ABSTRACT OF THE DISSERTATION

Essays on Bayesian Inference in Financial Economics

By XIANGHUA LIU

Dissertation Director:

Hiroki Tsurumi

This dissertation consists of three essays on Bayesian inference in financial economics. The first essay explores the impact of discretization errors on the parametric estimation of continuous-time financial models. Euler and other discretization schemes cause discretization errors in solving stochastic differential equations. The empirical impact of these discretization errors on estimating two continuous-time financial models is investigated by using Monte Carlo experiments to compare the "exact" estimator and "Euler" estimator for the Euler scheme. The primary finding is that reducing the discretization interval to reduce the discretization error does not necessarily improve the performance of the estimators. This implies that discretization schemes may yield reliable results when the sampling interval is regularly small and shortening the discretization intervals or using data augmentation techniques may be redundant in practice.

The second essay examines the identification problem in state-space models under the Bayesian framework. Underidentifiability causes no real difficulty in the Bayesian approach in that a legitimate posterior distribution might be achieved for unidentified parameters when appropriate priors are imposed. When estimating unidentified pa-

rameters, Markov chain Monte Carlo algorithms may yield misleading results even if the algorithms seem to converge successfully. In addition, the identification problem does really not matter when the prediction of state-space models instead of parameter estimation is concerned.

The third essay extensively studies credit risk models using Bayesian inference. Bayesian inference is conducted and Markov chain Monte Carlo (MCMC) algorithms are developed for three popular credit risk models. Empirical results show that these three models in which the same PD (probability of default) can be estimated using different information may yield quite different results. Motivated by the empirical results about credit risk model uncertainty, I propose a "combined" Bayesian estimation method to incorporate information from different datasets and model structure for estimating the PD. This new approach provides an insight in dealing with two practical problems, model uncertainty and data insufficiency, in credit risk management.

Acknowledgements

I would like to thank Hiroki Tsurumi, my advisor, for his many suggestions and constant support during this research. It was him who introduced me to the wonderful field of Bayesian econometrics and Markov chain Monte Carlo algorithms. Without his help, I would never have completed my dissertation.

I am also thankful to the members of my committee, Bruce Mizrach and John Landon-Lane, for their inspirational comments in either finance or econometrics. I would also like to thank my outside committee member Xing Zhou for his valuable input in the area of credit risk.

In addition, I greatly appreciate the useful comments and suggestions from the Rutgers seminar and my job markets talks. Special thanks go to Dorothy Rinaldi for her consistent help and support.

Finally I leave my last thanks to my wife Mengli and my daughter Allie, for their not falling sleep when listening to my job market paper presentation.

Dedication

To my wife Mengli and my daughter Allie.

Table of Contents

List of Tables

List of Figures

Introduction

Financial economics has become one of the most challenging areas in modern economics. The complexity of financial modeling problems leaves many unanswered questions for researchers to explore. Bayesian inference equipped with Markov chain Monte Carlo (MCMC) algorithms has been shown to be very useful to deal with complicated financial models and their statistical inference. My dissertation aims to investigate some issues related to the application of Bayesian inference in financial economics and provide some insights about the usefulness of Bayesian inference and MCMC methods in this area.

Motivated by my research in the last chapter, the first two chapters discuss two important fundamental topics in financial econometrics: discretization and identification. First, discretization is a popular way to deal with continuous-time financial models. Although researchers have been aware of the discretization errors brought by the discretization schemes, there has been little literature about the impact of the discretization errors on parameter estimation of continuous-time models. The main contribution of this chapter is the empirical investigation on the magnitude of the impact. Some important findings of the empirical analysis based on some Monte carlo experiments indicate that this impact is often less significant than expected for some regular cases and it is redundant and even harmful for parameter estimation

to try to reduce the discretization errors by increasing the sampling frequency and decreasing the sampling interval. These results may be explained by Florens-Zmirou (1989) and Yoshida (1992) while they are likely to be neglected in practice.

Identification is also a very important topic in econometrics. The chapter first reviews the literature on this topic in the perspective of classical and Bayesian econometrics. Bayesians have different treatments to the identification problem as Lindly (1971) claim that "underidentifiability causes no real difficulty in the Bayesian approach". This chapter is focused on discussing the impact of the identification problem on the estimation of state-space models that become very popular in financial modeling. It is not unusual that financial practioners are not aware that state-space models are generally unidentified. The main purpose of this chapter is to find what happen to parameter estimation, model prediction and MCMC convergency if the state-space model is unidentified. A main conclusion based on some simulation analysis is that unidentified parameters do not always stop the MCMC algorithms for estimating these parameters from converging. However, the converged values could be misleading estimation results if unidentified parameters exist. This result has very useful implication for the practical use of MCMC algorithms. An additional result from this chapter is that identification really does not have impact on the model prediction of state-space models.

Credit risk modeling and estimation is an exploding research area for these years. Its importance has been evidenced by the recent financial crisis. Its attractiveness lies in the fact that default behaviors are so difficult to characterize that there is no agreement about the "perfect" credit risk models and a large amount of varied statistical models are competing. In this paper I first want to show the usefulness of

Bayesian inference and MCMC methods in the estimation of a variety of credit risk models. Credit risk models inherit the complexity of preceding financial models and their statistical inference. It is natural to consider the MCMC methods to deal with those difficulties as before. An empirical study using real data shows different credit risk models may yield widely different results. Instead of selecting the "best" as in the traditional statistical inference, I propose a way to combine the estimation results from different models to incorporate information from different data sources and model structures. This new approach can be very useful when we face some practical issues such as data insufficiency and model uncertainty in credit risk analysis, and may shred some light on the future development of credit risk management.

Chapter 1

Discretization Errors in Estimating Continuous-Time Financial Models via MCMC Methods

1.1 Introduction

Continuous-time models have become inevitable in the field of modern financial theory. Relative to the discrete counterparts, which were developed earlier, continuous-time models are mathematically elegant, but computationally complicated. Generally an Ito process is used to model the continuous path of a financial variable $X(t)$ as:

$$dX(t;\theta) = \mu(X(t);\theta)dt + \sigma(X(t);\theta)dW(t), \tag{1.1}$$

where $\mu(X(t);\theta)$ is the drift term, $\sigma(X(t);\theta)$ is the diffusion term, $W(t)$ is a standard Brownian motion (or Wiener process) and θ is the parameter vector defined on a compact set Θ. A renowned example is the Black-Scholes (1973)'s geometric Brownian

motion model, which is applied to model stock prices. The term structure theory of interest rates is another field where various continuous-time stochastic processes are assumed to be the true driver underlying the dynamics of the instantaneous interest rate (or spot rate). Vasicek (1977) and Cox, Ingersoll and Ross (1985) (hereafter CIR) are two most popular single-factor term structure models of interest rates. In the two models, the spot rate is modelled to follow continuous-time Ornstein-Uhlenbeck process and square-root diffusion (Bessel) process respectively.

In single-factor models, the asset return is uniquely determined by one state variable. Recent empirical evidences including Longstaff and Schwartz (1992) and Pearson and Sun (1994) show that a single factor model is not sufficient to model the dynamics of interest rates and adding more factors will dramatically improve the fit of the term structure model to real data. Two- or three-factor models (even with jumps in both return and volatility processes) have replaced single-factor models in modelling equity return and term structure of interest rates. One of the most recent developments in continuous-time financial models is the affine jump-diffusion model, which assumes the drift, diffusion and jump intensity have affine structure to yield a closed-form solution for some asset price, by Kan and Duffie (1996) and Duffie, Pan and Singleton (2002).

The common feature of above models is that the evolution of the financial variables has a continuous path, although this assumption is never evidenced in the real world. Concerning statistical inference on these continuous-time models, the first problem arising is that only discretely sampled data can be obtained in practice. Lo (1988) discusses the maximum likelihood estimation (MLE) of continuous-time models with discretely sampled data. To obtain the likelihood function of the samples, a closed-

form solution of the transition density function of the continuous-time stochastic process is indispensable. Following the statement of Lo (1988), suppose the data $\mathbf{X} = (X_0, X_1, \ldots, X_n)$ are sampled at $n+1$ discrete time points $t_0, t_1, \ldots, t_n$, where $X_i = X(t_i)$. Then, from the Markovian property of the Ito process (1), the likelihood function of the parameter vector θ can be written as:

$$\begin{aligned} l(\theta; \mathbf{X}) &= f(X_0, X_1, \ldots, X_n) \\ &= f(X_0) \prod_{i=1}^{n} f(X_i | X_{i-1}), \end{aligned} \tag{1.2}$$

where $f(X_i|X_{i-1})$ is the transition density function, which satisfies the Fokker-Planck partial differential equation (PDE). For instance, transition densities of the Ornstein-Uhlenbeck process in Vasicek's model can be solved to be Gaussian. Then the likelihood function of discrete samples is fully analytically specified.

Unfortunately, Vasicek's model is only one of few exceptions and most stochastic differential equations (SDE) are not explicitly solvable. In those cases, MLE is not straightly feasible. To solve this problem, econometricians generally consider two approaches: the first is to adopt some distribution-free estimation methods. A popular method used in financial econometrics is the method of moments, particularly Generalized Method of Moments (GMM) proposed by Hansen (1982). For example, Chan, Karolyi, Longstaff and Sanders (CKLS) (1992) use GMM to empirically estimate and compare available constant elasticity of volatility models of the term structure of interest rates. Some extended moments methods including Efficient Method of Moments (EMM) by Gallant and Tauchen (1996) and Simulated Method of Moments (SMM) by Duffie and Singleton (1993) are quite popular in this field now. In another track, Aït-Sahalia (1996) considers nonparametric approaches to the estimation of diffusion processes, which allows flexible, nonparametric estimation of the drift and

diffusion functions.

The second approach is to approximate the transition densities of continuous-time processes. Kloeden and Platen (1992) provide an excellent overview on the discretization schemes for SDEs. The simplest discretization method is the Euler scheme, which is just the first-order stochastic Taylor expansion of an Ito process. The Milstein scheme, the second-order stochastic Taylor expansion and other higher-order expansions are also discussed in the book. These discretization schemes all approximate the transition densities by appropriate Gaussian distributions. The higher-order schemes generally provide more accurate approximation or faster convergence to the *exact* solution of the SDE. It can be shown that the approximation schemes will converge to the diffusion processes when the discrete time intervals converge to zero and other conditions are satisfied. Discretization errors exist when the intervals are nonzero in practice. There is a large amount of literature including Kloeden and Platen (1992) on the magnitude of discretization errors on approximating the continuous-time stochastic process with the discretization schemes. However, there is relatively less research about the impact of these discretization errors on the parameter estimation of continuous-time models. Florens-Zmirou (1989) presents the conditions under which the estimator of the drift parameters in a diffusion process converges to the true value when the discretization schemes are used. He also finds the analytical forms for the asymptotical bias of the estimators for a fixed discretization interval.

In this paper, I aim to investigate the effect of discretization schemes, particular the first-order Euler scheme on the estimation of popular continuous-time financial models such as Vasicek's model and CIR model. Although the sizes of discretization errors can be studied analytically, the magnitude of the empirical effect of discretiza-

tion on parameter estimation in diffusion processes has rarely been explored. Furthermore, this paper specializes in the effect of discretization errors on the Bayesian estimation. Bayesian inference associated with the MCMC sampling methods now is widely applied to the estimation of financial models. Bayesian inference is equivalent to MLE approach in the case of non-informative priors, while MCMC methods have been shown to be a powerful tool to deal with high-dimensional parameters and non-standard distributions, which are often common settings in financial models. Johannes and Polson (2002) present an overview of the general procedures to estimate the diffusion processes using the MCMC approach in application to finance. Due to the restriction of the Bayesian methodology, the joint density function of discretely sampled data have to be fully specified before the posterior densities are obtained. Thus, discretization schemes are often used in the Bayesian estimation of financial models when the associated SDEs have no closed-form solutions. Eraker (2001) and Elerian, Chib, Shephard (2001) use the Euler scheme to estimate single-factor diffusion models using MCMC methods. Eraker, Johannnes and Polson (2003) and Eraker (2004) estimate affine jump-diffusion models using MCMC methods. The models they use are actually Euler discretization of continuous-time models. To reduce the discretization error, Elerian, Chib and Shephard (2001) suggest using data augmentation to "fill in" the intervals between discrete observations. They show that the data augmentation technique dramatically improve the performance of the parameter estimators when the original sampling interval is large.

The paper is organized as follows: In the second section, the overview of stochastic Taylor expansion and Euler scheme is provided, and the impact of Euler approximation on estimating continuous-time models is also studied analytically. In the

following two sections, Monte Carlo experiments are conducted under single-factor Vasicek's model and CIR model. In Vasicek's model, a Gibbs sampler can provide the "exact" estimates. For CIR model, the Metropolis-Hastings (MH) algorithm has to be used to sampled from a posterior density comprising of non-central χ^2 density functions. The reason why only these two single-factor models are chosen is that the exact transition densities in these models have closed-form. Not only the exact data generation from the continuous-time model can be done, but also the "exact" estimation is feasible so that we can compare it with the result of Euler approximation. The last section concludes.

1.2 Overview of Discretization Schemes

1.2.1 Stochastic Taylor Expansion

Kloeden and Platen (1992) provide a comprehensive treatment on the numerical solutions of SDEs. In essence, the discretization schemes on diffusion processes are the stochastic Taylor expansion with different orders. The Ito process in (1.1) has a formal expression as

$$X(t) = X(0) + \int_0^t \mu(X(s))ds + \int_0^t \sigma(X(s))dW(s). \tag{1.3}$$

A general stochastic Taylor expansion formula for a functional of the Ito process is

$$\begin{aligned} f(X(t)) &= f(X(0)) + c_1(X(0)) \int_0^t ds + c_2(X(0)) \int_0^t dW(s) \\ &\quad + c_3(X(0)) \int_0^t \int_0^{s_1} dW(s_2)dW(s_1) + R \end{aligned} \tag{1.4}$$

with coefficients

$$\begin{aligned} c_1(X(0)) &= \mu(X(0))f'(X(0)) + \frac{1}{2}\sigma^2(X(0))f''(X(0)) \\ c_2(X(0)) &= \sigma(X(0))f'(X(0)) \\ c_3(X(0)) &= \sigma(X(0))[\sigma(X(0))f''(X(0)) + \sigma'(X(0))f'(X(0))]. \end{aligned}$$

Here the remainder R consists of higher order multiple stochastic integrals. The stochastic Taylor formula can be thought of a generalization of both the deterministic Taylor formula and the Ito lemma. A proof is provided in Appendix A.

If we let $f(x) = x$ and consider a time interval from t to $t + \Delta$ [1] , then the stochastic Taylor formula is reduced to

$$\begin{aligned} X(t+\Delta) &= X(t) + \mu(X(t))\int_t^{t+\Delta} ds + \sigma(X(t))\int_t^{t+\Delta} dW(s) \\ &\quad + \sigma^2(X(t))\int_t^{t+\Delta}\int_t^{s_2} dW(s_1)dW(s_2) + R, \end{aligned} \tag{1.5}$$

where R is the expansion reminder consisting of higher order multiple stochastic integrals. By truncating the stochastic Taylor expansion, we can form discretization schemes for a SDE. Keeping only the first-order terms in the stochastic Taylor expansion, we can obtain the Euler approximation, the simplest Taylor approximation of an Ito process:

$$\widehat{X}_{t+\Delta} = \widehat{X}_t + \mu(\widehat{X}_t)\Delta + \sigma(\widehat{X}_t)\Delta W_{t+\Delta} \tag{1.6}$$

where $\widehat{X}_t$ denotes the discrete approximation of the continuous-time process $X(t)$. Furthermore, if we include the second-order terms, we obtain the Milstein scheme

$$\widehat{X}_{t+\Delta} = \widehat{X}_t + \mu(\widehat{X}_t)\Delta + \sigma(\widehat{X}_t)\Delta W_{t+\Delta} + \frac{1}{2}\sigma^2(\widehat{X}_t)(\Delta W_{t+\Delta}^2 - \Delta). \tag{1.7}$$

[1] we only consider the case of equal interval for simplicity

Note that the additional term is from the double Wiener integral, which can be computed from the Wiener increment $\Delta W_{t+\Delta}$ since

$$\int_t^{t+\Delta} \int_t^{s_1} dW(s_2)dW(s_1) = \frac{1}{2}(\Delta W_{t+\Delta}^2 - \Delta),$$

using the Ito's Lemma. (The proof is also referred to Appendix A.)

1.2.2 Discretization Bias and Convergency

Suppose that $\{\widehat{X}_t\}_{t=0}^T$ is the Euler discretization of a continuous-time path $\{X(t)\}_{t=0}^T$, the discretization bias of Euler approximation can be defined as

$$\begin{aligned} E(|\widehat{X}_T - X(T)|) &= E(|\int_0^T \mu(\widehat{X}_s)ds - \mu(X(T))\int_0^T ds) \\ &\quad + E(\int_0^T \sigma(\widehat{X}_s)dW(s) - \sigma(X(T))\int_0^T dW(s)|), \end{aligned} \tag{1.8}$$

at the final time instance T, given $\widehat{X}_0 = X(0)$. A nonzero bias generally exists when the drift and diffusion terms are not constant.

It is clear that the discretization bias converges to zero when the discretization interval goes to zero. In the other hand, we often want to show the discretized process $\widehat{X}_t$ converges in the strong sense with order γ ($\gamma > 0$) if there exists a finite constant K such that

$$E(|\widehat{X}_t - X(t)|) \leq K\Delta^\gamma \tag{1.9}$$

for any discretization interval Δ.

It can be shown that the Euler approximation converges with strong order of 0.5, which means the discretization bias is $O(\Delta^{0.5})$, under Lipschitz and bounded growth conditions on the drift and diffusion. And the Milstein scheme converges with strong order of 1.0 under the similar assumptions. Generally speaking, we can obtain more

accurate approximation by adding additional integrals from the stochastic Taylor expansion. Such integrals contain additional information about the sample paths of the Wiener process over the discretization intervals. For example, the Taylor approximation with strong convergence order of 1.5 can be obtained by including four more complicated integrals.

1.2.3 Impact of Euler Discretization on Statistical Inference

When the discretization interval converges to zero, the discretization schemes will converges to the "true" continuous-time stochastic processes. Thus, it is natural to preclude that the impact of discretization errors on parameter estimation will also decay. Florens-Zmirou (1989) and Yoshida (1992) show that the MLE $\hat{\theta}$ of the drift parameters θ in diffusion processes

$$dX(t) = \mu(X(t);\theta)dt + \sigma(X(t))dW(t), \tag{1.10}$$

is consistent when the discretization interval goes to zero and other conditions are satisfied as

$$\hat{\theta} \to \theta \qquad \text{if} \qquad \Delta \to 0, N \to \infty, N\Delta \to \infty,$$

where N is the number of discrete observations.

Since the discretization interval Δ is not zero in practice, $\hat{\theta}$ is generally asymptotically biased. Florens-Zmirou (1989) shows that, when the sample interval Δ is a nonzero constant, the MLE $\widehat{\theta}_\Delta$ in the Euler scheme for (1.10) does not converge to the true value of the parameter θ. The asymptotical bias is a function of the discretization Δ.

For instance, in Vasicek's model, the estimators of parameters a, b and c using Euler approximation can be shown to be inconsistent since they converge to values

that are not identical to the true values as long as the discretization interval $\Delta \neq 0$,

$$\widehat{a}_{euler} \rightarrow \kappa_0\theta_0\Delta \neq \theta_0(1 - e^{-\kappa_0\Delta}) = a_0,$$
$$\widehat{b}_{euler} \rightarrow 1 - \kappa\Delta \neq e^{-\kappa_0\Delta} = b_0,$$
$$\widehat{c}^2_{euler} \rightarrow \sigma^2\Delta \neq \frac{\sigma_0^2}{2\kappa_0}(1 - e^{-2\kappa_0\Delta}) = c_0^2$$

where $\kappa_0, \theta_0, \sigma_0$ or a_0, b_0, c_0 are true parameter values. This will be discussed in detail in the next section.

Although the MLE for the discretization scheme is asymptotically biased and the magnitude of the asymptotical bias may have an analytical form, it is worthwhile empirically exploring the impact of the discretization errors on the estimators using the discretization schemes, particularly compared with the finite-sample variances of the estimators. The following questions can be addressed in the empirical study:

- How is the performance of the estimators for the Euler scheme affected by the length of the discretization interval? Or what size of the sampling interval will cause an significant discretization errors to parameter estimation?

- Will and how much will the parameter estimation using the Euler scheme be improved by reducing the discretization intervals?

The second question also has practical implication to Bayesian inference using the data augmentation technique proposed by Elerian, Chib and Shephard (2001).

1.3 Euler Discretization under Vasicek's Model

1.3.1 The Model

Vasicek (1977) introduced the first popular continues-time term structure model of interest rates. In Vasicek's model, the spot rate $r(t)$ follows an Ornstein-Uhlenbeck process characterized by an SDE

$$dr(t) = \kappa(\theta - r(t))dt + \sigma dW(t), \tag{1.11}$$

where the parameter θ represents the long-term equilibrium interest rate level, the parameter κ controls the adjusting speed of spot rates to the long-term level, and the parameter σ reflects the volatility. Intuitively speaking, the process will drift up when the current spot rate is below the long-term level θ and will drift down when the current spot rate is above it. So the Ornstein-Uhlenbeck process captures mean-reversion, the stylized fact of interest rates. However, Vasicek's model has a drawback that the probability that the spot rate drops below zero is positive under the Gaussian transition density assumption. The interest rate is never negative in reality.

