}) \\
&= \sum_{r,s=1}^{N} J_k^{rs} \log \hat{a}_{sr} + R(a),
\end{aligned}$$

where $R(a)$ is independent of the $\hat{a}_{sr}$.

Therefore,

$$\begin{aligned}
L(\hat{\theta}) &= E_\theta[\log \frac{dP_{\hat{\theta}}}{dP_\theta}|\mathcal{Y}_k] \\
&= E_\theta[\sum_{r,s=1}^{N} J_k^{rs} \log \hat{a}_{sr} + R(a)|\mathcal{Y}_k] \\
&= \sum_{r,s=1}^{N} E_\theta[J_k^{rs} \log \hat{a}_{sr}|\mathcal{Y}_k] + E_\theta[R(a)|\mathcal{Y}_k] \\
&= \sum_{r,s=1}^{N} \log \hat{a}_{sr} E_\theta[J_k^{rs}|\mathcal{Y}_k] + E_\theta[R(a)|\mathcal{Y}_k] \\
&= \sum_{r,s=1}^{N} \log \hat{a}_{sr} \hat{J}_k^{rs} + \hat{R}(a).
\end{aligned}$$

The optimal estimate of $\hat{a}_{ji}$ is, therefore, the value which solves the following maximization problem:

$$\max_{\hat{a}_{ji}} \sum_{r,s=1}^{N} \log \hat{a}_{sr} \hat{J}_k^{rs} + \hat{R}(a)$$

subject to

$$\sum_{s=1}^{N} \hat{a}_{sr} = 1.$$

The Lagrangian is

$$\mathcal{L} = \sum_{r,s=1}^{N} \hat{J}_k^{rs} \log \hat{a}_{sr} + \hat{R}(a) + \lambda\Big(\sum_{s=1}^{N} \hat{a}_{sr} - 1\Big).$$

Differentiating in $\hat{a}_{ji}$ and λ and equating the derivatives to 0 gives:

$$\frac{1}{\hat{a}_{ji}} \hat{J}_k^{ij} + \lambda = 0$$

$$\sum_{s=1}^{N} \hat{a}_{sr} = 1.$$

Recall $\sum_{s=1}^{N} \hat{a}_{sr} = 1$ and note $\sum_{s=1}^{N} J_k^{rs} = O_k^r$ so that $\sum_{s=1}^{N} \hat{J}_k^{rs} = \hat{O}_k^r$. We then have $\lambda = -\hat{O}_k^r$ so that $\hat{a}_{ji} = \frac{\hat{J}_k^{ij}}{\hat{O}_k^r}$. Now, since $\hat{J}_k^{ij} = \frac{\sigma(J^{ij})_k}{\langle q_k, 1\rangle}$, and $\hat{O}_k^i = \frac{\sigma(O^i)_k}{\langle q_k, 1\rangle}$, we have

$$\hat{a}_{ji} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}. \quad \Box$$

Proof of **Lemma 1.9.**

Using Bayes' Theorem, we have

$$\begin{aligned} P_{\hat{\theta}}(Y_k = f_j | \mathcal{G}_{k-1} \vee \{X_k\}) &= E_{\hat{\theta}}[\langle Y_k, f_j\rangle | \mathcal{G}_{k-1} \vee \{X_k\}] \\ &= \frac{E_\theta[\Lambda_k \langle Y_k, f_j\rangle | \mathcal{G}_{k-1} \vee \{X_k\}]}{E_\theta[\Lambda_k | \mathcal{G}_{k-1} \vee \{X_k\}]} \\ &= \frac{E_\theta[\lambda_k \langle Y_k, f_j\rangle | \mathcal{G}_{k-1} \vee \{X_k\}]}{E_\theta[\lambda_k | \mathcal{G}_{k-1} \vee \{X_k\}]}. \end{aligned}$$

Now, the denominator is

$$E_\theta[\sum_{i=1}^{N}\sum_{j=1}^{M}(\frac{\hat{c}_{ji}}{c_{ji}})\langle X_k, e_i\rangle\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee\{X_k\}] =$$

$$= \sum_{i=1}^{N} E_\theta[\sum_{j=1}^{M}(\frac{\hat{c}_{ji}}{c_{ji}})\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee\{X_k\}]\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N} E_\theta[\sum_{j=1}^{M}(\frac{\hat{c}_{ji}}{c_{ji}})\langle Y_k, f_j\rangle|X_k = e_i]\langle X_k, e_i\rangle$$

(by the Markov property)

$$= \sum_{i=1}^{N}\Big(\sum_{j=1}^{M}(\frac{\hat{c}_{ji}}{c_{ji}})E_\theta[\langle Y_k, f_j\rangle|X_k = e_i]\Big)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\Big(\sum_{j=1}^{M}(\frac{\hat{c}_{ji}}{c_{ji}})c_{ji}\Big)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\Big(\sum_{j=1}^{M}\hat{c}_{ji}\Big)\langle X_k, e_i\rangle = \sum_{i=1}^{N}\langle X_k, e_i\rangle = 1.$$

The numerator is

$$E_\theta[(\sum_{i=1}^{N}\sum_{j=1}^{M}(\frac{\hat{c}_{ji}}{c_{ji}})\langle Y_k, f_j\rangle\langle X_k, e_i\rangle)\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee\{X_k\}] =$$

$$= E_\theta[\sum_{i=1}^{N}(\frac{\hat{c}_{ji}}{c_{ji}})\langle Y_k, f_j\rangle\langle X_k, e_i\rangle|\mathcal{G}_{k-1}\vee\{X_k\}]$$

$$= \sum_{i=1}^{N}(\frac{\hat{c}_{ji}}{c_{ji}})E_\theta[\langle Y_k, f_j\rangle|\mathcal{G}_{k-1}\vee\{X_k\}]\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}(\frac{\hat{c}_{ji}}{c_{ji}})E_\theta[Y_k = f_j|X_k = e_i]\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}(\frac{\hat{c}_{ji}}{c_{ji}})P_\theta(Y_k = f_j|X_k = e_i)\langle X_k, e_i\rangle$$

$$= \sum_{i=1}^{N}\hat{c}_{ji}\langle X_k, e_i\rangle = \langle \hat{C}X_k, f_j\rangle.$$

This means that

$$P_{\hat{\theta}}(Y_k = f_j | \mathcal{G}_{k-1} \vee \{X_k\}) = P_{\hat{\theta}}(Y_k = f_j | X_k = e_i) = \hat{c}_{ji}. \quad \square$$

Proof of **Lemma 1.10.**

$$\frac{dP_{\hat{\theta}}}{dP_\theta}|\mathcal{G}_k = \prod_{l=0}^{k}(\sum_{r=1}^{N}\sum_{s=1}^{M}(\frac{\hat{c}_{sr}}{c_{sr}})\langle X_l, e_r\rangle\langle Y_l, f_s\rangle)$$

so

$$\begin{aligned}\log\frac{dP_{\hat{\theta}}}{dP_\theta} &= \sum_{l=0}^{k}\log(\sum_{r=1}^{N}\sum_{s=1}^{M}(\frac{\hat{c}_{sr}}{c_{sr}})\langle X_l, e_r\rangle\langle Y_l, f_s\rangle) \\ &= \sum_{l=0}^{k}\sum_{r=1}^{N}\sum_{s=1}^{M}\langle X_l, e_r\rangle\langle Y_l, f_s\rangle(\log\hat{c}_{sr} - \log c_{sr}) \\ &= \sum_{r=1}^{N}\sum_{s=1}^{M}T_k^{rs}\log\hat{c}_{sr} + R(c),\end{aligned}$$

where $R(c)$ is independent of the $\hat{c}_{sr}$. Therefore,

$$\begin{aligned}L(\hat{\theta}) &= E_\theta[\log\frac{dP_{\hat{\theta}}}{dP_\theta}|\mathcal{Y}_k] \\ &= E_\theta[\sum_{r=1}^{N}\sum_{s=1}^{M}T_k^{rs}\log\hat{c}_{sr} + R(c)|\mathcal{Y}_k] \\ &= \sum_{r=1}^{N}\sum_{s=1}^{M}\log\hat{c}_{sr}E_\theta[T_k^{rs}|\mathcal{Y}_k] + E_\theta[R(c)|\mathcal{Y}_k] \\ &= \sum_{r=1}^{N}\sum_{s=1}^{M}\log\hat{c}_{sr}\hat{T}_k^{rs} + \hat{R}(c).\end{aligned}$$

The optimal estimate of $\hat{c}_{ji}$ is, therefore, the value which solves the following maximization problem:

$$\max_{\hat{c}_{ji}}\sum_{r=1}^{N}\sum_{s=1}^{M}\log\hat{c}_{sr}\hat{T}_k^{rs} + \hat{R}(c)$$

subject to

$$\sum_{s=1}^{M} \hat{c}_{sr} = 1.$$

The Lagrangian is

$$\mathcal{L} = \sum_{r=1}^{N}\sum_{s=1}^{M} \hat{T}_k^{rs} \log \hat{c}_{sr} + \hat{R}(c) + \lambda\Big(\sum_{s=1}^{M} \hat{c}_{sr} - 1\Big).$$

Differentiating in $\hat{c}_{ji}$ and λ and equating the derivatives to 0 gives:

$$\frac{1}{\hat{c}_{ji}}\hat{T}_k^{ij} + \lambda = 0$$

$$\sum_{s=1}^{M} \hat{c}_{sr} = 1.$$

Recall that $\sum_{s=1}^{M} \hat{c}_{sr} = 1$ and $\sum_{s=1}^{M} T_k^{rs} = O_{k+1}^r$. so that $\sum_{s=1}^{M} \hat{T}_k^{rs} = \hat{O}1_k^r$. We then have $\lambda = -\hat{O}1_k^r$ so that $\hat{c}_{ji} = \frac{\hat{T}_k^{ij}}{\hat{O}1_k^r}$. Now, since $\hat{T}_k^{ij} = \frac{\sigma(T^{ij})_k}{\langle q_k, 1\rangle}$, and $\hat{O}1_k^i = \frac{\sigma(O1^i)_k}{\langle q_k, 1\rangle}$, we have

$$\hat{c}_{ji} = \frac{\sigma(T^{ij})_k}{\sigma(O1^i)_k}.$$

The result follows from Remark 1.2. □

Proof of **Lemma 1.11.**

$$\begin{aligned}
v_k^i &= \bar{E}[\bar{\Lambda}_{k+1.T} | \mathcal{Y}_{0.T} \vee \{X_k = e_i\}] \\
&= \bar{E}[\bar{\Lambda}_{k+2.T}\bar{\lambda}_{k+1} | \mathcal{Y}_{0.T} \vee \{X_k = e_i\}] \\
&= \bar{E}[\bar{\Lambda}_{k+2.T}(M\sum_{j=1}^{M}\langle CX_{k+1}, f_j\rangle\langle Y_{k+1}, f_j\rangle) | \mathcal{Y}_{0.T} \vee \{X_k = e_i\}] \\
&= \sum_{l=1}^{N} \bar{E}[\bar{\Lambda}_{k+2.T}(M\sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)\langle X_{k+1}, e_l\rangle | \mathcal{Y}_{0.T} \vee \{X_k = e_i\}] \\
&= \sum_{l=1}^{N} \bar{E}[\bar{\Lambda}_{k+2.T}\langle X_{k+1}, e_l\rangle | \mathcal{Y}_{0.T} \vee \{X_k = e_i\}](M\sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)
\end{aligned}$$

$$= \sum_{l=1}^{M} \bar{E}[\langle X_{k+1}, e_l\rangle \bar{E}[\bar{\Lambda}_{k+2,T}|\mathcal{Y}_{0,T} \vee \{X_k = e_i\} \vee \{X_k = e_l\}]|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]$$

$$\times (M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} \bar{E}[\langle X_{k+1}, e_l\rangle v_{k+1}^l|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}](M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

(by the Markov property and the definition of v)

$$= \sum_{l=1}^{N} \bar{E}[\langle X_{k+1}, e_l\rangle|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} \bar{E}[\langle AX_k + V_{k+1}, e_l\rangle|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} \bar{E}[\langle AX_k, e_l\rangle|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$+ \sum_{l=1}^{N} \bar{E}[\langle V_{k+1}, e_l\rangle|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} \bar{P}(X_{k+1} = e_l|\mathcal{Y}_{0,T} \vee \{X_k = e_i\})v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$+ \sum_{l=1}^{N}\langle \bar{E}[V_{k+1}|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}], e_l\rangle v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} \bar{P}(X_{k+1} = e_l|X_k = e_i)v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$+ \sum_{l=1}^{N}\langle \bar{E}[V_{k+1}|X_k = e_i], e_l\rangle v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} a_{li} v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$+ \sum_{l=1}^{N}\langle \bar{E}[\bar{E}[V_{k+1}|\mathcal{F}_k]|X_k = e_i], e_l\rangle v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle)$$

$$= \sum_{l=1}^{N} a_{li} v_{k+1}^l(M \sum_{j=1}^{M} c_{jl}\langle Y_{k+1}, f_j\rangle).$$

It follows that $v_k = A'B(Y_{k+1})v_{k+1}$, as required. $\square$

Proof of **Lemma 1.12.**

Consider v^j_{T-1}:

$$\begin{aligned}
v^j_{T-1} &= \bar{E}[\bar{\Lambda}_{T,T}|\mathcal{Y}_{0,T} \vee \{X_{T-1} = e_j\}] \\
&= \bar{E}[\bar{\lambda}_T|\mathcal{Y}_{0,T} \vee \{X_{T-1} = e_j\}] \\
&= \bar{E}[M\sum_{l=1}^{M}\langle CX_T, f_l\rangle\langle Y_T, f_l\rangle|\mathcal{Y}_{0,T} \vee \{X_{T-1} = e_j\}] \\
&= M\sum_{j=1}^{M}\bar{E}[\langle CX_T, f_l\rangle|\mathcal{Y}_{0,T} \vee \{X_{T-1} = e_j\}]\langle Y_T, f_l\rangle \\
&= M\sum_{j=1}^{M}\bar{E}[\sum_{i=1}^{N}c_{li}\langle X_T, e_i\rangle|\mathcal{Y}_{0,T} \vee \{X_{T-1} = e_j\}]\langle Y_T, f_l\rangle \\
&= \sum_{i=1}^{N}\bar{E}[\langle X_T, e_i\rangle|\mathcal{Y}_{0,T} \vee \{X_{T-1} = e_j\}](M\sum_{l=1}^{M}c_{li}\langle Y_T, f_l\rangle) \\
&= \sum_{i=1}^{N}\bar{E}[\langle X_T, e_i\rangle|\{X_{T-1} = e_j\}](M\sum_{l=1}^{M}c_{li}\langle Y_T, f_l\rangle) \\
&= \sum_{i=1}^{N}\bar{P}(X_T = e_i|X_{T-1} = e_j)(M\sum_{l=1}^{M}c_{li}\langle Y_T, f_l\rangle) \\
&= \sum_{i=1}^{N}a_{ij}(M\sum_{l=1}^{M}c_{li}\langle Y_T, f_l\rangle).
\end{aligned}$$

It follows that $v_{T-1} = A'B(Y_T)\mathbf{1}$. $\square$

Proof of **Theorem 1.4.**

$$\bar{E}[\bar{\Lambda}_{0,T}X_k|\mathcal{Y}_{0,T}] = \sum_{i=1}^{N}\bar{E}[\bar{\Lambda}_{0,T}\langle X_k, e_i\rangle X_k|\mathcal{Y}_{0,T}] = \sum_{i=1}^{N}\bar{E}[\bar{\Lambda}_{0,T}\langle X_k, e_i\rangle|\mathcal{Y}_{0,T}]e_i.$$

Consider the i-th component:

$$\begin{aligned}
\bar{E}[\bar{\Lambda}_{0,T}\langle X_k, e_i\rangle|\mathcal{Y}_{0,T}] &= \bar{E}[\bar{\Lambda}_{0,k}\bar{\Lambda}_{k+1,T}\langle X_k, e_i\rangle|\mathcal{Y}_{0,T}] \\
&= \bar{E}[\bar{\Lambda}_{0,k}\bar{E}[\bar{\Lambda}_{k+1,T}|\mathcal{Y}_{0,T} \vee \{X_k = e_i\}]\langle X_k, e_i\rangle|\mathcal{Y}_{0,T}] \\
&= \bar{E}[\bar{\Lambda}_{0,k}v_k^i\langle X_k, e_i\rangle|\mathcal{Y}_{0,T}] \\
&= \bar{E}[\bar{\Lambda}_{0,k}\langle X_k, e_i\rangle|\mathcal{Y}_{0,T}]v_k^i \\
&= \langle \bar{E}[\bar{\Lambda}_{0,k}X_k|\mathcal{Y}_{0,T}], e_i\rangle v_k^i \\
&= q_k^i v_k^i,
\end{aligned}$$

where $q_k^i := \langle \bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_k], e_i\rangle$.

Therefore, $\bar{E}[\bar{\Lambda}_{0,T}X_k|\mathcal{Y}_{0,T}] = \sum_{i=1}^N q_k^i v_k^i e_i = \text{diag}\,(q_k \cdot v_k')$. □

Proof of **Theorem 1.5.**

$$\langle \sigma(J^{ij}X)_{k+1}, v_{k+1}\rangle =$$

$$= \langle B(Y_{k+1})A\sigma(J^{ij}X)_k + (M\sum_{s=1}^M c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle a_{ji}e_j, v_{k+1}\rangle$$

(by Theorem 2.2)

$$\begin{aligned}
&= \langle B(Y_{k+1})A\sigma(J^{ij}X)_k, v_{k+1}\rangle + \langle (M\sum_{s=1}^M c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle a_{ji}e_j, v_{k+1}\rangle \\
&= \langle B(Y_{k+1})A\sigma(J^{ij}X)_k, v_{k+1}\rangle + (M\sum_{s=1}^M c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle a_{ji}\langle v_{k+1}, e_j\rangle \\
&= \langle \sigma(J^{ij}X)_k, A'B(Y_{k+1})v_{k+1}\rangle + (M\sum_{s=1}^M c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle\langle v_{k+1}, e_j\rangle a_{ji} \\
&= \langle \sigma(J^{ij}X)_k, v_k\rangle + (M\sum_{s=1}^M c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle\langle v_{k+1}, e_j\rangle a_{ji}.
\end{aligned}$$

That is,

$$\langle \sigma(J^{ij}X)_{k+1}, v_{k+1}\rangle - \langle \sigma(J^{ij}X)_k, v_k\rangle = (M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle\langle v_{k+1}, e_j\rangle a_{ji}.$$

Since $J_0^{ij} = 0$ and $v_T = \mathbf{1}$,

$$\sum_{k=0}^{T-1}[\langle \sigma(J^{ij}X)_{k+1}, v_{k+1}\rangle - \langle \sigma(J^{ij}X)_k, v_k\rangle] = \langle \sigma(J^{ij}X)_T, v_T\rangle - \langle \sigma(J^{ij}X)_0, v_0\rangle$$
$$= \langle \sigma(J^{ij}X)_T, v_T\rangle = \langle \sigma(J^{ij}X)_T, \mathbf{1}\rangle = \sigma(J^{ij})_T.$$

Hence,

$$\begin{aligned}\sigma(J^{ij})_T &= \sum_{k=0}^{T-1} a_{ji}(M\sum_{s=1}^{M} c_{sj}\langle Y_{k+1}, f_s\rangle)\langle q_k, e_i\rangle\langle v_{k+1}, e_j\rangle \\ &= \sum_{k=1}^{T} a_{ji}(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle\langle v_k, e_j\rangle. \quad \square\end{aligned}$$

Proof of **Corollary 1.1.**

Since $\sigma(O^i)_T = \sum_{j=1}^{N}\sigma(J^{ij})_T$, we have

$$\begin{aligned}\sigma(O^i)_T &= \sum_{j=1}^{N}\sigma(J^{ij})_T = \sum_{j=1}^{N} a_{ji}\sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle\langle v_k, e_j\rangle(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle) \\ &= \sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle\sum_{j=1}^{N} a_{ji}\langle v_k, e_j\rangle(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle) \\ &= \sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle\sum_{j=1}^{N} v_k^j a_{ji}(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle) \\ &= \sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle\langle A'B(Y_{k+1})v_k, e_i\rangle \\ &= \sum_{k=1}^{T}\langle q_{k-1}, e_i\rangle\langle v_{k-1}, e_i\rangle. \quad \square\end{aligned}$$

Proof of **Theorem 1.6.**

$$\langle \sigma(T^{ij}X)_{k+1}, v_{k+1}\rangle =$$

$$= \langle B(Y_{k+1})A\sigma(T^{ij}X)_k + Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle e_i, v_{k+1}\rangle$$

(by Theorem 2.3)

$$= \langle B(Y_{k+1})A\sigma(T^{ij}X)_k, v_{k+1}\rangle + \langle Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle e_i, v_{k+1}\rangle$$

$$= \langle B(Y_{k+1})A\sigma(T^{ij}X)_k, v_{k+1}\rangle + Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle\langle v_{k+1}, e_i\rangle$$

$$= \langle \sigma(T^{ij}X)_k, A'B(Y_{k+1})v_{k+1}\rangle + Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle\langle v_{k+1}, e_i\rangle$$

$$= \langle \sigma(T^{ij}X)_k, v_k\rangle + Mc_{ji}\langle Y_{k+1}, f_j\rangle\langle Aq_k, e_i\rangle\langle v_{k+1}, e_i\rangle.$$

That is,

$$\langle \sigma(T^{ij}X)_{k+1}, v_{k+1}\rangle - \langle \sigma(T^{ij}X)_k, v_k\rangle = Mc_{ji}\langle Y_{k+1}, f_j\rangle)\langle Aq_k, e_i\rangle\langle v_{k+1}, e_j\rangle.$$

