

MODELS IN FINANCE AND MEDICINE USING BAYESIAN INFERENCE

by

YOKO IBUKA

A Dissertation submitted to the  
Graduate School-New Brunswick  
Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Economics

written under the direction of

Professor Hiroki Tsurumi

and

Professor Louise B. Russell

and approved by

Hiroki Tsurumi

Louise B. Russell

John Landon-Lane

Edwin J. Green

New Brunswick, New Jersey

May, 2008

## ABSTRACT OF THE DISSERTATION

Models in Finance and Medicine Using Bayesian Inference

By YOKO IBUKA

Dissertation Directors:

Hiroki Tsurumi

and

Louise B. Russell

The purpose of this dissertation is to analyze three models in medicine and finance using Bayesian inference with the Markov chain Monte Carlo method. The model in medicine addresses cost-effectiveness analysis using *copulas*, and the two models in finance include discrete-time asset pricing models and a short-term interest rate model with stochastic volatility.

The first chapter develops the model that allows dependence between cost and effectiveness using copulas in cost-effectiveness analysis. The model was applied with sample of adults from the NHANES I Epidemiologic Follow-up Study, assuming a log-normal distribution for cost and a Weibull distribution for effectiveness as the marginals. Cost-effectiveness analysis is conducted for two types of patients using the estimated posterior densities of parameters regarding the hypothetical intervention for hypertension. A simulation based on Bayesian predictive densities is also performed to analyze cost and

effectiveness at an individual patient level. The empirical result indicated a negative dependence between measures of effectiveness and cost.

The second chapter conducts a Bayesian analysis of discrete-time asset pricing model. The chapter particularly discusses the *naive discretization* problem, which arises from using discrete-time data to estimate continuous-time models. Our results using generated data showed that the naive discretization would not work well when data generating process is unknown, when the data is sampled at low frequency, and averaged data is used.

The final chapter develops a Bayesian analysis of a short-term interest rate model with stochastic volatility. The model was developed based on the CKLS model (Chan et al. 1992). We constructed MCMC algorithms suitable for the model based on the Jacquire, Polson and Rossi(1994) algorithm. The empirical results with the 3-month Treasury constant maturity rate suggested that there was high autocorrelation in volatility of the error terms. Finally, the developed model was compared with the model with a GARCH error, using Bayesian predictive densities. The predictive densities obtained by CKLS with stochastic volatility have wider variance than the ones from CKLS-GARCH, and the realized value did not fall in the support of the predicted values for the CKLS-GARCH model because of the tight variance in prediction.

## ACKNOWLEDGEMENT

I would like to acknowledge and extend my heartfelt gratitude to the following persons and institution that have made the completion of my dissertation possible.

- My dissertation advisors, Profs Hiroki Tsurumi and Louise B. Russell, for their guidance, encouragement, understanding and patience all through my graduate study at Rutgers. Their wisdom and knowledge have always inspired me.
- My dissertation committee members, Profs John Landon-Lane and Edwin J. Green, for their valuable comments on the earlier draft of the chapters of my dissertation.
- Faculties in the Department of Economics, for their education.
- Yoshida Scholarship Foundation, Tokyo, Japan, for their financial support for my graduate study between 2002 and 2005.
- Dr. Sahoko Kaji, for her guidance and encouragement ever since I met economics.
- Dorothy Rinaldi, for her encouragement and assistance on my life in graduate school.
- My colleagues in the Department of Economics, for discussion on my research and/or my life in general.
- All my friends in New Jersey and Japan, for being there for me. I have to give a special mention for the support given by Kaveh Akram and Mizuho Saito.
- My parents, Nobuko and Nobuo Ibuka, my sister Yasuko Ohashi and my brother in law Keita Ohashi, who have always believed in me.

## DEDICATION

To my mother and father.

## TABLE OF CONTENTS

|  |    |
|--|----|
| Abstract   | ii |
| Acknowledgement  | iv |
| Dedication   | v  |
| List of Tables   | ix |
| List of Figures  | xi |
| Part I: Introduction   | 1  |
| Part II: Chapter 1 A Bayesian Approach to Modeling Dependence in Cost-Effectiveness Analysis | 5  |
| 1.1 Introduction   | 5  |
| 1.2 The Model  | 7  |
| 1.3 Bayesian Estimation  | 12 |
| 1.4 Cost-Effectiveness Analysis  | 13 |
| 1.5 An Example   | 14 |
| 1.5.1 Data   | 16 |
| 1.5.2 Estimation   | 17 |
| 1.5.3 Estimation Result  | 18 |
| 1.5.4 A Simulation   | 21 |
| 1.6 Discussion   | 37 |
| Chapter 2: Bayesian Estimation of Discrete Time Asset Pricing Models                         | 39 |
| 2.1 Introduction   | 39 |
| 2.2 Continuous Asset Pricing Models  | 42 |