A main reason why Vasicek's model is one of the most popular term structure models is that the Ornstein-Uhlenbeck process in the Vasicek model is a Gaussian process which is relatively easy to handle in statistical inference. This allows us to do "exact" estimation on Vasicek's model since both the marginal and transition densities are fully specified.

Solving the SDE (1.11), an explicit solution of the spot rate at any time t is obtained as

$$r(t) = r(0)e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma e^{-\kappa t} \int_0^t e^{\kappa s} dW(s). \tag{1.12}$$

For the purpose of inference, we are interested in the density of $r(s)$ conditional on $r(t)$ $(t < s)$, which is normal with mean and variance given by

$$E(r(s)|r(t)) = r(t)e^{-\kappa(s-t)} + \theta(1 - e^{-\kappa(s-t)}) \quad (1.13)$$

$$Var(r(s)|r(t)) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(s-t)}). \quad (1.14)$$

1.3.2 Exact Bayesian Inference Using Gibbs Sampler

Suppose that we observe discretely sampled spot rates $\{r_i\}_{i=0}^n$ at time points $\{t_i\}_{i=0}^n$. Assume that the time intervals between every two samples are a constant, $\Delta = t_i - t_{i-1}$ [2] . for $i = 1, 2, ...,$ and $t_0 = 0$. We use a subscript i to denote the sample realized at time $t(i)$, i.e., $r_i = r(t_i) = r(i\Delta)$. If the spot rates $\{r_i\}_{i=0}^n$ are sampled from Vasicek's model, the conditional density of r_i on r_{i-1} is normal with mean and variance given by

$$E(r_i|r_{i-1}) = r_{i-1}e^{-\kappa\Delta} + \theta(1 - e^{-\kappa\Delta}) \quad (1.15)$$

$$Var(r_i|r_{i-1}) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta}). \quad (1.16)$$

Since we know the full joint distribution of all samples, the statistical inference on the model is straightforward. The likelihood function is the product of normal density functions as

$$l(\kappa, \theta, \sigma) = \prod_{i=1}^{n} p(r_i|r_{i-1}) \quad (1.17)$$

$$= \prod_{i=1}^{n} \phi(r_i; E(r_i|r_{i-1}), Var(r_i|r_{i-1})) \quad (1.18)$$

[2] we assume that Δ is measured in year since the interest rate is usually measured in year. $\Delta = 1/12$ for monthly data, $\Delta = 1/50$ for weekly data and $\Delta = 1/250$ for daily data, etc.

where $\phi(x; a, b)$ denotes the normal density function of random variable x with mean a and variance b. We adopt flat priors for parameters as

$$p(\kappa) \propto c; p(\theta) \propto c; p(\sigma) \propto \sigma^{-1}$$

By the Bayes's rule, the joint posterior density of three parameters are

$$p(\kappa, \theta, \sigma | data) \propto \prod_{i=1}^{n} \phi(r_i; E(r_i|r_{i-1}), Var(r_i|r_{i-1})) \cdot \sigma^{-1}. \tag{1.19}$$

It is clear that the Bayesian posterior mean using the flat priors is equivalent to the maximum likelihood estimator (or OLS) in this model.

To obtain the marginal posterior densities of a parameter, we have to integrate the joint density over other parameters. Gibbs sampler is a numerical integration method, which iteratively samples parameters from their fully conditional densities.

To implement the Gibbs sampler algorithm, we first consider the Vasicek's model as a linear Gaussian regression model:

$$r_{i+1} = a_{exact} + b_{exact} r_i + c_{exact} \varepsilon_{i+1} \tag{1.20}$$

where $\varepsilon_{i+1} \sim N(0, 1)$, and the new parameters are linked to the original parameters as

$$a_{exact} = \theta(1 - e^{-\kappa\Delta}); b_{exact} = e^{-\kappa\Delta}; c^2_{exact} = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta}) \tag{1.21}$$

Instead of sampling the original interested parameters, we use the Gibbs sampler to sample three transformed parameter a_{exact}, b_{exact} and c_{exact}, because Bayesian inference and Gibbs sampler in a linear Gaussian regression model is straightforward. After obtaining the MCMC samples for three transformed parameters, we can easily recover three original parameters. The Gibbs sampler algorithm for the "exact" model is as follows:

Step 1 Set the initial values $a_{exact}^{(0)}, b_{exact}^{(0)}$ and $c_{exact}^{(0)}$;

Step 2 Draw $\beta^{(j)} = (a_{exact}^{(j)} b_{exact}^{(j)})'$ from a normal distribution

$$N(\hat{\beta}_{ols}, c_{exact}^{2(j-1)}(X'X)^{-1})$$

where $\hat{\beta}_{ols} = (X'X)^{-1}X'Y$, Y and X are respectively the dependent variable observations vector and independent variables observations matrix in the linear model (1.20) and $c_{exact}^{(j-1)}$ is the $j-1$-th draw of the parameter c_{exact};

Step 3 Draw $c_{exact}^{2(j)}$ from an inverted gamma distribution

$$IG(\frac{N}{2}, \left[\frac{1}{2}\sum_{i=1}^{n}(r_i - a_{exact}^{(j)} - b_{exact}^{(j)} r_{i-1})^2\right]^{-1})$$

Step 4 Recover three original interested parameters by

$$\kappa^{(j)} = -\frac{ln(b_{exact}^{(j)})}{\Delta}; \theta^{(j)} = \frac{a_{exact}^{(j)}}{1 - b_{exact}^{(j)}}; \sigma^{2(j)} = \frac{2\kappa^{(j)} c_{exact}^{2(j)}}{1 - 2b_{exact}^{(j)}};$$

Step 5 Iterate the procedure by increasing j.

1.3.3 Asymptotic Discretization Bias of Euler Approximation

When we apply the Euler scheme to Vasicek's model, an approximation of the SDE (1.11) is:

$$r_{i+1} - r_i = \kappa(\theta - r_i)\Delta + \sigma\sqrt{\Delta}\varepsilon_{i+1} \tag{1.22}$$

which can be expressed as

$$r_{i+1} = a_{euler} + b_{euler} r_i + c_{euler}\varepsilon_{i+1} \tag{1.23}$$

where

$$a_{euler} = \kappa\theta\Delta; b_{euler} = 1 - \kappa\Delta; c^2_{euler} = \sigma^2\Delta. \tag{1.24}$$

The discretized model is also a linear Gaussian model. It is easy to show that

$$a_{euler} \to a_{exact}; b_{euler} \to b_{exact}; c_{euler} \to c_{exact}$$

when $\Delta \to 0$. This means the Euler discretization converge to the true continuous process when the sampling interval is infinitely small. At the same time, this shows the asymptotic discretization bias of Euler discretization on Vasicek's model always exists when the sampling interval is nonzero.

To show the inconsistency, suppose we obtain the point estimates $\hat{a}, \hat{b}$ and $\hat{c}$, which converge to the true parameters in (1.21) as

$$\begin{aligned}
\hat{a} &\to \theta_0(1 - e^{-\kappa_0\Delta}), \\
\hat{b} &\to e^{-\kappa_0\Delta}, \\
\hat{c}^2 &\to \frac{\sigma_0^2}{2\kappa_0}(1 - e^{-2\kappa_0\Delta}),
\end{aligned}$$

where κ_0, θ_0 and σ_0 are true parameter values. Then we recover the estimates of original parameters κ, θ and σ by using the discretized model as

$$\begin{aligned}
\hat{\kappa}_{euler} &= \frac{1 - \hat{b}}{\Delta} \to \frac{1 - e^{-\kappa_0}}{\Delta} \neq \kappa_0, \\
\hat{\theta}_{euler} &= \frac{\hat{a}}{1 - \hat{b}} \to \frac{\theta_0(1 - e^{-\kappa_0\Delta})}{1 - e^{-\kappa_0\Delta}} = \theta_0, \\
\hat{\sigma}^2_{euler} &= \frac{\hat{c}^2}{\Delta} \to \sigma_0^2 \frac{1 - e^{-2\kappa_0\Delta}}{2\kappa_0\Delta} \neq \sigma_0^2.
\end{aligned}$$

We can find that both estimators for κ and σ are inconsistent, while only the estimator for θ is not introduced the bias by the Euler discretization.

Consider a second-order Taylor expansion of the exponential function $e^{-\kappa\Delta}$:

$$e^{-\kappa\Delta} = 1 - \kappa\Delta + \frac{1}{2}(\kappa\Delta)^2 + o(\Delta^2). \tag{1.25}$$

Then the asymptotic discretization bias for parameters κ and σ^2 in Vasicek's model can be numerically measured as

$$\begin{aligned} Bias(\kappa) &= plim(\hat{\kappa}_{euler}) - \kappa_0 = \frac{1 - e^{-\kappa_0}}{\Delta} - \kappa_0 \\ &= \frac{1}{2}\kappa_0^2\Delta + o(\Delta), \end{aligned} \tag{1.26}$$

$$\begin{aligned} Bias(\sigma^2) &= plim(\hat{\sigma^2}_{euler}) - \sigma_0^2 = \sigma_0^2\frac{1 - e^{-2\kappa_0\Delta}}{2\kappa_0\Delta} - \sigma_0^2 \\ &= \sigma_0^2\kappa_0\Delta + o(\Delta). \end{aligned} \tag{1.27}$$

We can conclude that the discretization bias could be negligible relative to the value of parameters when the discretization interval Δ is sufficiently small.

1.3.4 Monte Carlo Experiments

Although the numerical explanation of Euler discretization bias is quite clear in Vasicek's model, it is still necessary to investigate the effect of discretization on the empirical analysis. In this section, we compare the accuracy of the estimators using Euler discretization on the continuous-time model to that of an "exact" estimator, which is feasible in Vasicek's model.

Since the continuous-time model is assumed to be the true model, we generate simulated data using the exact transition distribution of the continuous-time Vasicek's model characterized by Equation (1.20) and (1.21). The Bayesian inference using discretely sampled data is implemented on two models: one model is the "exact" model, in which the likelihood function is the product of the exact transition density

functions of the continuous-time model, the other model is the discretized model using Euler approximation, in which the transition densities are derived from the Euler discretization of the continuous-time model.

Gibbs sampler is implemented to obtain the full posterior distributions and the posterior means are reported as the point estimators of parameters. Using the flat priors, the posterior mean is equivalent to the maximum likelihood estimator in Vasicek's model. Theoretically, the "exact" estimator is consistent, while the "Euler" estimator is asymptotically biased. Moreover, the smaller the sampling interval is, the closer the "Euler" estimator should be to the "exact" estimator. We are interested in the finite-sample performance of the "Euler" estimator compared to the "exact" estimator.

Two sets of Monte Carlo experiments are conducted. In the first set, the sample size is fixed to be $1,000$. The estimates for sampling intervals $\Delta = 10, 5, 1, 1/4, 1/20, 1/100$ are reported and the impact of discretization interval on estimation is evaluated. In the second set, the whole sampling period is fixed to be 40 years. We generate data according to three popular sampling intervals $\Delta = 1/12$, $\Delta = 1/50$ and $\Delta = 1/250$, responding to monthly, weekly and daily sampling. Then sample sizes are different, being 480, $2,000$ and $1,0000$ respectively, for three intervals. It seems that the second set of experiments can better mimic the practical case. In reality, we often face the choice of different data sets with the same sampling period and different sampling frequencies. If we increase sampling frequency or "fill" the discretization intervals, the sample size will increase.

Furthermore, we need a criterion to measure the accuracy of estimators. Mean absolute deviation (MAD) is used in this section to compare the performance of

Table 1.1: Comparison of MADs: fixed sampling size, different sampling intervals

	$\Delta = 10$		$\Delta = 5$		$\Delta = 1$	
Parameter	Euler	Exact	Euler	Exact	Euler	Exact
κ	.4004	.2093	.3165	.0687	.1065	.0344
θ	.0215	.0215	.0217	.0217	.0440	.0440
σ	.5436	.1732	.4385	.0565	.1501	.0217
	$\Delta = 1/4$		$\Delta = 1/20$		$\Delta = 1/100$	
Parameter	Euler	Exact	Euler	Exact	Euler	Exact
κ	.0554	.0641	.1290	.1343	.5453	.5516
θ	.0842	.0842	.1806	.1806	.4259	.4259
σ	.0375	.0189	.0142	.0178	.0241	.0275

Note: the true values of parameters are

$$\kappa = 0.5, \theta = 4.0, \sigma = 0.8$$

"Euler" and "exact" estimators. The MAD of a parameter θ is defined as

$$MAD(\theta) = \frac{1}{R}\sum_{i=1}^{R} |\hat{\theta}^{(i)} - \theta|,$$

where $\hat{\theta}^{(i)}$ is the point estimate of parameter θ at the i-th replication. A large value of MAD indicates a poor performance of the point estimate. 300 replications are made to compute the MADs. In each replication, $6,000$ MCMC samples are drawn and first $1,000$ are "burned". The results of MADs are reported in Table 1.1 and 1.2.

The values of MAD are reported in Table 1.1 and 1.2. First, we compare the values of MAD of "exact" estimates and "Euler" estimates in Table 1. When the sampling intervals are "too large", say $\Delta = 10, 5$ and 1. The "Euler" estimators show large discretization bias since their MADs are significantly bigger than those of

Table 1.2: Comparison of MADs: fixed sampling period, different sampling intervals

	$\Delta = 1/12$		$\Delta = 1/50$		$\Delta = 1/250$	
	$N = 480$		$N = 2,000$		$N = 10,000$	
Parameter	Euler	Exact	Euler	Exact	Euler	Exact
κ	.1460	.1576	.1577	.1607	.1641	.1647
θ	.2051	.2051	.1967	.1967	.2138	.2138
σ	.0236	.0357	.0102	.0120	.0043	.0045

Note: the true values of parameters are

$$\kappa = 0.5, \theta = 4.0, \sigma = 0.8$$

"exact" estimators. When the intervals become smaller, such as $\Delta = 1/4$, the "Euler" estimates do not performs worse than the "exact" estimators any more and are even closer to the truth than "exact" estimates in many settings. This contradiction to the argument of discretization bias should be due to the finite-sample variances.

Another interesting finding is that the MADs for both "exact" and "Euler" estimators are getting worse when the sampling intervals become extremely small. This result seems to be a contradiction to our theory in the first place because the shorter the discretization interval is, the closer discretization schemes should be to the true diffusion process. If we go back and check the conditions for the consistency of the MLE for the diffusion process. In Florens-Zmirou (1989) and Yoshida (1992), the necessary conditions for the consistency of the MLE include $\Delta \to 0$, $N \to \infty$ and $N\Delta \to \infty$. That means, when the discretization interval goes to zero, the sample size of discrete observations must increase at a faster speed than the speed at which the interval shrinks. In this experiment setting, the sample size does not increase when

the discretization interval decreases. When this trend persists, neither estimators converge to the true values.

In Table 1.2, we also find no evidence that the "exact" estimator dominates the "Euler" estimator when the sampling intervals are as small as monthly. For both parameters κ and σ, Euler scheme yields lower MAD using all three discretization intervals. The result suggests that monthly sampling interval is already short enough to provide satisfying approximation to Vasicek's model in these experiments, and the attempt to reduce the discretization errors by reducing the sampling interval may not yield worthwhile rewards. In that sense, a higher-order discretization scheme or a data augmentation procedure for Euler approximation might be redundant when the sampling interval is regularly small.

Meanwhile, we do not observe the obvious decrease of discretization bias when we shorten the length of the sampling interval Δ. As seen in the table, the accuracy of estimate of κ is getting worse when the sampling interval decreases from 1/12 to 1/50 to 1/250. The performance of the estimate of θ is improved first when Δ drops to 1/50, then becomes poorer when the interval is shorten to 1/250. The evolution of the estimate of σ is the same as that of θ in this particular experiment. Lo (1988) claims that the maximum likelihood estimator of the drift parameter of a Wiener process is inconsistent when the sampling size goes to zero at the same time that the total sampling period is fixed, which is exactly our second experiment design. This might explain the poor performance of both "exact" and "Euler" estimators for an extremely short interval. This result further shows that using data augmentation might deteriorate the estimation in some cases.

1.4 Euler Discretization Under CIR Model

1.4.1 The Model

CIR model is another best-known term structure model of interest rates. In their seminal paper, the spot rate is modelled to follow a square-root (Bessel) process as

$$dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dW(t). \tag{1.28}$$

Compared with the Vasicek's model, the CIR model keeps the mean-reverting characteristic, but the volatility is not constant, but depends on the spot rate $r(t)$. The property of this process seems to be consistent with observed styled facts of nominal interest rates: the interest rate is less volatile for low levels than high levels of the rate. Moreover, the nominal interest rate cannot be negative in the CIR model, which is a major advantage relative to the Vasicek's model.

The CIR model is probably the most popular term structure model both among academia and financial industry. Its popularity stems from the fact that it is the most tractable model of a positive mean reverting process. But undoubtedly it is more computationally complicated than Vasicek's model. The SDE characterizing the CIR square-root diffusion has no explicitly solution, though the transition density of the process has a closed-form expression. With the original contribution of Feller (1951), CIR (1985) show that the density of $r(s)$ conditioned on $r(t)$ can be evaluated as

$$p_{\chi^2}(r(s)|r(t)) = ce^{-u-v}(\frac{v}{u})^{q/2}I_q(2\sqrt{uv}) \tag{1.29}$$

where

$$c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa(s-t)})}, u = cr(t)e^{-\kappa(s-t)}, v = cr(s), q = \frac{2\kappa\theta}{\sigma^2} - 1$$

and $I_q(\cdot)$ is the modified Bessel function [3] of the first kind of order q, which is defined as

$$I_q(z) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(q+k+1)} \left(\frac{z}{2}\right)^{2k+q}.$$

This noncentral χ^2 distribution has degrees of freedom $2(q+1)$ and a non-centrality parameter $2u$. Furthermore, the conditional mean and variance of $r(s)$ are

$$E(r(s)|r(t)) = \theta(1 - e^{-\kappa(s-t)}) + e^{-\kappa(s-t)} r(t) \tag{1.30}$$

$$Var(r(s)|r(t)) = \sigma^2 \kappa^{-1} [r(t)(e^{-\kappa(s-t)} - e^{-2\kappa(s-t)}) + (\theta/2)(1 - e^{-\kappa(s-t)})^2] \tag{1.31}$$

1.4.2 Exact Bayesian Inference Using Metropolis-Hastings Algorithm

Again, suppose we have a set of sampled spot rates $\{r_i\}_{i=0}^n$. Based on the transition density of the CIR process, we can obtain the joint posterior density of three parameters:

$$p(\kappa, \theta, \sigma | data) \propto \prod_{i=1}^{n} p_{\chi^2}(r_i | r_{i-1}) \cdot \sigma^{-1}, \tag{1.32}$$

where $p_{\chi^2}(r_i|r_{i-1})$ is evaluated by equation (1.29).

The complication of the noncentral χ^2 distribution causes the analytical difficulty when we try to make statistical inference on the CIR model. Besides applying the methods of moments, a discretization that approximates it using a normal distribution seems a feasible approach. Chen and Scott (1995) propose an approximation as:

$$p(r_i|r_{i-1}) \approx \phi(r_i; E(r_i|r_{i-1}), Var(r_i|r_{i-1}))$$

[3] I used the GAUSS procedure `mbesseli` to evaluate the modified Bessel function of the first order. The fragility of this procedure to large argument limits the choice of parameter values and sampling intervals. I only chose relatively large intervals to do Monte Carlo experiments in this paper.

where $E(r_i|r_{i-1})$ and $Var(r_i|r_{i-1})$ are the functions in Equation (1.30) and (1.31). This approximation is based on a fact that a normal density function is likely to be close to a non-central χ^2 density function with similar mean and variance. The comparison between a non-central χ^2 density function and two kinds of normal approximation is shown in Figure 3. The real line represents the density function of r_i conditioned on $r_{i-1} = 8$ given a CIR model and some parameter values. The dotted line is derived from an Euler discretization approximation on the CIR model. And the dashed line is based on another normal approximation of the CIR transition density function, by using a normal density with the same mean and variance as the CIR density. The latter approach obviously provides a more accurate approximation than the former. However, the difference in the approximation accuracy is almost illegible in this graph.

This approximation motivates another feasible approach to estimating the CIR model. A Metropolis-Hasting (MH) algorithm allows to sample from an arbitrary "exact" distribution by drawing from a feasible proposal distribution first if the "exact" density function can be evaluated numerically. A normal approximation is a natural proposal density function for sampling from the noncentral chi-square distribution.

Fruhwirth-Schnatter and Geyer (1996) use MH to estimate the CIR state-space model. They call the MCMC algorithm *Metropolis within Gibbs*. The proposal density function they use is the normal approximation proposed by Chen and Scott (1993). Here I also use the MH within Gibbs sampler, but the proposal density function I adopt is simpler. If we use Chen and Scott's proposal, the next draw must be dependent on the previous draw. Instead, I select the Euler approximation as the proposal.

Therefore, the algorithm I use here is actually an independent MH algorithm.