Since $T_0^{ij} = 0$ and $v_T = 1$,

$$\sum_{k=0}^{T-1}[\langle \sigma(T^{ij}X)_{k+1}, v_{k+1}\rangle - \langle \sigma(T^{ij}X)_k, v_k\rangle] = \langle \sigma(T^{ij}X)_T, v_T\rangle - \langle \sigma(T^{ij}X)_0, v_0\rangle$$
$$= \langle \sigma(T^{ij}X)_T, v_T\rangle = \langle \sigma(T^{ij}X)_T, 1\rangle = \sigma(T^{ij})_T.$$

Hence,

$$\sigma(T^{ij})_T = \sum_{k=0}^{T-1} Mc_{ji}\langle Y_{k+1}, f_j\rangle)\langle Aq_k, e_i\rangle\langle v_{k+1}, e_i\rangle$$
$$= \sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle)\langle Aq_{k-1}, e_i\rangle\langle v_k, e_i\rangle. \quad \square$$

Proof of **Corollary 1.2.**

Since $\sigma(O1^i)_T = \sum_{j=1}^M \sigma(T^{ij})_T$, we have

$$\sigma(O1^i)_T = \sum_{j=1}^{M} \sigma(T^{ij})_T = \sum_{j=1}^{M}\sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle\langle v_k, e_i\rangle$$
$$= \sum_{k=1}^{T}\langle v_k, e_i\rangle\langle Aq_{k-1}, e_i\rangle(M\sum_{j=1}^{M} c_{ji}\langle Y_k, f_j\rangle). \quad \square$$

Proof of **Lemma 1.14.**

$$\sigma(J^{ij})_{T+1}$$
$$= a_{ji}\sum_{k=1}^{T+1}(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle e_j' A'B(Y_{k+1})\cdots A'B(Y_T)A'B(Y_{T+1})\mathbf{1}$$
$$= a_{ji}\sum_{k=1}^{T}(M\sum_{s=1}^{M} c_{sj}\langle Y_k, f_s\rangle)\langle q_{k-1}, e_i\rangle e_j' A'B(Y_{k+1})\cdots A'B(Y_T)\mathbf{1}$$
$$+ a_{ji}(M\sum_{s=1}^{M} c_{sj}\langle Y_{T+1}, f_s\rangle)\langle q_T, e_i\rangle e_j'\mathbf{1}$$
$$= \Gamma_T' A'B(Y_{T+1}) + a_{ji}(M\sum_{s=1}^{M} c_{sj}\langle Y_{T+1}, f_s\rangle)\langle q_T, e_i\rangle e_j'$$
$$= \Gamma_{T+1}'\mathbf{1}.$$

The result follows. $\square$

Proof of **Lemma 1.15.**

$$\sigma(T^{ij})_{T+1} = \sum_{k=1}^{T+1} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle e_i' A'B(Y_{k+1})\cdots A'B(Y_T)A'B(Y_{T+1})\mathbf{1}$$
$$= \sum_{k=1}^{T} Mc_{ji}\langle Y_k, f_j\rangle\langle Aq_{k-1}, e_i\rangle e_i' A'B(Y_{k+1})\cdots A'B(Y_T)A'B(Y_{T+1})\mathbf{1}$$

$$+ Mc_{ji}\langle Y_{T+1}, f_j\rangle\langle Aq_T, e_i\rangle e_i'\mathbf{1}$$

$$= H_T' A' B(Y_{T+1})\mathbf{1} + Mc_{ji}\langle Y_{T+1}, f_j\rangle\langle Aq_T, e_i\rangle e_i'\mathbf{1}$$

$$= H_{T+1}'\mathbf{1}.$$

The result follows. □

Chapter 2

HMM Implementation Results

1. INTRODUCTION

A Markov Chain Model is often used to describe the dynamics of a firm's credit rating as an indicator of the likelihood of default. The rating labels, from the highest rating of AAA/Aaa to the lowest rating of C and then the default D, are taken to be the states of the process. The transition probability matrix gives probabilities of rating migration from one state to another within a unit of time, such as a quarter or a year. The dynamics are stationary so that the probability of a transition from one rating category to the next does not change over time. Taking powers of the transition probability matrix allows for predicting the probability of degradation in credit quality or even default within any time frame. The key assumption behind the Markov chain representation of credit rating evolution is the Markov property, which implies that the rating process should have no memory of its past behaviour so that prior rating changes should have no predictive power for the direction of future rating changes. As documented in Carty and Fons [5], Carty and Lieberman [6], and Lando and Skodeberg [24], the credit rating process seems to exhibit "momentum" or "rating drift:" a firm recently upgraded (downgraded) is more likely to be upgraded (downgraded).

We propose that the observed rating process is corrupted by what we may call "noise." We assume that the credit rating evolution can be described by a Markov chain but that we do not observe this Markov chain directly. Rather, it is hidden in "noisy" observations represented by the posted credit ratings. In this chapter we implement the Hidden Markov Model (HMM) described in Chapter 1 to a data set of Standard & Poor's credit ratings. The *outcome* of the HMM is a probability distribution for a "true" rating at time k given the observed ratings up to and including time k, and estimates of the parameters of the model, namely the elements of the transition matrix and the probabilities of observing a particular rating given a "true" rating.

2. THE RATINGS DATA

Here we describe some of the aspects of debt ratings obtained from the Standard & Poor's COMPUSTAT database, as relevant for our subsequent implementation of the Hidden Markov Model.

2.1. Basic Properties

Our analysis takes advantage of the Standard & Poor's COMPUSTAT database, which contains rating histories for 1,301 obligors over the period 1985-1999.

The universe of obligors is mainly large U.S. and Canadian corporate institutions. The obligors include industrials, utilities, insurance companies, banks and other financial institutions and real estate companies.

To capture credit quality dynamics, the creditworthiness of obligors must be assessed, as credit events typically concern a firm as a whole. Unfortunately, published ratings typically focus on individual bond issues. Therefore, Standard & Poor's implement a number of transformations. Prior to September 1, 1998, the company level rating is taken to be the highest issue level rating that the company has on its senior secured debt. When a company does not have senior secured debt issues, the implied senior rating is used.[1] The last point is worth elaborating. We interpret credit ratings as indicators of the chance of default and likelihood of migration to a different (lower or higher) rating class. However there are clearly differences in rating between senior and subordinated debt in recognition of differences in anticipated recovery rate in case of default. It is certainly true that senior debt obligations may be satisfied in full during bankruptcy procedures while subordinated debt is paid off only partially. The anticipated recovery rate for subordinated debt is lower and this type of debt is given a lower rating, which then reflects recovery rate differences in addition to the likelihood of default. Since we are not interested in recovery rate differences, only the most senior credit rating is used as a proxy for the company level rating. As of September 1, 1998, all ratings

[1]Standard & Poor's assign an implied senior rating when a company applies for a subordinated rating based on the issuance of subordinated debt only.

in the COMPUSTAT dataset are Standard & Poor's issuer credit ratings.

The COMPUSTAT database provides annual ratings. Every year each of the rated obligors is assigned to one of the Standard and Poor's 8 rating categories, ranging from AAA (highest rating) to CCC (lowest rating) as well as D (payment in default) and the NR (not rated) state.

We have a total of 19,515 firm-years in our sample. However, only 34% of those observations are "non-zero," i.e. correspond to a firm with one of the 8 rating labels in a given year. The remaining 66% of observations represent transitions to the so-called NR (not rated) status. Transitions to NR are discussed in detail in the next section. Approximately 85% of non-zero ratings are range from B to A. The median rating is BB, the highest non investment-grade rating. Approximately 1% of the observed "non-zero" ratings are AAA and 2% are defaults. The most common rating is B, two rating categories above default, which accounts for 25.5% of the "non-zero" observations.

2.2. Treatment of Transitions to "Not Rated" Status

Not every issuer has been assigned a rating for each of the 15 years between 1985 and 1999. As a result, the COMPUSTAT dataset contains many transitions to the NR (not rated) status. The majority of rating withdrawals occur when a firm's only outstanding issue is paid off or its debt issuance program matures. However, transitions to NR may be due to other reasons as well, such as failure to pay the requisite fee to Standard & Poor's. Unfortunately, the details of individual transitions to NR are not known. In particular, we do not know whether a deterioration of credit quality known only to the obligor has led the issuer to decide to bypass an agency rating. In other words, we do not know whether a given transition to NR is "benign" or "bad."

The industry standard calls for removing transitions to NR from the dataset. The procedure depends on whether transitions to NR are considered "negative information" or "non-information." Regardless of how the NR category is interpreted, probability transitions to NR are distributed among other states.

Our main objective is to utilize as many rating transitions as possible. However, most firms experience few rating changes. Moreover, when credit ratings do change, they usually stay within a fairly narrow range, a few consecutive rating categories. Transitions to NR pose a problem as well since only a small subset of rated obligors have been assigned a rating for the same number of consecutive years (84 firms rated over 1988-1998).

Therefore, contrary to the industry standard for the estimation of transition matrices, we retain the NR category in our dataset. As a result, we have 15 years of rating history for all 1,301 obligors.

3. IMPLEMENTATION RESULTS

As mentioned before, firms generally experience few rating changes within a narrow range. Implementation of the HMM algorithm based on a rating history of one company is then problematic. To overcome this difficulty, we apply the algorithm to an aggregate of firms in the dataset rather than an individual company. This allows for more observed transitions between rating categories which makes inference possible. Specifically, instead of estimating the distribution and parameters for the Markov chain X_k^l for each firm l, we estimate the distribution and parameters for $\sum_{l=1}^{L} X_k^l$ given the additivity of all stochastic processes discussed in Chapter 1. This approach is appealing for other reasons as well. By considering aggregate rating processes for a particular industry, we can estimate parameters of the model, namely matrices A and C, specific to that industry. These parameter estimates can then be used to obtain a distribution for the signal process $\{X_k\}$ for a particular firm from the industry via Theorem 1.1. This way we use only the most relevant information in the estimation. Note that given the estimates of A and C, Theorem 1.1 can also be used to make predictions with regards to the evolution of a particular company's rating.

Each credit rating category, 8 in total, was identified with a unit vector in $\mathbb{R}^8$. The initial values of the model parameters were as follows. Matrix A was taken to be the July 1998 historical transition matrix based on Standard & Poor's credit ratings obtained from J.P. Morgan's $CreditMetrics^{TM}$ dataset. The matrix is given in Appendix IIA. Note that this transition matrix does not include transitions to the NR category. Consequently, the NR category was treated as "negative information" and combined with the default state D. Matrix C was arbitrarily taken to be the 8×8 tri-diagonal matrix. The matrix is given in Appendix IIB. Non-zero entries of this matrix are interpreted as follows. The probability of the observed rating agreeing with the "true" rating is assumed to be 0.5 for all rating categories. The probability of the observed rating being one notch higher than the "true" rating is 0.3 (0.5 for default state D), and the probability of the observed rating being one notch lower than the "true" rating is 0.2 (0.5 for AAA). Note that when $C = I$, processes X (the signal) and Y (the observations) are identical, i.e. there is no "noise" in the system. Given the relatively short time period, parameter estimates were updated with the arrival of every new observation for the 1,301 firms in the data set using the formulae given in Chapter 1. Repetition of the estimation procedures ensures that the model and estimates improve with each iteration. Simulation results are presented in Appendix IIC. We report the estimated parameters of the model, namely matrices $\hat{A}$ and $\hat{C}$, as well as the aggregate variance/covariance matrices for the two martingale increments, V and W, in the semimartingale representation of X and Y, respectively.

Consider first the aggregate variance estimates for the semimartingale increments W in the semimartingale representation $Y_k = CX_k + W_k$. The variances are generally small and decrease significantly with the arrival of each new observation, which confirms the "self-tuning" property of the model. There is an improvement in the quality of the estimates with each successive pass through the data.

Consider now the estimated transition matrix $\hat{A}$. Entries above the diagonal correspond to rating upgrades and those below the diagonal to rating downgrades.

We see that non-zero transition probabilities are concentrated and highest on the diagonal: obligors are most likely to maintain their current rating. The second largest probability is usually on the off-diagonal. This confirms the observation that rating agencies usually do not change a company's rating by more than one category at a time. Downgrades seem to be more common than upgrades, except for the BBB category. For firms rated BBB, an upgrade seems more likely than a downgrade. BBB firms therefore tend to hold on to their investment-grade status. Note that for Enron, maintaining an investment-grade rating was one of the conditions for the success of a proposed merger with Dynergy inc., which eventually did not succeed. Finally, the lower the initial rating, the greater the probability of transition to the $NR + D$. Note also that this probability is estimated as zero for the two highest rating classes, AAA and AA, and virtually zero for A. In other words, once the highest-rated firms enter the data set, the remain rated until 1999. This probability is nearly 15% for BBB and increases to 38.5% for CCC. It is then the lower-rated firms who disappear from the data set. Possible reasons are: bankruptcy, maturing debt followed by no new issues possibly because of concerns over credit quality, or opting for no rating in anticipation of unfavourable rating assessment.

Recall that in general matrix C describes the relationship between the signal process X and the observation process Y. In particular, non-zero entries above the diagonal indicate that the observed rating may be higher than the "true" credit rating. In this case, the estimated matrix $\hat{C}$ is tri-diagonal, with the highest probabilities generally on the main diagonal. This suggests that if there is "noise" in the rating system, it is mostly confined to the neighbouring rating categories. Note also that for A and $NR + D$, the probability on the diagonal is estimated to be close to one, which suggests that the observed credi rating may agree with the "true" rating. For BBB and B, the estimates suggest that the observed rating may be higher than the "true" rating. For AAA, AA and CCC, the observed rating may be lower than the "true" rating. Note that for AAA, the probability of observing AA is estimated as one, which suggests that Standard & Poor's may be reluctant to upgrade firms to the highest rating AAA. Overall however, the results are therefore inconclusive with regards to the overall quality of the Standard & Poor's rating system. Longer rating histories may be required to accurately capture the rating dynamics.

Our results seem to suggest that the rating process may be influenced by the fact that it is often crucial for a borrower to maintain investment-grade rating. We have therefore reclassified all firms in the sample as investment grade, speculative grade or default/NR and then applied the HMM algorithm to the new data set. The results presented in Appendix IID confirm that investment-grade firms do generally hold on to their status, but there is an estimated 28% probability of downgrade to speculative-grade status. However, for speculative-grade firms, the probability of upgrade to investment-grade status is virtually zero. Speculative-grade firms tend to maintain their status or disappear from the data set because of either default or withdrawn rating. Estimated matrix C suggests that rating agencies may be reluctant to upgrade firms to investment-grade status, which results in estimated probability of 16% that the observed rating is speculative-grade when the "true" credit quality is investment-grade. For speculative-grade firms, the estimated matrix C confirms our earlier observation that these firms tend to disappear from the data set quickly, perhaps because they choose to have their rating withdrawn in anticipation of unfavourable news. However, we conclude as before that longer rating histories and further analysis may be required to verify our results.

Appendix IIA

Standard & Poor's Historical Transition Matrix (July 1998)

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	0.908	0.006	0.001	0.000	0.000	0.000	0.002	0.000
AA	0.083	0.909	0.024	0.003	0.001	0.001	0.000	0.000
A	0.007	0.077	0.913	0.059	0.006	0.002	0.004	0.000
BBB	0.001	0.006	0.052	0.875	0.077	0.004	0.012	0.000
BB	0.001	0.001	0.007	0.050	0.812	0.069	0.027	0.000
B	0.000	0.001	0.002	0.011	0.084	0.835	0.117	0.000
CCC	0.000	0.000	0.000	0.001	0.010	0.039	0.645	0.000
D	0.000	0.000	0.001	0.002	0.010	0.049	0.193	1.000

Appendix IIB

Initial Matrix C

	AAA	AA	A	BBB	BB	B	CCC	D
AAA	0.500	0.300	0.000	0.000	0.000	0.000	0.000	0.000
AA	0.500	0.500	0.300	0.000	0.000	0.000	0.000	0.000
A	0.000	0.200	0.500	0.300	0.000	0.000	0.000	0.000
BBB	0.000	0.000	0.200	0.500	0.300	0.000	0.000	0.000
BB	0.000	0.000	0.000	0.200	0.500	0.300	0.000	0.000
B	0.000	0.000	0.000	0.000	0.200	0.500	0.300	0.000
CCC	0.000	0.000	0.000	0.000	0.000	0.200	0.500	0.500
D	0.000	0.000	0.000	0.000	0.000	0.000	0.200	0.500

Appendix IIC
HMM Implementation Results

The following pages present the output of a computer program written to implement the estimation procedures from Chapter 1.

For each of the passes through the data set, we are given the estimates for matrices A and C, as well as $VarV_k$ and $VarW_k$. Recall that $a_{ji} = P(X_{k+1} = e_j | X_k = e_i)$ and $c_{ji} = P(Y_k = f_j | X_k = e_i)$. For matrix A, probabilities above the diagonal correspond to rating upgrades, and those below the diagonal to rating downgrades. For matrix C, entries above the diagonal correspond to the probability that the observed rating is higher than the "true" rating. A non-zero entry below the diagonal means that the observed rating is lower than the "true" rating.

Pass 1

Estimated matrix A

0.855	0.004	0.000	0.000	0.000	0.000	0.000	0.000
0.124	0.870	0.016	0.002	0.001	0.001	0.000	0.000
0.017	0.115	0.925	0.064	0.007	0.002	0.002	0.000
0.001	0.008	0.049	0.869	0.080	0.004	0.005	0.000
0.002	0.001	0.006	0.045	0.765	0.057	0.009	0.000
0.000	0.002	0.002	0.009	0.071	0.618	0.035	0.000
0.000	0.001	0.000	0.003	0.024	0.086	0.577	0.000
0.000	0.000	0.002	0.009	0.052	0.232	0.372	1.000

Estimated matrix C

0.095	0.036	0.000	0.000	0.000	0.000	0.000	0.000
0.905	0.566	0.217	0.000	0.000	0.000	0.000	0.000
0.000	0.399	0.638	0.418	0.000	0.000	0.000	0.000
0.000	0.000	0.145	0.396	0.260	0.000	0.000	0.000
0.000	0.000	0.000	0.187	0.512	0.345	0.000	0.000
0.000	0.000	0.000	0.000	0.228	0.641	0.130	0.000
0.000	0.000	0.000	0.000	0.000	0.015	0.012	0.006
0.000	0.000	0.000	0.000	0.000	0.000	0.858	0.994

Aggregate variance/covariance matrix for V

20.981	-17.845	-2.542	-0.213	-0.323	-0.004	-0.032	-0.022
-17.845	39.177	-19.023	-1.582	-0.281	-0.294	-0.110	-0.041
-2.542	-19.023	42.100	-16.556	-2.267	-0.711	-0.296	-0.705
-0.213	-1.582	-16.556	40.868	-16.356	-2.597	-1.200	-2.364
-0.323	-0.281	-2.267	-16.356	47.816	-14.621	-4.685	-9.283
-0.004	-0.294	-0.711	-2.597	-14.621	56.525	-12.230	-26.067
-0.032	-0.110	-0.296	-1.200	-4.685	-12.230	56.897	-38.345
-0.022	-0.041	-0.705	-2.364	-9.283	-26.067	-38.345	76.826

Aggregate variance/covariance matrix for W

17.613	-15.281	-2.332	0.000	0.000	0.000	0.000	0.000
-15.281	83.818	-62.738	-5.799	0.000	0.000	0.000	0.000
-2.332	-62.738	122.249	-44.309	-12.869	0.000	0.000	0.000
0.000	-5.799	-44.309	89.996	-31.346	-8.541	0.000	0.000
0.000	0.000	-12.869	-31.346	88.042	-43.225	-0.601	0.000
0.000	0.000	0.000	-8.541	-43.225	65.591	-1.296	-12.530
0.000	0.000	0.000	0.000	-0.601	-1.296	4.624	-2.727
0.000	0.000	0.000	0.000	0.000	-12.530	-2.727	15.257

Pass 2

Estimated matrix A

0.911	0.006	0.000	0.000	0.000	0.000	0.000	0.000
0.080	0.892	0.015	0.002	0.001	0.000	0.000	0.000
0.007	0.097	0.941	0.069	0.005	0.001	0.001	0.000
0.000	0.004	0.037	0.871	0.061	0.002	0.001	0.000
0.001	0.000	0.003	0.036	0.748	0.039	0.003	0.000
0.000	0.001	0.001	0.006	0.070	0.555	0.014	0.000
0.000	0.001	0.000	0.003	0.032	0.098	0.555	0.000
0.000	0.000	0.003	0.013	0.083	0.305	0.427	1.000

Estimated matrix C

0.018	0.006	0.000	0.000	0.000	0.000	0.000	0.000
0.982	0.652	0.135	0.000	0.000	0.000	0.000	0.000
0.000	0.343	0.795	0.477	0.000	0.000	0.000	0.000
0.000	0.000	0.070	0.394	0.198	0.000	0.000	0.000
0.000	0.000	0.000	0.130	0.607	0.269	0.000	0.000
0.000	0.000	0.000	0.000	0.195	0.727	0.046	0.000
0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.954	1.000

Aggregate variance/covariance matrix for V

12.370	-28.933	-3.645	-0.285	-0.398	-0.007	-0.049	-0.035
-28.933	29.560	-36.096	-2.606	-0.412	-0.417	-0.197	-0.076
-3.645	-36.096	37.212	-32.923	-3.687	-1.061	-0.458	-1.442
-0.285	-2.606	-32.923	34.396	-28.065	-4.213	-2.078	-5.094
-0.398	-0.412	-3.687	-28.065	38.139	-24.767	-8.816	-19.809
-0.007	-0.417	-1.061	-4.213	-24.767	41.658	-19.910	-47.807
-0.049	-0.197	-0.458	-2.078	-8.816	-19.910	43.493	-68.883
-0.035	-0.076	-1.442	-5.094	-19.809	-47.807	-68.883	66.321

Aggregate variance/covariance matrix for W

3.224	-18.189	-2.648	0.000	0.000	0.000	0.000	0.000
-18.189	62.608	-120.515	-7.722	0.000	0.000	0.000	0.000
-2.648	-120.515	109.456	-85.767	-22.776	0.000	0.000	0.000
0.000	-7.722	-85.767	70.473	-53.841	-13.138	0.000	0.000
0.000	0.000	-22.776	-53.841	62.092	-72.829	-0.687	0.000
0.000	0.000	0.000	-13.138	-72.829	37.878	-1.531	-15.971
0.000	0.000	0.000	0.000	-0.687	-1.531	0.432	-2.838
0.000	0.000	0.000	0.000	0.000	-15.971	-2.838	3.552

Pass 3

Estimated matrix A

0.943	0.009	0.000	0.000	0.000	0.000	0.000	0.000
0.054	0.898	0.011	0.002	0.001	0.000	0.000	0.000
0.003	0.089	0.960	0.096	0.005	0.001	0.000	0.000
0.000	0.003	0.024	0.840	0.040	0.001	0.001	0.000
0.000	0.000	0.001	0.027	0.676	0.022	0.001	0.000
0.000	0.000	0.000	0.005	0.069	0.443	0.006	0.000
0.000	0.001	0.000	0.006	0.056	0.123	0.549	0.000
0.000	0.000	0.003	0.024	0.152	0.411	0.443	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.759	0.045	0.000	0.000	0.000	0.000	0.000
0.000	0.241	0.939	0.585	0.000	0.000	0.000	0.000
0.000	0.000	0.015	0.356	0.132	0.000	0.000	0.000
0.000	0.000	0.000	0.059	0.716	0.193	0.000	0.000
0.000	0.000	0.000	0.000	0.152	0.806	0.025	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.975	1.000

Aggregate variance/covariance matrix for V

8.493	-36.705	-4.253	-0.339	-0.432	-0.009	-0.060	-0.047
-36.705	23.928	-51.281	-3.290	-0.488	-0.489	-0.302	-0.110
-4.253	-51.281	35.946	-50.518	-4.723	-1.308	-0.662	-2.514
-0.339	-3.290	-50.518	31.398	-34.971	-5.292	-3.175	-9.077
-0.432	-0.488	-4.723	-34.971	32.297	-31.113	-13.612	-32.913
-0.009	-0.489	-1.308	-5.292	-31.113	28.702	-24.966	-63.707
-0.060	-0.302	-0.662	-3.175	-13.612	-24.966	35.588	-93.202
-0.047	-0.110	-2.514	-9.077	-32.913	-63.707	-93.202	58.422

Aggregate variance/covariance matrix for W

0.061	-18.247	-2.651	0.000	0.000	0.000	0.000	0.000
-18.247	38.171	-158.470	-7.879	0.000	0.000	0.000	0.000
-2.651	-158.470	76.374	-119.164	-27.794	0.000	0.000	0.000
0.000	-7.879	-119.164	46.548	-65.096	-14.877	0.000	0.000
0.000	0.000	-27.794	-65.096	32.699	-89.238	-0.704	0.000
0.000	0.000	0.000	-14.877	-89.238	19.707	-1.602	-17.458
0.000	0.000	0.000	0.000	-0.704	-1.602	0.117	-2.867
0.000	0.000	0.000	0.000	0.000	-17.458	-2.867	1.516

Pass 4

Estimated matrix A

0.955	0.011	0.000	0.000	0.000	0.000	0.000	0.000
0.042	0.880	0.006	0.002	0.001	0.000	0.000	0.000
0.003	0.106	0.975	0.145	0.006	0.001	0.000	0.000
0.000	0.002	0.016	0.779	0.025	0.001	0.000	0.000
0.000	0.000	0.000	0.020	0.572	0.012	0.000	0.000
0.000	0.000	0.000	0.006	0.068	0.343	0.004	0.000
0.000	0.001	0.000	0.010	0.088	0.147	0.547	0.000
0.000	0.000	0.003	0.038	0.240	0.496	0.448	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.825	0.017	0.000	0.000	0.000	0.000	0.000
0.000	0.175	0.980	0.745	0.000	0.000	0.000	0.000
0.000	0.000	0.003	0.233	0.083	0.000	0.000	0.000
0.000	0.000	0.000	0.023	0.799	0.183	0.000	0.000
0.000	0.000	0.000	0.000	0.118	0.815	0.023	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.977	1.000

Aggregate variance/covariance matrix for V

7.134	-43.191	-4.815	-0.384	-0.453	-0.012	-0.068	-0.057
-43.191	22.949	-67.008	-3.771	-0.529	-0.549	-0.428	-0.137
-4.815	-67.008	39.026	-70.305	-5.558	-1.550	-0.982	-4.066
-0.384	-3.771	-70.305	30.856	-38.526	-6.115	-4.451	-13.965
-0.453	-0.529	-5.558	-38.526	24.854	-34.715	-18.084	-45.240
-0.012	-0.549	-1.550	-6.115	-34.715	16.868	-27.896	-72.916
-0.068	-0.428	-0.982	-4.451	-18.084	-27.896	29.223	-113.292
-0.057	-0.137	-4.066	-13.965	-45.240	-72.916	-113.292	48.104

Aggregate variance/covariance matrix for W

0.009	-18.255	-2.652	0.000	0.000	0.000	0.000	0.000
-18.255	24.850	-183.298	-7.894	0.000	0.000	0.000	0.000
-2.652	-183.298	48.279	-140.625	-29.785	0.000	0.000	0.000
0.000	-7.894	-140.625	26.163	-69.259	-15.401	0.000	0.000
0.000	0.000	-29.785	-69.259	14.567	-97.644	-0.712	0.000
0.000	0.000	0.000	-15.401	-97.644	10.060	-1.636	-18.554
0.000	0.000	0.000	0.000	-0.712	-1.636	0.067	-2.892
0.000	0.000	0.000	0.000	0.000	-18.554	-2.892	1.121

Pass 5

Estimated matrix A

0.964	0.014	0.000	0.000	0.000	0.000	0.000	0.000
0.034	0.869	0.003	0.002	0.000	0.000	0.000	0.000
0.002	0.114	0.981	0.186	0.007	0.001	0.000	0.000
0.000	0.002	0.012	0.720	0.018	0.000	0.000	0.000
0.000	0.000	0.000	0.018	0.463	0.008	0.000	0.000
0.000	0.001	0.000	0.008	0.079	0.312	0.004	0.000
0.000	0.001	0.000	0.013	0.116	0.155	0.546	0.000
0.000	0.000	0.003	0.054	0.317	0.524	0.449	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.869	0.011	0.000	0.000	0.000	0.000	0.000
0.000	0.131	0.987	0.873	0.000	0.000	0.000	0.000
0.000	0.000	0.002	0.113	0.077	0.000	0.000	0.000
0.000	0.000	0.000	0.014	0.785	0.201	0.000	0.000
0.000	0.000	0.000	0.000	0.139	0.798	0.024	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.976	1.000

Aggregate variance/covariance matrix for V

6.270	-48.851	-5.357	-0.422	-0.466	-0.014	-0.075	-0.065
-48.851	21.291	-82.050	-4.116	-0.551	-0.611	-0.568	-0.157
-5.357	-82.050	37.896	-89.312	-6.188	-1.828	-1.392	-6.052
-0.422	-4.116	-89.312	28.251	-40.487	-6.855	-5.703	-18.873
-0.466	-0.551	-6.188	-40.487	15.626	-36.738	-21.007	-53.295
-0.014	-0.611	-1.828	-6.855	-36.738	9.957	-29.593	-78.071
-0.075	-0.568	-1.392	-5.703	-21.007	-29.593	22.350	-129.214
-0.065	-0.157	-6.052	-18.873	-53.295	-78.071	-129.214	36.055

Aggregate variance/covariance matrix for W

0.009	-18.264	-2.652	0.000	0.000	0.000	0.000	0.000
-18.264	17.939	-201.220	-7.901	0.000	0.000	0.000	0.000
-2.652	-201.220	28.554	-150.184	-30.857	0.000	0.000	0.000
0.000	-7.901	-150.184	11.622	-71.027	-15.689	0.000	0.000
0.000	0.000	-30.857	-71.027	7.787	-102.586	-0.717	0.000
0.000	0.000	0.000	-15.689	-102.586	6.144	-1.656	-19.449
0.000	0.000	0.000	0.000	-0.717	-1.656	0.048	-2.915
0.000	0.000	0.000	0.000	0.000	-19.449	-2.915	0.918

Pass 6

Estimated matrix A

0.970	0.017	0.000	0.000	0.000	0.000	0.000	0.000
0.028	0.871	0.003	0.001	0.000	0.000	0.000	0.000
0.002	0.108	0.982	0.196	0.005	0.001	0.000	0.000
0.000	0.002	0.011	0.680	0.014	0.000	0.000	0.000
0.000	0.000	0.000	0.019	0.390	0.007	0.000	0.000
0.000	0.001	0.000	0.010	0.088	0.306	0.004	0.000
0.000	0.001	0.000	0.018	0.134	0.156	0.545	0.000
0.000	0.000	0.003	0.075	0.369	0.529	0.450	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.898	0.010	0.000	0.000	0.000	0.000	0.000
0.000	0.102	0.987	0.904	0.000	0.000	0.000	0.000
0.000	0.000	0.003	0.082	0.089	0.000	0.000	0.000
0.000	0.000	0.000	0.014	0.735	0.202	0.000	0.000
0.000	0.000	0.000	0.000	0.177	0.796	0.023	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.976	1.000

Aggregate variance/covariance matrix for V

5.671	-53.970	-5.856	-0.452	-0.474	-0.016	-0.081	-0.071
-53.970	18.558	-94.971	-4.371	-0.566	-0.680	-0.733	-0.172
-5.856	-94.971	32.072	-104.433	-6.670	-2.122	-1.830	-8.370
-0.452	-4.371	-104.433	23.285	-41.808	-7.523	-6.881	-23.586
-0.474	-0.566	-6.670	-41.808	8.346	-37.712	-22.474	-57.375
-0.016	-0.680	-2.122	-7.523	-37.712	6.062	-30.608	-81.111
-0.081	-0.733	-1.830	-6.881	-22.474	-30.608	16.166	-141.109
-0.071	-0.172	-8.370	-23.586	-57.375	-81.111	-141.109	26.067

Aggregate variance/covariance matrix for W

0.001	-18.265	-2.652	0.000	0.000	0.000	0.000	0.000
-18.265	13.896	-215.106	-7.910	0.000	0.000	0.000	0.000
-2.652	-215.106	20.370	-155.858	-31.667	0.000	0.000	0.000
0.000	-7.910	-155.858	6.766	-71.915	-15.884	0.000	0.000
0.000	0.000	-31.667	-71.915	4.519	-105.404	-0.720	0.000
0.000	0.000	0.000	-15.884	-105.404	3.675	-1.668	-20.098
0.000	0.000	0.000	0.000	-0.720	-1.668	0.034	-2.934
0.000	0.000	0.000	0.000	0.000	-20.098	-2.934	0.667

Pass 7

Estimated matrix A

0.974	0.019	0.000	0.000	0.000	0.000	0.000	0.000
0.025	0.856	0.002	0.001	0.000	0.000	0.000	0.000
0.002	0.120	0.984	0.207	0.005	0.001	0.000	0.000
0.000	0.002	0.010	0.658	0.012	0.000	0.000	0.000
0.000	0.000	0.000	0.020	0.362	0.007	0.000	0.000
0.000	0.001	0.000	0.012	0.093	0.310	0.004	0.000
0.000	0.002	0.000	0.020	0.141	0.156	0.545	0.000
0.000	0.000	0.003	0.082	0.387	0.527	0.451	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.895	0.009	0.000	0.000	0.000	0.000	0.000
0.000	0.105	0.989	0.907	0.000	0.000	0.000	0.000
0.000	0.000	0.003	0.078	0.096	0.000	0.000	0.000
0.000	0.000	0.000	0.015	0.704	0.202	0.000	0.000
0.000	0.000	0.000	0.000	0.200	0.796	0.024	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.976	1.000

Aggregate variance/covariance matrix for V

5.263	-58.677	-6.374	-0.472	-0.482	-0.018	-0.087	-0.075
-58.677	17.817	-107.598	-4.579	-0.577	-0.755	-0.914	-0.181
-6.374	-107.598	28.221	-116.358	-7.050	-2.389	-2.212	-10.493
-0.472	-4.579	-116.358	18.052	-42.719	-8.036	-7.772	-27.171
-0.482	-0.577	-7.050	-42.719	4.327	-38.164	-23.148	-59.263
-0.018	-0.755	-2.389	-8.036	-38.164	3.692	-31.209	-82.892
-0.087	-0.914	-2.212	-7.772	-23.148	-31.209	11.098	-149.474
-0.075	-0.181	-10.493	-27.171	-59.263	-82.892	-149.474	17.754

Aggregate variance/covariance matrix for W

0.000	-18.265	-2.652	0.000	0.000	0.000	0.000	0.000
-18.265	12.770	-227.868	-7.918	0.000	0.000	0.000	0.000
-2.652	-227.868	17.540	-160.012	-32.291	0.000	0.000	0.000
0.000	-7.918	-160.012	4.736	-72.374	-15.999	0.000	0.000
0.000	0.000	-32.291	-72.374	2.655	-106.973	-0.722	0.000
0.000	0.000	0.000	-15.999	-106.973	2.159	-1.676	-20.565
0.000	0.000	0.000	0.000	-0.722	-1.676	0.026	-2.950
0.000	0.000	0.000	0.000	0.000	-20.565	-2.950	0.483

Pass 8

Estimated matrix A

0.977	0.022	0.000	0.000	0.000	0.000	0.000	0.000
0.021	0.850	0.002	0.001	0.000	0.000	0.000	0.000
0.002	0.123	0.985	0.213	0.004	0.000	0.000	0.000
0.000	0.002	0.009	0.635	0.011	0.000	0.000	0.000
0.000	0.000	0.000	0.023	0.359	0.007	0.000	0.000
0.000	0.001	0.000	0.014	0.098	0.322	0.004	0.000
0.000	0.002	0.000	0.023	0.141	0.154	0.545	0.000
0.000	0.000	0.003	0.091	0.387	0.517	0.450	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.898	0.009	0.000	0.000	0.000	0.000	0.000
0.000	0.102	0.989	0.911	0.000	0.000	0.000	0.000
0.000	0.000	0.002	0.072	0.100	0.000	0.000	0.000
0.000	0.000	0.000	0.018	0.688	0.201	0.000	0.000
0.000	0.000	0.000	0.000	0.212	0.797	0.026	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.974	1.000

Aggregate variance/covariance matrix for V

4.882	-63.021	-6.878	-0.488	-0.489	-0.020	-0.092	-0.077
-63.021	16.123	-118.939	-4.747	-0.588	-0.832	-1.088	-0.187
-6.878	-118.939	24.024	-125.672	-7.381	-2.641	-2.542	-12.445
-0.488	-4.747	-125.672	13.980	-43.410	-8.442	-8.445	-29.882
-0.489	-0.588	-7.381	-43.410	2.556	-38.399	-23.483	-60.210
-0.020	-0.832	-2.641	-8.442	-38.399	2.416	-31.576	-83.969
-0.092	-1.088	-2.542	-8.445	-23.483	-31.576	7.568	-155.157
-0.077	-0.187	-12.445	-29.882	-60.210	-83.969	-155.157	12.378

Aggregate variance/covariance matrix for W

0.000	-18.265	-2.652	0.000	0.000	0.000	0.000	0.000
-18.265	11.429	-239.289	-7.926	0.000	0.000	0.000	0.000
-2.652	-239.289	14.938	-163.004	-32.815	0.000	0.000	0.000
0.000	-7.926	-163.004	3.344	-72.647	-16.070	0.000	0.000
0.000	0.000	-32.815	-72.647	1.760	-107.935	-0.723	0.000
0.000	0.000	0.000	-16.070	-107.935	1.382	-1.681	-20.909
0.000	0.000	0.000	0.000	-0.723	-1.681	0.021	-2.965
0.000	0.000	0.000	0.000	0.000	-20.909	-2.965	0.359

Pass 9

Estimated matrix A

0.981	0.027	0.000	0.000	0.000	0.000	0.000	0.000
0.018	0.849	0.001	0.001	0.000	0.000	0.000	0.000
0.001	0.120	0.987	0.217	0.004	0.000	0.000	0.000
0.000	0.002	0.008	0.623	0.010	0.000	0.000	0.000
0.000	0.000	0.000	0.025	0.374	0.007	0.000	0.000
0.000	0.001	0.000	0.015	0.104	0.346	0.005	0.000
0.000	0.002	0.000	0.024	0.137	0.149	0.546	0.000
0.000	0.000	0.003	0.094	0.371	0.497	0.449	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.902	0.007	0.000	0.000	0.000	0.000	0.000
0.000	0.098	0.992	0.909	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.072	0.107	0.000	0.000	0.000
0.000	0.000	0.000	0.020	0.676	0.197	0.000	0.000
0.000	0.000	0.000	0.000	0.218	0.801	0.030	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.970	1.000

Aggregate variance/covariance matrix for V

4.556	-67.068	-7.357	-0.500	-0.495	-0.023	-0.098	-0.079
-67.068	14.142	-128.654	-4.882	-0.598	-0.908	-1.243	-0.192
-7.357	-128.654	20.087	-133.092	-7.668	-2.871	-2.810	-14.132
-0.500	-4.882	-133.092	10.834	-43.940	-8.758	-8.928	-31.819
-0.495	-0.598	-7.668	-43.940	1.737	-38.550	-23.679	-60.766
-0.023	-0.908	-2.871	-8.758	-38.550	1.741	-31.826	-84.684
-0.098	-1.243	-2.810	-8.928	-23.679	-31.826	5.185	-158.984
-0.079	-0.192	-14.132	-31.819	-60.766	-84.684	-158.984	8.729

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	9.802	-249.088	-7.929	0.000	0.000	0.000	0.000
-2.652	-249.088	12.099	-164.885	-33.235	0.000	0.000	0.000
0.000	-7.929	-164.885	2.129	-72.841	-16.122	0.000	0.000
0.000	0.000	-33.235	-72.841	1.279	-108.600	-0.724	0.000
0.000	0.000	0.000	-16.122	-108.600	0.995	-1.685	-21.183
0.000	0.000	0.000	0.000	-0.724	-1.685	0.021	-2.980
0.000	0.000	0.000	0.000	0.000	-21.183	-2.980	0.290

Pass 10

Estimated matrix A

0.984	0.031	0.000	0.000	0.000	0.000	0.000	0.000
0.015	0.849	0.001	0.001	0.000	0.000	0.000	0.000
0.001	0.115	0.987	0.215	0.003	0.000	0.000	0.000
0.000	0.001	0.008	0.602	0.009	0.000	0.000	0.000
0.000	0.000	0.000	0.031	0.392	0.008	0.000	0.000
0.000	0.001	0.000	0.019	0.109	0.368	0.006	0.000
0.000	0.002	0.000	0.026	0.132	0.146	0.547	0.000
0.000	0.000	0.003	0.105	0.355	0.478	0.447	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.908	0.008	0.000	0.000	0.000	0.000	0.000
0.000	0.092	0.991	0.911	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.064	0.106	0.000	0.000	0.000
0.000	0.000	0.000	0.025	0.676	0.197	0.000	0.000
0.000	0.000	0.000	0.000	0.218	0.801	0.035	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.965	1.000

Aggregate variance/covariance matrix for V

4.303	-70.892	-7.810	-0.511	-0.502	-0.026	-0.104	-0.080
-70.892	12.362	-136.855	-4.992	-0.608	-0.984	-1.382	-0.195
-7.810	-136.855	17.000	-139.072	-7.946	-3.104	-3.046	-15.753
-0.511	-4.992	-139.072	8.704	-44.388	-9.028	-9.306	-33.327
-0.502	-0.608	-7.946	-44.388	1.393	-38.665	-23.818	-61.163
-0.026	-0.984	-3.104	-9.028	-38.665	1.419	-32.015	-85.218
-0.104	-1.382	-3.046	-9.306	-23.818	-32.015	3.708	-161.605
-0.080	-0.195	-15.753	-33.327	-61.163	-85.218	-161.605	6.686

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	8.866	-257.953	-7.930	0.000	0.000	0.000	0.000
-2.652	-257.953	10.468	-166.101	-33.622	0.000	0.000	0.000
0.000	-7.930	-166.101	1.411	-72.994	-16.163	0.000	0.000
0.000	0.000	-33.622	-72.994	1.076	-109.135	-0.725	0.000
0.000	0.000	0.000	-16.163	-109.135	0.799	-1.688	-21.405
0.000	0.000	0.000	0.000	-0.725	-1.688	0.020	-2.997
0.000	0.000	0.000	0.000	0.000	-21.405	-2.997	0.237

Pass 11

Estimated matrix A

0.984	0.031	0.000	0.000	0.000	0.000	0.000	0.000
0.015	0.828	0.001	0.001	0.000	0.000	0.000	0.000
0.001	0.136	0.988	0.211	0.003	0.000	0.000	0.000
0.000	0.002	0.008	0.596	0.008	0.000	0.000	0.000
0.000	0.000	0.000	0.037	0.422	0.009	0.001	0.000
0.000	0.002	0.000	0.022	0.118	0.406	0.008	0.000
0.000	0.002	0.000	0.027	0.123	0.139	0.549	0.000
0.000	0.000	0.003	0.106	0.326	0.447	0.442	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.894	0.009	0.000	0.000	0.000	0.000	0.000
0.000	0.106	0.991	0.895	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.076	0.106	0.000	0.000	0.000
0.000	0.000	0.000	0.030	0.676	0.197	0.000	0.000
0.000	0.000	0.000	0.000	0.218	0.801	0.045	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.955	1.000