Next, let us consider the Euler discretization of the CIR model:

$$r_{i+1} - r_i = \kappa(\theta - r_i)\Delta + \sigma\sqrt{r_i\Delta}\varepsilon_{i+1}. \tag{1.33}$$

The transition density function in the discretized model is:

$$p(r_{i+1}|r_i) = \phi(r_i; \kappa\theta\Delta + (1 - \kappa\Delta)r_i, \sigma^2\Delta r_i). \tag{1.34}$$

The Euler approximation provides simple proposal densities for sampling parameters from their full conditional posterior densities. Take the parameter κ as example, the conditional posterior density is:

$$p(\kappa|\cdot, data) \propto \prod_{i=1}^{n} p_{\chi^2}(r_i|r_{i-1}|\cdot), \tag{1.35}$$

where $|\cdot$ denotes conditional on all other parameters. Accordingly, the proposal density at the j-th draw can be chosen as:

$$q(\kappa^{(j)}) = \prod_{i=1}^{n} \phi(r_i; \kappa\theta\Delta + (1 - \kappa\Delta)r_{i-1}, \sigma^2\Delta r_{i-1}|\cdot). \tag{1.36}$$

Finally, the Metropolis-Hastings algorithm within Gibbs sampler is implemented in the following procedures:

Step 1 Set initial values $\kappa^{(0)}$, $\theta^{(0)}$ and $\sigma^{(0)}$;

Step 2 Draw current κ' from its proposal density characterized by the discretized model (1.33). Evaluate the function values of true density and proposal density at the $j-1$ and j-th draws. Accept κ' as κ^j with a probability

$$\rho = min\{1, \frac{p(\kappa'|\cdot)}{p(\kappa^{(j-1)}|\cdot)} \cdot \frac{q(\kappa^{(j-1)})}{q(\kappa')}\};$$

Step 3 Follow the same procedure to draw $\theta^{(j)}$ and $\sigma^{(j)}$ based on the independent MH algorithm;

Step 4 Iterate the procedure by increasing j.

1.4.3 Monte Carlo Experiments

First we assume the CIR square-root process is the true data generating process. We follow the simulation method discussed in Johnson and Kotz (1970) and Robert and Casella (1999) to generate noncentral χ^2 random variables: a noncentral chi-square distributed variable $\chi^2_\nu(\delta)$ with a degree of freedom $\nu > 1$ [4] can be written as a sum of two other random variables:

$$\chi^2_\nu(\delta) = \chi^2_1(\delta) + \chi^2_{\nu-1} \tag{1.37}$$

where $\chi^2_{\nu-1}$ is a central χ^2 variable, which can be generated by using a gamma random variable generator, and $\chi^2_1(\delta)$ is a noncentral χ^2 random variable with 1 degree of freedom, which can be generated using a standard normal variable $Y \sim N(0,1)$ and a constant as

$$(Y + \sqrt{\delta})^2 \sim \chi^2_1(\delta)$$

Different from Vasicek's model, CIR model requires a MH step to obtain MCMC estimates. We are also interested in the impact of the introduction of MH step on the accuracy of Euler discretization. Both the numerical evaluation of noncentral χ^2 density function and the implementation of hybrid MH algorithm within Gibbs sampler has dramatic need of computing resources. The Monte Carlo experiment of

[4] The random variable generation when $\nu < 1$ can be done by using the Poisson mixture of central chi-square variables. We do not consider this case in the experiments

Table 1.3: Empirical Performance of Euler Discretization under CIR Model

		Replication 1	Replication 2	Replication 3
Euler	κ	0.5277	0.5045	0.5679
	θ	3.9770	3.7556	3.8489
	σ	0.7785	0.7708	0.7987
Exact	κ	0.5262	0.5079	0.5699
	θ	3.9159	3.9358	3.9431
	σ	0.7811	0.7735	0.8073
Accept	κ	0.9686	0.9520	0.9483
	θ	0.9203	0.8786	0.8833
	σ	0.7466	0.7413	0.5363

Note: the true values of parameters are $\kappa = 0.5, \theta = 4.0, \sigma = 0.8$, the sampling interval $\Delta = 1/12$ and the sample size is $1,000$

estimating the CIR model is very time-consuming. For the CIR model, I no longer use MAD as the criterion of performance evaluation. Instead, I randomly report the results of three experiments, and compare the performance of two approaches in three experiments.

We expect that the "exact" estimation by using Metropolis-Hasting algorithm achieve more accurate results since Metropolis-Hasting algorithm "reselects" samples from the Euler discretization to some extent. However, the practical performance is not guaranteed based on our findings on the Vasicek's model. In Table 1.3, the Bayesian posterior mean is still used as the point estimator. Their distances to true

values are studied. The standard deviation and autocorrelation of MCMC samples are ignored in the report. We also report the acceptance rates for sampling three parameters since they can also be used to check the accuracy of Euler approximation. It is remarkable that the acceptance rate for parameter σ is relatively low, which is consistent with the poor estimation performance for σ. This phenomenon is reasonable because the diffusion term in the CIR model is heteroscedastic and much more complicated than that in the Vasicek's model.

The comparison is made based on the results presented in Table 1.3. There is no strong evidence supporting the advantage of "exact" estimation over the Euler discretization approach. The dominance relationship is vague based on those experiment results. Basically the estimates from two approaches are very close to each other. The differences for parameters κ and σ are less than .01. These experiments testify that the Euler discretization may provide reliable estimators for the CIR square-root process when the sampling interval is reasonably small.

1.5 Concluding Remarks

Continuous-time stochastic processes have been widely used in modern financial theory. But the continuous-time setting brings great difficulties in estimation. Discretization of continuous-time models is often inevitable since we cannot obtain the closed-form likelihood function of discretely sampled data in most cases. Euler scheme is the most popular discretization method, but the estimator using the Euler discretization scheme haven been proven to be asymptotically biased in general. In this paper, by empirically comparing the accuracy of the "exact" estimator and the estimator using the Euler approximation, we find that Euler approximation provides reliable esti-

mates relative to the "exact" estimator when the discretization interval is regularly small. In those cases, a higher-order discretization scheme or a data augmentation to Euler scheme might be redundant and even "be more of a hindrance than a help" for improving the accuracy of estimates. This result is consistent for either a Gaussian model or a non-Gaussian model and either when only a Gibbs sampler is used or when an MH step is added.

Restricted by the availability of closed-form solution of the models, only two single-factor models are studied in this paper. More complicated and successful multi-factor models (even with jumps) have become the mainstream today. The extension to multi-factor models with jumps should be considered, though it is not easy to find a multi-factor jump-diffusion model which can be exactly generated and estimated.

Chapter 2

Identification of State-Space Models: A Bayesian Perspective

2.1 Introduction and Literature Review

2.1.1 Definition and Early Development of Identification

The identification problem in econometrics was initiated by Koopmans (1949). Koopmans supplemented the traditional procedure of statistical inference, in which a population is inferred from a sample, by another step from the population to the structure of the model. Although statistical inference from the observations to the parameters of the joint distribution of the observations is feasible, the step from that distribution to the parameters of the structural models often fails. Then an identification problem arises.

The terminology of identification (or its equivalence, identifiability) was first applied to simultaneous equations model of econometrics, and was extended to general

scientific inference problems in the 50s and 60s. Koopmans (1949) and Koopmans and Reiersol (1950) introduce the identification problem to econometrics and discuss this with the structural equation system in details. There is a vector y_t of endogenous variables, say, price and quantity, and a vector x_t of exogenous variables, say, income. We can write the econometric supply-demand model in a structural form as:

$$\Gamma y_t = Bx_t + \varepsilon_t, \varepsilon_t \sim N(0, \Sigma). \tag{2.1}$$

But statistical inference is only made from the observations to the joint distribution of the observations characterized by the following reduced-form model:

$$y_t = \Pi x_t + v_t, v_t \sim N(0, \Omega), \tag{2.2}$$

where $\Pi = \Gamma^{-1}B$ and $\Omega = \Gamma^{-1}\Sigma\Gamma^{-1'}$. The following step to transform the joint distribution represented by parameters (Π, Ω) to the joint distribution represented by (Γ, B, Σ) is sometimes not trivial. If we can find more than one set of (Γ, B, Σ) to yield the same joint distribution, these sets are called observationally equivalent by Koopmans. Structural parameters (Γ, B, Σ) are identified when there is only one set of (Γ, B, Σ) that can be obtained from (Π, Ω). Identification on the structural parameters can be obtained by imposing restrictions such as $\Psi(\Gamma, B, \Sigma) = 0$, for instance, the exclusive restrictions on the simultaneous equations model.

Not only in the simultaneous equations model or in econometric models, identification problem also lies in general statistical inference. A simple example is that an observable random variable y is modelled to be affected by two independent errors as $y_i = \mu + u_i + v_i, u_i \sim N(0, \sigma_u^2), v_i \sim N(0, \sigma_v^2), cov(u_i, v_i) = 0$. The distribution of y only depends on $\sigma_u^2 + \sigma_v^2$ given μ. Obviously infinitely many pairs of (σ_u^2, σ_v^2) will generate the same value of the sum $\sigma_u^2 + \sigma_v^2$ and hence the same joint distribution

of observations. F.M.Fisher(1966,1977) treats the identification problem in a general statistical setting and initiates the separation of identification from a branch of economics, though identification has been an important topic remaining in econometrics and contributes one chapter by Hsiao(1983) in the Handbook of Econometrics.

Koopmans and Reiersol (1950) define identification in terms of model structures. A *structure* $S = (F, \phi)$ covers both parametric and nonparametric specifications. A vector of observed random variables y follow a conditional probability distribution $f(y|S)$, which is uniquely *generated* by the structure S. A model $\mathcal{S}$ is a set of possible structures to be investigated. Then identification problem can be formalized in a pair of definitions as follows:

- Definition 1: Two structures S and S^* are observationally equivalent if $f(y|S) = f(y|S^*)$ for all y.
- Definition 2: A structure S in $\mathcal{S}$ is identifiable if there is no other structure S in $\mathcal{S}$ which is observationally equivalent.

Rothenberg (1971) follows Koopmans and Reiersol (1950) and gives a little different definition, in which only parametric structures are considered. Rothenberg uses a parameter point to represent the whole structure. Suppose that α is the parameter vector in a parameter space A. Identification of parametric models can be defined as:

- Definition 1: Two parameter points α^1 and α^2 are said to be observationally equivalent if $f(y, \alpha^1) = f(y, \alpha^2)$ for all y.
- Definition 2: A parameter point α^0 in A is said to be identifiable if there is no other α in A which is observationally equivalent.

Rothenberg's definition is more popular than Koopmans's because he is more specific about the structure and model so that testing the identification of a structure becomes feasible. Rothenberg furthermore reaches a criterion that local identifiability of the parameters is equivalent to non-singularity of the information matrix.

2.1.2 Bayesian View on Identification

In the classical econometrics, lack of identification matters because statistical inference about structural parameters is impeded unless deterministic restrictions are imposed. But Bayesians claim that "underidentifiability causes no real difficulty in the Bayesian approach"(Lindly,1971). Raiffa and Schlaifa (1961) make clear the fact that a proper prior distribution implies a proper posterior distribution. Rothenberg (1973) claims that it is possible for a parameter to be "identifiable" even though the data are completely irrelevant. Zellner(1971) also argues that uncertain prior information can be used to solve identification problem without resorting to exact restrictions that are generally necessary in the sample theory framework.

A simple example to illustrate this point of view is about the multicollinearity problem in a linear regression model as follows:

$$Y = X\beta + \varepsilon \tag{2.3}$$

where X is a $n \times k$ matrix and β is a $k \times 1$ vector of unknown parameters. The OLS estimator for this linear regression model is $\hat{\beta} = (X'X)^{-1}X'Y$. There exists a multicollinearity problem about the inference of β if X is not full-column rank, which implies some explanatory variables are perfectly correlated. Multicollinearity causes difficulty in implementing the OLS estimation of β since the matrix $X'X$ is singular, hence the inverse does not exist. Multicollinearity can be also interpreted

as an identification problem in that more than one set of values of β yields the same value of the likelihood function. Because X is not full-column rank, it is possible to find a $k \times 1$ vector α such that $X\alpha = 0$. Then $\beta + c\alpha$ is observationally equivalent to β for any constant c, since:

$$X(\beta + c\alpha) = X\beta + c \cdot X\alpha = X\beta \tag{2.4}$$

The traditional solution to multicollinearity is to impose restrictions to the parameter vector β, for example, dropping one explanatory variables which is perfectly correlated to the other, which is equivalent to assigning the value zero to one parameter, or imposing some inequality restrictions as in a ridge regression.

β is unidentified in the presence of multicollinearity. However, multicollinearity causes no real difficulty in Bayesian inference of this linear model. A proper prior distribution is assigned to β as $p(\beta) \sim N(b_0, V_0^{-1})$. Suppose the variance of the error ε is known as σ^2. Let $\tau = \sigma^{-2}$ and $\hat{\beta} = (X'X)^{-1}X'Y$. Then the posterior distribution of β is normal with mean μ and variance V as follows:

$$\begin{aligned} \mu &= (X'X\tau + V_0)^{-1}(X'X\tau\hat{\beta} + V_0 b_0) \\ V &= X'X\tau + V_0 \end{aligned} \tag{2.5}$$

If the prior distribution is proper and V_0 is non-singular, the matrix $X'X\tau + V_0$ is possibly invertible even if $X'X$ is not. With a probabilistic restriction instead of a deterministic restriction, we obtain a properly concentrated posterior distribution for β and β seems "identifiable" from the Bayesian view.

In this example, a parameter that is not identifiable in the viewpoint of a frequentist could be "identifiable" in the viewpoint of a Bayesian. But Bayesians also may not be able to distinguish between infinite possible values of a parameter when the

parameter has a diffuse or flat posterior distribution. This happens when the parameters is unidentified in the Koopmans-Rothenberg definition and a diffuse or flat prior is used. A question arises here: does Bayesian theory need a different definition of identification from the classical one? Or equivalently, is identification a property of the likelihood only, or both the likelihood and the prior. Morale (1971) defines that a model is "identified" if the posterior density of parameters is not flat. In that sense, a parameter is still possibly "identifiable" from the Bayesian approach even if it is not identifiable from the classical approach.

Although this question is still unsolved, most Bayesians agree that the Koopmans-Rothenberg definition is meaningful in that an experiment is valuable if it can yield information leading a better decision. The identification problem can be also explained by the uninformativeness of the data. When the parameter is unidentifiable from the classical approach, data gives no guide to the inference of the parameter and the statistical inference based on this model structure in which the parameter is unidentified changes no opinion about this parameter. Drèze (1972) used the following theorem to characterize identification in terms of the informativeness of observations:

- Theorem: If $p(y|\beta_i)$ is the same for all i, then $p(\beta_i|y)$ is equal to $p(\beta_i)$, and the observation y is noninformative, when this property holds for all y, the β_i's are observationally equivalent.

This theorem can be simply proved by using the Bayes rule:

$$p(\beta_i|y) = \frac{p(\beta_i)p(y|\beta_i)}{\sum_i p(\beta_i)p(y|\beta_i)} = \frac{p(\beta_i)}{\sum_i p(\beta_i)} = p(\beta_i) \tag{2.6}$$

since $p(y|\beta_i)$ is the same for all i and $\sum_i p(\beta_i) = 1$. The theorem indicates that when the parameter is unidentified, the data is uninformative to this parameter and the prior dominates the likelihood and affects the posterior despite of infinite data.

It has been widely accepted that defining identification as a property of the prior is misleading and there is no necessity to redefine identification in the Bayesian approach. Kadane (1975) demonstrated that "identification is a property of the likelihood function and is the same whether considered classically or from the Bayesian approach". This remark claims that identification has nothing to do with the prior and an attention should be paid to the model structure even if a concentrated posterior can always be easily obtained with an informative prior, as Wald said "we must take care in interpreting 'making sense"'.

Identification problem and its Bayesian treatment have been extensively discussed in a variety of statistical or econometrical models such as generalized linear models (Gelfand and Sahu (1999)), structural VAR models (Sims and Zha (1998)), etc.. In this paper, we focus on a set of dynamic time series models, state-space models, which have been widely used in financial econometrics. Next section introduces state-space models and discusses the identification problem on state-spaces models caused by a linear transformation on state variables. The third section lists examples on the identification problem in a specific class of financial models with the state-space representation. The fourth section explores the empirical effects of the identification problem on the Bayesian inference and the convergence of MCMC algorithms. The fifth Section studies the impacts of the identification problem on the prediction of state-space models. The last section summarizes.

2.2 State-Space Models: Linear Transformation and Identification

The state-space modeling representation is very attractive to structural modeling of macroeconomic and financial time series because it assumes that the dynamics of economic time series is derived from the dynamics of some unobserved state variables that have sound economic meaning. A typical linear state-space model is given by the following system of equations:

$$\begin{aligned} y_t &= Hx_t + Gz_t + u_t \\ x_t &= Fx_{t-1} + v_t \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} E(u_t' u_t) &= Q \\ E(v_t' v_t) &= R \end{aligned}$$

where F, G, H, Q and R are matrices of parameters. It is possible to observe the realizations of y_t, while x_t is unobservable and has to be estimated based on the information of observed y_t and model structure.

Hamilton (1994) states that "in the absence of restrictions on F, H, G, Q and R, the parameters of the state-space representation are unidentified - more than one set of values for the parameters can give rise to the identical value of the likelihood function, and the data give us no guide for choosing among these". Hannan (1971) states that F, G, H, Q, R and x_t are identified only up to a linear transformation $x_t \to Tx_t, F \to TFT^{-1}, G \to TG, H \to HT^{-1}, Q \to TQ, R \to TRT^{-1}$, where T is a non-singular constant matrix.

The identification problem of state-space models caused by the linear transformation is illustrated by using a simple scalar state-space model without other exogenous variables as follows:

$$\begin{aligned} y_t &= ax_t + u_t \\ x_t &= bx_{t-1} + v_t \end{aligned} \tag{2.8}$$

where $u_t \sim N(0, c^2)$ and $v_t \sim N(0, d^2)$, for all t, are independent. All variables y_t, x_t and parameters a, b, c, d are scalars for simplicity of illustration.

Let $a^* = a/k$, $d^* = kd$ and $x_t^* = kx_t$ for an arbitrary constant k. We can show that $f(y_t|a, b, c, d) = f(y_t|a^*, b, c, d^*)$, hence (a, d) and (a^*, d^*) are observationally equivalent and parameters a and d are unidentified.

To show this, we evaluate the likelihood function with respect to these two sets of parameter values. The likelihood function of (2.8) can be written as:

$$L(a, b, c, d) = f(y_1, y_2, \cdots, y_n) \tag{2.9}$$

$$= \prod_{t=2}^{n} f(y_t|y^{t-1}) \tag{2.10}$$

where $y^{t-1} = \{y_1, y_2, \cdots, y_{t-1}\}$. The individual conditional likelihood function $f(y_t|y^{t-1})$ is often obtained by estimating unobserved state variables x_t using the Kalman filter. The Kalman filter is an algorithm to sequentially update the projection of the unobserved state variables. Given the Kalman filter estimate of the state variables, the conditional distribution of y_t is normal:

$$y_t|y^{t-1} \sim N(a\hat{x}_{t|t-1}, a^2 P_{t|t-1} + c^2) \tag{2.11}$$

where $\hat{x}_{t|t-1} = \hat{E}(x_t|y^{t-1})$ is the optimal linear forecast of x_t based on the observations of y up to $t-1$ and $P_{t|t-1} = \hat{E}(x_t - \hat{x}_{t|t-1})^2$ is the associated mean square errors of

the forecasts. The Kalman filter calculates the forecasts of x_t recursively as:

$$\begin{aligned} \hat{x}_{t+1|t} &= b\hat{x}_{t|t-1} + \frac{abP_{t|t-1}}{a^2P_{t|t-1}+c^2}(y_t - a\hat{x}_{t|t-1}) \quad (2.12) \\ P_{t+1|t} &= b^2(P_{t|t-1} - \frac{a^2P_{t|t-1}^2}{a^2P_{t|t-1}+c^2}) + d^2 \quad (2.13) \end{aligned}$$

Next we want to show that all individual conditional likelihood functions ($t = 1, \ldots, n$) in (2.11) keep unchanged if we replace the original values of parameters (a, b, c, d) with a new set values (a^*, b, c, d^*) . Let us start at $t = 1$. We initiate the Kalman filter using the stationary distribution of the state variable:

$$\begin{aligned} \hat{x}_{1|0} &= 0 \quad (2.14) \\ P_{1|0} &= \frac{d^2}{1-b^2} \end{aligned}$$

The initial conditional distribution is $y_1 \sim N(0, \frac{a^2d^2}{1-b^2} + c^2)$. This distribution does not change as we replace (a, d) with (a^*, d^*) because $a^*d^* = ad$. After the replacement, the initial forecast of state variable $\hat{x}_{1|0}$ at $t = 1$ is still zero. However, the MSE $P_{1|0}$ changes as $P^*_{1|0} = k^2P_{1|0}$.

Then we move to next time point $t = 2$. The Kalman filter yields the prediction of x_2 given the information up to time 1:

$$\begin{aligned} \hat{x}_{2|1} &= b\hat{x}_{1|0} + \frac{abP_{1|0}}{a^2P_{1|0}+c^2}(y_1 - a\hat{x}_{1|0}) \quad (2.15) \\ P_{2|1} &= b^2(P_{1|0} - \frac{a^2P_{1|0}^2}{a^2P_{1|0}+c^2}) + d^2 \end{aligned}$$

and $y_2|x_1, y_1 \sim N(a\hat{x}_{2|1}, a^2P_{2|1} + c^2)$. After we replace (a, d) with (a^*, d^*), both the estimate and variance of x_2 change as follows:

$$\begin{aligned} \hat{x}^*_{2|1} &= k\hat{x}_{2|1} \quad (2.16) \\ P^*_{2|1} &= k^2P_{2|1}. \end{aligned}$$

But the conditional distribution or the individual likelihood function does not change because $N(a^*\hat{x}^*_{2|1}, a^{*2}P^*_{2|1} + c^2) = N(a\hat{x}_{2|1}, a^2P_{2|1} + c^2)$.