Aggregate variance/covariance matrix for V

3.980	-74.369	-8.285	-0.520	-0.510	-0.029	-0.109	-0.081
-74.369	12.027	-145.066	-5.098	-0.618	-1.070	-1.515	-0.198
-8.285	-145.066	15.908	-144.124	-8.216	-3.344	-3.249	-17.209
-0.520	-5.098	-144.124	7.169	-44.772	-9.259	-9.587	-34.433
-0.510	-0.618	-8.216	-44.772	1.211	-38.772	-23.931	-61.481
-0.029	-1.070	-3.344	-9.259	-38.772	1.276	-32.177	-85.665
-0.109	-1.515	-3.249	-9.587	-23.931	-32.177	2.736	-163.444
-0.081	-0.198	-17.209	-34.433	-61.481	-85.665	-163.444	5.170

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	9.016	-266.967	-7.932	0.000	0.000	0.000	0.000
-2.652	-266.967	10.500	-167.236	-33.973	0.000	0.000	0.000
0.000	-7.932	-167.236	1.313	-73.134	-16.198	0.000	0.000
0.000	0.000	-33.973	-73.134	0.972	-109.615	-0.725	0.000
0.000	0.000	0.000	-16.198	-109.615	0.723	-1.691	-21.609
0.000	0.000	0.000	0.000	-0.725	-1.691	0.023	-3.016
0.000	0.000	0.000	0.000	0.000	-21.609	-3.016	0.224

Pass 12

Estimated matrix A

0.985	0.034	0.000	0.000	0.000	0.000	0.000	0.000
0.013	0.810	0.001	0.001	0.000	0.000	0.000	0.000
0.001	0.150	0.988	0.202	0.002	0.000	0.000	0.000
0.000	0.002	0.007	0.592	0.008	0.000	0.000	0.000
0.000	0.000	0.000	0.044	0.459	0.010	0.001	0.000
0.000	0.002	0.000	0.026	0.126	0.454	0.011	0.000
0.000	0.003	0.000	0.028	0.113	0.130	0.554	0.000
0.000	0.000	0.003	0.107	0.292	0.406	0.434	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.888	0.010	0.000	0.000	0.000	0.000	0.000
0.000	0.112	0.990	0.872	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.092	0.106	0.000	0.000	0.000
0.000	0.000	0.000	0.036	0.679	0.200	0.000	0.000
0.000	0.000	0.000	0.000	0.215	0.798	0.061	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.939	1.000

Aggregate variance/covariance matrix for V

3.669	-77.527	-8.768	-0.528	-0.520	-0.033	-0.114	-0.082
-77.527	11.063	-152.648	-5.194	-0.630	-1.163	-1.635	-0.200
-8.768	-152.648	14.607	-148.579	-8.496	-3.599	-3.430	-18.579
-0.528	-5.194	-148.579	6.200	-45.128	-9.469	-9.808	-35.287
-0.520	-0.630	-8.496	-45.128	1.148	-38.884	-24.031	-61.759
-0.033	-1.163	-3.599	-9.469	-38.884	1.244	-32.333	-86.078
-0.114	-1.635	-3.430	-9.808	-24.031	-32.333	2.113	-164.774
-0.082	-0.200	-18.579	-35.287	-61.759	-86.078	-164.774	4.248

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	8.821	-275.786	-7.935	0.000	0.000	0.000	0.000
-2.652	-275.786	10.294	-168.371	-34.313	0.000	0.000	0.000
0.000	-7.935	-168.371	1.316	-73.278	-16.233	0.000	0.000
0.000	0.000	-34.313	-73.278	0.960	-110.091	-0.726	0.000
0.000	0.000	0.000	-16.233	-110.091	0.720	-1.694	-21.817
0.000	0.000	0.000	0.000	-0.726	-1.694	0.031	-3.043
0.000	0.000	0.000	0.000	0.000	-21.817	-3.043	0.235

Pass 13

Estimated matrix A

0.987	0.038	0.000	0.000	0.000	0.000	0.000	0.000
0.011	0.792	0.001	0.001	0.000	0.000	0.000	0.000
0.001	0.161	0.988	0.198	0.002	0.000	0.000	0.000
0.000	0.002	0.007	0.591	0.008	0.000	0.000	0.000
0.000	0.000	0.000	0.049	0.478	0.010	0.001	0.000
0.000	0.003	0.001	0.029	0.131	0.481	0.015	0.000
0.000	0.003	0.000	0.028	0.107	0.125	0.559	0.000
0.000	0.000	0.003	0.106	0.274	0.383	0.426	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.862	0.005	0.000	0.000	0.000	0.000	0.000
0.000	0.138	0.994	0.859	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.101	0.106	0.000	0.000	0.000
0.000	0.000	0.000	0.040	0.679	0.200	0.000	0.000
0.000	0.000	0.000	0.000	0.215	0.797	0.076	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.924	1.000

Aggregate variance/covariance matrix for V

3.322	-80.349	-9.235	-0.537	-0.532	-0.039	-0.121	-0.082
-80.349	9.858	-159.330	-5.289	-0.644	-1.276	-1.765	-0.201
-9.235	-159.330	13.302	-152.690	-8.778	-3.870	-3.601	-19.897
-0.537	-5.289	-152.690	5.593	-45.448	-9.658	-9.988	-35.978
-0.532	-0.644	-8.778	-45.448	1.102	-39.002	-24.126	-62.020
-0.039	-1.276	-3.870	-9.658	-39.002	1.260	-32.490	-86.485
-0.121	-1.765	-3.601	-9.988	-24.126	-32.490	1.760	-165.794
-0.082	-0.201	-19.897	-35.978	-62.020	-86.485	-165.794	3.697

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	6.941	-282.725	-7.936	0.000	0.000	0.000	0.000
-2.652	-282.725	8.373	-169.483	-34.634	0.000	0.000	0.000
0.000	-7.936	-169.483	1.291	-73.422	-16.266	0.000	0.000
0.000	0.000	-34.634	-73.422	0.946	-110.571	-0.727	0.000
0.000	0.000	0.000	-16.266	-110.571	0.723	-1.697	-22.022
0.000	0.000	0.000	0.000	-0.727	-1.697	0.038	-3.077
0.000	0.000	0.000	0.000	0.000	-22.022	-3.077	0.239

Pass 14

Estimated matrix A

0.988	0.040	0.000	0.000	0.000	0.000	0.000	0.000
0.010	0.767	0.001	0.001	0.000	0.000	0.000	0.000
0.001	0.183	0.989	0.196	0.002	0.000	0.000	0.000
0.000	0.003	0.007	0.591	0.008	0.000	0.000	0.000
0.000	0.000	0.001	0.054	0.503	0.011	0.002	0.000
0.000	0.004	0.001	0.032	0.141	0.524	0.025	0.000
0.000	0.003	0.000	0.027	0.099	0.118	0.573	0.000
0.000	0.000	0.003	0.100	0.247	0.347	0.400	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.840	0.005	0.000	0.000	0.000	0.000	0.000
0.000	0.160	0.995	0.859	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.097	0.109	0.000	0.000	0.000
0.000	0.000	0.000	0.044	0.673	0.199	0.000	0.000
0.000	0.000	0.000	0.000	0.219	0.799	0.116	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.883	1.000

Aggregate variance/covariance matrix for V

2.969	-82.814	-9.704	-0.545	-0.547	-0.045	-0.126	-0.083
-82.814	8.838	-165.378	-5.377	-0.658	-1.391	-1.873	-0.202
-9.704	-165.378	12.258	-156.537	-9.064	-4.154	-3.755	-21.067
-0.545	-5.377	-156.537	5.146	-45.753	-9.841	-10.138	-36.541
-0.547	-0.658	-9.064	-45.753	1.084	-39.132	-24.218	-62.262
-0.045	-1.391	-4.154	-9.841	-39.132	1.306	-32.664	-86.899
-0.126	-1.873	-3.755	-10.138	-24.218	-32.664	1.488	-166.599
-0.083	-0.202	-21.067	-36.541	-62.262	-86.899	-166.599	3.197

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	6.549	-289.274	-7.937	0.000	0.000	0.000	0.000
-2.652	-289.274	7.726	-170.342	-34.952	0.000	0.000	0.000
0.000	-7.937	-170.342	1.040	-73.566	-16.302	0.000	0.000
0.000	0.000	-34.952	-73.566	0.969	-111.077	-0.727	0.000
0.000	0.000	0.000	-16.302	-111.077	0.799	-1.700	-22.278
0.000	0.000	0.000	0.000	-0.727	-1.700	0.063	-3.137
0.000	0.000	0.000	0.000	0.000	-22.278	-3.137	0.316

Pass 15

Estimated matrix A

0.989	0.044	0.000	0.000	0.000	0.000	0.000	0.000
0.009	0.749	0.001	0.000	0.000	0.000	0.000	0.000
0.001	0.189	0.984	0.127	0.001	0.000	0.000	0.000
0.000	0.004	0.008	0.550	0.006	0.000	0.000	0.000
0.000	0.001	0.001	0.084	0.514	0.011	0.002	0.000
0.000	0.008	0.001	0.051	0.144	0.538	0.031	0.000
0.000	0.007	0.000	0.040	0.097	0.115	0.582	0.000
0.000	0.000	0.005	0.148	0.239	0.336	0.385	1.000

Estimated matrix C

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.000	0.845	0.009	0.000	0.000	0.000	0.000	0.000
0.000	0.155	0.991	0.763	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.162	0.109	0.000	0.000	0.000
0.000	0.000	0.000	0.075	0.673	0.199	0.000	0.000
0.000	0.000	0.000	0.000	0.219	0.799	0.136	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.863	1.000

Aggregate variance/covariance matrix for V

2.650	-84.963	-10.146	-0.554	-0.575	-0.056	-0.136	-0.083
-84.963	7.561	-170.313	-5.471	-0.681	-1.582	-2.040	-0.204
-10.146	-170.313	12.796	-160.398	-9.560	-4.686	-4.008	-23.344
-0.554	-5.471	-160.398	5.501	-46.150	-10.081	-10.327	-37.251
-0.575	-0.681	-9.560	-46.150	1.507	-39.289	-24.326	-62.560
-0.056	-1.582	-4.686	-10.081	-39.289	1.800	-32.860	-87.372
-0.136	-2.040	-4.008	-10.327	-24.326	-32.860	1.639	-167.314
-0.083	-0.204	-23.344	-37.251	-62.560	-87.372	-167.314	4.474

Aggregate variance/covariance matrix for W

0.000	-18.266	-2.652	0.000	0.000	0.000	0.000	0.000
-18.266	6.994	-296.265	-7.940	0.000	0.000	0.000	0.000
-2.652	-296.265	8.731	-171.617	-35.416	0.000	0.000	0.000
0.000	-7.940	-171.617	1.565	-73.808	-16.348	0.000	0.000
0.000	0.000	-35.416	-73.808	1.374	-111.744	-0.728	0.000
0.000	0.000	0.000	-16.348	-111.744	1.015	-1.703	-22.576
0.000	0.000	0.000	0.000	-0.728	-1.703	0.080	-3.213
0.000	0.000	0.000	0.000	0.000	-22.576	-3.213	0.373

Appendix IID

HMM Implementation Results for the Modified Sample

The following pages present the implementation results for the modified Standard & Poor's rating sample, where all firms were reclassified as investment-grade, speculative grade or default/NR. As in Appendix IIC, for each of the passes through the data set, we are given the estimates for matrices A and C, as well as $VarV_k$ and $VarW_k$.

Initial matrix A

0.900	0.020	0.000
0.100	0.900	0.000
0.000	0.080	1.000

Initial matrix C

0.500	0.300	0.000
0.500	0.500	0.500
0.000	0.200	0.500

Pass 1

Estimated matrix A

0.852	0.012	0.000
0.148	0.869	0.000
0.000	0.119	1.000

Estimated matrix C

0.601	0.230	0.000
0.399	0.255	0.165
0.000	0.515	0.835

Aggregate variance/covariance for V

60.050	-59.413	-0.636
-59.413	104.280	-44.866
-0.636	-44.866	45.502

Aggregate variance/covariance for W

167.976	-115.697	-52.279
-115.697	240.447	-124.751
-52.279	-124.751	177.030

Pass 2

Estimated matrix A

0.887	0.011	0.000
0.113	0.831	0.000
0.000	0.157	1.000

Estimated matrix C

0.654	0.149	0.000
0.346	0.189	0.079
0.000	0.662	0.921

Aggregate variance/covariance for V

42.506	-101.137	-1.419
-101.137	99.426	-102.568
-1.419	-102.568	58.485

Aggregate variance/covariance for W

128.274	-203.542	-92.707
-203.542	179.297	-216.202
-92.707	-216.202	131.879

Pass 3

Estimated matrix A

0.902	0.008	0.000
0.098	0.784	0.000
0.000	0.208	1.000

Estimated matrix C

0.746	0.073	0.000
0.254	0.124	0.035
0.000	0.803	0.965

Aggregate variance/covariance for V

33.080	-133.552	-2.084
-133.552	99.219	-169.373
-2.084	-169.373	67.470

Aggregate variance/covariance for W

82.227	-264.964	-113.513
-264.964	118.167	-272.947
-113.513	-272.947	77.551

Pass 4

Estimated matrix A

0.899	0.005	0.000
0.101	0.756	0.000
0.000	0.239	1.000

Estimated matrix C

0.867	0.034	0.000
0.133	0.080	0.022
0.000	0.886	0.978

Aggregate variance/covariance for V

29.583	-162.752	-2.467
-162.752	93.163	-233.336
-2.467	-233.336	64.347

Aggregate variance/covariance for W

41.805	-297.808	-122.474
-297.808	69.485	-309.588
-122.474	-309.588	45.602

Pass 5

Estimated matrix A

0.890	0.003	0.000
0.110	0.747	0.000
0.000	0.250	1.000

Estimated matrix C

0.936	0.025	0.000
0.065	0.067	0.018
0.000	0.908	0.982

Aggregate variance/covariance for V

28.135	-190.660	-2.694
-190.660	83.708	-289.135
-2.694	-289.135	56.026

Aggregate variance/covariance for W

21.227	-313.212	-128.296
-313.212	44.748	-338.932
-128.296	-338.932	35.166

Pass 6

Estimated matrix A

0.884	0.002	0.000
0.116	0.741	0.000
0.000	0.257	1.000

Estimated matrix C

0.953	0.023	0.000
0.048	0.061	0.014
0.000	0.916	0.986

Aggregate variance/covariance for V

25.982	-216.491	-2.845
-216.491	74.134	-337.438
-2.845	-337.438	48.453

Aggregate variance/covariance for W

14.896	-323.478	-132.926
-323.478	34.665	-363.331
-132.926	-363.331	29.029

Pass 7

Estimated matrix A

0.880	0.002	0.000
0.121	0.736	0.000
0.000	0.262	1.000

Estimated matrix C

0.955	0.023	0.000
0.045	0.060	0.013
0.000	0.918	0.988

Aggregate variance/covariance for V

23.715	-240.099	-2.952
-240.099	65.370	-379.200
-2.952	-379.200	41.870

Aggregate variance/covariance for W

12.479	-332.103	-136.780
-332.103	30.162	-384.868
-136.780	-384.868	25.391

Pass 8

Estimated matrix A

0.875	0.002	0.000
0.125	0.734	0.000
0.000	0.265	1.000

Estimated matrix C

0.953	0.023	0.000
0.047	0.061	0.012
0.000	0.916	0.988

Aggregate variance/covariance for V

21.475	-261.492	-3.034
-261.492	57.494	-415.301
-3.034	-415.301	36.183

Aggregate variance/covariance for W

11.234	-339.940	-140.177
-339.940	28.051	-405.083
-140.177	-405.083	23.612

Pass 9

Estimated matrix A

0.870	0.002	0.000
0.130	0.734	0.000
0.000	0.264	1.000

Estimated matrix C

0.950	0.026	0.000
0.050	0.067	0.012
0.000	0.907	0.988

Aggregate variance/covariance for V

19.459	-280.885	-3.101
-280.885	50.568	-446.476
-3.101	-446.476	31.242

Aggregate variance/covariance for W

10.599	-347.213	-143.503
-347.213	28.046	-425.856
-143.503	-425.856	24.098

Pass 10

Estimated matrix A

0.859	0.002	0.000
0.141	0.734	0.000
0.000	0.265	1.000

Estimated matrix C

0.943	0.028	0.000
0.057	0.072	0.013
0.000	0.899	0.987

Aggregate variance/covariance for V

18.118	-298.949	-3.155
-298.949	45.245	-473.657
-3.155	-473.657	27.235

Aggregate variance/covariance for W

10.231	-354.292	-146.655
-354.292	28.201	-446.978
-146.655	-446.978	24.274

Pass 11

Estimated matrix A

0.849	0.002	0.000
0.151	0.737	0.000
0.000	0.262	1.000

Estimated matrix C

0.934	0.033	0.000
0.066	0.083	0.015
0.000	0.884	0.986

Aggregate variance/covariance for V

16.550	-315.452	-3.202
-315.452	40.336	-497.490
-3.202	-497.490	23.880

Aggregate variance/covariance for W

10.146	-361.257	-149.835
-361.257	30.618	-470.630
-149.835	-470.630	26.833

Pass 12

Estimated matrix A

0.827	0.002	0.000
0.173	0.742	0.000
0.000	0.256	1.000

Estimated matrix C

0.917	0.039	0.000
0.083	0.099	0.018
0.000	0.862	0.982

Aggregate variance/covariance for V

15.694	-331.103	-3.244
-331.103	36.640	-518.478
-3.244	-518.478	21.031

Aggregate variance/covariance for W

10.609	-368.485	-153.217
-368.485	35.916	-499.319
-153.217	-499.319	32.070

Pass 13

Estimated matrix A

0.798	0.002	0.000
0.202	0.745	0.000
0.000	0.253	1.000

Estimated matrix C

0.897	0.043	0.000
0.104	0.109	0.021
0.000	0.848	0.979

Aggregate variance/covariance for V

14.610	-345.676	-3.281
-345.676	33.583	-537.488
-3.281	-537.488	19.047

Aggregate variance/covariance for W

10.495	-375.585	-156.612
-375.585	38.651	-530.870
-156.612	-530.870	34.945

Pass 14

Estimated matrix A

0.774	0.002	0.000
0.226	0.753	0.000
0.000	0.246	1.000

Estimated matrix C

0.874	0.050	0.000
0.126	0.126	0.028
0.000	0.824	0.972

Aggregate variance/covariance for V

12.719	-358.360	-3.317
-358.360	29.933	-554.738
-3.317	-554.738	17.285

Aggregate variance/covariance for W

10.231	-382.258	-160.169
-382.258	46.890	-571.087
-160.169	-571.087	43.775

Pass 15

Estimated matrix A

0.718	0.002	0.000
0.282	0.755	0.000
0.000	0.244	1.000

Estimated matrix C

0.830	0.052	0.000
0.170	0.132	0.029
0.000	0.816	0.971

Aggregate variance/covariance for V

11.399	-369.728	-3.348
-369.728	27.278	-570.649
-3.348	-570.649	15.942

Aggregate variance/covariance for W

9.663	-388.470	-163.620
-388.470	48.522	-613.398
-163.620	-613.398	45.761

Chapter 3

Kalman Filtering Model

1. INTRODUCTION

One of the challenges in implementing the Hidden Markov Model of credit rating evolution over time is the nature of the rating data. In general, the model requires many observed transitions between rating categories. Since individual firms experience few rating changes within a narrow range, it is necessary to consider an aggregate of firms in the data set rather than an individual company. Another characteristic of the rating data is that the dataset contains many transitions to the NR (not rated) status. Accounting for the information content of transitions to and from NR poses a challenge, as the specific reasons behind the "missing" ratings are not known. A potential solution to this problem comes in the form of a continuous time approximation to the rating dynamics. By considering a Gaussian approximation to the evolution of an average company rating over time, we transform the discrete time model with a Markov Chain observed in martingale noise to a linear

Kalman filtering problem. In this context, the "true" average credit rating is a noisy signal observed through Gaussian noise. There are benefits to this approach. First, we can draw from a large body of existing research into Kalman filtering and its applications. Second, we are able to use all available ratings and account for transitions to and from the NR status implicitly by considering the dynamics of the average rating. As before, we use the EM algorithm to estimate parameters of the model.

2. GAUSSIAN APPROXIMATION TO CREDIT RATING DYNAMICS

Recall from Chapter 1 that X a discrete-time, finite-state, time homogeneous Markov chain with the state space $\{e_1, e_2, \ldots, e_N\}$, $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T \in \mathbb{R}^N$. Its semimartingale representation is

$$X_{k+1} = AX_k + V_{k+1}, \quad k = 0, 1, \ldots,$$

where V_{k+1} is a martingale increment with $E[V_{k+1}|\mathcal{F}_k] = 0 \in \mathbb{R}^N$. Recall that $\mathcal{F}_k = \sigma\{X_0, X_1, \ldots, X_k\}$ is the σ-*field* containing all the information about the process X up to and including time k. Let $p_k = (p_1, \ldots, p_N)' = E[X_k]$. Then, $p_{k+1} = Ap_k = A^{k+1}p_0 \in \mathbb{R}^N$ and $Var V_k = E[V_k V_k'] = \text{diag}\,(Ap_{k-1}) - A\,\text{diag}\,(p_{k-1})A'$.