By deduction, we can show the individual likelihood function is the same for (a, b, c, d) and (a^*, b, c, d^*) at $t = 3, 4, \ldots . n$. Finally, our conclusion is that different values for the parameter set (a, b, c, d) can give rise to identical values for the likelihood function. (a, b, c, d) and (a^*, b, c, d^*) are observationally equivalent and, particularly, the parameters a and d are not identified. We observe that such observationally equivalent model structures are obtained by a linear transformation on the state variable x_t. In the transformation, parameters b and c are not involved, which means we can not find different values of b or c that are observationally equivalent. So b and c are identified.

Identification can be achieved by imposing restriction on those matrices of parameters to make them have canonical form. This process is often called parameter normalization. For example, we can let $d = 1$ in the scalar model (2.8). Then new model structure is identifiable. This is shown in the fourth section based on the convergence of MCMC chains for the parameters.

2.3 Application in Finance: Identification of Affine Term Structure and Reduced-Form Credit Risk Models

The identification problem in state-space models has import implication in financial modeling since state-space models have been widely used in asset pricing. A typical example is term structure of interest rates. A zero-coupon bond that pays $1 at the

maturity T has a price at time t

$$P_t = E_t^Q[exp(-\int_t^T r_s ds)] \tag{2.17}$$

where Q is the risk-neutral probability measure. An m-factor affine term structure model (ATSM), proposed by Duffie and Kan(1996), assumes that the instantaneous short rate r_t is an affine function of a number of factors $x_1, x_2, \ldots, x_m$. Let a vector $X = (x_1, x_2, \ldots, x_m)$, then

$$r_t = \delta_0 + \delta' X_t = \delta_0 + \sum_{i=1}^{m} \delta_i x_{it} \tag{2.18}$$

And an ATSM assumes the factors follow "affine diffusion" as

$$X_t = K(\Theta - X_t)dt + \Sigma\sqrt{S_t}dW_t \tag{2.19}$$

where S_t is a diagonal matrix with the i-th diagonal element given by

$$[S_t]_{ii} = \alpha_i + \beta_i' S_t. \tag{2.20}$$

The bond prices P_t and short rates r_t are equivalently observable in practice, while the factors X_t are not. The ATSM has a state-space representation if measurement errors are introduced to equation (2.18). Based on the conclusion drawn in the previous section, the parameter vector $(K, \Theta, \Sigma, \{\alpha_i\}_{i=1}^m, \{\beta_i\}_{i=1}^m)$ is generally not identified without certain restrictions. It is possible to find a linear transformation to the latent state variables and parameters such that the dynamic of r_t is not affected. Dai and Singleton (2000) refer to such linear transformations as "invariant transformations". An invariant transformation for an m-factor ATSM is defined by a nonsingular $m \times m$ matrix L and a $m \times 1$ vector h such that $X_t \to LX_t + h, \delta_0 \to \delta_0 - \delta' L^{-1} h, \delta \to L'^{-1}\delta, K \to LKL^{-1}, \Theta \to L\Theta + h, \Sigma \to L\Sigma, \alpha_i \to \alpha_i - \beta_i' L^{-1} h, \beta_i \to L'^{-1}\beta_i$ for

$i = 1, \ldots, m$. A set of normalizations on parameter matrices and vectors are also proposed to achieve an identified m-factor ATSM model. For example, some α_i must be assigned zero, some β_i must be assigned one, some diagonal elements of Σ must also be normalized, etc..

Next we consider a reduced-form credit risk model, which is a special case of a two-factor ATSM model, to illustrate the identification problem of ATSM models. Lando (1998) and Duffie and Singleton (1999) present a valuation model for defaultable corporate bonds, which has been a benchmark for reduced-form credit risk models. In this model, the time-t price of a defaultable zero-coupon bond that pays a dollar at expiration date T unless it defaults before T is

$$P_t = E_t^Q \left\{ \exp \left[- \int_t^T (r_s + h_s(1 - L_s)) ds \right] \right\} \tag{2.21}$$

where r_t is the instantaneous default-free interest rate, h_t is the default probability or the hazard rate for default, and L_t is the recovery rate at time t. It is very common to assume both default-free rate and default probability follow some affine diffusion process, for example, a Cox-Ingersoll-Ross process in Duffee (1999):

$$\begin{aligned} dr_t &= \kappa_1(\theta_1 - r_t)dt + \sigma_1 \sqrt{r_t} dW_{1t} \\ dh_t &= \kappa_2(\theta_2 - h_t)dt + \sigma_2 \sqrt{h_t} dW_{2t} \end{aligned} \tag{2.22}$$

where W_{1t} and W_{2t} are standard Wiener processes and might be correlated. There are different treatments for the recovery rate L_t in practice. Duffee assumes that $L_t = 0$ for all t for the simplicity of estimation. Zero recovery rate implies that the bond holder is not able to recover any of the loss when default happens, which is obviously not realistic since most bonds have some forms of collateral in practice. A practical way to deal with L_t is to estimate a constant recovery rate from historical

data and replace L_t with the estimate in equation (2.21). Both ways are actually imposing restrictions on L to make the model identifiable. A wrong try would be a joint estimation of the recovery rate and default probability when the recovery rate is an unknown parameter, because the following two-factor ATSM is not identified:

$$\begin{aligned} P_t &= E_t^Q \left\{ \exp\left[-\int_t^T (r_s + h_s(1-L))ds \right] \right\} \\ dr_t &= \kappa_1(\theta_1 - r_t)dt + \sigma_1\sqrt{r_t}dW_{1t} \\ dh_t &= \kappa_2(\theta_2 - h_t)dt + \sigma_2\sqrt{h_t}dW_{2t} \end{aligned} \tag{2.23}$$

An invariance transformation as $h_t \to ch_t(1-L) \to (1-L)/c, \theta_2 \to c\theta_2, \sigma_2 \to \sqrt{c}\sigma_2$ would keep P_t unchanged.

Empirical evidences have shown that there exists significant negative correlation between default-free rate r_t and default probability h_t, which makes the above reduced-form model inadmissible in practice. Lando (2002) propose a transformation to above reduced-form credit risk model to incorporate the negative correlation. He assumes $h_t = \alpha + \beta r_t$, where β measures the correlation between $h_(t)$ and r_t. Assuming a constant recovery, the two-factor ATSM is reduced to a one-factor model as:

$$P_t = E_t^Q \left\{ \exp\left[-\int_t^T (\alpha L + (1+\beta L)r_s)ds \right] \right\}. \tag{2.24}$$

It is obvious that this one-factor ATSM is not identified again. Neither α nor β is identified because we can find an invariance transformation on r_t. All above examples show that it is relatively easy to specify an unidentified model when the asset pricing model has a state-space form.

2.4 Bayesian Identification and Parameter Estimation

Estimation of state-space models is generally not trivial because unobserved state variables must be estimated with parameters together. The Kalman filter is the traditional way to deal with a linear, normal state-space model. The extend Kalman filter might apply when the state-space model is nonlinear or nonnormal. But this is not guaranteed in general cases. MCMC methods have been shown to be suitable for the estimation of state-space models even if they are highly nonlinear and nonnormal. Using a single-move (Carlin, Polson and Stoffer, 1992) or multi-move (Carter and Kohn, 1994) Gibbs sampler, we are able to draw random samples from the posterior distributions of latent variables as well as other structural parameters.

In Introduction, we concluded that "identification problem does not cause real problem to Bayesian inference" if an informative prior is used. A legitimate posterior distribution is still attainable and MCMC chains may still converge to a stationary distribution even if some parameters are unidentified. Regarding the state-space models, some implications from this conclusion include:

1. MCMC convergence can be achieved and statistical inference based on a legitimate posterior distribution can be made for an unidentified state-space model, for which the classical inference often collapses;

2. The MCMC chain for an unidentified parameter might not converge if a flat prior is used.

There implications also arise some questions such as:

1. When a parameter in a state-space model is unidentified, does the stationary distribution of the MCMC chain "make sense" even if the chain always converges to something when the prior is informative?

2. Is the posterior distribution for an unidentified parameter exactly the same or very similar as the prior distribution, or equivalently, don't data provide any information on a parameter that is unidentified in a state-space model?

3. Is the MCMC convergence of other identified parameters affected when the convergence fails on the unidentified parameters?

4. Can we use the MCMC convergence as a criterion for checking the identification of a state-space model?

We try to answer these questions in the following by using a numerical example.

Model (2.8) is used as the data-generating process, in which the values of the parameters are: $a = 0.7, b = 0.8, c = 0.3, d = 0.2$. Using these values, we generate the observations on y_t for a sample size 300. We use a single-move Gibbs sampler to draw MCMC samples of four structural parameters (a, b, c, d), monitor the convergence of MCMC draws and analyze the posterior distributions. Totally $10,000$ MCMC samples are drawn for each parameter, and first $2,000$ are "burned out".

Figure 2.1 displays the MCMC draws when improper priors are used for four parameters $p(a), p(b) \sim const$, $p(c) \sim c^{-1}$ and $p(d) \sim d^{-1}$. Horizontal dished lines represent true values $a = 0.7, b = 0.8, c = 0.3, d = 0.2$. Only the chains for identified parameters b and c converge. The posterior means of b and c are 0.7955 and 0.3066. It can clearly seen that the MCMC chains for parameters a and d do not converge at all after $10,000$ draws, while b and c are more likely to have stationary posterior

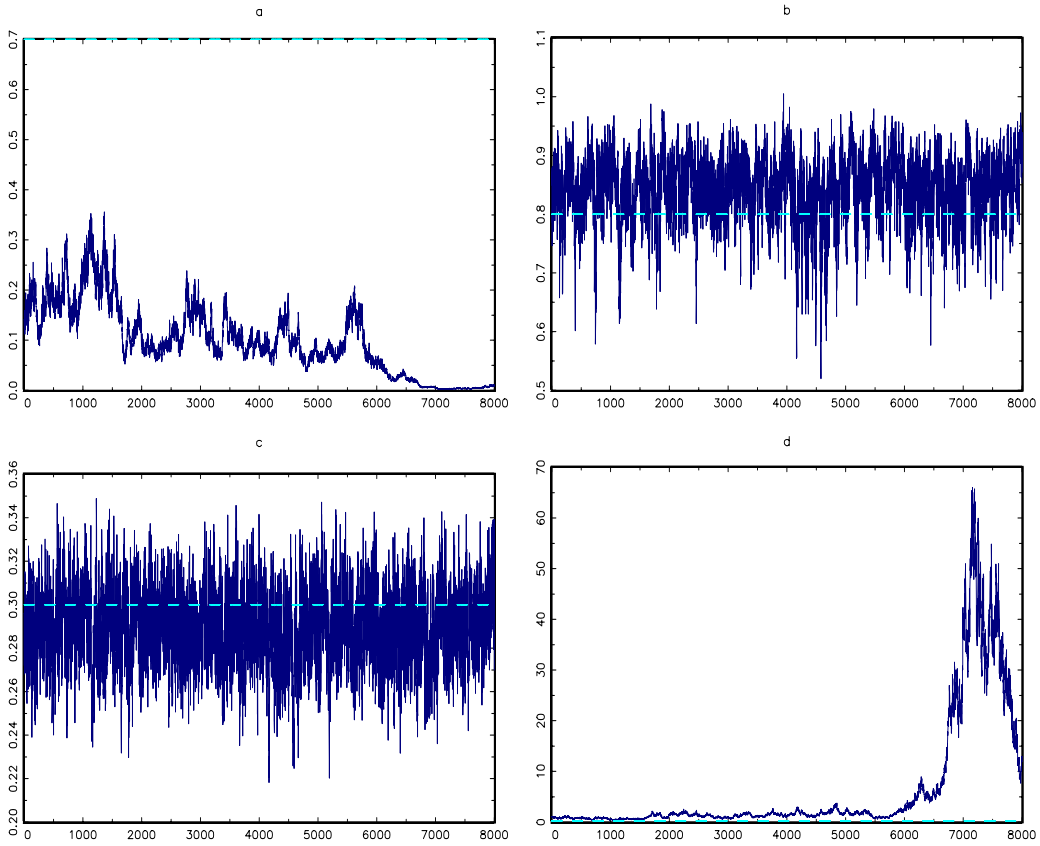

Figure 2.1: MCMC convergence of four parameters when priors are improper

distributions concentrated at the true values. The posterior means of b and c are 0.7955 and 0.3066, very close to the true values. These results are expected since it has been shown in the second section that b and c are identified but a and d are not. When improper priors are used on unidentified parameters, concentrated posterior distributions are still not achievable and Bayesian inference has troubles as well as the classical approach. Good news to question 3 above is that the MCMC convergence for identified parameters b and c seems unaffected by two unidentified parameters a and d. Both MCMC chains converge to stationary posterior distributions that are well concentrated at the true values, although the other two chains do not converge at all. The Gibbs sampler based on conditional posterior distributions works well even if some marginal posterior distributions are improper. Another interesting finding that should be reported is that the serial correlations for unidentified parameters a and d are very high, 97% and 99%, and even persistent when we choose a longer interval to compute correlations. The serial correlations for identified parameters b and c are also high, but die out soon when the interval becomes longer. This result has been reported by some author as an evidence whether a parameter is identified or not.

Next we explore what will happen if priors are proper and informative. Four conjugate priors are chosen as $a, b \sim N(10, 10^2)$ and $c^2, d^2 \sim IG(3, 0.1)$. Prior distributions are chosen to be quite far away from the true values of the parameters, which generate the data, so that we can compare the effects of priors and data on posteriors. The plot of MCMC draws is presented in Figure **??**. Now all four MCMC chains successfully pass the Geweke's Zg Convergency test, which indicate legitimate posteriors exist even for unidentified parameters when informative priors are used, which is consistent with our previous argument about Bayesian identification.

But a question still remains: do Bayesian inference results make any sense? We report that the posterior means for (a, b, c, d) are $0.0382, 0.9250, 0.4317, 1.6453$ respectively. It seems that the posterior distributions of two identified parameters b and c "make more sense" because they are roughly concentrated at the true values that govern the likelihood of generated data, considering the effects of "unreasonable" priors. The posterior means of two unidentified parameters a and d are not close to the true values, neither to the prior means. We have to admit that identification problem causes serious troubles since Bayesian inference does not supply too much useful information to improve our decision on these two unidentified parameters. This result implies that it has to be kept in mind for modelers and researchers that identification must be diagnosed even if their MCMC algorithms converge perfectly and a concentrated Bayesian posterior can be obtained. Without confidence about model identification, Bayesian parameter estimation and MCMC algorithms might be a "trap".

Drèze'e theorem states that the posterior distribution is the exactly the same as the prior distribution for an unidentified parameter when the data is totally uninformative. In this state-space model we can see the likelihood function is not absolutely flat and data still provides some information for the inference of unidentified parameters. Since the priors cannot perfectly dominate the posteriors, the posterior distributions of unidentified parameters are still a combination of information from both the prior distributions and the data. However, compared with the identified parameters, Bayesian inference on unidentified parameters "makes less sense" because the data is less relevant to them.

Finally we propose an answer to the question 4 above: it is possible to use the

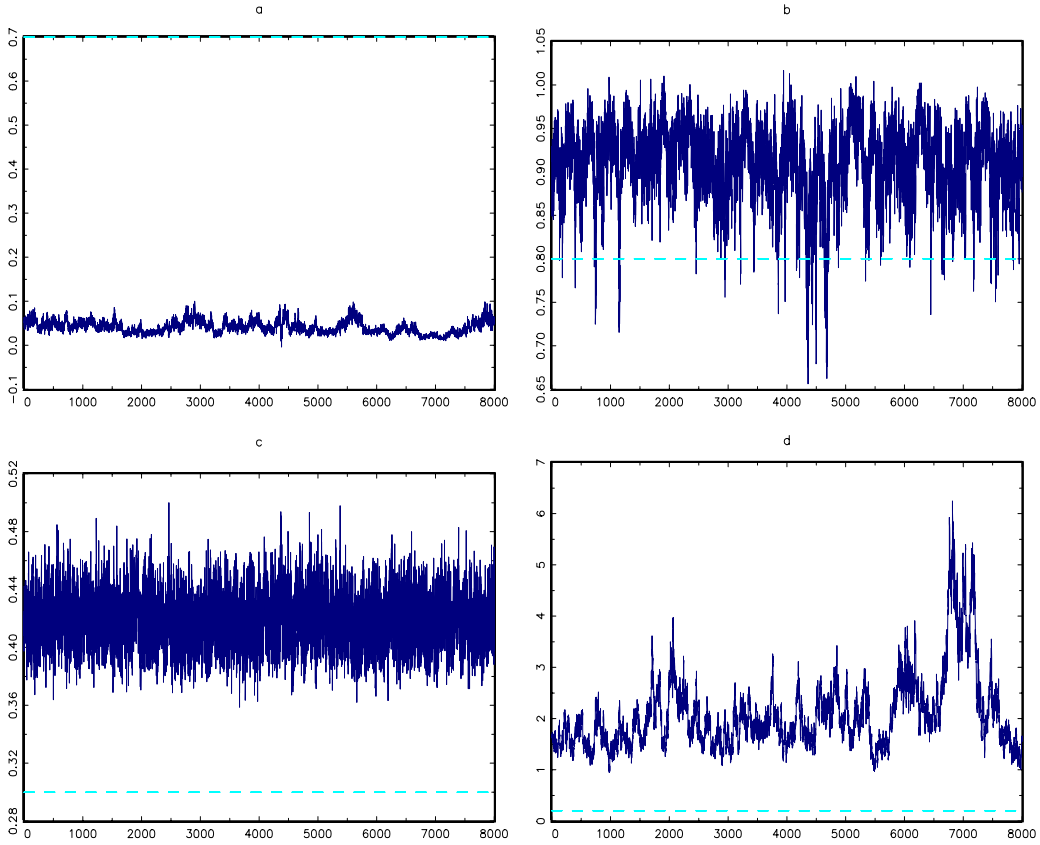

Figure 2.2: MCMC draws of four parameters when priors are proper and informative

MCMC convergence as a diagnostic tool for checking the identification of some structural models including state-space models. As long as we use proper informative priors, Bayesian inference always yields legitimate posterior distributions, and the MCMC chain must converge to a stationary distribution. If we observe that the MCMC chains do not converge in a state-space model, this is a strong sign for us that some parameters are unidentified in this model.

2.5 Bayesian Identification and Prediction

In many cases, we are interested in prediction rather than parameter estimation in research. Although identification still remains a problem in parameter estimation, it is unclear whether the identification problem imposes important impact on the prediction of a state-space model.

We have already shown that a legitimate Bayesian inference is possible for an unidentified state-space model given informative priors. Thus, it is not a problem to obtain a legitimate predictive posterior distribution of the observed variable y_t even if the state-space model is unidentified.

To evaluate the predictive performance, we compare the predictive posterior distribution of an unidentified state-space model and that of an equivalent but identified model. Normalization, or rescaling, can eliminate the overparameterization and, therefore solve the identification problem. For example, the state-space model in (2.8) becomes identified when the parameter d is normalized:

$$\begin{aligned} y_t &= a^* x_t^* + u_t \\ x_t^* &= b x_{t-1}^* + v_t, \end{aligned} \tag{2.25}$$

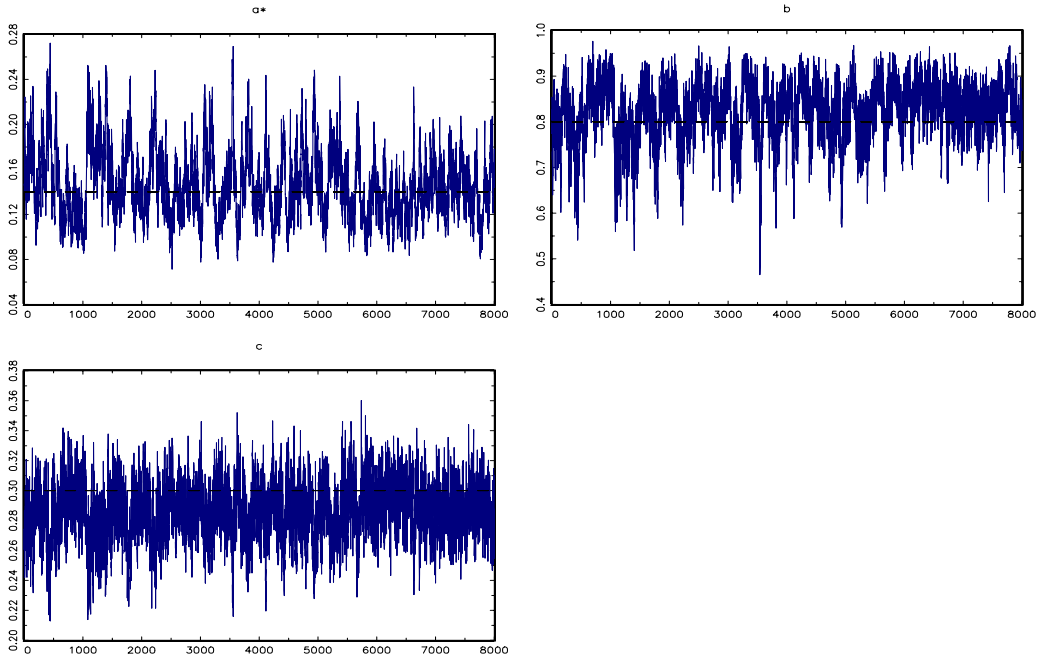

Figure 2.3: MCMC draws of three parameters in the identified state-space model when priors are improper

and $u_t \sim N(0, c^2)$ and $v_t \sim N(0, 1)$, where $a^* = a * d$ and $x_t^* = x_t/d$. Only three parameters (a, b, c) are left after normalization. Model (2.25) is equivalent to model (2.8) except that some parameters and latent variables are rescaled. This "canonical" model is identified. We simulate numerical data from this model and run a Gibbs sampler using noninformative priors. MCMC draws for three parameters are displayed in Figure 2.3. All three chains converge to the stationary distribution well concentrated at true values even if priors are improper.

We generate time series of y_t from both models: identified and unidentified, with two sets of true values of parameters. Note that two paths are almost the same since two models are equivalent. We proceed to estimate both models using the Bayesian

approach and generate the MCMC draws of the parameters from their posterior distribution. Our goal is to find the one-step-ahead prediction y_{T+1} MCMC algorithms allow us to easily obtain random draws from the posterior predictive distribution of y_{T+1} given the MCMC draws from the posterior distribution of the parameters and latent variables. Since model (2.25) is identified, we take its posterior predictive distribution as the benchmark, and investigate that how different the posterior predictive distribution of the unidentified model (2.23) is from the benchmark. The difference can be treated as a measure of the impact of the identification problem on the prediction of this state-space model.

Posterior distributions of parameters in the unidentified model (2.23) are sensitive to prior information even if the sample size is large. So a sensitivity analysis is conducted by using three different prior distributions:

- Prior 1: $a, b \sim N(0,1), c, d \sim IG(3,1)$ and $a^* \sim N(0, 0.4)$ [1]
- Prior 2: $a, b \sim N(10,1), c, d \sim IG(10,1)$ and $a^* \sim N(3, 0.4)$
- Prior 3: $a, b \sim N(10,1), c, d \sim IG(3,0.1)$ and $a^* \sim N(20, 70)$

The results are displayed in Table (2.1). Both the first and second set of priors are dominated by the data since the posterior means in the identified model are close to the true values used to generate the data. The third set of priors has a nonneglectable impact on the posterior distribution since the posterior means in the identified model are relatively far away from the true values. Obviously all three sets of priors have significant impacts on the posterior distributions in the unidentified models, which is

[1] Note that $a^* = a \cdot d$. Here the prior mean and variance of a^* match those of a and d

Table 2.1: Sensitivity Analysis of Prior Distribution on Parameter Estimates

Panel A: The Identified Model				
		Posterior Mean and Std using Different Priors		
Parameter	True Value	Prior 1	Prior 2	Prior 3
a^*	0.14	0.131(0.025)	0.136(0.025)	0.078(0.021)
b	0.8	0.846(0.054)	0.854(0.051)	0.916(0.040)
c	0.3	0.304(0.017)	0.292(0.016)	0.419(0.019)
Panel B: The Unidentified Model				
		Posterior Mean and Std using Different Priors		
Parameter	True Value	Prior 1	Prior 2	Prior 3
a	0.7	0.177(0.059)	0.453(0.112)	0.041(0.014)
b	0.8	0.857(0.051)	0.825(0.060)	0.922(0.038)
c	0.3	0.307(0.016)	0.290(0.016)	0.419(0.019)
d	0.2	0.782(0.305)	0.336(0.054)	1.999(0.691)

expected. Note that the MCMC algorithms converge under the unidentified model with three sets of informative priors.

The posterior predictive distributions of y_{T+1} in the identified and unidentified models are plotted and compared in Figure 2.4. We can clearly see that the unidentified model yields almost the same prediction as the identified parameter does when different informative priors are used. It can be concluded that the identification problem "does not matter" for the prediction while it may cause misleading estimates for parameters.

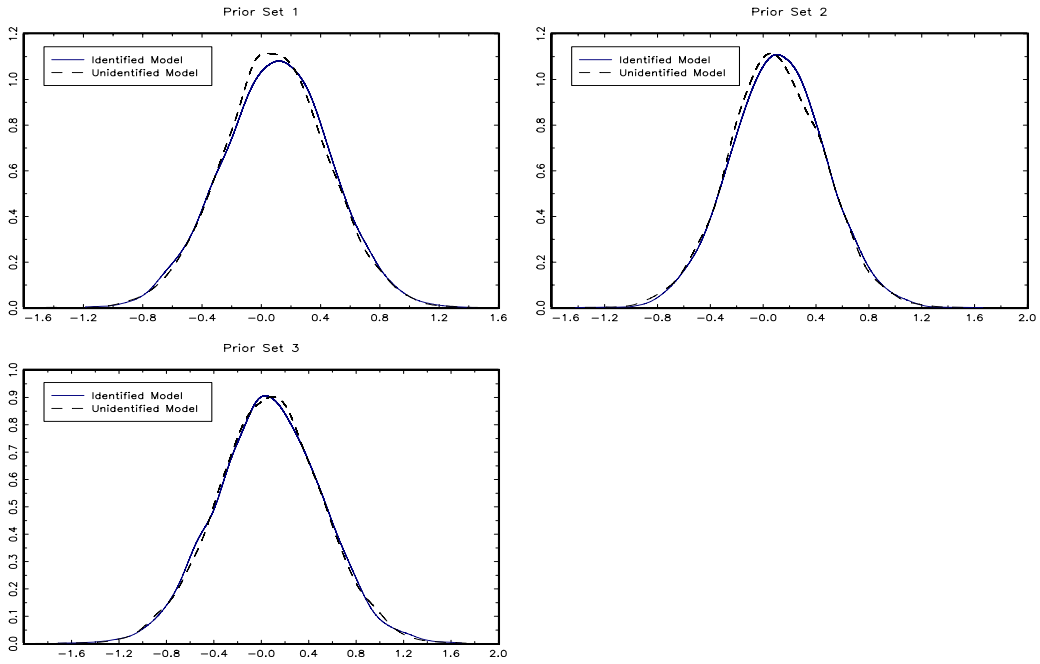

Figure 2.4: The posterior predictive distributions of y_{T+1} in the identified and unidentified models

2.6 Concluding Remarks

State-space models have been widely used in financial econometrics. But modelers often fail to realize that state-space models without restrictions are generally unidentified. Bayesian inference and MCMC algorithms have been frequently used to estimate state-spaces models including their application in finance. But researchers might ignore the identification problem in the models when they only focus on the convergence of their MCMC algorithms. It is the advantage of Bayesian inference over the classical inference that identification problem causes no real trouble in Bayesian inference because a legitimate posterior distribution can always be attained when priors are proper and informative even if the parameter is unidentified. However, Bayesian treatment to identification problem could be the disadvantage, even a "trap" for researchers. Their Bayesian estimates of parameters might yield results that do not "make sense" even if their MCMC algorithms converge perfectly in the presence of unidentified parameters. This paper suggests high caution be kept for researchers to use MCMC algorithms to estimate state-space models. Model specification should always be checked about potential identification problem before Bayesian inference can be implemented effectively. We also found that, if the prediction instead of parameter estimation is the main concern for the researcher, the identification problem really causes no problems to Bayesian prediction in state-space models.

Chapter 3

Bayesian Inference of Credit Risk Models

3.1 Introduction

While the concept of "credit risk" is as old as banking itself, credit risk emerged as a significant risk management issue during the 1990s. A series of financial crisis in the 1990s and more recent Subprime Crisis evidences the importance of credit risk measurement and management in the financial industry. As the introduction of financial mathematics to this area, practitioners have become more and more relied on quantitative model to measure credit risk and price the securities subject to default such as corporate bonds, mortgage-backed securities and complex credit derivatives. Theoretical research and practical application of quantitative credit risk models have been exploding in recent years. Duffie and Singleton (2003) and Lando (2004) provide comprehensive treatments of the theoretical and practical foundation of credit risk modeling. The famous Basel II Accord that allows the banks to develop their credit

risk models for PD (Probability of Default) and other measures also stimulates the quantitative research in the area of credit risk.

However, there is not an unanimous theoretical framework to model credit risk so far. There are a large number of credit risk models that have been developed. The credit risk models used extensively in the practice of credit risk management and asset pricing range from the relatively simple credit rating transition matrix to some complicated continuous-time models such as the Merton (or KMV) model. The examples of them include the approach to use external credit ratings by Moody's or S&P to estimate the PD(Probability of Default) (Lando and Skodeberg (2002), Nickell et al.(2000)), the credit scoring models that link the likelihood of defaults to a set of debtor characteristics (Altman (1968), Blume et al.(1998), Shumway (2001)), the structural credit risk models originated by Merton (1974) and extended by Black and Cox (1996), Leland and Toft (1996), Collin-Dufresne and Goldstein (2001) and so on.

The first concern facing the users of credit risk models is the validation of quantitative models, which includes the issues such as how well the models fit the real data, how accurately the models can predict the defaults, what are the pros and cons of the competing models, which model should be selected among a number of alternatives, etc.. Extensive empirical work has been conducted by academic and practitioners. Although there are too many to list, some important work are listed as follows: Lando and Skodeberg (2002), Nickell et al.(2000), Amato and Furfine (2004) studies how to incorporate systematic variables to improve the accuracy of the unconditional credit rating transition matrix; Beaver (1966), Blume et al.(1998)), Shumway (2001) compare the performance of the credit scoring models using various statistical techniques

and different covariates; Eom et. al. (2003) reveals that the Merton model is not able to generate credit spreads as high as those observed spreads, hence underestimates the PD, while most of the other structural models predict spreads that are too high on average.

This paper is about the empirical analysis of different credit risk models. Since there are too many credit risk models to study, we classify the credit risk models of interest into three groups: the rating migration models, the credit scoring models, and the asset value models. Although there are other classification of credit risk models available in the literature, our classification is based on the information that the models use to estimate the PD. The rating migration models in Lando and Skodeberg (2002) and Nickell et al.(2000) use the historical credit rating data to estimate the PDs of debtors with different ratings. The credit scoring models in Altman (1968), Blume et al.(1998) and Shumway (2001) supplement the credit rating data with the financial ratios or other debtor characteristics. The asset value models extended from Merton (1994) all utilize the information contained in the time series of the debtors' stock prices to estimate the PDs. Based on this classification, we can see that previous empirical research on credit risk models is generally limited in one class. For example, the empirical work on the credit scoring models focuses on the select of covariates or statistical models to improve the credit scoring system. The comparison of asset value models in Eom et. al. (2003) explores the effects of more complicated structures on the predictive power of the Merton model. We call this "vertical" analysis or comparison of credit risk models. The existing research rarely compare the credit risk models across the classes we define above. Thus, the first task of this paper is to study and compare the credit risk models "horizontally", i.e.,

the models in different classes.

Another feature distinguishing our work from others is that we follow the Bayesian approach to make inference of the PD. Although this is not the first attempt to introduce Bayesian inference in credit risk models and some researchers and practitioners have realized that Bayesian inference can be very useful in credit risk model validation, the literature on the Bayesian analysis in the area of credit risk is much more scanty than the literature on the classical approach. Most of the previous work only focus on assigning priors or adding parameter uncertainty in a specific credit risk model [1] . In this paper, we make full use of Bayesian inference for three different credit risk models. We derive the posterior distribution of the PD either by a closed-form density function or the Markov chain Monte Carlo (MCMC) sampling methods.

The three credit risk models we discuss in this paper have been extensively explored by using other statistical methods, for example, nonparametric estimation for the rating migration model, maximum likelihood estimation for the credit scoring model and calibration for the asset value model. We claim that, compared to other methods, the Bayesian inference equipped with MCMC methods has a number of advantages in the statistical analysis of credit risk models: First, the Bayesian inference can supplement the data analysis with the prior information. The incorporation of prior information is useful in the practice of credit risk analysis because risk analysts often need to use subjective judgement or personal opinion to adjust the results

[1] McNeil and Wendin (2007) apply a generalized linear mixed model to credit risk and estimate the model using a Gibbs sampler. Gossl (2005) jointly estimates the PD and default correlation for a credit risk portfolio in a Bayesian approach. Kiefer (2006) asserts how to incorporate expert opinion as priors for low-default portfolios with little historical default information. By using the Bayesian inference, Kadam and Lenk (2007) allow issuer heterogeneity in credit rating migration despite of data sparsity

from quantitative analysis. Secondly, Bayesian inference can easily produce the full distribution of the PD, instead of several moments or intervals estimated in the classical approach. Practitioners have realized that, since the distribution of the PD is often highly skewed, the information in the distribution will be much more useful than point estimates. Thirdly, many credit risk models are extremely complicated. MCMC methods have been shown to be suitable for dealing with highly nonlinear or nonnormal model structures, high dimensional parameter space including unobserved latent variables, which are common in credit risk models. Admittedly, MCMC methods are computation-intensive and the Bayesian estimation of the PD more time consuming than other methods. However, continuous advances in computing power mitigate these drawbacks.

Another advantage to use the Bayesian inference for credit risk models is that we can use the Bayes rule to incorporate the information contained in different data sets under different credit risk models for estimating a PD. In this paper we propose a "combined" Bayesian estimation method for the PD, which is actually a special case of using informative priors in the Bayesian inference. The basic idea is that we can use the PD estimate from one credit risk model as the prior in the second credit risk model given the data sets used in these two models are different. This new method to estimate the PD is motivated by the two severe issues in the practical implementation of quantitative models in credit risk management: Data insufficiency and model uncertainty. If we stick to one specific credit risk model in practice, it is often difficult or impossible to collect sufficient data to make reliable statistical inference. Furthermore, the existing credit risk models are highly simplified models and tend to over- or under-estimate the PD systematically. This combined estimation

can incorporate information for the PD from different sources, and tends to alleviate the data insufficiency problem. Secondly, the combined estimator of the PD is likely to moderate the PD estimators under different classes of credit risk models.

The paper is structured as follows: The second section introduce three credit risk models that are benchmark for the three classes: rating migration models, credit scoring models and asset value models. Bayesian inference for three models is conducted and, if it is necessary, MCMC algorithms are developed for specific models. Empirical analysis using the real data is following in the third section. The PD of the same obligor is estimated under three different models using different data sets. The posterior distributions of the PD under different models are compared. In the fourth section, we introduce the motivation and implementation of combining PD estimates from different models. Empirical results in the third section are used to illustrate the methodology. The last section concludes and discusses the possible extensions.

3.2 Credit Risk Models and Bayesian Inference

There are a large number of credit risk models available and the number is increasing every year. Following the classification method in introduction, we study three classes of credit risk models that utilize different information to estimate the PD of a specific debt obligor. In each class, we focus on a representative (or benchmark) model and conduct the Bayesian inference on these models. The posterior distributions of the PD in these models are derived either directly or by using the MCMC methods.

Before we proceed to a specific credit risk model, we define the default of the

obligor i, during a time period, as a binary variable D_i:

$$D_i = \begin{cases} 1 & \text{if obligor } i \text{ defaults} \\ 0 & \text{otherwise} \end{cases}$$

which is often assumed to follow a Bernoulli distribution

$$D_i \sim Bernoulli(PD_i)$$

where $PD_i = Prob\{D_i = 1\} = 1 - Prob\{D_i = 0\}$ is the default probability of obligor i.

3.2.1 A Rating Migration Model

In rating migration models, historical migrations of credit ratings are used to estimate the PDs. A comprehensive treatment of this class of credit risk models is referred to Lando and Skodeberg (2002).

Suppose that we are interested in estimating the PD of an obligor rated i (for example, AAA) at time t. N_i obligors are rated i a certain time period, say 1 year, ago. Then we count how many of them default (downgraded to rating D) by the end of the time period, say $N_{iD} = \sum_{l=1}^{N_i} D_{il}$. The average annual PD of the obligors with the rating i is computed by simply taking the ratio

$$\hat{PD}_i = \frac{N_{iD}}{N_i}. \tag{3.1}$$

This estimator is first a nonparametric estimator as the Kaplan-Meier estimator. Meanwhile, it can also be shown to be the maximum likelihood estimator of the PD given the Bernoulli distribution of defaults.

The average PDs for all ratings can be computed by the same way. Finally we can obtain a transition matrix with elements PD_{ij} representing the transition probability

Table 3.1: One-year Credit Rating Transition Matrix 05/2004-05/2005 (%)

	AAA	AA	A	BBB	BB	B	CCC	CC	Default	NA
AAA	90.00	10.00	0	0	0	0	0	0	0	0
AA	0	87.95	7.23	1.20	0	0	0	0	0	3.62
A	0	0	92.49	4.92	0	0	0	0	0	2.59
BBB	0	0.16	1.44	92.47	3.85	0.32	0	0	0	1.76
BB	0	0	0	4.85	87.44	4.41	0.22	0	0	3.08
B	0	0	0.50	0.50	8.73	72.57	3.24	0.50	1.50	12.47
CCC	0	0	0	0	1.75	14.04	56.14	0	19.30	8.77
CC	0	0	14.29	0	0	28.57	0	0	14.29	42.86

Note: NA represents the firms that do not have ratings at the end of the time peiord

from the rating i to the rating j over one year. The PDs are a part of the transition matrix. We use S&P historical credit rating data to compute the one-year credit rating transition matrix from 05/2004 to 05/2005, which is presented in Table 3.1. Since there is no exogenous variables involved with the estimation of this transition matrix, this matrix is also called the unconditional transition matrix in comparison with the conditional transition matrix sensitive to business cycles or other systematic factors. An important example for conditional credit rating transition matrix is Nickell et al.(2000), who report that the cyclicality and sensitivity of credit ratings to business cycles and estimate the credit rating transition matrix conditioned on some macroeconomic variables.

Bayesian inference for the PDs in the unconditional credit rating transition matrix is straightforward. We observe that there are N_i obligors with credit rating i at the

beginning of the observation period. At the end of the observation period, say 1 year, $N_{iD}(=\sum_{l=1}^{N_i} D_{il})$ of them default and the rest did not. Since we assume that the default of the obligors follows a $Bernoulli(PD_i)$ distribution, the joint distribution of observed defaults, or the likelihood, is proportional to

$$\begin{aligned}\prod_{l=1}^{N_i} p(D_{il}|PD_i) &= \prod_{l=1}^{N_i} PD_i^{D_{il}}(1-PD_i)^{1-D_{il}} \\ &= PD_i^{\sum_{l=1}^{N_i} D_{il}}(1-PD_i)^{N_i-\sum_{l=1}^{N_i} D_{il}} \\ &= PD_i^{N_{iD}}(1-PD_i)^{N_i-N_{iD}}\end{aligned}$$

If a noninformative prior is used, the posterior density function of PD_i is

$$p(PD_i|data) \sim PD_i^{N_{iD}}(1-PD_i)^{N_i-N_{iD}}, \tag{3.2}$$

which is a $Beta(N_{iD}+1, N_i-N_{iD}+1)$ distribution. The posterior mean of PD_i with a noninformative prior is $\frac{N_{iD}+1}{N_i+2}$, which is approximately equal to the maximum likelihood estimator when the sample sizes N_i and N_{iD} are large.

We can also use a conjugate prior, $Beta(a_0, b_0)$. Then the posterior density function of PD is

$$p(PD_i|data) \sim PD_i^{N_{iD}+a_0-1}(1-PD_i)^{N_i-N_{iD}+b_0-1}, \tag{3.3}$$

which is a $Beta(N_{iD}+a_0, N_i-N_{iD}+b_0)$ distribution. The sampling from a Beta distribution is straightforward and MCMC algorithms are not necessary.

The elicitation of PD's prior distribution has practical meanings to the practitioners in the industry of credit risk management. In practice, credit risk analyst often subjectively adjust the PD estimates from the quantitative credit risk models when they feel the estimates are obviously not reasonable. Actually the subjective

judgement or adjustment can be more scientifically incorporated in the statistical inference under the Bayesian approach because Bayesian inference is in essence a subjective decision theory. For Bayesians, expert opinions take the role of priors, which can be combined with the information in the data through the Bayes rule. Then personal judgement evolves to the elicitation of prior distribution of PD. PD is a quantity falling in an interval between 0 and 1. Therefore, a candidate for the prior distribution of PD can be a Beta distribution:

$$p(PD) \sim Beta(a_0, b_0)$$

with mean $a_0/(a_0+b_0)$ and variance $a_0b_0/(a_0+b_0)^2(a_0+b_0+1)$. Experts can express their opinion about PD by controlling the mode and dispersion of the prior Beta distribution.

3.2.2 A Credit Scoring Model

One of the limitations for rating-based models is the assumption that a group of heterogenous obligors have identical default probabilities if they have the same credit ratings. It is easy to argue that the obligors with the same ratings still differ in their credit qualities. Credit scoring models, initiated by Altman (1968), links the PDs of different obligors (even with the same ratings) to the obligor-specific information. In credit scoring models, the PD of an obligor depends on a set of obligor characteristics, for example, financial ratios representing their insolvency, liquidity and profitability.

Researchers and practitioners have the flexibility to select combination of covariates that "the best" explain the default behaviors of all obligors. A renowned example about the financial ratios used to predict the default of a firm is Altman's Z score that combines five financial ratios: working capital to total assets, retained earning to

total assets, EBIT to total assets, market value of equity to book value of total debt, and sales to total assets. We choose these five covariates in our "benchmark" credit scoring model. There are different possible ways to model the dependence between a binary variable and a set of covariates. Altman (1968) uses the linear discriminant analysis. S&P's CreditModel uses the neural networks and supporting vector machine. The most popular and easily interpreted model to link the credit score with the default is the probit (or logit) model.