We suppose that we do not observe X directly. Rather, we observe a process Y with state space $\{f_1, f_2, \ldots, f_M\}$, $f_j = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^M$. The semimartingale representation of Y is

$$Y_k = CX_k + W_k, \quad k = 0, 1, \ldots,$$

where W is a martingale increment with $E[W_k|\mathcal{G}_{k-1} \vee \{X_k\}] = 0 \in \mathbb{R}^M$ and

$$\mathcal{G}_k = \sigma\{X_0, \ldots, X_k, Y_0, \ldots, Y_k\}$$

represents possible histories of both processes X and Y. We have

$$VarW_k = E[W_k W_k'] = E[(Y_k - CX_k)(Y_k - CX_k)'] = \text{diag}\,(Cp_k) - C\,\text{diag}\,(p_k)\,C'.$$

Recall that we take the process X to represent a firm's "true" credit quality and Y to be the noisy observations given by the posted credit labels. However, individual firms generally experience few rating changes and within a narrow range, and so we consider an aggregate of firms in our data set, described in Chapter 2, rather than an individual company. The number of observed rating changes is further limited by many transitions to the NR (not rated) status, precise reasons for which are generally not known. We shall therefore consider an average "true" rating and an average observed rating label and then apply Kalman filtering to the Gaussian analogs of these averaged processes. Specifically, suppose that we have rating data for L firms, and so we consider L independent signal processes X, with L independent observation processes Y such that for $l = 1, 2, \ldots, L$,

$$X_{k+1}^l = AX_k^l + V_{k+1}^l$$
$$Y_k^l = CX_k^l + W_k^l.$$

Write

$$X_k = \frac{1}{L}\sum_{l=1}^{L} X_k^l, \quad V_k = \frac{1}{L}\sum_{l=1}^{L} V_k^l$$

and

$$Y_k = \frac{1}{L}\sum_{l=1}^{L} Y_k^l, \quad W_k = \frac{1}{L}\sum_{l=1}^{L} W_k^l.$$

Then, the dynamics of the averaged X and Y are $X_{k+1} = AX_k + V_{k+1}$, and $Y_k = CX_k + W_k$, respectively.

Following Krichagina, Lipster and Rubinovich [22] we note that an optimal linear filter for a non-Gaussian process is the optimal linear filter for a Gaussian analog of the original process, i.e. for a Gaussian process with the same mathematical expectation and correlation. The Gaussian analog of the averaged process X is then the vector process

$$x_{k+1} = Ax_k + v_{k+1},$$

where v_k, $k = 1, 2, \ldots$, is the sequence of independent Gaussian vectors with $E[v_k] = 0$ and covariance matrix $Q_k = \text{diag}\,(Ap_{k-1}) - A\,\text{diag}\,(p_{k-1})A'$, and x_0 is a Gaussian vector independent of v_k with $E[x_0] = p_0$, and covariance Q_0 such that $Q_0(i, i) = p_0^i(1 - p_0^j)$, $Q_0(i, j) = -p_0^i p_0^j$ for $i \neq j$. Similarly, the Gaussian analog of the averaged process Y is

$$y_k = Cx_k + w_k,$$

where w_k, $k = 1, 2, \ldots$, is the sequence of independent Gaussian vectors with $E[w_k] = 0$ and covariance matrix $R_k = \text{diag}\,(Ap_{k-1}) - A\,\text{diag}\,(p_{k-1})A'$, independent of X. The Kalman filter for this dynamical system is described next.

3. SQUARE-ROOT KALMAN FILTERING ALGORITHM

Assume that state and observation processes are given by the linear dynamics

$$x_{k+1} = Ax_k + v_{k+1} \in \mathbb{R}^N,$$

$$y_k = Cx_k + w_k \in \mathbb{R}^M,$$

where A and C are matrices of appropriate dimensions, v_k and w_k are normally distributed with means zero and respective covariance matrices Q_k and R_k, assumed nonsingular.

Write

$$\mathcal{Y}_k = \sigma\{y_0, y_1, \ldots, y_k\}$$

and define

$$\hat{x}_{k,k-1} = E[x_k|\mathcal{Y}_{k-1}] \qquad \textit{(a priori state estimate)}$$

$$\Sigma_{k,k-1} = E[(x_k - \hat{x}_{k,k-1})(x_k - \hat{x}_{k,k-1})'|\mathcal{Y}_{k-1}] \qquad \textit{(a priori error covariance)}$$

$$\hat{x}_{k,k} = E[x_k|\mathcal{Y}_k] \qquad \textit{(a posteriori state estimate)}$$

$$\Sigma_{k,k} = E[(x_k - \hat{x}_{k,k})(x_k - \hat{x}_{k,k})'|\mathcal{Y}_k] \qquad \textit{(a posteriori error covariance)}$$

The Kalman filter uses a form of feedback control: the filter estimates the process state at some time and then obtains feedback in the form of noisy measurements. The equations for the Kalman filter therefore fall into two groups: *time update* equations and *measurement update* equations. The time update equations are responsible for projecting forward in time the current state and error covariance estimates to obtain *a priori* estimates for the next time step. The measurement update equations are responsible for the feedback, i.e. for incorporating a new measurement into the *a priori* estimate to obtain an improved *a posteriori* estimate. After each time and measurement update pair, the process is repeated with the previous *a posteriori* estimates used to predict the new *a priori* estimates. Kalman filter is therefore recursive, which is one of its very appealing features.

The time update equations are, $k = 1, 2, \ldots$:

$$\hat{x}_{k,k-1} = A\hat{x}_{k-1,k-1}$$

$$\Sigma_{k,k-1} = A\Sigma_{k-1,k-1}A' + Q_k.$$

The measurement update equations are, $k = 1, 2, \ldots$:

$$K_k = \Sigma_{k,k-1}C'(C\Sigma_{k,k-1}C' + R_k)^{-1}$$

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k(y_k - C\hat{x}_{k,k-1})$$

$$\Sigma_{k,k} = (I - K_kC)\Sigma_{k,k-1}$$

Note that the only time-consuming operation in the Kalman filtering process is the computation of the *Kalman gain matrix* K_k. In the actual implementation of the filter, it is important to be able to compute K_k efficiently, preferably without directly inverting the matrix $(C\Sigma_{k,k-1}C' + R_k)$ at each time step. We therefore turn to the so called *square-root algorithm*, which only requires inversion of triangular matrices and improves the computational accuracy by working with the square root of possibly very large or very small numbers.

First we note without proof the following result from linear algebra:

Lemma 3.1. *For any positive definite symmetric marix A, there is a unique lower trangular matrix $A^{1/2}$ such that $A = A^{1/2}(A^{1/2})'$. More generally, for any $n \times (n+p)$ matrix A, there is an $n \times n$ matrix $\tilde{A}$ such that $\tilde{A}\tilde{A}' = AA'$.* $\square$

$A^{1/2}$ has the property of being a "square root" of A, and since it is lower triangular, its inverse can be computed more efficiently. The factorization of a matrix into the product of a lower triangular matrix and its transpose is usually done by a scheme known as *Cholesky decomposition*, described in Appendix IIIA.

Define $H_k := (C\,\Sigma_{k,k-1}C' + R_k)^{1/2}$. Let $S_{0,0} = Q_0^{1/2}$ and $S_{k,k-1}$ be the square root of the matrix $(AS_{k-1,k-1}Q_{k-1}^{1/2})(AS_{k-1,k-1}Q_{k-1}^{1/2})'$, and

$$S_{k,k} = S_{k,k-1}(I - S'_{k,k-1}C'(H'_k)^{-1}(H_k + R_k^{1/2})^{-1}CS_{k,k-1})$$

for $k = 1, 2, \ldots$. The auxiliary matrices $S_{k,k-1}$ and $S_{k,k}$ are also square roots, although they are not necessarily lower trangular nor positive definite:

Theorem 3.1. $S_{0,0}S'_{0,0} = \Sigma_{0,0}$, *and for* $k = 1, 2, \ldots,$

$$S_{k,k-1}S'_{k,k-1} = \Sigma_{k,k-1}$$

$$S_{k,k}S'_{k,k} = \Sigma_{k,k}.$$

Proof See Appendix IIIB.

The *square-root Kalman filtering algorithm* can then be stated as follows:

(i) Compute $S_{0,0} = Q_0^{1/2}$.

(ii) For $k = 1, 2, \ldots$, compute $S_{k,k-1}$, a square root of the matrix

$$(AS_{k-1,k-1}Q_{k-1}^{1/2})(AS_{k-1,k-1}Q_{k-1}^{1/2})',$$

and the matrix

$$H_k = (C\,S_{k,k-1}S'_{k,k-1}C' + R_k)^{1/2},$$

and then compute

$$S_{k,k} = S_{k,k-1}(I - S'_{k,k-1}C'(H'_k)^{-1}(H_k + R_k^{1/2})^{-1}CS_{k,k-1}).$$

(iii) Compute $\hat{x}_0 = p_0$, and for $k = 1, 2, \ldots$, using the information from (ii), compute

$$K_k = S_{k,k-1}S'_{k,k-1}C'(H'_k)^{-1}H_k^{-1}.$$

Then compute

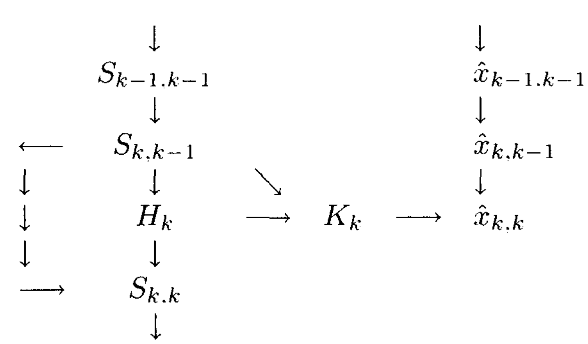
$$\hat{x}_{k,k-1} = A\hat{x}_{k-1,k-1}$$

and

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k(y_k - C\hat{x}_{k,k-1}).$$

We again remark that we only have to invert triangular matrices, and in addition, these matrices are square roots of ones that might have very small or very large entries. The algorithm is illustrated in the following figure:

Figure 3.1.



In some cases state constraints may have to be incorporated into the structure of the Kalman filter. When x_k is taken to represent the average rating, its components should add up to one. The required state constraint is then $\mathbf{1}'x_k = 1$ and we have the following:

Proposition 3.1. *Given the unconstrained state estimate $\hat{x}_{k,k}$, the constrained state estimate is $\tilde{x}_{k,k} = \hat{x}_{k,k} - \Sigma_{k,k}\mathbf{1}(\mathbf{1}'\Sigma_{k,k}\mathbf{1})^{-1}(\mathbf{1}'\hat{x}_{k,k} - 1)$.*

Proof. See Simon and Chia [30].

4. PARAMETER ESTIMATION

Consider the following time-invariant Gaussian state-space model:

$$x_{k+1} = Ax_k + v_{k+1} \in \mathbb{R}^N, \qquad v_k \sim N(0, Q)$$

$$y_k = Cx_k + w_k, \qquad w_k \sim N(0, R),$$

with initial state mean μ and initial state covariance Σ. We are interested in computing the MLE of the parameter $\theta = (\mu, \Sigma, A, C, R, Q; \mathbf{1}'A = \mathbf{1}', \mathbf{1}'C = \mathbf{1}', \mathbf{1}'\mu = 1)$ given the observation sequence $y_1, \ldots, y_T$. Note that we require matrices A and C to have columns that add up to one. We use the EM algorithm described in Chapter 1 to compute the MLE.

Step 1 *(E-step)*: The joint log likelihood of the complete data $x_0, x_1, \ldots, x_T$, and

$y_1, \ldots, y_T$ can be written in the form

$$\begin{aligned}\log L = &-\frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_0-\mu)'\Sigma^{-1}(x_0-\mu)\\ &-\frac{T}{2}\log|Q| - \frac{1}{2}\sum_{k=1}^{T}(x_k - Ax_{k-1})'Q^{-1}(x_k - Ax_{k-1})\\ &-\frac{T}{2}\log|R| - \frac{1}{2}\sum_{k=1}^{T}(y_k - Cx_k)'R^{-1}(y_k - Cx_k)\end{aligned}$$

so that

$$\begin{aligned}\mathcal{Q}(\theta,\hat{\theta}_j) = &-\frac{1}{2}\log|\Sigma| - \frac{1}{2}E_{\hat{\theta}_j}\left[(x_0-\mu)'\Sigma^{-1}(x_0-\mu)|\mathcal{Y}_T\right]\\ &-\frac{T}{2}\log|Q| - \frac{1}{2}E_{\hat{\theta}_j}\left[\sum_{k=1}^{T}(x_k - Ax_{k-1})'Q^{-1}(x_k - Ax_{k-1})|\mathcal{Y}_T\right]\\ &-\frac{T}{2}\log|R| - \frac{1}{2}E_{\hat{\theta}_j}\left[\sum_{k=1}^{T}(y_k - Cx_k)'R^{-1}(y_k - Cx_k)|\mathcal{Y}_T\right]\\ &+E_{\hat{\theta}_j}[R(\hat{\theta}_j)|\mathcal{Y}_T],\end{aligned}$$

where $\hat{\theta}_j$ denotes the parameter estimate at the j-th iteration and the term $R(\hat{\theta}_j)$ does not involve θ. Define:

$$\begin{aligned}\hat{x}_{k.T} &= E_{\hat{\theta}_j}[x_k|\mathcal{Y}_T]\\ \Sigma_{k.T} &= E_{\hat{\theta}_j}[(x_k - \hat{x}_{k.T})(x_k - \hat{x}_{k.T})'|\mathcal{Y}_T]\\ \Sigma^T_{k.k-1} &= E_{\hat{\theta}_j}[(x_k - \hat{x}_{k.T})(x_{k-1} - \hat{x}_{k-1.T})'|\mathcal{Y}_T].\end{aligned}$$

Note that the random vector $\hat{x}_{k.k}$ is the usual Kalman filter estimator, whereas $\hat{x}_{k.T}$ is the minimum mean square error smoothed estimator of x_k based on all of the observed data.

Set also

$$F = \sum_{k=1}^{T} (\Sigma_{k-1,T} + \hat{x}_{k-1,T}\hat{x}'_{k-1,T})$$
$$G = \sum_{k=1}^{T} (\Sigma^{T}_{k,k-1} + \hat{x}_{k,T}\hat{x}'_{k-1,T})$$
$$U = \sum_{k=1}^{T} (\Sigma_{k,T} + \hat{x}_{k,T}\hat{x}'_{k,T}).$$

Taking the expectations, we can rewrite $\mathcal{Q}(\theta, \hat{\theta}_j)$ as

$$\begin{aligned}\mathcal{Q}(\theta, \hat{\theta}_j) = &- \frac{1}{2}\log|\Sigma| - \frac{1}{2} tr\left[\Sigma^{-1}(\Sigma_{0,T} + (\hat{x}_{0,T} - \mu)(\hat{x}_{0,T} - \mu)')\right] \\ &- \frac{T}{2}\log|Q| - \frac{1}{2} tr\left[Q^{-1}(U - GA' - AG' + AFA')\right] \\ &- \frac{T}{2}\log|R| - \frac{1}{2} tr\left[R^{-1}\sum_{k=1}^{T}((y_k - C\hat{x}_{k,T})(y_k - C\hat{x}_{k,T})' + C\,\Sigma_{k,T}C')\right] \\ &+ E_{\hat{\theta}_j}[R(\hat{\theta}_j)|\mathcal{Y}_T].\end{aligned}$$

In order to calculate $\hat{x}_{k,T}$ and $\Sigma_{k,T}$, one performs the set of backward recursions for $k = T, T-1, \ldots, 1$ (*cf.* Shumway and Stoffer [29]):

$$\hat{x}_{k-1,T} = \hat{x}_{k-1,k-1} + \Sigma_{k-1,k-1}A'\Sigma^{-1}_{k,k-1}(\hat{x}_{k,T} - A\hat{x}_{k-1,k-1})$$

$$\Sigma_{k-1,T} = \Sigma_{k-1,k-1} + \Sigma_{k-1,k-1}A'\,\Sigma^{-1}_{k,k-1}(\Sigma_{k,T} - \Sigma_{k,k-1})(\Sigma_{k-1,k-1}A'\,\Sigma^{-1}_{k,k-1})'.$$

The covariance $\Sigma^{T}_{k,k-1}$ can be calculated using the backwards recursions (*cf.* Shumway and Stoffer [29])

$$\begin{aligned}\Sigma^{T}_{k-1,k-2} = &\Sigma_{k-1,k-1}(\Sigma_{k-1,k-2}A'\Sigma^{-1}_{k-1,k-2})'\Sigma_{k-1,k-1}A'\Sigma^{-1}_{k,k-1} \\ &\times (\Sigma^{T}_{k,k-1} - A\Sigma_{k-1,k-1})(\Sigma_{k-2,k-2}A'\Sigma^{-1}_{k-1,k-2})'\end{aligned}$$

for $k = T, T-1, \ldots, 2$, where

$$\Sigma^T_{T,T-1} = (I - K_T C) A \, \Sigma_{T-1,T-1}.$$

Step 2 *(M-step)* To implement the M-step, i.e. to compute

$$\hat{\theta}_{j+1} = \text{argmax}_{\theta \in \Theta} \mathcal{Q}(\theta, \hat{\theta}_j),$$

we set the derivatives $\partial \mathcal{Q} / \partial \theta = 0$.

Proposition 3.2. *Define* $Z = \sum_{k=1}^T y_k \hat{x}'_{k,T}$. *The EM parameter estimates are given by*

$$\hat{A} = GF^{-1} + Q\mathbf{1}(\mathbf{1}'Q\mathbf{1})^{-1}(\mathbf{1}' - \mathbf{1}'GF^{-1}) \qquad \textit{State transition matrix}$$

$$\hat{Q} = \frac{1}{T}(U - GA' - AG' + AFA') \qquad \textit{State noise covariance}$$

$$\hat{C} = ZU^{-1} + R\mathbf{1}(\mathbf{1}'R\mathbf{1})^{-1}(\mathbf{1}' - \mathbf{1}'ZU^{-1}) \qquad \textit{Observation matrix}$$

$$\hat{R} = \frac{1}{T}\sum_{k=1}^{T}[(y_k - C\hat{x}_{k,T})(y_k - C\hat{x}_{k,T})' + C\,\Sigma_{k,T}C'] \qquad \textit{Observation noise covariance}$$

$$\hat{\mu} = \hat{x}_{0,T} - \Sigma\mathbf{1}(\mathbf{1}'\Sigma\mathbf{1})^{-1}(\mathbf{1}'\hat{x}_{0,T} - 1) \qquad \textit{Initial state mean}$$

$$\hat{\Sigma} = \Sigma_{0,T} - \hat{x}_{0,T}\hat{x}'_{0,T} \qquad \textit{Initial state covariance.}$$

Proof. In Appendix IIIB.

5. APPLICATION

The Kalman filtering model described above was applied in an example as follows. We assumed that the state and observation processes are given by linear dynamics

$$x_{k+1} = Ax_k + v_{k+1} \in \mathbb{R}^2,$$

$$y_k = Cx_k + w_k \in \mathbb{R}^2,$$

where A and C are matrices of dimension two, v_k and w_k are normally distributed with means zero and covariance matrices Q and R, respectively. We generated 100 observations, namely 100 "true" state vectors and 100 observations of the "true" state. The first 50 observations were used to estimate the parameters of the model: state matrix A, observation matrix C, state error covariance Q and state error covariance R. Recall that the EM algorithm is self-tuning and guarantees that log-likelihood increase monotonically with every iteration. In this case, log-likelihood converged after only seven passes through the data. The true and estimated parameters for this HMM are shown in Appendix IIIC. Note that estimated parameters are very close to true parameters. In order to illustrate the predictive performance of the algorithm, we used the Kalman filtering model with estimated parameters to predict the state given the second half of the observations. We then plotted predicted versus actual state. The plot is given in Appendix IIIC and shows that predictions are quite accurate.