In a probit credit scoring model, the default of a firm i still follows a Bernoulli distribution, while the PD depends on a linear combination of k explanatory variables $X = (x_1, x_2, \cdots, x_k)$, instead of being a constant in rating-based models, as

$$Prob\{D_i = 1\} = \Phi(X_i\beta), \tag{3.4}$$

where $\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_k)'$ and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

Albert and Chib (1993) propose a Gibbs sampler to estimate probit models using data augmentation. They re-write the probit model using some latent variables as

$$D_i^* = \beta X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,1) \tag{3.5}$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

where the latent variables D_i^* are continuous variables instead of binary variables as D_i. To obtain a maximum likelihood estimator of β, we have to integrate latent variables D_i^* out of the likelihood. However, D_i^* can be sampled as well as the parameters β if the data augmentation technique is used in the Gibbs sampler.

To implement the Gibbs sampler using data augmentation, we have to specify the full conditional posterior distributions of β and D_i^*. Given all latent variables

$D^* = (D_1^*, D_2^*, \cdots, D_N^*)'$ and data $D = (D_1, D_2, \cdots, D_N)'$, the conditional posterior distribution of β is straightforward from a linear regression model:

$$p(\beta|D^*, D) \sim N(\hat{\beta}, V), \tag{3.7}$$

where $\hat{\beta} = (X'X)^{-1}X'D^*$ and $V = (X'X)^{-1}$ are just the OLS estimates when the priors are noninformative.

It can be shown that D_i^* follows a truncated normal distribution given β and D_i:

$$\begin{aligned} D_i^*|D_i = 1, \beta &= D_i^*|D_i^* \leq 0, \beta \sim N(X_i\beta, 1)I[0, \infty) \\ D_i^*|D_i = 0, \beta &= D_i^*|D_i^* < 0, \beta \sim N(X_i\beta, 1)I(-\infty, 0). \end{aligned} \tag{3.8}$$

Sampling from a truncated normal distribution is not trivial. Different MCMC algorithms are available such as accept-reject and Metropolis-Hastings algorithms. An efficient algorithm using inverse CDF transformation was developed by Devroye(1986), which is based on a simple fact that the CDF of a continuous random variable has a uniform $[0, 1]$ distribution. Define $\varepsilon_i = D_i^* - X_i\beta$, which has an truncated $N(0, 1)$ distribution and its CDF must follow a $U[0, 1]$ distribution:

$$\begin{aligned} F(\varepsilon_i|D_i = 1, \beta) &= \frac{\Phi(\varepsilon_i) - \Phi(-X_i\beta)}{1 - \Phi(-X_i\beta)} \sim U[0, 1] \\ F(\varepsilon_i|D_i = 0, \beta) &= \frac{\Phi(\varepsilon_i)}{\Phi(-X_i\beta)} \sim U[0, 1]. \end{aligned} \tag{3.9}$$

Specifically, the Gibbs sampler using data augmentation and inverse CDF transformation is carried out in the following steps:

Step 1: Choose an initial value for $\beta^{(0)}$.

Step 2: Draw u_i from the uniform [0,1] distribution. Generate ε_i as

$$\varepsilon_i = \begin{cases} \Phi^{-1}(u_i(1 - \Phi(-X_i\beta)) + \Phi(-X_i\beta)) & \text{if} \quad D_i = 1 \\ \Phi^{-1}(u_i\Phi(-X_i\beta)) & \text{if} \quad D_i = 0 \end{cases}$$

and the latent variable $D_i^* = X_i\beta + \varepsilon_i$ for $i = 1, 2, \cdots, N$.

Step 3: Draw $\beta^{(1)}$ from

$$N((X'X)^{-1}X'D^*, (X'X)^{-1}).$$

Step 4: Repeat Step 2 through Step 3 for M iterations.

The Gibbs sampler will yield random draws from the posterior distribution of β.

Note that PD is not directly a parameter in the credit scoring model. Instead, PD_i is a function of the parameter β and the explanatory variable X_i as

$$PD_i = \Phi(X_i\beta).$$

The random draws of PD_i of our interest can be obtained by directly computing $\Phi(X_i\beta)$ using the MCMC draws from the joint posterior distribution of β and the values of X_i.

3.2.3 An Asset Value Model

Merton (1974) is the foundation for all structural credit risk models. We call the class of credit risk models based on Merton (1974) the asset value models because the default probabilities are implicit in the movements of the market values of the assets in these models.

Following Merton (1974), we assume that a firm, which is the debt issuer, has a simple capital structure: a single homogeneous class of debt and the residual claim (equity). At time t, the market values of its asset, liability(bond) and equity(stock) are denoted V_t, B_t and S_t. The following accounting equation must hold:

$$V_t = B_t + S_t. \tag{3.10}$$

At the meantime we assume that the market value of the assets follows a geometric Brownian motion as:

$$dV_t = \mu V_t dt + \sigma V_t dW_t, \tag{3.11}$$

where μV_t is the physical drift, σV_t is the physical diffusion and W_t is a standard Brownian motion under the physical probability space $(\Omega, \mathcal{F}, \mathcal{P})$. If the maturity of the debt is T, the asset value at T will be

$$V_T = V_t exp[(\mu - \frac{1}{2}\sigma^2)(T-t) + \sigma(W_T - W_t)]$$

which follows a lognormal distribution as:

$$log(V_T) - log(V_t)|\mathcal{F}_t \sim N((\mu - \frac{1}{2}\sigma^2)(T-t), \sigma(T-t))$$

The only liability of the firm is assumed to be a zero-coupon bond with face value F with a constant risk-free interest rate r.

Then the default motivation is modeled. Merton assumes that, acting in the best interests of stockholders at the maturity, the firm must either pay the promised payment F to the bondholders when the residual claim is positive, i.e. $V_T - F \geq 0$, or choose default on debt and liquid its asset to repay the bondholders when its asset cannot even cover the promised payment of debt, i.e. $V_T - F \geq 0$. Then at the maturity the value of the bond is

$$\begin{aligned} B_T &= min(F, V_T) \\ &= F - max(0, F - V_T). \end{aligned} \tag{3.12}$$

At the same time, the value of equity at the maturity is

$$\begin{aligned} S_T &= V_T - min(F, V_T) \\ &= max(V_T - F, 0), \end{aligned} \tag{3.13}$$

which exactly matches the expiration value of a European call option on a non-dividend common stock where V_t corresponds to the value of the underlying stock and F corresponds to the strike price. The Black-Scholes formula gives a solution for the market value of equity at time t as

$$\begin{aligned} S_t &= E_t^Q\big\{exp[-r(T-t)] \cdot max(V_t - F, 0)\big\} \\ &= V_t \cdot \Phi(d_1) - F \cdot exp[-r(T-t)] \cdot \Phi(d2) \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} d_1 &= \frac{log(V_t/F) + (r + 1/2\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned}$$

and $\Phi(\cdot)$ is the cumulative distribution function for a standard normal distribution.

The Merton model was originally developed for pricing a defaultable corporate bond. We can also use it to find the implicit $n-$year default probability of the debt issuer, which is the probability that the time $t+n$ asset value of the firm falls below the default threshold value:

$$\begin{aligned} PD_t &= Prob\{V_{t+n} < F | \mathcal{F}_t\} \\ &= \Phi(-\frac{ln(V_t) + (\mu - 1/2\sigma^2)(n) - lnF}{\sigma\sqrt{n}}). \end{aligned} \tag{3.15}$$

where the fraction in the bracket is called the DD (Distance to Default) in the Moody's KMV model, and the PD calculate using the Merton's model is called the EDF (Expected Default Frequency), i.e. $EDF = \Phi(-DD)$.

The assumption about the default behavior in the Merton model is criticized to be unrealistic. The first empirical work for the Merton model by Jones et. al. (1984)

finds that the Merton model is not able to generate credit spreads as high as those observed spreads, hence underestimates the PD. A large number of structural credit risk models extend the Merton model by allowing more complicated assumptions about the default threshold process: Black and Cox (1996) use an exponentially increasing function for the default process; Leland and Toft (1996) claim that the default threshold is determined exogenously by the optimal capital structure of the firm; Collin-Dufresne and Goldstein (2001) assume a continuous-time stochastic process for the default threshold. A comprehensive project by Eom et. al. (2003) reveals that most of the other structural models predict spreads that are too high on average. The empirical research for structural credit risk models has been historically limited because of the difficulties in statistical methods. Early literature and industry often rely on informal statistical analysis based on calibration techniques. Duan et. al. (2002) and Ericsson and Reneby (2002) use the maximum likelihood method (MLE) for structural credit risk models and argue that MLE has several advantages over the calibration method.

In this section, we conduct a Bayesian inference of the Merton model. To implement the Merton model and compute the PD, we need the asset value V_t and its drift μ and volatility σ as inputs. However, V_t are not observable in practice and have to be estimated with the parameters together. MCMC methods provide a unified framework to estimate unobserved latent variables with structural parameters together.

First we introduce measurement errors to the stock prices since the realized values could deviate from the theoretical values. Then the Merton model has a state-space

model form as

$$\begin{aligned} S_t &= C^{BS}(V_t, \sigma) + \varepsilon_t \\ dV_t &= \mu V_t dt + \sigma V_t dW_t, \end{aligned} \tag{3.16}$$

in which the equity values S_t are observed with errors $\varepsilon_t \sim N(0, \sigma_e^2)$, and $C^{BS}(V_t, \sigma)$ is the Black-Scholes formula in equation (3.14) with the underlying asset value V_t and volatility σ and other parameters including the risk-free interest rate, strike price and maturity are observed or defined.

This is a nonlinear state-space model. There are two approaches to apply a Gibbs sampler to draw unobserved latent variables in a state-space model: The single-move Gibbs sampler by Carlin, Polson and Stoffer (1992) and the multi-move Gibbs sampler by Carter and Kohn (1994). In this section we use the single-move Gibbs sampler with the Metropolis-Hastings algorithm in each Gibbs step to estimate this nonlinear state-space model.

Bayesian inference in the state space model is aimed at deriving the joint posterior distribution of parameters $\theta = (\mu, \sigma, \sigma_e)$ and latent variables $\mathbf{V} = (V_1, \cdots, V_n)$ given the data $\mathbf{S} = (S_1, \cdots, S_n$, which can be obtained by the Bayes rule as

$$\begin{aligned} p(\theta, \mathbf{V}|\mathbf{S}) &\propto p(\mathbf{S}|\mathbf{V}, \theta) p(\mathbf{V}, \theta) \\ &\propto p(\mathbf{S}|\mathbf{V}, \theta) p(\mathbf{V}|\theta) p(\theta) \end{aligned} \tag{3.17}$$

where $p(\mathbf{S}|\mathbf{V}, \theta)$ is determined by the measurement equation in (3.16), $p(\mathbf{V}|\theta)$ is determined by the state equation in (3.16), and $p(\theta)$ is the priors of the parameters. Generally there is no analytical expression for this posterior density if the state space model is nonlinear and nonnormal. MCMC algorithms including Gibbs sampler and Metropolis-Hastings algorithm can be used to draw random samples from this joint

posterior distribution. The details of MCMC algorithms are presented in the appendix.

With MCMC draws from the joint posterior distribution of all parameters and latent variables, the draws of one-year PD at time t can be obtained by computing

$$PD_t = \Phi(-\frac{log(V_t/F) + (\mu - 1/2\sigma^2)}{\sigma}), \tag{3.18}$$

where PD in the asset value model is a function of both model parameters and a latent variable. We can recall that PD is a parameter in the rating migration model and a function of a set of model parameters and covariates in the credit spread model.

3.3 Empirical Analysis

We implement the Bayesian inference on three models in the last section, empirically estimate the PDs for a specific obligor using real data, and finally compare the three posterior distributions of the same PD.

3.3.1 Data

Since three models use different data sets. We have a number of data sets to introduce in the following.

Credit ratings and financial statement information are from COMPUSTAT, which is a database of financial, statistical and market information covering publicly traded companies in the US and Canada over past 20 years. It provides annual or quarterly data from financial statements and supplemental data. It also contains the monthly long-term debt ratings since 1985. S&P Long Term Domestic Issuer Credit Rating range from AAA to D. A bond with a rating BBB or above is known as an investment

grade bond; one with a rating BB or lower is known as an speculation grade bond, or junk bond. In an attempt to refine these ratings further, S&P now on occasion assigns a + or − to its ratings to indicate that the bond is at the upper or lower end of the rating category. Here we ignore + or −, which means AAA+, AAA and AAA− are classified to the rating AAA. By doing this, we can try to avoid insufficient observations for some ratings. We choose the one-year time period between the end of May 2004 and the end of May 2005. During this period, 2032 have valid ratings from S&P among the 9854 firms.

For the credit scoring model, we adopt the five financial ratios used in the renowned Altman's Z score:

1. working capital/total assets
2. retained earning/total assets
3. EBIT/total assets
4. market value of equity/book value of total debt
5. sales/total assets

Some firms in the Compustat dataset do not have financial statement information we need. The sample size in the credit scoring model is reduced to 1972 after we drop the firms with no information or extraordinary observations. Among 1972 firms, 14 firms default during the year.

We choose a company, Honeywell International Inc.(NYSE symbol: HON), to conduct the empirical analysis of the asset value model. Daily stock prices of Honeywell are collected from CRSP. The daily equity values S_t are computed by multiplying

the quoted daily stock prices by the number of the common shares (preferred stocks are added if available). The number of common shares and preferred stock values are obtained from COMPUSTAT. We need to use stock prices to estimate important parameters such as the instantaneous asset value and its drift and volatility. Since we are interested in PD at the end of May 2005, we collect the daily stock prices from 01/01/2005 to 05/31/2005 and the number of outstanding common shares and total liability at the end of 2004. The risk free rate is set to be 3-month T-bill rate at 05/31/2005 of 2.8%. The time to maturity of liability is set to be 5 years [2] .

For the purpose of comparison, we collect the credit default swap (CDS) prices of Honeywell from Datastream. CDS is a financial contract in which one party makes periodic payment to exchange for the protection from the other party when the reference entity defaults. The price of the CDS, represented as a percentage of the notional amount, can be treated as the investors' expectation about the PD of the reference entity. Although the CDS market is not very liquid and the prices may be a distorted indicator of the true PD, we use the CDS price as a criterion to compare the posterior distributions of the PD of Honeywell from three different credit risk models. The CDS price is helpful to identify which model tends to overestimate or underestimate the PD.

3.3.2 Results

First we estimate the rating migration model. Table 3.2 presents how many obligors in each rating default during 06/2004-05/2005. The point estimates of PD are computed using the maximum likelihood method. Using noninformative (uniform [0,1]) prior

[2]It is the industry practice to set the maturity of the hybrid liability to be 5 years. We check different values for the maturity and find the estimation results are robust to the time to maturity

Table 3.2: Number of Defaults during 05/2004-05/2005

	AAA	AA	A	BBB	BB	B	CCC	CC
Number of Obligors (05/2004)	20	83	386	624	454	401	57	7
Defaults (05/2005)	0	0	0	0	0	6	11	1
Point estimate of PD (05/2005)	0	0	0	0	0	1.50%	19.30%	14.29%

distribution, we can also find the posterior distribution for any PD associated with a particular rating, which follows a Beta distribution. For example, the posterior distribution of the PD of an AAA-rated obligor is $Beta(1, 21) \propto (1 - PD)^{20}$, and the posterior distribution of the PD of a CCC-rated obligor is $Beta(12, 47) \propto PD^{11}(1 - PD)^{46}$. Honeywell was rated A by S&P during the year. So the posterior distribution of its PD follows a Beta(1,387) distribution, which is plotted in Figure 3.1.

Next, the credit scoring model using the Altman's Z score variables yields the estimates of five coefficients for the financial ratios. We draw 5,000 samples using the Gibbs sampler from their posterior distributions. The first 2,000 are "burned". All MCMC chains converged very quickly. The posterior means and standard deviations are presented in Table 3.3. It can be seen that four of five explanatory variables are significant, and three of them are significantly negative, which is reasonable since those variables should be low when default happens. Retained earning/total assets is not quite significant and the coefficient of Sales/total assets seems to contradict our expectation. With the MCMC draws from the posterior distribution of the parameters, we can easily obtain MCMC draws for the PD of any firm by computing the function $PD_i = \Phi(X_i\beta)$. These MCMC draws are from the posterior distributions of the PDs. The posterior distribution of PD of Honeywell using the credit scoring

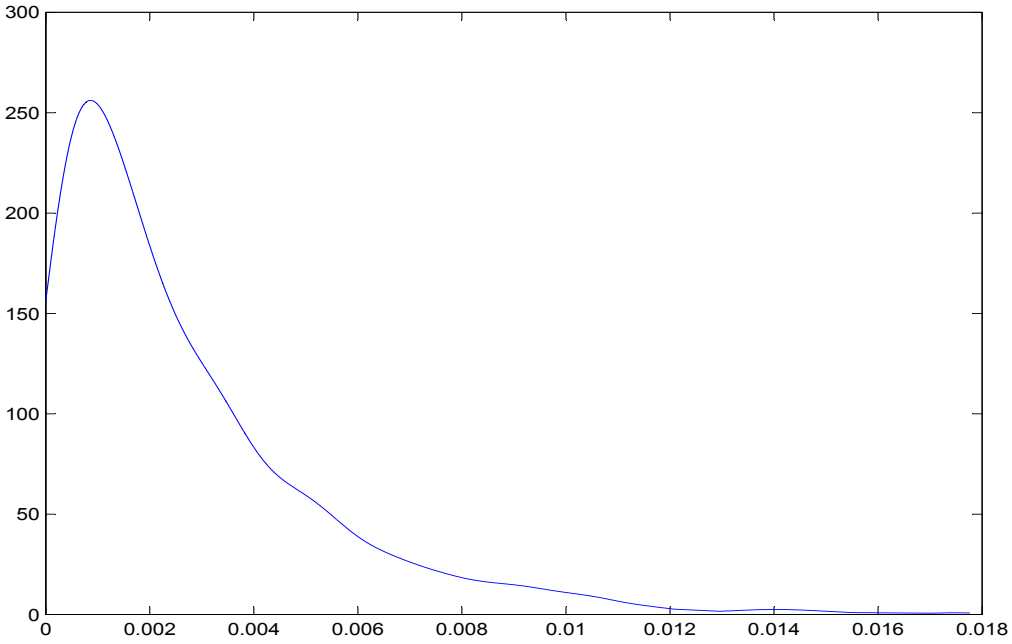

Figure 3.1: Posterior distribution of PD of Honeywell, Beta(1,387), using the rating migration model

Table 3.3: Estimates of The Credit Scoring Model

Coefficient	Posterior Mean	Std Deviation
constant	-2.1272	0.1167
β_1 : working capital/total assets	-0.0180	0.0045
β_2 : retained earning/total assets	0.0007	0.0017
β_3 : EBIT/total assets	-0.0503	0.0121
β_4 : market value of equity/book value of total debt	-0.0100	0.0047
β_5 : sales/total assets	0.0015	0.0009

model is displayed in Figure 3.2.

Last, we estimate the asset value model based on Merton (1974) using the daily stock prices of Honeywell in 2005. We still burn the first 2000 draws and keep 3000 draws for all parameters and latent variables. Again with the MCMC draws from the posterior distribution of the parameters and the latent variable, we can obtain the draws for the PD based on Equation . The posterior pdf of the PD of Honeywell is plotted in Figure 3.3.

Finally we display three posterior distributions of Honeywell's PD at the end of May 2005 in Figure 3.4. The vertical line in three figures represents the CDS price, 6.8 bp, of Honeywell on May 31 2009, which can be used here to evaluate how far three distributions are away from one another. From the figures, we can observe that the posterior density corresponding to the rating migration model has the highest peak, above 200, among the three, and the smallest variance. The posterior mode is near to the CDS price. The posterior density corresponding to the credit scoring model has a lower peak, below 100, and is less skewed. But its posterior mode or mean,

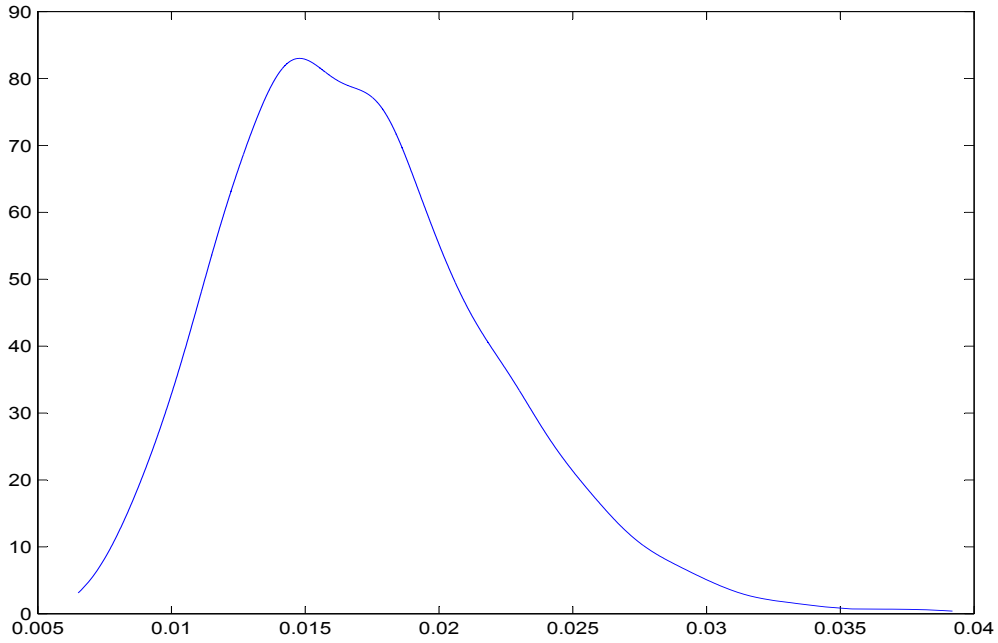

Figure 3.2: Posterior distribution of PD of Honeywell, using the credit scoring model

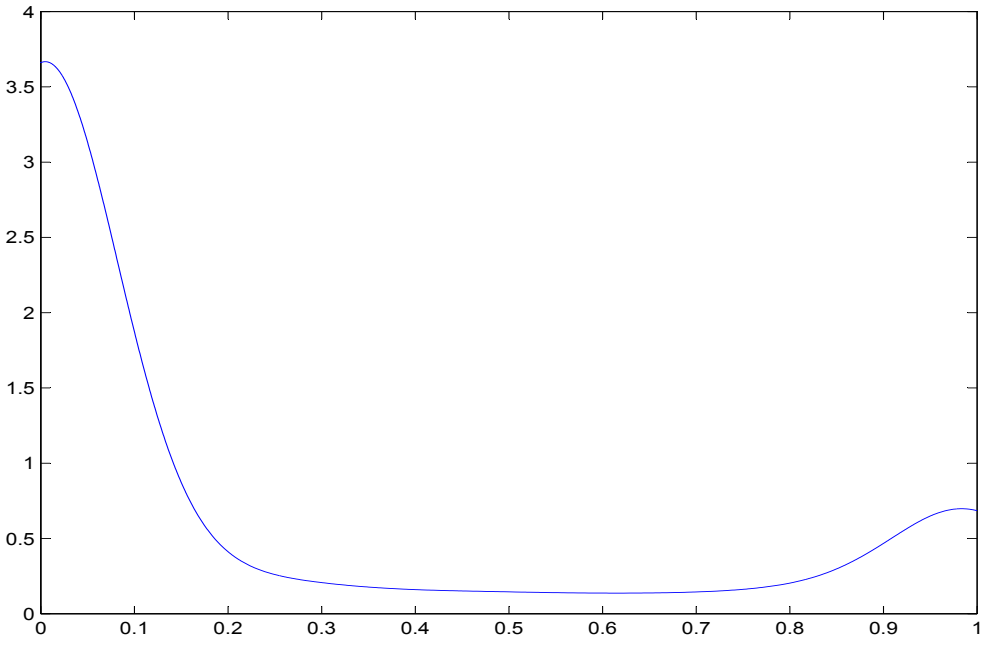

Figure 3.3: Posterior distribution of PD of Honeywell using the asset value model

even the majority of the distribution, is well above the CDS price. We conclude that the credit scoring model tends to overestimate the PD of Honeywell compared to the other models. The posterior density corresponding to the asset value model is very flat and the peak is below 4. As a result, it has a very fat tail and large variance. This posterior distribution seems to be much less informative about the PD than the other two. This should be due to a much smaller sample size and more complicated state-space model structure used by the asset value model. Cross-sectional data are used in both the rating migration and credit scoring model and the sample size is over one thousand, while the time series data used by the asset value model only contains around one hundred daily stock prices. Moreover, the asset value model tends to underestimate the PD since the posterior mode is the nearest to zero among the three. This result coincides the finding by Jones et. al. (1984) and Eoms et. al. (2003).

3.4 A Combined Bayesian Estimation of PD

The empirical results in the last section show that different credit risk models often yield largely different estimation results for the same PD. A natural way to deal with contradicted results between competing models is to select the "best" credit risk model, which fits the data in terms of in-sample fitting or out-of-sample prediction. However, the direct model comparison based on model fits is not feasible in the case of our "horizontal" comparison of credit risk models. For example, in a credit scoring model, the dependent variable is credit ratings, while stock prices or corporate bond prices is the predicted variable in an asset value model. It does not make any sense to compare the mean square errors of credit ratings and those of stock

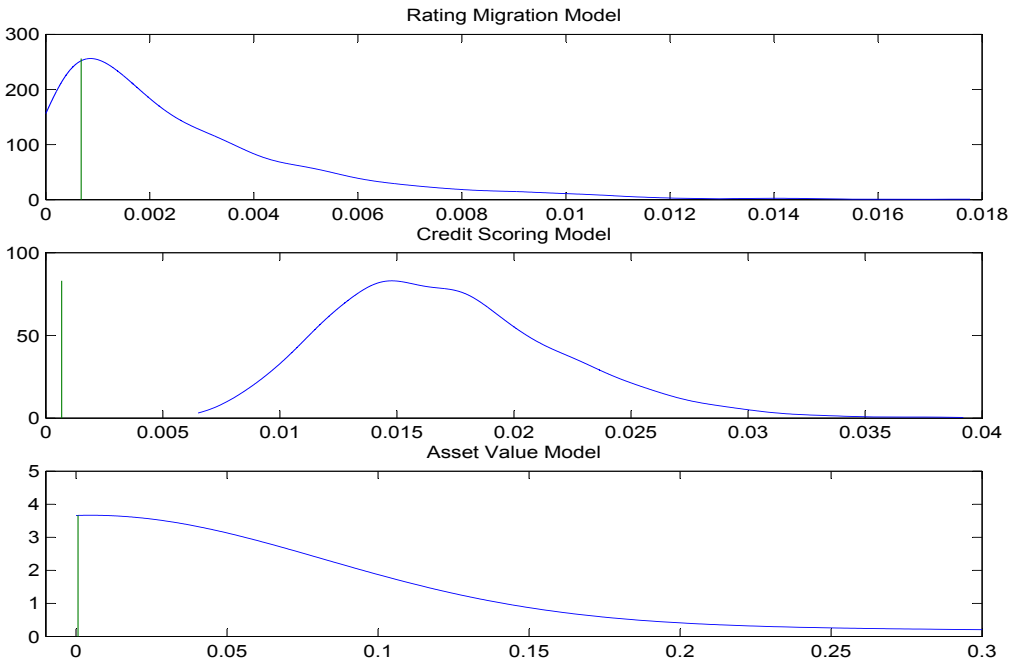

Figure 3.4: Comparison of three posterior distributions of Honeywell's PD using three different credit risk models

prices if we compare these two models using some criterion based on the mean square errors. More importantly, these two types of credit risk models describe the economic reasoning behind the defaults in the different ways. Thus, both of them could be correctly specified and it is not necessary to accept one model and reject the other. In that sense, a statistical hypothesis testing between these two models might not be appropriate.

The above issue can be referred as the model uncertainty problem in the practice of quantitative credit risk modeling. Practitioners in the banking industry often complaint that it is difficult to find a satisfactory quantitative model that can consistently explain historical defaults and make reliable prediction. This practical issue is even exacerbated by another problem called data insufficiency. Data insufficiency is referred to the cases in which it is difficult or impossible for the users of the credit risk models to find sufficient data to make reliable statistical inference on the models. First, default events are naturally infrequent, particularly for the obligors with low PDs. Therefore it is often not likely to observe enough historical defaults for the obligers with short credit history. A typical example is in the credit rating transition matrix in Table (3.1). For the obligors with rating above B, the PDs are estimated to be zero. These results obviously contradicting our commonsense should be due to the data insufficiency. Moreover, many credit risk models rely on financial information such as balance sheets and market prices of financial assets, which are frequently not available in the sense of sufficient observations.

Motivated by these two practical issues, we propose a combined estimation of PD using the Bayesian inference. The methodology of this combined estimation is simply based on the Bayes rule: We use the PD estimates from one model using

one data set as the prior information in the second model using a different data set. This combination will yield a posterior distribution of PD which incorporate different sources of information of multiple model structures. It is reasonable to believe this approach is useful to reduce the data insufficiency and model uncertainty problem.

Suppose that we are interested in combining the PD estimates from the credit scoring model M_1 and the asset value model M_2 that are studies previously. These two models are using different data sets to estimate the PD of the same obligor. M_1 uses a set of historical credit ratings, D_1, while M_2 uses D_2, a set of the time series of stock prices. Both models may suffer from the data sufficiency problem because some firms may not have a long history of externally assigned credit ratings or publicly listed stocks. Meanwhile two models are not necessarily competing models since they are explaining different economic relations around the same default behavior. A model selection to reject one of them is not appropriate. Let us consider to combine the PD estimates from these two models to yield a "combined" posterior distribution of PD. The empirical analysis in the last section outputs the posterior distributions, more precisely MCMC draws, of the PD of Honeywell from two models: $p_1(PD|D_1)$ and $p_2(PD|D_2)$. Through the Bayes rule, we can derive the "combined" posterior distribution of the PD as:

$$p_{12}(PD|D_1, D_2) \propto p_2(D_2|PD)p_1(PD|D_1), \tag{3.19}$$

assuming that D_1 and D_2 are conditionally independent given PD, where the subscripts 1 and 2 for the posterior distributions denote that the density functions depend on corresponding model structures. This result can be consecutively extended to incorporate more data sets and models as long as the conditional independence assumption can be satisfied. Note that the conditional independence assumption pre-

vents us from combining the rating migration model and the credit scoring model since their data sets are overlapped.

Generally we are not likely to find a closed-form posterior density function $p_{12}(PD|D_1, D_2)$. For the practical implementation, we need to resort to numerical solutions. Here we introduce a sampling techniques proposed by Rubin (1987), which is often called the Sampling/Importance Resampling (SIR). SIR is based on the importance sampling algorithm but involved with a resampling step. Let us illustrate this algorithm by using our example to sample from $p_{12}(PD|D_1, D_2)$ as:

$$p_{12}(PD|D_1, D_2) \propto p_2(D_2|PD)p_1(PD|D_1). \tag{3.20}$$

In the importance sampling algorithm, $p_{12}(PD|D_1, D_2)$ is called the target distribution, $p_1(PD|D_1)$ is called the proposal distribution, and $p_2(D_2|PD)$ is the importance function.

To obtain random draws from the target distribution $p_{12}(PD|D_1, D_2)$, SIR requires two conditions to be satisfied: First, drawing directly from the proposal distribution is feasible, which is satisfied in our example since we have MCMC draws from $p_2(D_2|PD)$. Second, the importance function can be numerically evaluated, which is also satisfied in our example. By using the Bayes formula, we can derive that

$$p_2(D_2|PD) \propto p_2(PD|D_2)/p_2(PD), \tag{3.21}$$

in which $p_2(PD|D_2)$ can be numerically evaluated by using any kernel density estimation technique since we have MCMC draws from this distribution.

Finally, SIR is implemented in the following steps:

Step 1: Draw PD^j for $j = 1, \cdots, m$ from the proposal distribution $p_1(PD|D_1)$.

Step 2: Evaluate the importance function $\tilde{w}^j = p_2(D_2|PD)$ at each PD^j, and compute weights as

$$w^j = \frac{\tilde{w}^j}{\sum_{i=1}^{m} \tilde{w}^i}$$

.

Step 3: Resample m draws with replacement from the multinomial distribution $\{PD^j\}_{j=1}^m$ with probabilities $\{w^j\}_{j=1}^m$.

The m draws we obtain are approximately from the posterior distribution $p_{12}(PD|D_1, D_2)$. Importance sampling requires that the proposal density function should be close to the target density. When this condition is not met, we need increase the size of independent draws, m, to reduce the approximation error. Theoretically the order of the combination (which one is the proposal distribution) does not affect the "combined" posterior distribution if only noninformative priors are used for the PD. However, the choice of the proposal density may have a large impact on the accuracy of numerical approximation. This effect is expected to fade when the number of MCMC draws is large enough.

Figure 3.5 shows how we implement the SIR algorithm and combine two posterior distributions of the PD of Honeywell from the credit scoring model and the asset value model. Since the posterior distribution from the asset value model is much flatter than that from the credit scoring model, we expect that the credit scoring model will dominate the combination. Therefore, the posterior distribution using the credit scoring model is more appropriate to be selected as the proposal distribution.

We can observe that the posterior distribution based on the credit scoring model almost assign the equal probabilities to the two sides of the center of the distribution

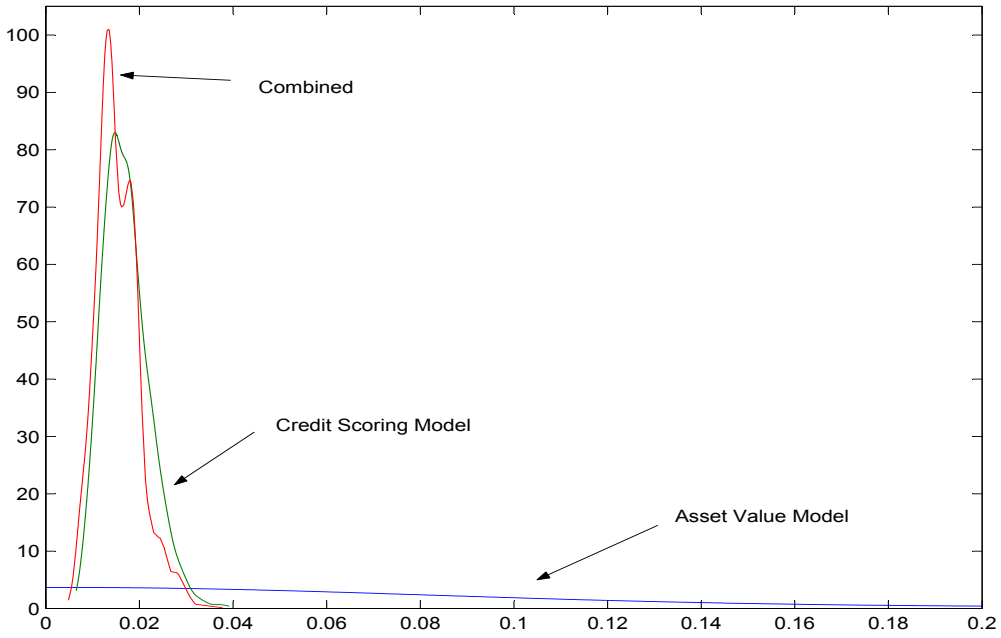

Figure 3.5: Combination of two posterior distributions of Honeywell's PD from the credit scoring model and the asset value model

since the distribution is roughly normally distributed. When the information of stock prices comes in the analysis, we are informed by the asset value model that the PD is more likely to be around zero than to be large. Then the values of the importance function, determined by the asset value model, adjust the proposal distribution and assigned higher weights to the left half side of the proposal density function than to the right half side. As a result, the left half side of the proposal density function shifts upwards and the right half side shifts downwards to reach the "combined" posterior distribution. It is also observed that the combination seems to have some difficulties in determining which model dominates in the area of the peak of the proposal density. It is expected that the mode of the "combined" distribution is between, actually a weighted average of, two "pre-combined" distributions. Moreover, the combination makes the posterior distribution more concentrated in terms of a smaller variance. This result is also reasonable because more information is incorporated. The estimation of the PD can benefit from this combination in that the decision maker is more confident about what range the PD should fall in.

3.5 Concluding Remarks

The research in the area of credit risk models is expected to be growing in the future because of the current credit crisis. However, it is probably a impossible task to find a credit risk models to perfectly explain the default behaviors. The competition between different credit risk models will exist for a long time. In this paper, we empirically explore and compare three classes of credit risk models, which are actively discussed by academics and used by practitioners. Our comparison is different from previous research in that the comparison is not limited to the same class of models using the

same information. We call it "horizontal" comparison because the competing models use different information and model comparison criteria for model fits or prediction are inappropriate. Our comparison is based on the posterior distributions of the same PD under three different credit risk models. Bayesian inference and MCMC methods have been shown to be powerful to in dealing with high-dimensional parameters, unobserved latent variables and highly complicated model structures in the credit scoring and asset value models. With the MCMC methods, we can easily derive the full posterior distributions of the PD instead of some point or interval estimates that can be obtained under the classical approach. Furthermore, we utilize the advantage of the Bayesian inference again, and propose a new estimation method for PD. This so-called combined estimation method pools the information from different sources for the estimation of the PD by using the PD estimates from one model as the prior information for another model. The combined estimation is motivated by the attempt to reduce two important practical issues in the implementation of credit risk models: data insufficiency and model uncertainty.

The "horizontal" comparison and combination between different classes of credit risk models can be extended to more complicated model structures in the future direction. For example, the recent frontier for the credit scoring models is the proportional hazard models that allow the time-varying PD. We can consider to compare and combine this kind of model with Collin-Dufresne and Goldstein (2001) or its extension. Other possible sources for estimating default probabilities include the market prices of corporate bonds and CDS (credit default swaps). The combination is possible to be extended to include these new information.

Appendix A

Proof of the Stochastic Taylor Formula

Consider the SDE characterizing the Ito process

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t). \tag{A.1}$$

Let f be a function of $X(t)$. Applying the Ito lemma, we can obtain

$$df(X(t)) = \{\mu(X(t))f'(X(t)) + \frac{1}{2}\sigma^2(X(t))f''(X(t))\}dt + \sigma(X(t))f'(X(t))dW(t). \tag{A.2}$$

Defining

$$\begin{aligned} L^0 &= \mu\frac{\partial}{\partial X} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial X^2} \\ L^1 &= \sigma\frac{\partial}{\partial X} \end{aligned}$$

then

$$df(X(t)) = L^0 f(X(t))dt + L^1 f(X(t))dW(t). \tag{A.3}$$

Let $f(x) = \mu(x)$, then

$$\mu(X(t)) = \mu(X(0)) + \int_0^t L^0\mu(X(s))ds + \int_0^t L^1\mu(X(t))dW(t). \tag{A.4}$$

Similarly, choose $f(x) = \sigma(x)$, then

$$\sigma(X(t)) = \sigma(X(0)) + \int_0^t L^0\sigma(X(s))ds + \int_0^t L^1\sigma(X(t))dW(t). \tag{A.5}$$

Substituting Equations (49) and (50) into (47) leads to

$$\begin{aligned}
f(X(t)) = f(X(0)) &+ \int_0^t \Big\{ \big[\mu(X(0)) + \int_0^{s_1} L^0\mu(X(s_2))ds_2 + \int_0^{s_1} L^1\mu(X(s_2))dW(s_2)\big] f'(X(s_1)) \\
&+ \frac{1}{2}\big[\sigma(X(0)) + \int_0^{s_1} L^0\sigma(X(s_2))ds_2 + \int_0^{s_1} L^1\sigma(X(s_2))dW(s_2)\big]^2 f''(X(s_1))\}ds_1 \\
&+ \int_0^t \{\sigma(X(0)) + \int_0^{s_1} L^0\sigma(X(s_2))ds_2 + \int_0^{s_1} L^1\sigma(X(s_2))dW(s_2)\Big\} f'(X(s_1))dW(s_1) \\
&= f(X(0)) + \big[\mu(X(0))f'(X(0)) + \frac{1}{2}\sigma^2(X(0))\big]\int_0^t ds + \sigma(X(0))f'(X(0))\int_0^t dW(s) \\
&+ \big[\sigma^2(X(0))f''(X(0)) + \sigma(X(0))\sigma'(X(0))f'(X(0))\big]\int_0^t\int_0^{s_1} dW(s_2)dW(s_1) + R_7
\end{aligned} \tag{A.6}$$

which is exactly the stochastic Taylor expansion formula in (4).

Furthermore, the above double Wiener integral is evaluated by the following derivation

$$\begin{aligned}
\int_0^t\int_0^{s_1} dW(s_2)dW(s_1) &= \int_0^t [W(s_1) - W(0)]dW(s_1) \\
&= \int_0^t W(s_1)dW(s_1) - W(0)\int_0^t dW(s_1) \\
&= \frac{1}{2}\int_0^t [dW^2(s_1) - ds_1] \\
&= \frac{1}{2}(W^2(t) - t)
\end{aligned} \tag{A.7}$$

since $dW^2 = 2WdW + dt$ by Ito's lemma.