Appendix IIIA

Lower-Triangular Cholesky Decomposition

Let $A = (a_{ij})_{1 \le i,j \le n}$ be a symmetric, positive-definite matrix. A lower-triangular matrix $L = (l_{ij})_{1 \le i,j \le n}$ such that $A = LL'$ is obtain through the following algorithm:

i. $l_{11} = \sqrt{a_{11}}$.

ii. For $i = 2, \ldots, n$,

$$l_{j1} = a_{j1}/l_{11}.$$

iii. For $i = 2, \ldots, n$; $j = i+1, \ldots, n$,

$$l_{ii} = \sqrt{\left(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2\right)};$$

$$l_{ji} = \left(a_{ji} - \sum_{k=1}^{i-1} l_{jk} l_{ik}\right) \Big/ l_{ii}.$$

Appendix IIIB

Proofs of Results in Chapter 3

Proof of **Theorem 3.1.**

The first statement is trivial since $\Sigma_{0,0} = Q_0$. The other relationships can be proved by induction. Suppose that $S_{k-1,k-1}S'_{k-1,k-1} = \Sigma_{k-1,k-1}$. Then it follows immediately that $S_{k,k-1}S'_{k,k-1} = \Sigma_{k,k-1}$ using the relation between $\Sigma_{k,k-1}$ and $\Sigma_{k-1,k-1}$ in the Kalman filtering process. We verify that the second relationship holds for k using the first relationship for the same k. Since $C\Sigma_{k,k-1}C' = H_kH'_k - R_k$ so that

$$
\begin{aligned}
&(H'_k)^{-1}(H_k + R_k^{1/2})^{-1} + [(H_k + R_k^{1/2})']^{-1}H_k^{-1} \\
&\quad - (H'_k)^{-1}(H_k + R_k^{1/2})^{-1}C\Sigma_{k,k-1}C'[(H_k + R_k^{1/2})']^{-1}H_k^{-1} \\
&= (H'_k)^{1/2}(H_k + R_k^{1/2})^{-1}[H_k(H_k + R_k^{1/2})' + (H_k + R_k^{1/2})H'_k \\
&\quad - H_kH'_k + R_k][(H_k + R_k^{1/2})']^{-1}H_k^{-1} \\
&= (H'_k)^{-1}(H_k + R_k^{1/2})^{-1}[H_kH'_k + H_k(R_k^{1/2})' \\
&\quad + R_k^{1/2}H'_k + R_k][(H_k + R_k^{1/2})']^{-1}H_k^{-1}
\end{aligned}
$$

$$= (H_k')^{-1}(H_k + R_k^{1/2})^{-1}(H_k + R_k^{1/2})(H_k + R_k^{1/2})'[(H_k + R_k^{1/2})']^{-1}H_k^{-1}$$

$$= (H_k')^{-1}H_k^{-1} = (H_k H_k')^{-1}.$$

it follows from the first relationship that

$$\begin{aligned}
S_{k.k}S_{k.k}' &= S_{k.k-1}[I - S_{k.k-1}'C'(H_k')^{-1}(H_k + R_k^{1/2})^{-1}C\,S_{k.k-1}] \\
&\quad \times [I - S_{k.k-1}'C'[(H_k + R_k^{1/2})']^{-1}H_k^{-1}C\,S_{k.k-1}]S_{k.k-1}' \\
&= S_{k.k-1}[I - S_{k.k-1}'C'(H_k')^{-1}(H_k + R_k^{1/2})^{-1}C\,S_{k.k-1} \\
&\quad - S_{k.k-1}'C'[(H_k + R_k^{1/2})]^{-1}H_k^{-1}C\,S_{k.k-1} \\
&\quad + S_{k.k-1}'C'(H_k')^{-1}(H_k + R_k^{1/2})^{-1}C\,S_{k.k-1}S_{k.k-1}'C' \\
&\quad \times [(H_k + R_k^{1/2})']^{-1}H_k^{-1}C\,S_{k.k-1}]S_{k.k-1}' \\
&= \Sigma_{k.k-1} - \Sigma_{k.k-1}C'[(H_k')^{-1}(H_k + R_k^{1/2})^{-1} + [(H_k + R_k^{1/2})']^{-1}H_k^{-1} \\
&\quad - (H_k')^{-1}(H_k + R_k^{1/2})^{-1}C\,\Sigma_{k.k-1}C'[(H_k + R_k^{1/2})']^{-1}H_k^{-1}]C\,\Sigma_{k.k-1} \\
&= \Sigma_{k.k-1} - \Sigma_{k.k-1}C'(H_k H_k')^{-1}C\,\Sigma_{k.k-1} \\
&= \Sigma_{k.k}. \quad \Box
\end{aligned}$$

Proof of **Proposition 3.2.** The problem to solve is

$$\begin{aligned}
\underset{A.C.Q.R.\mu.\Sigma}{\text{maximize}} &-\frac{1}{2}\log|\Sigma| - \frac{1}{2}tr\left[\Sigma^{-1}(\Sigma_{0.T} + (\hat{x}_{0.T} - \mu)(\hat{x}_{0.T} - \mu)')\right] \\
&-\frac{T}{2}\log|Q| - \frac{1}{2}tr\left[Q^{-1}(U - GA' - AG' + AFA')\right] \\
&-\frac{T}{2}\log|R| - \frac{1}{2}tr\left[R^{-1}\sum_{k=1}^{T}((y_k - C\hat{x}_{k.T})(y_k - C\hat{x}_{k.T})' + C\,\Sigma_{k.T}C')\right]
\end{aligned}$$

where

$$F = \sum_{k=1}^{T} (\Sigma_{k-1.T} + \hat{x}_{k-1.T}\hat{x}'_{k-1.T})$$
$$G = \sum_{k=1}^{T} (\Sigma^{T}_{k.k-1} + \hat{x}_{k.T}\hat{x}'_{k-1.T})$$
$$U = \sum_{k=1}^{T} (\Sigma_{k.T} + \hat{x}_{k.T}\hat{x}'_{k.T}).$$

subject to

$$\mathbf{1}'A = \mathbf{1}'$$

$$\mathbf{1}'C = \mathbf{1}'$$

$$\mathbf{1}'\mu = 1.$$

The Lagrangian is

$$\begin{aligned}\mathcal{L} = &-\frac{1}{2}\log|\Sigma| - \frac{1}{2} tr\left[\Sigma^{-1}(\Sigma_{0.T} + (\hat{x}_{0.T} - \mu)(\hat{x}_{0.T} - \mu)')\right] \\ &- \frac{T}{2}\log|Q| - \frac{1}{2} tr\left[Q^{-1}(U - GA' - AG' + AFA')\right] \\ &- \frac{T}{2}\log|R| - \frac{1}{2} tr\left[R^{-1}\sum_{k=1}^{T}((y_k - C\hat{x}_{k.T})(y_k - C\hat{x}_{k.T})' + C\,\Sigma_{k.T}C')\right] \\ &+ E_{\hat{\theta}_j}[R(\hat{\theta}_j)|\mathcal{Y}_T] \\ &+ tr\Lambda'_1(\mathbf{1}'A - \mathbf{1}') + tr\Lambda'_2(\mathbf{1}'C - \mathbf{1}') + \Lambda_3(\mathbf{1}'\mu - 1).\end{aligned}$$

Differentiating in A and Λ_1 and setting the derivatives to zero gives

$$Q^{-1}G - Q^{-1}AF + \mathbf{1}\Lambda_1 = 0 \qquad (1)$$

and

$$\mathbf{1}'A = \mathbf{1}'. \qquad (2)$$

From equation (1) we have $A = GF^{-1} + Q\mathbf{1}\Lambda_1 F^{-1}$. Substituting into (2) we obtain $\Lambda_1 F^{-1} = (\mathbf{1}'Q\mathbf{1})^{-1}(\mathbf{1}' - \mathbf{1}'GF^{-1})$. It follows that

$$A = GF^{-1} + Q\mathbf{1}(\mathbf{1}'Q\mathbf{1})^{-1}(\mathbf{1}' - \mathbf{1}'GF^{-1}).$$

Differentiating in Q and setting the derivative to zero gives

$$\frac{T}{2}Q - \frac{1}{2}(U - GA' - AG' + AFA') = 0$$

and

$$Q = \frac{1}{T}(U - GA' - AG' + AFA').$$

Differentiating in C and Λ_2 and setting the derivatives to zero gives

$$R^{-1}Z - CR^{-1}U + \mathbf{1}\Lambda_2 = 0 \tag{3}$$

and

$$\mathbf{1}'C = \mathbf{1}'. \tag{4}$$

From equation (3) we have $C = (R^{-1}Z + \mathbf{1}\Lambda_2)U^{-1}R$. Substituting into (4) we obtain $\Lambda_2 = (\mathbf{1}' - \mathbf{1}'R^{-1}ZU^{-1}R)R^{-1}U$. It follows that

$$C = R^{-1}ZU^{-1}R + \mathbf{1}(\mathbf{1}' - \mathbf{1}'R^{-1}ZU^{-1}R).$$

Differentiating in R and setting the derivative to zero gives

$$\frac{T}{2}R - \frac{1}{2}\sum_{k=1}^{T}[(Y_k - C\hat{x}_{k.T})(y_k - C\hat{x}_{k.T})' + C\Sigma_{k.T}C'] = 0$$

and

$$R = \frac{1}{T}\sum_{k=1}^{T}[(Y_k - C\hat{x}_{k.T})(y_k - C\hat{x}_{k.T})' + C\Sigma_{k.T}C'].$$

Differentiating in μ and setting the derivatives to zero gives

$$\Sigma^{-1}(\hat{x}_{0.T} - \mu) + \Lambda_3 \mathbf{1} = 0$$

and

$$\mathbf{1}'\mu = 1,$$

so that $\mu = \hat{x}_{0.T} - \Sigma\mathbf{1}(\mathbf{1}'\Sigma\mathbf{1})^{-1}(\mathbf{1}'\hat{x}_{0.T} - 1)$.

Differentiating in Σ and setting the derivative to zero gives

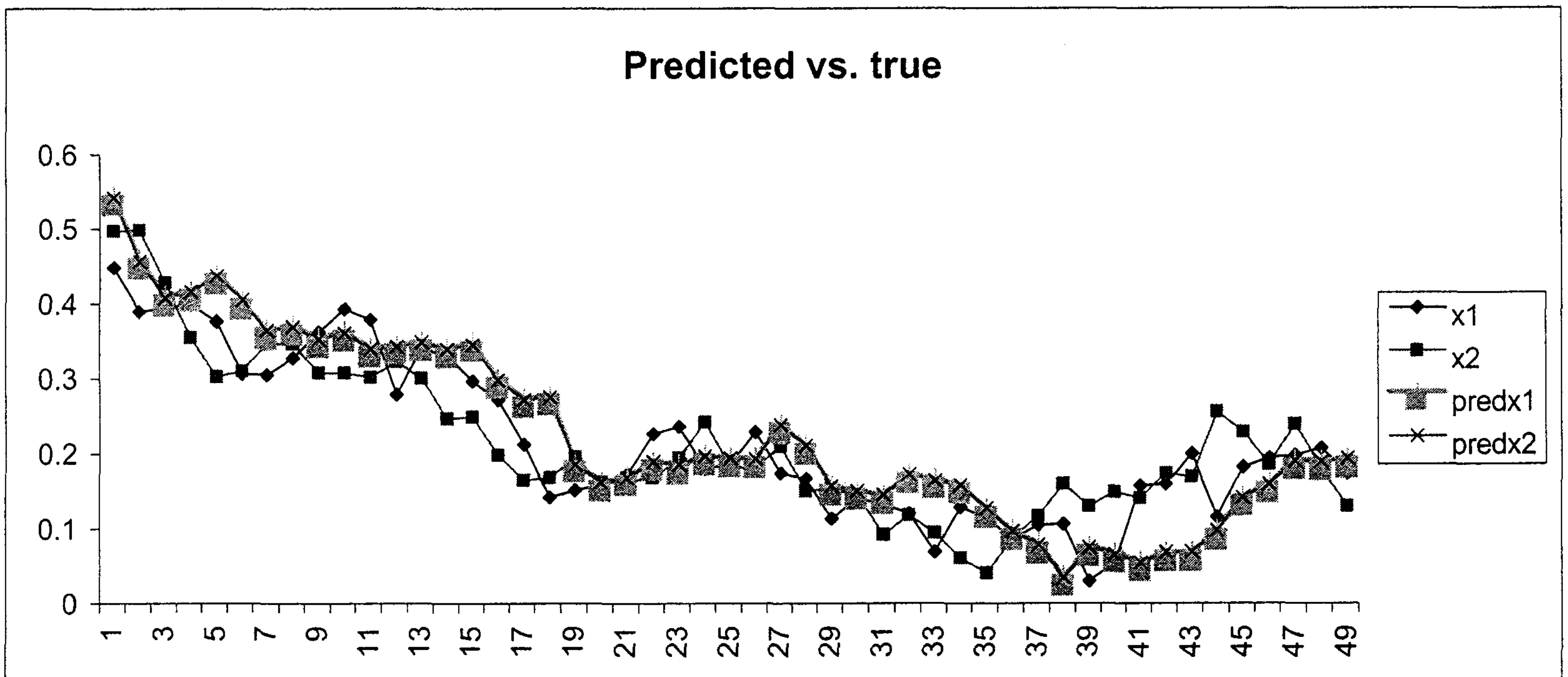
$$\frac{1}{2}\Sigma - \frac{1}{2}(\Sigma_{0.T} - \hat{x}_{0.T}\hat{x}'_{0.T}) = 0$$

so that $\Sigma = \Sigma_{0.T} - \hat{x}_{0.T}\hat{x}'_{0.T}$.

Appendix IIIC - Results for Kalman Filtering Example

Estimated A		**True A**		**Estimated C**		**True C**	
0.6949	0.3073	0.7	0.3	0.7452	0.678	0.9	0.5
0.3051	0.6927	0.3	0.7	0.2548	0.322	0.1	0.5

Estimated Q		**True Q**		**Estimated R**		**True R**	
0.0012	0	0.001	0	0.0116	0	0.01	0
0	0.0012	0	0.001	0	0.0081	0	0.01



Chapter 4

Modelling Default Risk

1. INTRODUCTION

A variety of approaches to valuation of default risk have been presented in the literature and implemented by practitioners. In this chapter we follow the so-called *reduced-form approach*, where default is an unpredictable event governed by a *hazard process* defined in terms of *intensity* λ. If a bondholder receives a payoff only if there is no default before maturity. the bond is priced as if it were default-free by replacing the risk-free short rate r with $r + \lambda$ (See for example Duffie and Singleton [9] or Lando [23]). Following Elliott. Jeanblanc and Yor [16]. and Bielecki. Rutkowski [3]. we make precise the technical conditions under which default acts as a change of interest rate. First we consider the case of only one source of information. namely the time when the default appears. and then discuss valuation of defaultable bonds for investors who also observe the issuer's credit quality represented by their rating. We show that if the default time admits an intensity λ. the value of a zero-recovery defaultable bond is its value as if it were risk-free adjusted by the probability of no default before maturity given in terms of λ. This probability is also calculated explicitly in terms of the issuer's credit quality when the credit quality is added to the investor's information set. We extend our model to allow for a rebate payment at default conditional on the issuer's credit quality. Finally, we discuss valuation of defaultable bonds when "true" credit quality is not observed directly but through noisy credit ratings.

2. THE MARKOV CHAIN MODEL OF CREDIT QUALITY EVOLUTION

Consider a probability space $(\Omega, \mathcal{G}, Q)$ where, for pricing purposes, Q is an *equivalent martingale measure.*

Suppose $\{X_t\}$, $t \geq 0$, is a finite-state *Markov Chain* on $(\Omega, \mathcal{G}, Q)$, with state space $S = \{s_1, s_2, \ldots, s_N\}$. Without loss of generality, we identify the points in S with the unit vectors $\{e_1, e_2, \ldots, e_N\}$, where $e_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}_N$.

Denote by E_Q the expectation under the measure Q. The distribution of X_t is then $p_t := E[X_t] = (p_t^1, p_t^2, \ldots p_t^N)$, where $p_t^i = Q(X_t = e_i) = E[\langle X_t, e_i \rangle]$. We suppose this dis- tribution evolves according to the Kolmogorov equation

$$\frac{dp_t}{dt} = Ap_t,$$

where A is the "Q-matrix," such that $A = (a_{ji})$, $1 \leq i, j \leq N$, $a_{ji} \geq 0$ for all $i \neq j$, $\sum_{j=1}^{N} a_{ji} = 0$. The default state e_N is assumed to be *absorbing*, so the last column of

A has all zeros. Denote by $\{\mathcal{F}_t\}$ the right-continuous, complete filtration generated by the process X. For $s \leq t$, write $\Phi(s,t) = \exp(A(t-s))$ for the transition matrix associated with A, so that

$$\frac{d\Phi(s,t)}{dt} = A\Phi(s,t)$$

and

$$E[X_t|\mathcal{F}_s] = E[X_t|X_s] = \Phi(s,t)X_s.$$

$\Phi(s,t)$ is then a $N \times N$ matrix whose (j,i)-th entry specifies the probability of X being in state j at time t given that the chain is in state i at time s.

Lemma 4.1.

$$V_t := X_t - X_0 - \int_0^t AX_r\,dr$$

is a (vector) $\mathcal{F}_t$-martingale under Q.

Proof. See Appendix IV.

The semi-martingale representation of the Markov Chain X is, therefore,

$$X_t = X_0 + \int_0^t AX_r\,dr + V_t$$

or, in differential form,

$$dX_t = AX_t\,dt + dV_t.$$

As in the previous chapters, we shall suppose that the Markov Chain X described above represents the evolution of credit quality over time, so that N states of X correspond to N credit rating categories. $\Phi(s,t)$ is then the transition matrix and A is its generator matrix.

3. THE DEFAULT TIME

We are interested in modelling how a default time might depend on credit quality. Let τ be an $\mathbb{R}_+$-valued random variable on $(\Omega, \mathcal{G}, Q)$ representing the *default time*, such that $Q(\tau = 0) = 0$ and $Q(t < \tau) > 0$ for all $t > 0$. Denote by $\{\eta_t\}$ the increasing *default process* defined as $\eta_t := 1\!\!1_{\tau \leq t}$. Each sample path is then equal to 0 before random time τ, and it equals 1 for $t \geq \tau$. Let $\{\mathcal{H}_t\}$ denote the natural filtration of η, with $\mathcal{H}_t := \sigma(\eta_u, u \leq t)$ generated by the sets $\{t < \tau\}$ and $\{\tau \leq s\}$ for $s \leq t$, and set $\mathcal{H}_\infty = \sigma(\eta_u : u \in \mathbb{R}_+)$. Note that $\{\mathcal{H}_t\}$ is the smallest filtration such that τ is an $\{\mathcal{H}_t\}$-stopping time. $\mathcal{H}_t$ represents information whether default has occurred at or before time t. In other words, we observe default as it occurs. We note the following important result:

Proposition 4.1. (Dellacherie) *If Y is any integrable, $\mathcal{G}$-measurable random variable, then*

$$E_Q[Y|\mathcal{H}_t] = 1\!\!1_{\tau \leq t} E_Q[Y|\mathcal{H}_\infty] + 1\!\!1_{t<\tau} \frac{E_Q[Y 1\!\!1_{t<\tau}]}{Q(t < \tau)}.$$

In particular,

$$E_Q[Y|\mathcal{H}_t] 1\!\!1_{t<\tau} = 1\!\!1_{t<\tau} \frac{E_Q[Y 1\!\!1_{t<\tau}]}{Q(t < \tau)};$$

and if Y is $\sigma(\tau)$-measurable - that is, $Y = h(\tau)$ - then

$$E_Q[Y|\mathcal{H}_t] = 1\!\!1_{\tau \leq t} h(\tau) + 1\!\!1_{t<\tau} \frac{E_Q[h(\tau) 1\!\!1_{t<\tau}]}{Q(t < \tau)}. \quad \square$$

Let $F(t) = Q(\tau \leq t)$ be the right-continuous distribution function of τ. If F is differentiable, then τ admits a density $f = F'$ and we have the following:

Lemma 4.2. *The process* $\eta_t - \int_0^{\tau \wedge t} \frac{f(u)}{1-F(u)}\, du = \eta_t - \int_0^t \mathbb{1}_{u \leq \tau} \frac{f(u)}{1-F(u)}\, du$ *is an* $\{\mathcal{H}_t\}$*-martingale.*

Proof. See Appendix IV.

Let $\lambda(s) = \frac{f(s)}{1-F(s)}$. Then, solving the ODE $F'(s) = (1 - F(s))\lambda(s)$, we obtain

$$F(s) = Q(\tau \leq s) = 1 - \exp\left(-\int_0^s \lambda(u)\, du\right).$$

An increasing function $\Gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ given by the formula $1 - F(t) = \exp(-\Gamma(t))$ is called a *hazard function* of τ. In the present setting we have $\Gamma(t) = \int_0^t \lambda(u)\, du$ and λ is called the $\{\mathcal{H}_t\}$*-intensity* or the *hazard rate* of τ. In fact, the intensity of the random time τ is often defined as an $\{\mathcal{H}_t\}$-adapted, non-negative process λ for which $\eta_t - \int_0^{\tau \wedge t} \lambda(s)\, ds$ is a $\{\mathcal{H}_t\}$-martingale. In the present setting, τ admits the deterministic intensity $\lambda(t) = \frac{f(t)}{1-F(t)}$.

Remark Note that if τ is exponentially distributed with parameter γ, $F(s) = 1 - \exp(-\gamma t)$ and the intensity of τ is constant: $\lambda(t) = \gamma$ for all $t \in \mathbb{R}_+$.

Corollary 4.1. *For an agent with information* $\mathcal{H}_t$, *the probability of no default before time* $T > t$ *is*

$$Q(T < \tau | \mathcal{H}_t) = \mathbb{1}_{t < \tau} \exp\left(-\int_t^T \lambda(s)\, ds\right).$$

Proof. Appendix IV.

Suppose that there is another source of information represented by a filtration $\{\mathcal{H}'_t\}$. Consider the enlarged filtration $\{\mathcal{E}_t\} = \{\mathcal{H}_t \vee \mathcal{H}'_t\}$, with $\mathcal{E}_\infty = \sigma(\tau) \vee \mathcal{H}'_\infty$.

As before, τ is a random default time defined on the filtration $(\Omega, \mathcal{G}, Q)$. For any $t \in \mathbb{R}_+$, we now write $F(t) = Q(\tau \leq t|\mathcal{H}'_t)$. We define the $\{\mathcal{H}'_t\}$ *-hazard process* of τ under Q through the formula $1 - F(t) = \exp(-\Gamma(t))$.

We have the following generalization of Proposition 4.1:

Proposition 4.2. *Let Y be any integrable, $\mathcal{E}_T$-measurable random variable, then for any $t < T$*

$$E_Q[\mathbb{1}_{t<\tau}Y|\,\mathcal{E}_t] = \mathbb{1}_{t<\tau}\frac{E_Q[\mathbb{1}_{t<\tau}Y|\mathcal{H}'_t]}{Q(t<\tau|\mathcal{H}'_t)} = \mathbb{1}_{t<\tau}E_Q[\mathbb{1}_{t<\tau}Y\exp(\Gamma(t))|\mathcal{H}'_t]$$

and if Y is $\sigma(\tau)$-measurable – that is, $Y = h(\tau)$ – then

$$E_Q[\mathbb{1}_{t<\tau\leq T}\,h(\tau)|\,\mathcal{E}_t] = \mathbb{1}_{t<\tau}\,E_Q[\int_t^T h(u)\exp(\Gamma(t) - \Gamma(u))\,d\Gamma(u)|\mathcal{H}'_t].$$

Proof. Appendix IV.

Corollary 4.2.

$$E_Q[\mathbb{1}_{T<\tau}|\,\mathcal{E}_t] = \mathbb{1}_{t<\tau}E_Q[\exp(\Gamma(t) - \Gamma(T))|\mathcal{H}'_t].$$

Proof. Appendix IV.

Assume that the $\{\mathcal{H}'_t\}$-hazard process Γ is absolutely continuous and that $\Gamma(t) = \int_0^t \lambda(u)\,du$, for some $\{\mathcal{H}'_t\}$-measurable process λ, referred to as the *stochastic intensity* of the random time τ.