Appendix B

MCMC Algorithms for the Asset Value Model based on Merton (1974)

Suppose that observed data consist of a sequence of stock prices of the obligor $\mathbf{S} = \{S_{i\Delta}\}_{i=1}^{n}$ within equal time interval Δ before the maturity date T. Before we proceed, let us reparameterize the model (3.16) as

$$S_t = C(v_t, \sigma) + \varepsilon_t, \tag{B.1}$$
$$v_t = v_{t-\Delta} + \mu^* \Delta + \sigma\sqrt{\Delta}\eta_t,$$

where $v_t = ln(V_t)$, $\mu^* = \mu - 1/2\sigma^2$, $\eta_t \sim N(0,1)$, and the function

$$C(v_t, \sigma) = exp(v_t)\Phi(d_1) - exp[-r(T-t)]\Phi(d_2).$$

The joint posterior density function of a set of latent variables $\mathbf{v} = \{v_{i\Delta}\}_{i=1}^{n}$ and

parameters $\mathbf{\Theta} = (\sigma_e, \mu^*, \sigma)$ can be derived using the Bayes rule as:

$$\begin{aligned} p(\mathbf{v}, \mathbf{\Theta}|\mathbf{S}) &\propto p(\mathbf{S}|\mathbf{v}, \mathbf{\Theta}) \cdot p(\mathbf{v}|\mathbf{\Theta}) \cdot p(\mathbf{\Theta}) \\ &\propto \prod_{i=1}^{n} p(S_{i\Delta}|v_{i\Delta}, \sigma, \sigma_e) \cdot \prod_{i=1}^{n} p(v_{i\Delta}|v_{(i-1)\Delta}, \mu, \sigma) \cdot p(\mathbf{\Theta}) \\ &\propto \prod_{i=1}^{n} \frac{1}{\sigma_e} exp\big[- \frac{(S_{i\Delta} - C(v_{i\Delta}, \sigma))^2}{2\sigma_e^2} \big] \cdot \prod_{i=1}^{n} \frac{1}{\sigma} exp\big[- \frac{(v_{i\Delta} - v_{(i-1)\Delta} - \mu\Delta)^2}{2\sigma^2\Delta} \big] \cdot p(\mathbf{\Theta}). \end{aligned} \tag{B.2}$$

We use a Gibbs sampler to draw samples from the marginal posterior distribution of each hyperparameter. To implement the Gibbs sampler, we need to obtain fully conditional posterior density functions for all hyperparameters. For simplicity, we assume independent noninformative priors [1] as

$$p(\mu^*) \sim c, p(\sigma^2) \sim \sigma^{-2}, p(\sigma_e^2) \sim \sigma_e^{-2}.$$

Their fully conditional posterior density functions are derived as

- The conditional posterior distribution of μ^* is normal:

$$\begin{aligned} p(\mu^*|\cdot) &\propto p(\mathbf{v}|\mu^*, \sigma) \cdot p(\mu) \\ &\propto \prod_{i=1}^{n} exp\big[- \frac{(v_{i\Delta} - v_{(i-1)\Delta} - \mu\Delta)^2}{2\sigma^2\Delta} \big] \\ &\sim N(\frac{1}{n\Delta} \sum_{i=1}^{n} (v_{i\Delta} - v_{(i-1)\Delta}), \frac{\sigma^2}{n}), \end{aligned} \tag{B.3}$$

  and the conditional posterior distribution of σ_e^2 is inverted Gamma:

$$\begin{aligned} p(\sigma_e^2|\cdot) &\propto p(\mathbf{S}|\sigma_e, \sigma, \mathbf{v}) \cdot p(\sigma_e^2) \\ &\propto \prod_{i=1}^{n} \frac{1}{\sigma_e} exp\{ - \frac{[S_{i\Delta} - C(v_{i\Delta}, \sigma)]^2}{2\sigma_e^2} \} \cdot \frac{1}{\sigma_e^2} \\ &\sim IG(\frac{n}{2}, \frac{1}{2} \sum_{i=1}^{n} [S_{i\Delta} - C(v_{i\Delta}, \sigma)]^2). \end{aligned} \tag{B.4}$$

[1] we can also use conjugate priors $\mu^* \sim N, \sigma^2 \sim IG, \sigma_e^2 \sim IG$

- The conditional posterior distribution of σ^2 is nonstandard:

$$p(\sigma^2|\dot{)} \propto p(\mathbf{S}|\mathbf{v},\sigma)\cdot p(\mathbf{v}|\sigma)\cdot p(\sigma^2) \tag{B.5}$$

$$\propto \prod_{i=1}^{n} exp\{-\frac{[S_{i\Delta}-C(v_{i\Delta},\sigma)]^2}{2\sigma_e^2}\}\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{\Delta}}exp\{-\frac{[v_{i\Delta}-v_{(i-1)\Delta}-\mu\Delta]^2}{2\sigma^2\Delta}\}\frac{1}{\sigma^2} \tag{B.6}$$

where the second part of the posterior kernel is an inverted Gamma, but the first part is not recognized as the kernel of any standard distribution. I propose a Metropolis-Hasting algorithm to draw from this nonstandard distribution. Define $\phi=\sigma^2$, then the posterior density function of ϕ is

$$p(\phi|\cdot) \propto \prod_{i=1}^{n} exp\{-\frac{[S_{i\Delta}-C(\phi)]^2}{2\sigma_e^2}\}\prod_{i=1}^{n}\phi^{1/2}\cdot exp\{-\frac{[v_{i\Delta}-v_{(i-1)\Delta}-\mu\Delta]^2}{2\Delta}\phi\}\cdot\phi, \tag{B.7}$$

where

$$C(\phi)=exp(v_t)\Phi(d_1)-exp[-r(T-t)]\Phi(d_2)$$

is a nonlinear function of ϕ, since

$$d_1=\frac{[v_{i\Delta}+(r+\frac{1}{2\phi})(T-i\Delta)]\sqrt{\phi}}{\sqrt{T-i\Delta)}}$$

and

$$d_2=\frac{[v_{i\Delta}+(r-\frac{1}{2\phi})(T-i\Delta)]\sqrt{\phi}}{\sqrt{T-i\Delta)}}.$$

We approximate the nonlinear function using a Taylor expansion around the last draw, say $\phi^{(j-1)}$, as

$$S_{i\Delta}-C(\phi)\approx S_{i\Delta}-C(\phi^{j-1})-(\phi-\phi^{(j-1)})C_\phi(\phi^{(j-1)}), \tag{B.8}$$

where $C_\phi(\phi^{(j-1)})$ is the partial derivative of C with respect to ϕ evaluated at $\phi^{(j-1)}$. Then the first part of the posterior density function in (B.7) can be

approximated using a normal kernel for ϕ with mean m and variance V where

$$\begin{aligned} m &= \phi^{(j-1)} + V \cdot \frac{\sum_{i=1}^{n}\{[S_{i\Delta} - C(\phi^{(j-1)})]C_{\phi}(\phi^{(j-1)})\}}{\sigma_e^2} \\ V &= \frac{\sigma_e^2}{\sum_{i=1}^{n} C_{\phi}^2(\phi^{(j-1)})}. \end{aligned} \tag{B.9}$$

Furthermore, we approximate this normal distribution using a Gamma distribution with the same mean and variance as

$$N(m, V) \approx Gamma(\frac{m^2}{V}, \frac{m}{V}). \tag{B.10}$$

Since the product of two Gamma kernels is still a Gamma kernel, we finally derive a Gamma proposal distribution for the Metropolis-Hasting algorithm:

$$q(\phi|\phi^{(j-1)}) \sim Gamma(\frac{m^2}{V} + \frac{n}{2}, \frac{m}{V} + \frac{1}{2\Delta}\sum_{i=1}^{n}[v_{i\Delta} - v_{(i-1)\Delta} - \mu\Delta]^2). \tag{B.11}$$

- The conditional posterior distribution of latent variable $v_{i\Delta}$ $i = (1, 2, \cdots, n)$ is also nonstandard:

$$p(v_{i\Delta}|\cdot) \propto p(S_{i\Delta}|v_{i\Delta}, \Theta) \cdot p(v_{(i+1)\Delta}|v_{i\Delta}, \Theta) \cdot p(v_{(i)\Delta}|v_{(i-1)\Delta}, \Theta) \tag{B.12}$$

$$\begin{aligned} &\propto exp\{-\frac{[S_{i\Delta} - C(v_{i\Delta})]^2}{2\sigma_e^2}\} \cdot exp\{-\frac{[v_{(i+1)\Delta} - v_{i\Delta} - \mu\Delta]^2}{2\sigma^2\Delta}\} \\ &\quad \cdot exp\{-\frac{[v_{i\Delta} - v_{(i-1)\Delta} - \mu\Delta]^2}{2\sigma^2\Delta}\}. \end{aligned} \tag{B.13}$$

The last two components of the product are normal kernels for $v_{i\Delta}$, but the first component is nonstandard. Again, we use the Taylor expansion around the last draw $v_{i\Delta}^{(j-1)}$ to approximate the first component as

$$S_{i\Delta} - C(v_{i\Delta}) \approx S_{i\Delta} - C(v_{i\Delta}^{(j-1)}) - (v_{i\Delta} - v_{i\Delta}^{(j-1)})C_v(v_{i\Delta}^{(j-1)}), \tag{B.14}$$

where $C_v(v_{i\Delta}^{(j-1)})$ is the partial derivative of C with respect to $v_{i\Delta}$ evaluated at $v_{i\Delta}^{(j-1)}$. Then we derive a normal proposal distribution for the Metropolis-Hasting algorithm:

$$q(v_{i\Delta}|v_{i\Delta}^{(j-1)}) \sim N(m_v, V_v), \tag{B.15}$$

where

$$\begin{aligned} m_v &= v_{i\Delta}^{(j-1)} + V_v\big[\frac{[S_{i\Delta} - v_{i\Delta}^{(j-1)}]C_v(v_{i\Delta}^{(j-1)})}{\sigma_e^2} + \frac{v_{(i+1)\Delta} + v_{(i-1)\Delta}}{\sigma^2\Delta}\big] \\ V_v^{-1} &= \frac{C_v^2(v_{i\Delta}^{(j-1)})}{\sigma_e^2} + \frac{2}{\sigma^2\Delta}. \end{aligned}$$

Note that $v_{n\Delta}$ and v_0 have slightly different posterior densities and proposal distributions.

The MCMC algorithms are implemented in the following steps:

Step 1 Set the initial values $\mu^{*(0)}, \sigma^{(0)}$, $\sigma_e^{(0)}$ and $\mathbf{v}^{(0)}$;

Step 2 At the j-th iteration, draw $\mu^{*(j)}$ from a normal distribution in (B.3);

Step 3 Draw $\sigma_e^{2(j)}$ from an inverted Gamma distribution in (B.4);

Step 4 Draw ϕ' from a proposal Gamma distribution in (B.11) and accept this draw and let $\phi^{(j)} = \phi'$ with a probability

$$\rho = min\left\{1, \frac{p(\phi'|\cdot)}{p(\phi^{(j-1)}|\cdot)} \cdot \frac{q(\phi^{(j-1)}|\phi')}{q(\phi'|\phi^{(j-1)})}\right\}$$

, and recover the draw of σ by letting $\sigma^{(j)} = 1/\sqrt{\phi^{(j)}}$;

Step 5 Draw $\{v_{i\Delta}^{(j)}\}_{i=1}^n$ iteratively under the single-move approach. At each move, draw $v'_{i\Delta}$ from a normal proposal distribution in (B.15) and accept it with a probability

$$\rho = min\left\{1, \frac{p(v'_{i\Delta}|\cdot)}{p(v_{i\Delta}^{(j-1)}|\cdot)} \cdot \frac{q(v_{i\Delta}^{(j-1)}|v'_{i\Delta})}{q(v'_{i\Delta}|v_{i\Delta}^{(j-1)})}\right\};$$

Step 6 Repeat Step 2 through Step 5 for M iterations.

Bibliography

[1] Albert, J. and S. Chib (1993), Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, 669C679.

[2] Aldrich, J. (2002), How Likelihood and Identification went Bayesian, *International Statistical Review*, 70, 1, 79-98.

[3] Altman, E. I. (1968), Financial Ratios, Discrimination Analysis and the Prediction of Corporate Bankrutcy, *Journal of Finance* 23, 589-609.

[4] Amato, J. D. and C. H. Furfine (2004), Are Credit Ratings Procyclical? *Journal of Banking and Finance* 28, 2641-2677.

[5] Basel Committee on Banking Supervision (2004), *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*, Bank for International Settlements.

[6] Black, F., and J. Cox (1976), Valuing Corporate Securities: Some Effects of Bond Indenture Provisions, *Journal of Finance*, 31, 351-367.

[7] Black, F. and M. Scholes (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637C659.

[8] Blume, M. E., F. Lim and A. C. MacKinlay (1998), The Declining Credit Quality of US Corporate Debt: Myth or Reality? *Journal of Finance* 53, 1389C1413.

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

[10] Carter, C. K. and R. Kohn (1994), On Gibbs Sampling for State Space Models, *Biometrika*, 81, 541-553

[11] Chan, K., A. Karolyi, F, Longstaff and A. Sanders (1992). An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance*, 47, 1209-1227

[12] Chen, R., R. and L. Scott (1993). Multi-factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman filter model, working paper, Rutgers University

[13] Collin-Dufresne, P. and R. Goldstein (2001), Do Credit Spreads Reflect Stationary Leverage Ratios, *Journal of Finance* 56, 1929-1957.

[14] Cox, J. C., J. E. Ingersoll and S. A. Ross (1985), A Theory of the Term Structure of Interest Rates, *Econometrica*,53,385-407.

[15] Dai, Q., and K. S. Singleton (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance*, 1943-1978.

[16] Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer, New York.

[17] Dreze, J. (1972), Econometrics and Decision Theory, *Econometrica*, 40, 1-17.

[18] Duan, J., G. Gauthier, J. Simonato and S. Zaanoun (2002), Maximum Likelihood Estimation of Structural Credit Spread Models - Deterministic and Stochastic Interest Rates, University of Toronto, working paper.

[19] Duffee, G. (1999), Estimating the Price of Default Risk, *Review of Financial Studies*, 12, 197-226.

[20] Duffie, D. and K. Singleton (1999), Modeling Term Structures of Defautable Bonds, *Review of Financial Studies*, 12, 687-720.

[21] Duffie, D. and K. Singleton (2003), *Credit Risk Pricing, Measurement and Management*, Princeton University Press, Princeton.

[22] Duffie, D. and R. Kan , 1996, A Yield-Factor Model of Interest Rates, Mathematical Finance, 6, 379-406.

[23] Duffie, Darrel, Jun Pan, and Kenneth J. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, Econometrica 68, 1343C1376.

[24] Dyrting, S. (2004), Evaluating the Noncentral Chi-Square Distribution for the Cox-Ingersoll-Ross Process, *Computational Economics*,24,35-50.

[25] Elerian, O., S. Chib and N. Shephard (2001), Likelihood Inference for Discretely Observed Nonlinear Diffusions, *Econometrica*, 69, 959-993.

[26] Eom, Y. H., J. Helwege, and J. Huang (2002), Structural Models of Corporate Bond Pricing: An Empirical Analysis, *Review of Financial Studies*17, 499-544.

[27] Eraker, B. (2001), MCMC Analysis of Diffusion Models with Application to Finance, *Journal of Business and Economic Statistics*, 19, 177-191.

[28] Eraker, Bjorn, Michael S. Johannes, and Nicholas G. Polson, 2003, The Impact of Jumps in Returns and Volatility, *Journal of Finance*, 53, 1269C1300.

[29] Eraker, Bjorn, 2004, Do Stock Prices and Volatility Jumps Reconciling Evidence from Spot and Option Prices, *Journal of Finance*, 59, 1367-1403.

[30] Ericsson, J. and J. Reneby (2002), Estimating Structural Bond Pricing Models, McGill University working paper.

[31] Feller, W. (1951), Two Singular Diffusion Problems, *Annals of Mathematics*, 54(1), 173-182

[32] Florens-Zmirou, D. (1989), Approximate Discrete-Time Schemes for Statistics of Diffusion Process, *Statistics*, 20, 547-557

[33] Fruhwirth-Schnatter, S. and A. LJ. Geyer (1996), Bayesian Estimation of Econmetric Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates Via MCMC Methods, working paper, Vienna University.

[34] Gelfand, A. E. and S. K. Sahu (1999), Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models, *Journal of the American Statistical Association*, 94, 247-253.

[35] Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering*, Springer, New York.

[36] Gossl, C. (2005), Predictions Based on Certain Uncertainties - A Bayesian Credit Portfolio Approach, working paper, HypoVereinsbank AG.

[37] Hamilton, J. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

[38] Hannan, E. J. (1971), The Identification and Parameterization of ARMAX and State Space Forms, *Econometrica*, 44, 713-723.

[39] Hsiao, C. (1983), Identification, In *Handbook of Econometrics*, Vol1, 224-283.

[40] Huang J. and M. Huang (2003), How Much of the Corporate-Treasury Yield Spread is Due to Credit Risk? A New Calibration Approach, working paper, Stanford University.

[41] Hull, J. and A. White (2000), Valuing Credit Default Swaps I: No Counterparty Default Risk, *Journal of Derivatives*, 8, 29-40.

[42] Jarrow, R., D. Lando, and S. Turnbull (1997), A Markov Model for the Term Structure of Credit Spreads, *Review of Financial Studies*, 10, 481-523.

[43] Jarrow, R. and S. Turnbull (1995), Pricing Options on Financial Securities Subject to Default Risk, *Journal of Finance*, 50, 53-86.

[44] Johannes, M and N.G. Polson (2004), MCMC Methods for Financial Econometrics. *Handbook of Financial Econometrics*, Y. Ait-Sahalia and L. Hansen eds.

[45] Johnson, N. and S. Kotz (1970), *Distributions in Statistics: Continuous Univariate Distribution*, Vol. 2, Wiley, New York

[46] Jones, E., Mason, S. and E. Rosenfeld (1984), Contingent Claims Analysis of Corporate Capital Structures: An Empirical Investigation, *Journal of Finance*, 39, 611-627.

[47] Kadam, A. and P. Lenk (2007), Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration, working paper, Cass Business School, London.

[48] Kadane, J. B. (1975), The Role of Identification in Bayesian Theory,In *Studies in Bayesian Econometrics and Statistics*, 175-191.

[49] Kiefer, N. M. (2006), Default Estimation for Low-Default Portfolios, working paper, Departement of Economics and Statistical Science, Cornell University.

[50] Kim, J., Ramaswamy, K. and S. Sundaresan (1993), Do Default Risk in Coupons Affect the Valuation of Corporate Bonds, *Financial Management*, 22, 117-131.

[51] Kloeden, P. E. and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Springer, New York.

[52] Koopmans, T. C. (1949), Identification Problems in Economic Model Construction, *Econometrica*, 17, 125-144.

[53] Koopmans, T. C. and O. Reiersol (1950), The Identification of Structural Characteristics, *Annals of Mathematical Statistics*, 21, 165-181.

[54] Lamoureux, C. G. and H. D. Witte (2002), Empirical Analysis of the Yield Curve: The Information in the Data Viewed through the Window of Cox, Ingersoll and Ross, *Journal of Finance*, 57, 1479-1520

[55] Lando, D. (1998), On Cox Processes and Credit Risk Securities, *Reviews of Derivatives Research*, 2, 99-120.

[56] Lando, D. (2002), *Credit Risk Modeling*, Princeton University Press, Princeton.

[57] Lando, D. (2004), *Credit Risk Modeling*, Princeton University Press, Princeton.

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

[59] Leland, H., and K. Toft (1996), Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads, *Journal of Finance*, 51, 987-1019.

[60] Lindley, D.V. (1971), *Bayesian Statistics: A Review*, SIAM, Philadelphia.

[61] Lo, A. W. (1988), Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data, *Econometric Theory*, 4, 237-241.

[62] Longstaff, F. A. and E. S. Schwartz (1995), A Simple Approach to Valuing Risky Fixed and Floating Rate Debt, *Journal of Finance*, 50, 789-820.

[63] McNeil, A. J. and J. P. Wendin (2007), Bayesian Inference for Generalized Linear Mixed Models of Portfolio Credit Risk, *Journal of Empirical Finance*, 14, 131-149.

[64] Melino, A. (1994), Estimation of Continous-Time Models in Finance, *Advances in Econometrics, Sixth World Congress, Volume II*, Cambridge University Press, Cambridge, 313-351.

[65] Merton, R. (1974), On the Pricing of Corporate Debt: the Risk Structure of Interest Rates, *Journal of Finance*, 29, 449-470.

[66] Morales, J. A. (1971), *Bayesian Full Information Structural Analysis*, Springer-Verlag, Berlin.

[67] Norman, K. B. (1997), Gaussian Estimation of Single-Factor Continuous Time Models of The Term Structure of Interest Rates, *Journal of Finance*, 52, 1695-1706

[68] Pearson, N.D. and T. Sun, 1994, Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model, *Journal of Finance*, 49, 1279-1304.

[69] Pluto, K. and D. Tasche (2005), Estimating Probabilities of Default for Low Default Portfolios, Deutsche Bundesbank.

[70] Raiffa, H. A. and R. Schlaifa (1961), *Applied Statistical Decision Theory*, Harvard University Press, Boston.

[71] Rannala, B. (2002), Identifiability of Parameters in Bayesian MCMC Inference of Phylogeny, *Systematic Biology*, 51(5), 754-760.

[72] Robert, C. and G. Casella (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

[73] Rothenberg, T. J. (1971), Identification in Parameteric Models, *Econometrica*, 39, 577-591.

[74] Rothenberg, T. J. (1973), *Efficient Estimation with a Prior Information*, Yale University Press, New Haven.

[75] Rubin, D. B. (1988), Using the SIR Algorithm to Simulate Posterior Distributions In *Bayesian Statistics*, edits J. M. Bernardo et al, Clarendon Press, Oxford. p395-420.

[76] Sims, C. A. and T. A. Zha (1998), Bayesian Methods for Dynamic Multivariate Models, International Economic Review, 39(4), 949-968.

[77] Shumway, T. (2001), Forecast Bankruptcy More Accurately: A Simpel Hazard Model, *Journal of Business*, 74, 101-124.

[78] UK Financial Services Authority (2007), Expert Group paper on Low Default Portfolios.

[79] Vasicek, O. A. (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188

[80] Walter, E. (1982), *Identifiability of State Space Models*, Springer-Verlag, New York.

[81] Yoshida, N. (1992), Estimation for Diffusion Processes from Discrete Observations, *Journal of Multivariate Analysis*, 41, 220-242.

[82] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John-Wiley, New York.

[83] Zhou, C. (2001), The Term Structure of Credit Spreads with Jump Risk, *Journal of Banking and Finance* 25, 2015-2040.

Curriculum Vita

Xianghua Liu

2003-2009	Ph.D. in Economics, Rutgers University
2001-2002	M.B.A., University of Illinois at Chicago
1995-1999	B.A. in Economics, Wuhan University

2009 Xianghua Liu, Liuling Li and Hiroki Tsurumi, "Bayesian Inference of Financial Models Using MCMC Algorithms", *Handbook of Quantitative Finance and Risk Management*, Edited by C.F. Lee and A.C. Lee.