Corollary 4.3. *For an agent with information $\mathcal{E}_t = \mathcal{H}_t \vee \mathcal{H}'_t$, the probability of no default before time $T > t$ is*

$$Q(T<\tau|\mathcal{E}_t) = E_Q[\mathbb{1}_{T<\tau}|\mathcal{E}_t] = \mathbb{1}_{t<\tau}E_Q\left[\exp\left(-\int_t^T \lambda(s)\,ds\right)|\mathcal{H}'_t\right]. \quad \square$$

We also have the following generalization of Lemma 4.2:

Lemma 4.3. *The process* $\eta_t - \int_0^{t\wedge\tau} \lambda(u)\, du = \eta_t - \Gamma(t \wedge \tau)$ *follows an* $\mathcal{E}_t$*-martingale.*

Proof. Appendix IV.

4. DEFAULTABLE CLAIMS

Suppose that there exists a riskless zero-coupon with a deterministic spot rate $r(s)$. Assume that this bond pays \$1 at its maturity T. The value of this bond at time $t < T$ is then given by

$$B(t, T) = \exp\left(- \int_t^T r(s)\, ds \right),$$

and its yield to maturity is

$$y(t, T) = -\frac{1}{T-t} \ln B(t, T).$$

As in the previous section, let τ be a default time with intensity λ. Suppose that there also is a defaultable zero-coupon bond whose face value is \$1, but that this amount is received at maturity T only if default has not occurred. Note that we assume zero recovery of face value in case of default. Hence, for such a defaultable zero-coupon bond the payoff is $\mathbb{1}_{T<\tau}$. We shall show in Section 5 how the model can be extended to account for a rebate payment at the time of default.

Lemma 4.4. *The time-t expected value of a defaultable zero-coupon bond which pays* $\mathbb{1}_{T<\tau}$ *at time* T*, for an agent who has information* $\mathcal{H}_t$ *regarding the occurrence of* *default is*

$$B^d(t,T) = E_Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{T<\tau}|\mathcal{H}_t\right]$$
$$= \mathbb{1}_{t<\tau}\exp\left(-\int_t^T (r(s)+\lambda(s))\,ds\right).$$

Proof. Appendix IV.

Note that the above formula indicates that the default acts as a change of interest rate. We have thus shown that within the present setting a defaultable zero-coupon bond may be valued as if it were default-free by replacing the risk-free short rate r with the default-adjusted rate $r+\lambda$.[1] Through this adjustment to the short rate we account for both the probability and timing of default, as well as for losses on default.

Note that

$$B^d(t,T) = \mathbb{1}_{t<\tau}B(t,T)\exp\left(-\int_t^T \lambda(s)\,ds\right) = B(t,T)Q(T<\tau|\mathcal{H}_t),$$

using Corollary 4.1. The time-t value of a defaultable zero-coupon bond is then equal to the value of a risk-free zero-coupon bond with the same maturity adjusted by the probability of no default before maturity.

For an arbitrary defaultable claim K with maturity T, its value is

$$E_Q\left[K\mathbb{1}_{T<\tau}\exp\left(-\int_t^T r(s)\,ds\right)|\mathcal{H}_t\right].$$

[1] This is not always the case in a more general setting, as shown in Elliott, Jeanblanc and Yor [16] and Bielecki, Rutkowski [3].

From Lemma 4.4, if K is independent of the default time τ, the time-t value is

$$\mathbb{1}_{t<\tau} \exp\left(-\int_t^T (r(s)+\lambda(s))\,ds\right) E_Q[K] = B^d(t,T)E_Q[K].$$

We can then write the time-t value of a defaultable coupon bond maturing at time T as the time-t value of a portfolio of defaultable claims K_i, $i=1,\ldots,n$, maturing at times $t < T_1 < T_2 < \cdots < T_n = T$ and corresponding to the remaining payments of expected coupon and maturity value:

$$B_C^d(t,T) = \sum_{i=1}^n E_Q[K_i]B^d(t,T_i),$$

which becomes

$$B_C^d(t,T) = C\sum_{i=1}^n B^d(t,T_i) + FB^d(t,T)$$

if bond cash flows, namely coupon payments C and maturity value F, are known with certainty.

We now assume that the probability of default depends on the issuer's credit quality represented by the process $\{X_t\}$. Hence we consider two sources of information, filtration $\{\mathcal{H}_t\}$ representing default information, and filtration $\{\mathcal{F}_t\}$ representing "true" credit quality information.

Lemma 4.5. *The time-t expected value of a defaultable zero-coupon bond which pays $\mathbb{1}_{T<\tau}$ at time T, for an agent who has information $\mathcal{E}_t = \mathcal{H}_t \vee \mathcal{F}_t$ regarding the occurrence of default and the issuer's credit quality is*

$$\begin{aligned} B^d(t,T) &= E_Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{T<\tau}|\mathcal{E}_t\right] \\ &= \mathbb{1}_{t<\tau}E_Q\left[\exp\left(-\int_t^T (r(s)+\lambda(s))\,ds\right)|X_t\right]. \end{aligned}$$

Proof. Appendix IV.

Recall that at any time the state X_t of the Markov Chain is one of the unit vectors e_i, $1 \leq i \leq N$. Consequently, any function of X_t, say $h(X_t)$, is given by a vector $(h_1, h_2, \ldots, h_N)'$, so that $h(X_t) = \langle h, X_t \rangle$. Let $L = \langle \lambda_1, \lambda_2, \ldots, \lambda_N \rangle$ be a vector of default intensities, each component for a different credit category. We thus suppose that default intensity is a function of X_t, that is $\lambda(t) = L(X_t) = \langle L, X_t \rangle$. Using Lemma 4.5, the time-t expected value of a zero-coupon defaultable bond is then

$$
\begin{aligned}
B^d(t,T) &= \mathbb{1}_{t<\tau}\, E_Q\left[\exp\left(-\int_t^T (r(s)+\lambda(s))\,ds\right)|X_t\right] \\
&= \mathbb{1}_{t<\tau}\, E_Q\left[\exp\left(-\int_t^T (r(s)+\langle L, X_s\rangle)\,ds\right)|X_t\right] \\
&= \mathbb{1}_{t<\tau} \exp\left(-\int_t^T r(s)\,ds\right) E_Q\left[\exp\left(-\int_t^T \langle L, X_s\rangle\,ds\right)|X_t\right] \\
&= \mathbb{1}_{t<\tau}\, B(t,T)\, E_Q\left[\exp\left(-\int_t^T \langle L, X_s\rangle\,ds\right)|X_t\right].
\end{aligned}
$$

We are interested in $E_Q\left[\exp\left(-\int_t^T \langle L, X_s\rangle\,ds\right)|X_t\right]$.

Define $\Gamma_{t,u} := \exp\left(-\int_t^u \langle L, X_s\rangle\,ds\right)$, so that $d\Gamma_{t,u} = -\langle L, X_u\rangle \Gamma_{t,u}\,du$. We shall work with a vector process $Z_{t,u} := X_u \Gamma_{t,u}$. Clearly, $\Gamma_{t,u} = \langle Z_{t,u}, \mathbf{1}\rangle$. Differentiating $Z_{t,u}$, we have the following Itô representation:

$$
\begin{aligned}
dZ_{t,u} = X_u d&\left[\exp\left(-\int_t^T \langle L, X_s\rangle\,ds\right)\right] + \exp\left(-\int_t^T \langle L, X_s\rangle\,ds\right) dX_u \\
&+ dX_u d\left[\exp\left(-\int_t^u \langle L, X_s\rangle\,ds\right)\right].
\end{aligned}
$$

Now,

$$d\left[\exp\left(-\int_t^u \langle L, X_s\rangle\, ds\right)\right] = -\exp\left(-\int_t^u \langle L, X_s\rangle\, ds\right)\langle L, X_u\rangle\, du,$$

and

$$\begin{aligned}\exp\left(-\int_t^u \langle L, X_s\rangle\, ds\right) dX_u &= \exp\left(-\int_t^u \langle L, X_s\rangle\, ds\right)(AX_u du + dV_u)\\ &= AZ_{t,u}du + \Gamma_{t,u}dV_u.\end{aligned}$$

Therefore, writing $\Psi = [A - \operatorname{diag} L]$, we have

$$\begin{aligned}dZ_{t,u} &= -Z_{t,u}\langle L, X_u\rangle du + AZ_{t,u}du + \Gamma_{t,u}dV_u\\ &= [A - \langle L, X_u\rangle I]Z_{t,u}du + \Gamma_{t,u}dV_u\\ &= [A - \operatorname{diag} L]Z_{t,u}du + \Gamma_{t,u}dV_u\\ &= \Psi Z_{t,u}du + \Gamma_{t,u}dV_u,\end{aligned}$$

and in integrated form,

$$Z_{t,T} = Z_{t,t} + \int_t^T \Psi Z_{t,u}du + \int_t^T \Gamma_{t,u}dV_u.$$

The expected value of this vector process given the state of the "true" credit quality process X is then

$$\begin{aligned}E_Q[Z_{t,T}|X_t] &= E_Q\left[Z_{t,t} + \int_t^T \Psi Z_{t,u}du + \int_t^T \Gamma_{t,u}dV_u|X_t\right]\\ &= Z_{t,t} + E_Q\left[\int_t^T \Psi Z_{t,u}du|X_t\right]\\ &= Z_{t,t} + \int_t^T E_Q[\Psi Z_{t,u}|X_t]\, du\\ &= X_t + \int_t^T \Psi E_Q[Z_{t,u}|X_t]\, du.\end{aligned}$$

Write $\hat{Z}_{t,u}$ for $E_Q[Z_{t,u}|X_t]$. The dynamics of $\hat{Z}_{t,u}$ are then $\hat{Z}_{t,T} = X_t + \Psi \int_t^T \hat{Z}_{t,u}\, du$. Solving this equation we obtain

$$\hat{Z}_{t,T} = E_Q[Z_{t,T}|X_t] = \exp(\Psi(T-t))X_t.$$

Since $\Gamma_{t,T} = \langle Z_{t,T}, \mathbf{1}\rangle$, we have that

$$E_Q[\Gamma_{t,T}|X_t] = \langle \exp(\Psi(T-t))X_t, \mathbf{1}\rangle.$$

The time-t expected value of a defaultable zero-coupon bond with maturity T can then be written as

$$B^d(t,T) = \mathbb{1}_{t<\tau} B(t,T)\langle \exp(\Psi(T-t))X_t, \mathbf{1}\rangle.$$

From Corollary 4.3 and Markov property, the probability of no default before maturity given default information up to time t and the issuer's "true" credit quality at time t is

$$\begin{aligned} Q(T<\tau|\mathcal{H}_t \vee X_t) &= \mathbb{1}_{t<\tau} E_Q\left[\exp\left(-\int_t^T \lambda(s)\, ds\right)|X_t\right] \\ &= \mathbb{1}_{t<\tau} E_Q[\Gamma_{t,T}|X_t] = \mathbb{1}_{t<\tau}\langle \exp(\Psi(T-t))X_t, \mathbf{1}\rangle. \end{aligned}$$

It follows that

$$B^d(t,T) = B(t,T)Q(T<\tau|\mathcal{H}_t \vee X_t),$$

and so the time-t expected value of a defaultable zero-coupon bond is again equal to the expected value of a default-free zero-coupon bond adjusted by the probability of no default before maturity. The yield to maturity for such a defaultable bond is

$$\begin{aligned} y^d(t,T) &= -\frac{1}{T-t}\ln B^d(t,T) \\ &= -\frac{1}{T-t}\ln(B(t,T)\langle \exp(\Psi(T-t)X_t, \mathbf{1}\rangle) \\ &= y(t,T) - \frac{1}{T-t}\ln\langle \exp(\Psi(T-t))X_t, \mathbf{1}\rangle. \end{aligned}$$

5. FRACTIONAL RECOVERY

Suppose again that there exists a riskless zero-coupon with a deterministic spot rate $r(s)$ and a defaultable zero-coupon bond with the same maturity T. As in the previous section, let τ be a default time with intensity $\lambda(t) = L(X_t) = \langle L, X_t\rangle$. Suppose also that the defaultable bond pays its face value in full at maturity if there is no default and a rebate at the default time if default appears before maturity. We assume that the fraction of face value paid at default also depends on the issuer's credit quality represented by the process $\{X_t\}$, i.e. we have $\delta(t) = D(X_t) = \langle D, X_t\rangle$. We are again working under the filtration $\{\mathcal{E}_t\} = \{\mathcal{H}_t \vee \mathcal{F}_t\}$. In the case when the default has not appeared at time t, the time-t value of the defaultable zero-coupon bond is

$$\begin{aligned} &\mathbb{1}_{t<\tau} E_Q\left[\exp\left(-\int_t^T r(s)ds\right)\mathbb{1}_{T<\tau}|\mathcal{E}_t\right] \\ &+ \mathbb{1}_{t<\tau} E_Q\left[\exp\left(-\int_t^\tau r(s)ds\right)\langle D, X_\tau\rangle \mathbb{1}_{\tau\le T}|\mathcal{E}_t\right] \\ &= B^d(t,T) + \mathbb{1}_{t<\tau} E_Q\left[\exp\left(-\int_t^\tau r(s)ds\right)\langle D, X_\tau\rangle \mathbb{1}_{\tau\le T}|\mathcal{E}_t\right]. \end{aligned}$$

In view of Proposition 4.2 and Markov property,

$$\mathbb{1}_{t<\tau} E_Q \left[\exp\left(-\int_t^\tau r(s)\,ds \right) \langle D, X_\tau \rangle \mathbb{1}_{\tau \le T} | \mathcal{E}_t \right] =$$

$$= \mathbb{1}_{t<\tau} E_Q \left[\int_t^T \exp\left(-\int_t^u \lambda(s)\,ds \right) \exp\left(-\int_t^u r(s)\,ds \right) \langle D, X_u \rangle \langle L, X_u \rangle \,du | X_t \right]$$

$$= \mathbb{1}_{t<\tau} E_Q \left[\int_t^T \exp\left(-\int_t^u \lambda(s)\,ds \right) \exp\left(-\int_t^u r(s)\,ds \right) \langle \Delta, X_u \rangle \,du | X_t \right],$$

where $\Delta = (\lambda_1 \delta_1, \ldots, \lambda_N \delta_N)$.

Define $\zeta_{t,v} := \int_t^v \exp\left(-\int_t^u (r(s) + \langle L, X_s \rangle\, ds\right) X_u\, du$. Then,

$$\langle \zeta_{t,v}, \Delta \rangle = \int_t^v \exp\left(-\int_t^u (r(s) + \langle L, X_s \rangle)\,ds \right) \langle \Delta, X_u \rangle \,du,$$

and we are interested in

$$E_Q \left[\int_t^T \exp\left(-\int_t^u \lambda(s)\,ds \right) \exp\left(-\int_t^u r(s)\,ds \right) \langle \Delta, X_u \rangle \,du | X_t \right]$$
$$= E_Q[\langle \zeta_{t,v}, \Delta \rangle | X_t].$$

Now,

$$Z_{t,v} := \frac{d\zeta_{t,v}}{dv} = \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) X_v$$

and writing $\Theta = [A - \operatorname{diag} L - r(v)I]$, we have

$$\begin{aligned}
dZ_{t,v} &= -(r(v) + \langle L, X_v \rangle) Z_{t,v}\,dv + \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) dX_v \\
&= -(r(v) + \langle L, X_v \rangle) Z_{t,v}\,dv + \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) (AX_v\,dv + dV_v) \\
&= -(r(v) + \langle L, X_v \rangle) Z_{t,v}\,dv + AZ_{t,v}\,dv + \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) dV_v \\
&= [A - \langle L, X_v \rangle I - r(v)I] Z_{t,v}\,dv + \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) dV_v \\
&= [A - \operatorname{diag} L - r(v)I] Z_{t,v}\,dv + \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) dV_v \\
&= \Theta Z_{t,v}\,dv + \exp\left(-\int_t^v (r(s) + \langle L, X_s \rangle)\,ds \right) dV_v,
\end{aligned}$$

and in integrated form,

$$Z_{t,T} = X_t + \int_t^T \Theta Z_{t,v}\, dv + \int_t^T \exp\left(-\int_t^v (r(s) + \langle L, X_s\rangle)\, ds\right) dV_v.$$

The expected value of this vector process given the state of the credit quality process X is then

$$E_Q[Z_{t,T}|X_t] = X_t + E_Q\left[\int_t^T \Theta Z_{t,v}\, dv | X_t\right] = X_t + \int_t^T \Theta E_Q[Z_{t,v}|X_t]\, dv.$$

Write $\hat{Z}_{t,v}$ for $E_Q[Z_{t,v}|X_t]$. Then $\hat{Z}_{t,T} = X_t + \Theta \int_t^T \hat{Z}_{t,v}\, dv$. Solving this equation we obtain

$$\hat{Z}_{t,T} = E_Q[Z_{t,T}|X_t] = \exp(\Theta(T-t))X_t.$$

It follows that

$$E_Q[\zeta_{t,T}|X_t] = \int_t^T \hat{Z}_{t,v}\, dv = \int_t^T \exp(\Theta(v-t))X_t\, dv.$$

The time-t expected value of a defaultable zero-coupon bond with maturity T can then be written as

$$\begin{aligned} B_\delta^d(t,T) &= B^d(t,T) + \mathbb{1}_{t<\tau} E_Q[\langle \zeta_{t,v}, \Delta\rangle | X_t] \\ &= B^d(t,T) + \mathbb{1}_{t<\tau} \langle E_Q[\zeta_{t,v}|X_t], \Delta\rangle \\ &= B^d(t,T) + \mathbb{1}_{t<\tau} \langle \int_t^T \exp(\Theta(v-t))X_t\, dv, \Delta\rangle. \end{aligned}$$

$B_\delta^d(t,T)$ is then the time-t value of a zero-recovery defaultable zero-coupon bond plus the time-t value of a rebate paid in case of default before maturity.

6. THE DEFAULT TIME AND THE HIDDEN MARKOV MODEL

We have shown that the time-t value of a defaultable zero-coupon bond that expires at time $T > t$ is

$$B^d(t,T) = \mathbb{1}_{t<\tau} B(t,T)\langle \exp(\Psi(T-t))X_t, \mathbf{1}\rangle,$$

where $B(t,T)$ is the time-t value of a default-free zero-coupon bond, $\Psi = [A - \operatorname{diag} L]$ and $L = \langle \lambda_1, \lambda_2, \ldots, \lambda_N\rangle$ is a vector of default intensities, one for each rating category.

Suppose now that the "true" credit quality $\{X_t\}$ is not observed directly. Rather, it is hidden in noisy observations $\{Y_t\}$ represented by the posted credit ratings. The quantity $B^d(t,T)$ can therefore be observed only in an "ideal" world where the "true" credit quality of a bond issue is readily available.

We shall suppose that rating observations are equally spaced, say annually or quarterly, as is the case with ratings posted by Standard & Poors or Moody's. Recall from Section 1 the continuous-time semi-martingale representation of the process X:

$$X_t = X_0 + \int_0^t AX_r\, dr + V_t.$$

However, in Chapter 1 we have developed a filtering algorithm, where both the state and observation processes have discrete dynamics. We shall therefore write $\Phi := e^{As}$, where s is the length of time between rating updates, such as a year or a quarter, and consider a discrete time version of the state process X,

$$X_t = \Phi X_{t-1} + V_t,$$

where V_t is an $(\mathcal{F}_t, Q)$-martingale increment. We also suppose, as in Chapter 1, that the observation process $\{Y_t\}$ has dynamics

$$Y_t = CX_t + W_t.$$

Let $\tilde{B}^d(t, T)$ be the time-t value of a defaultable zero-coupon bond for an agent who has information $\mathcal{H}_t \vee \mathcal{Y}_t$ regarding the occurrence of default and the issuer's credit rating. As in Chapter 1, we denote by $\mathcal{Y}_t$ the σ-algebra generated by all possible histories of the observation process Y up to and including time t. In other words, the agent knows whether default has occurred and observes the issuer's credit rating history but not the issuer's "true" credit quality.

Using Corollary 4.3,

$$\begin{aligned}
\hat{B}^d(t, T) &= E_Q\left[\mathbb{1}_{t<\tau} \exp\left(-\int_t^T (r(s) + \lambda(s))\, ds\right) |\mathcal{E}_t\right] \\
&= \mathbb{1}_{t<\tau}\, E_Q\left[\exp\left(-\int_t^T (r(s) + \langle L, X_s\rangle)\, ds\right) |\mathcal{Y}_t\right] \\
&= \mathbb{1}_{t<\tau} \exp\left(-\int_t^T r(s)\, ds\right) E_Q\left[\exp\left(-\int_t^T \langle L, X_s\rangle\, ds\right) |\mathcal{Y}_t\right] \\
&= \mathbb{1}_{t<\tau}\, B(t, T)\, E_Q\left[\exp\left(-\int_t^T \langle L, X_s\rangle\, ds\right) |\mathcal{Y}_t\right].
\end{aligned}$$

We are interested in $E_Q\left[\exp\left(-\int_t^T \langle L, X_s\rangle\, ds\right) |\mathcal{Y}_t\right]$.

Recall from Section 4 that $\Gamma_{t,u} := \exp\left(-\int_t^u \langle L, X_s\rangle\, ds\right)$ and the dynamics of the process $Z_{t,u} := X_u\Gamma_{t,u}$ are

$$Z_{t,T} = Z_{t,t} + \int_t^T \Psi Z_{t,u} du + \int_t^T \Gamma_{t,u} dV_u.$$

The expected value of this vector process given the possible rating histories $\mathcal{Y}_t$ is then

$$\begin{aligned}
E_Q[Z_{t,T}|\mathcal{Y}_t] &= E_Q\left[Z_{t,t} + \int_t^T \Psi Z_{t,u} du + \int_t^T \Gamma_{t,u} dV_u |\mathcal{Y}_t\right] \\
&= Z_{t,t} + E_Q\left[\int_t^T \Psi Z_{t,u} du |\mathcal{Y}_t\right] \\
&= Z_{t,t} + \int_t^T E_Q[\Psi Z_{t,u}|\mathcal{Y}_t]\, du \\
&= X_t + \int_t^T \Psi E_Q[Z_{t,u}|\mathcal{Y}_t]\, du.
\end{aligned}$$

Write $\tilde{Z}_{t,u}$ for $E_Q[Z_{t,u}|\mathcal{Y}_t]$. The dynamics of $\tilde{Z}_{t,u}$ are then $\tilde{Z}_{t,T} = X_t + \Psi \int_t^T \tilde{Z}_{t,u}\, du$. Solving this equation we obtain

$$\tilde{Z}_{t,T} = E_Q[Z_{t,T}|\mathcal{Y}_t] = \exp(\Psi(T-t))E_Q[X_t|\mathcal{Y}_t].$$

Since $\Gamma_{t,T} = \langle Z_{t,T}, \mathbf{1}\rangle$, we have that

$$E_Q[\Gamma_{t,T}|\mathcal{Y}_t] = \langle \exp(\Psi(T-t))E_Q[X_t|\mathcal{Y}_t], \mathbf{1}\rangle.$$

The time-t expected value of a defaultable zero-coupon bond with maturity T can then be written as

$$\tilde{B}^d(t,T) = \mathbb{1}_{t<\tau} B(t,T)\langle \exp(\Psi(T-t))E_Q[X_t|\mathcal{Y}_t], \mathbf{1}\rangle.$$

From Corollary 4.3, the probability of no default before maturity given default information up to time t and the issuer's credit rating history $\mathcal{Y}_t$ is

$$\begin{aligned}
Q(T<\tau|\mathcal{H}_t \vee \mathcal{Y}_t) &= \mathbb{1}_{t<\tau} E_Q\left[\exp\left(-\int_t^T \lambda(s)\, ds\right) |\mathcal{Y}_t\right] \\
&= \mathbb{1}_{t<\tau} E_Q[\Gamma_{t,T}|\mathcal{Y}_t] = \mathbb{1}_{t<\tau}\langle \exp(\Psi(T-t))E_Q[X_t|\mathcal{Y}_t], \mathbf{1}\rangle.
\end{aligned}$$

It follows that

$$\tilde{B}^d(t,T) = B(t,T)Q(T < \tau | \mathcal{H}_t \vee \mathcal{Y}_t),$$

and so the time-t expected value of a defaultable zero-coupon bond is again equal to the expected value of a default-free zero-coupon bond adjusted by the probability of no default before maturity. The latter is conditioned on the observed credit rating history. The yield to maturity for such a defaultable bond is

$$\begin{aligned} y^d(t,T) &= -\frac{1}{T-t}\ln\tilde{B}^d(t,T) \\ &= -\frac{1}{T-t}\ln(B(t,T)\langle\exp(\Psi(T-t)E_Q[X_t|\mathcal{Y}_t],\mathbf{1}\rangle) \\ &= y(t,T) - \frac{1}{T-t}\ln\langle\exp(\Psi(T-t))E_Q[X_t|\mathcal{Y}_t],\mathbf{1}\rangle. \end{aligned}$$

We can now use the filtering and parameter estimation algorithms from Chapter 1 to estimate $\hat{X}_t := E_Q[X_t|\mathcal{Y}_t]$ and then calculate the time-t expected value of a defaultable zero-coupon bond as

$$\tilde{B}^d(t,T) = \mathbb{1}_{t<\tau}B(t,T)\langle\exp(\Psi(T-t))\hat{X}_t,\mathbf{1}\rangle.$$

We have thus developed a model for calculating the value of a defaultable bond based on the best mean-square estimate of "true" credit quality given noisy observations represented by posted credit ratings.

Appendix IV

Proofs of Results in Chapter 4

Proof of **Lemma 4.1.** For $s \leq t$,

$$\begin{aligned} E[V_t - V_s|\mathcal{F}_s] &= E[X_t - X_s - \int_t^s AX_r\,dr|X_s] \\ &= \Phi(s,t)X_s - X_s - \int_s^t A\Phi(s,r)X_s\,dr \\ &= 0 \end{aligned}$$

since $\Phi(s,t) = \int_s^t A\Phi(s,r)\,dr$. $\square$

Proof of **Lemma 4.2.** From Proposition 4.1, for $s < t$

$$\begin{aligned} E_Q[\eta_t - \eta_s|\mathcal{H}_s] &= E_Q[\mathbb{1}_{s<\tau\leq t}|\mathcal{H}_s] \\ &= \mathbb{1}_{\tau\leq s}E_Q[\mathbb{1}_{s<\tau\leq t}|\mathcal{H}_\infty] + \mathbb{1}_{s<\tau}\frac{E_Q[\mathbb{1}_{s<\tau\leq t}\mathbb{1}_{s<\tau}]}{Q(s<\tau)} \\ &= \mathbb{1}_{s<\tau}\frac{E_Q[\mathbb{1}_{s<\tau\leq t}]}{Q(s<\tau)} \\ &= \mathbb{1}_{s<\tau}\frac{F(t)-F(s)}{1-F(s)}. \end{aligned}$$

If we denote

$$Z = \int_0^{\tau\wedge t} \frac{f(u)}{1-F(u)}\,du - \int_0^{\tau\wedge s} \frac{f(u)}{1-F(u)}\,du = \int_{\tau\wedge s}^{\tau\wedge t} \frac{f(u)}{1-F(u)}\,du$$
$$= \int_s^t \mathbb{1}_{u\le\tau} \frac{f(u)}{1-F(u)}\,du,$$

then clearly $Z = \mathbb{1}_{s<\tau} Z$. From Proposition 4.1 and then Fubini's Theorem we obtain

$$\begin{aligned}
E_Q[Z|\mathcal{H}_s] &= E_Q[\mathbb{1}_{s<\tau} Z|\mathcal{H}_s] \\
&= \mathbb{1}_{s<\tau} E_Q[\mathbb{1}_{s<\tau} Z|\mathcal{H}_\infty] + \mathbb{1}_{s<\tau} \frac{E_Q[\mathbb{1}_{s<\tau} Z]}{Q(s<\tau)} \\
&= \mathbb{1}_{s<\tau} \frac{E_Q[Z]}{Q(s<\tau)} = \mathbb{1}_{s<\tau} \frac{E_Q[\int_s^t \mathbb{1}_{u\le\tau} \frac{f(u)}{1-F(u)}\,du]}{1-F(s)} \\
&= \mathbb{1}_{s<\tau} \frac{\int_s^t \frac{f(u)}{1-F(u)} E_Q[\mathbb{1}_{u\le\tau}]\,du}{1-F(s)} = \mathbb{1}_{s<\tau} \frac{\int_s^t f(u)\,du}{1-F(s)} \\
&= \mathbb{1}_{s<\tau} \frac{F(t)-F(s)}{1-F(s)}.
\end{aligned}$$

The result follows. □

Proof of **Corollary 4.1.** From Proposition 4.1,

$$Q(T<\tau|\mathcal{H}_t) = E_Q[T<\tau|\mathcal{H}_t] = \mathbb{1}_{t<\tau} \frac{E_Q[\mathbb{1}_{T<\tau}]}{1-F(t)} = \frac{1-F(T)}{1-F(t)}. \quad \square$$

Proof of **Proposition 4.2.** The result follows if we show that

$$E_Q[\mathbb{1}_{t<\tau} Y Q(t<\tau|\mathcal{H}'_t)|\,\mathcal{E}_t] = E_Q[\mathbb{1}_{t<\tau} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]|\,\mathcal{E}_t].$$

Now, it is easy to see that for any $A \subset \mathcal{E}_t$, $\exists B \in \mathcal{H}'_t$ such that $A \cap \{t<\tau\} =$

$B \cap \{t < \tau\}$. Then,

$$\begin{aligned} E_Q[\mathbb{1}_{t<\tau} Y Q(t<\tau|\mathcal{H}'_t)|\,\mathcal{E}_t] &= \int_A \mathbb{1}_{t<\tau} Y Q(t<\tau|\mathcal{H}'_t)\,dQ \\ &= \int_{A\cap\{t<\tau\}} Y Q(t<\tau|\mathcal{H}'_t)\,dQ \\ &= \int_{B\cap\{t<\tau\}} Y Q(t<\tau|\mathcal{H}'_t)\,dQ \\ &= \int_B \mathbb{1}_{t<\tau} Y Q(t<\tau|\mathcal{H}'_t)\,dQ \end{aligned}$$

$$\begin{aligned} &= \int_B E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t] Q(t<\tau|\mathcal{H}'_t)\,dQ \\ &= \int_B E_Q[\mathbb{1}_{t<\tau} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]|\mathcal{H}'_t]\,dQ \\ &= \int_B \mathbb{1}_{t<\tau} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]\,dQ \\ &= \int_{B\cap\{t<\tau\}} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]\,dQ \\ &= \int_{A\cap\{t<\tau\}} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]\,dQ \\ &= \int_A \mathbb{1}_{t<\tau} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]\,dQ \\ &= E_Q[\mathbb{1}_{t<\tau} E_Q[\mathbb{1}_{t<\tau} Y|\mathcal{H}'_t]|\,\mathcal{E}_t], \end{aligned}$$

as required. Now,

$$E_Q[\mathbb{1}_{t<\tau\le T} h(\tau)|\,\mathcal{E}_t] = \mathbb{1}_{t<\tau} \exp(\Gamma(t)) E_Q[\mathbb{1}_{t<\tau\le T} h(\tau)|\mathcal{H}'_t].$$

To prove the last result it is enough to check that

$$E_Q[\mathbb{1}_{t<\tau\le T} h(\tau)|\mathcal{H}'_t] = E_Q\left[\int_t^T h(u)dF(u)|\mathcal{H}'_t\right]$$

for a piecewise constant function $h(u) = \sum_{i=0}^{n} h_i 1\!\!1_{t_i < u \leq t_{i+1}}$, where $t_0 = t < t_1 < \cdots < t_{n+1} = T$. We have

$$\begin{aligned} E_Q[1\!\!1_{t<\tau\leq T} h(\tau)|\mathcal{H}'_t] &= \sum_{i=0}^{n} E_Q[E_Q[h_i 1\!\!1_{t_i<\tau\leq t_{i+1}}|\mathcal{H}'_{t_{i+1}}]|\mathcal{H}'_t] \\ &= E_Q[\sum_{i=0}^{n} h_i(F(t_{i+1}) - F(t_i))|\mathcal{H}'_t] \\ &= E_Q\left[\sum_{i=0}^{n} \int_{t_i}^{t_i+1} h(u)\, dF(u)|\mathcal{H}'_t\right] \\ &= E_Q\left[\int_t^T h(u) dF(u)|\mathcal{H}'_t\right] \\ &= E_Q\left[\int_t^T h(u)\exp(-\Gamma(u))\, d\Gamma(u)|\mathcal{H}'_t\right]. \quad \square \end{aligned}$$

Proof of **Corollary 4.2.** Since $1\!\!1_{t<\tau} 1\!\!1_{T<\tau}$:

$$\begin{aligned} E_Q[1\!\!1_{T<\tau}|\,\mathcal{E}_t] &= E_Q[1\!\!1_{t<\tau} 1\!\!1_{T<\tau}|\,\mathcal{E}_t] \\ &= 1\!\!1_{t<\tau} E_Q[1\!\!1_{t<\tau} 1\!\!1_{T<\tau} \exp(\Gamma(t))|\mathcal{H}'_t] \text{ by Proposition 4.2} \\ &= 1\!\!1_{t<\tau} E_Q[1\!\!1_{T<\tau}|\mathcal{H}'_t] \exp(\Gamma(t)) \\ &= 1\!\!1_{t<\tau} E_Q[E_Q[1\!\!1_{T<\tau}|\mathcal{H}'_T]|\mathcal{H}'_t] \exp(\Gamma(t)) \\ &= 1\!\!1_{t<\tau} E_Q[Q(T < \tau|\mathcal{H}'_T) \exp(\Gamma(t))|\mathcal{H}'_t] \\ &= 1\!\!1_{t<\tau} E_Q[\exp(-\Gamma(T)) \exp(\Gamma(t))|\mathcal{H}'_t] \\ &= 1\!\!1_{t<\tau} E_Q[\exp(\Gamma(t) - \Gamma(T))|\mathcal{H}'_t]. \quad \square \end{aligned}$$

Proof of **Lemma 4.3.** First we show that $1\!\!1_{t<\tau}\exp(\Gamma(t))$ is an $\{\mathcal{E}_t\}$-martingale, i.e. that for $s \leq t$, $E_Q[1\!\!1_{t<\tau}\exp(\Gamma(t))|\mathcal{E}_s] = 1\!\!1_{s<\tau}\exp(\Gamma(s))$.

Using Proposition 4.2 and noting that $\mathbb{1}_{s<\tau}\mathbb{1}_{t<\tau} = \mathbb{1}_{t<\tau}$, we have

$$\begin{aligned}
E_Q[\mathbb{1}_{t<\tau}\exp(\Gamma(t))|\,\mathcal{E}_s] &= \mathbb{1}_{s<\tau}E_Q[\mathbb{1}_{t<\tau}\exp(\Gamma(t))|\mathcal{H}'_s]\exp(\Gamma(s)) \\
&= \mathbb{1}_{s<\tau}\exp(\Gamma(s))E_Q[E_Q[\mathbb{1}_{t<\tau}\exp(\Gamma(t))|\mathcal{H}'_t]|\mathcal{H}'_s] \\
&= \mathbb{1}_{s<\tau}\exp(\Gamma(s))E_Q[E_Q[\mathbb{1}_{t<\tau}|\mathcal{H}'_t]\exp(\Gamma(t))|\mathcal{H}'_s] \\
&= \mathbb{1}_{s<\tau}\exp(\Gamma(s))E_Q[(1-F(t))\exp(\Gamma(t))|\mathcal{H}'_s] \\
&= \mathbb{1}_{s<\tau}\exp(\Gamma(s))E_Q[\exp(-\Gamma(t))\exp(\Gamma(t))|\mathcal{H}'_s] \\
&= \mathbb{1}_{s<\tau}\exp(\Gamma(s)),
\end{aligned}$$

as required. With $L_t = (1-\eta_t)\exp(\Gamma(t)) = \mathbb{1}_{t<\tau}\exp(\Gamma(t))$,

$$\exp(-\Gamma(u))dL_u = (1-\eta_u)d\Gamma(u) - d\eta_u,$$

and $\int_0^t \exp(\Gamma(u))\,dL_u = \int_0^t d\eta_u - \int_0^t(1-\eta_u)\,d\Gamma(u) = \eta_t - \int_0^t \mathbb{1}_{u<\tau}\lambda(u)\,du = \eta_t - \int_0^{t\wedge\tau}\lambda(u)\,du$, which is an $\{\mathcal{E}_t\}$-martingale. □

Proof of **Lemma 4.4.** From Proposition 4.1,

$$\begin{aligned}
&E_Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{T<\tau}|\mathcal{H}_t\right] = \\
&= \mathbb{1}_{t<\tau}\frac{E_Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{T<\tau}\mathbb{1}_{t<\tau}\right]}{Q(t<\tau)} \\
&= \mathbb{1}_{t<\tau}\frac{E_Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{T<\tau}\right]}{Q(t<\tau)} \\
&= \mathbb{1}_{t<\tau}\frac{\exp\left(-\int_t^T r(s)\,ds\right)E_Q[\mathbb{1}_{T<\tau}]}{Q(t<\tau)}
\end{aligned}$$

$$= \mathbb{1}_{t<\tau} \frac{\exp\left(-\int_t^T r(s)\,ds\right) Q(T<\tau)}{Q(t<\tau)}$$

$$= \mathbb{1}_{t<\tau} \frac{\exp\left(-\int_t^T r(s)\,ds\right) \exp\left(-\int_0^T \lambda(s)\,ds\right)}{\exp\left(-\int_0^t \lambda(s)\,ds\right)}$$

$$= \mathbb{1}_{t<\tau} \exp\left(-\int_t^T (r(s)+\lambda(s))\,ds\right). \quad \square$$

Proof of **Lemma 4.5.** Using Corollary 4.3, we have

$$E_Q\left[\exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{T<\tau}|\mathcal{E}_t\right] = \exp\left(-\int_t^T r(s)\,ds\right)E_Q[\mathbb{1}_{T<\tau}|\mathcal{E}_t]$$

$$= \exp\left(-\int_t^T r(s)\,ds\right)\mathbb{1}_{t<\tau}E_Q\left[\exp\left(-\int_t^T \lambda(s)\,ds\right)|\mathcal{F}_t\right]$$

$$= \mathbb{1}_{t<\tau}E_Q\left[\exp\left(-\int_t^T (r(s)+\lambda(s))\,ds\right)|\mathcal{F}_t\right].$$

The result follows using Markov property. $\square$

Bibliography

[1] B.D.O. Anderson and J.B. Moore (1979), *Optimal Filtering*, Prentice-Hall.

[2] A. Bangia, F.X. Diebold and T. Schuermann (2000), "Rating Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing," working paper, Penn Institute for Economic Research.

[3] T.R. Bielecki and M. Rutkowski (2002), *Credit Risk: Modeling, Valuation and Hedging*, Springer-Verlag.

[4] G.J. Bierman (1977), *Factorization Methods for Discrete Sequential Estimation*, Academic Press.

[5] L.V. Carty and J.S. Fons (June 1994), "Measuring Changes in Corporate Credit Quality," *Journal of Fixed Income* (vol. 4, no. 1), pp. 27-41.

[6] L.V. Carty and D. Lieberman (1997), "Historical Default Rates of Corporate Bond Issuers, 1920-1996," *Moody's Investors Service.*

[7] C.K. Chui and G. Chen (1991), *Kalman Filtering with Real-Time Applications*, Springer-Verlag.

[8] I.D. Dichev and J.D. Piotroski (February 2001), "The Long-Run Stock Returns Following Bond Rating Changes," *The Journal of Finance* (vol. LVI, no. 1),

pp. 173-203.

[9] D. Duffie and K.J. Singleton (1999), "Modeling Term Structures of Defaultable Bonds," *The Review of Financial Studies* (vol. 12, no. 4), pp. 687-720.

[10] L.H. Ederington (November 1985), "Classification Models and Bond Ratings," *The Financial Review* (vol. 20, no. 4), pp. 237-262.

[11] R.J. Elliott (1994), "Exact Adaptive Filters for Markov Chains Observed in Gaussian Noise," *Automatica* (vol. 30, no. 9), pp. 1399-1408.

[12] R.J. Elliott (1997), "Improved Smoothers for Discrete Time HMM Parameter Estimation," working paper, University of Alberta and University of Adelaide.

[13] R.J. Elliott (Winter 1999), "Financial Filtering," Lecture Notes, University of Toronto.

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

[15] R.J. Elliott, W.C. Hunter and B.M. Jamieson (2001), "Financial Signal Processing: A Self Calibrating Model," *International Journal of Theoretical and Applied Finance* (vol. 22), pp. 567-584.

[16] R.J. Elliott, M. Jeanblanc, M. Yor (April 2000), "On Models of Default Risk," *Mathematical Finance* (vol. 10, no. 2), pp. 179-195.

[17] Z. Ghahramani and G.E. Hinton (1996), "Parameter Estimation for Linear Dynamical Systems," Technical Report CRG-TR-96-2, University of Toronto.

[18] R.A. Jarrow and S.M. Turnbull (March 1995), "Pricing Derivatives on Financial

Securities Subject to Credit Risk," *The Journal of Finance* (vol. L, no. 1), pp. 53-85.

[19] R.A. Jarrow, D. Lando and S.M. Turnbull (Summer 1997), "A Markov Model for the Term Structure of Credit Risk Spreads," *The Review of Financial Studies* (vol. 10, no. 2), pp. 481-523.

[20] R.S. Kaplan and G. Urwitz (April 1979), "Statistical Models of Bond Ratings," *The Journal of Business* (vol. 52, no. 2), pp. 231-261.

[21] D. Kliger and O. Sarig (December 2000), "The Information Value of Bond Ratings," *The Journal of Finance* (vol. LV, no. 6), pp. 2879-2902.

[22] N.V. Krichagina, R.Sh. Lipster and E.Ya. Rubinovich (Steklov Seminar 1994), "Kalman Filter for Markov Processes," in N.V. Krylov, R.Sh. Lipster and A.A. Novikov (eds.), *Statistics and Control of Stochastic Processes*, Optimization Software Inc., New York, pp. 197-213.

[23] D. Lando (1998), "On Cox Processes and Credit Risky Securities," *Review of Derivatives Research* (vol. 2, no. 2/3), pp. 99-120.

[24] D. Lando and T. Skodeberg (2002), "Analyzing Rating Transitions and Rating Drift with Continuous Observations," *The Journal of Banking and Finance* (vol. 26), pp. 423-444.

[25] J.R. Magnus and H. Neudecker (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

[26] G.J. McLachlan, T. Krishnan (1997), *The EM Algorithm and Extensions*, John

Wiley & Sons, Inc.

[27] J.P. Morgan (1997), *CreditMetrics^{TM} – Technical Document*, J.P. Morgan, New York.

[28] Standard & Poor's (2000), *Standard & Poor's COMPUSTAT (North America) User's Guide*, Standard & Poor's, Englewood Colorado.

[29] R.H. Shumway and D.S. Stoffer (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis* (vol. 3, no. 4), pp. 253-264.

[30] D. Simon and T. Chia (January 2002), "Kalman Filtering with State Equality Constraints," *IEEE Transactions on Aerospace and Electronic Systems* (vol. 39), pp. 128-136.