|  |    |
|--|----|
| 2.3 Numerical Examples with the Transition Functions   | 45 |
| 2.4 Numerical Examples Using the CKLS model  | 48 |
| 2.4.1 When $m$ is Known  | 48 |
| 2.4.4.1 Using All Data   | 48 |
| 2.4.4.2 Using Every $r$ -th Observation  | 51 |
| 2.4.2 When $m$ is Unknown  | 53 |
| 2.4.3 Eraker's Data Augmentation   | 55 |
| 2.4.4 When Averaged Data is Used   | 58 |
| 2.5 Empirical Evidence with Short-term Interest Rates  | 59 |
| 2.6 Concluding Remarks   | 63 |
| Chapter 3: A Bayesian Comparison in Short-term Interest Rate Models  | 65 |
| 3.1 Introduction   | 65 |
| 3.2 The Model and MCMC algorithms  | 67 |
| 3.3 Numerical Examples   | 70 |
| 3.4 Empirical Results, Predictive Densities and Comparison with GARCH model  | 75 |
| 3.5 Concluding Remarks   | 83 |
| Appendix A: Likelihood Function of the Joint Function of Survival and Distribution Functions                               | 84 |
| Appendix B: Derivation of the Transition Function of Vasicek Model   | 86 |
| Appendix C: MCMC Algorithm for Obtaining the Posterior p.d.f. of Parameters of Vasicek Model Using the Transition Function | 88 |
| Appendix D: Ait-Sahalia's Approximation Method of the Transition Functions   | 89 |

|  |     |
|--|-----|
| Appendix E: MCMC algorithms for Obtaining Posterior p.d.f. of Parameters | 91  |
| Appendix F: MCMC algorithm for Data Augmentation Proposed by Eraker      | 93  |
| Bibliography   | 96  |
| Curriculum Vita  | 100 |



## LIST OF TABLES

|   |    |
|---|----|
| 1.1 Descriptive Statistics: Sample from NHEFS   | 17 |
| 1.2 Posterior Statistics of Parameters  | 23 |
| 1.3 Means and Standard Deviations for Cost, Effectiveness and Incremental Cost<br>Effectiveness for a Group of Persons Aged 60 and No Chronic Condition (Low-risk<br>Persons) | 29 |
| 1.4 Means and Standard Deviations for Cost, Effectiveness and Incremental Cost and<br>Effectiveness for Two Types of an Individual  | 33 |
| 2.1 Posterior Statistics for Parameters with Different Values of $h$ with Vasicek Model:<br>Using the Transition Function   | 47 |
| 2.2 Posterior Statistics for Parameters with Various Values for $m$ : Using Discrete-time<br>Approximation  | 50 |
| 2.3 Posterior Statistics for Parameters after Sampling Every $r$ -th Interval When $m$ is<br>Known  | 53 |
| 2.4 Posterior Statistics for Parameter after Sampling Every $r$ -th Interval When $m$ is<br>Unknown   | 54 |
| 2.5 (a) Posterior Statistics for Parameters with Eraker's Data Augmentation: Native<br>Discretization   | 57 |
| 2.5 (b) Posterior Statistics for Parameters with Eraker's Data Augmentation: Exact Value<br>for $h$   | 57 |
| 2.6 Posterior Statistics for Parameters When Observations Are Averaged  | 59 |
| 2.7 Descriptive Statistics of Daily Effective Federal Funds Rate (1/1/1960-7/31/2004)   | 61 |

|  |    |
|--|----|
| 2.8 Empirical Results  | 61 |
| 3.1 Numerical Examples for the CKLS Model with Stochastic Volatility   | 74 |
| 3.2 Summary Statistics: weekly 3-month Treasury Bill Rate (01/08/1982-10/06/2006)  | 78 |
| 3.3 Estimation Results of the CKLS with stochastic volatility model and CKLS with<br>GARCH (1,1) model with 3 month Treasury bill rate | 80 |

## LIST OF FIGURES

|   |    |
|---|----|
| 1.1 Plot of Variables from NHEFS  | 19 |
| 1.2 Plot of Posterior Draws and Posterior pdf for the Dependence Parameter  | 24 |
| 1.3 Contour and Surface of Joint pdf of Effectiveness and Cost with Various Copulas   | 26 |
| 1.4 Cost-effectiveness Plane for Persons Aged 60 with No Chronic Conditions (Low-risk Persons)  | 30 |
| 1.5 Posterior Predictive Densities of Effectiveness and Cost for a Person Aged 60 with No Chronic Conditions (Low-risk Person) Under Treatment1 | 34 |
| 1.6 Posterior Predictive Densities of Incremental Effectiveness and Cost for a Person Aged 60 with No Chronic Conditions (Low-risk Person)      | 35 |
| 1.7 Comparison between Low-risk (Number of Chronic Condition 0) and High-risk Persons (Number of Chronic Conditions 4)                          | 36 |
| 2.1 Relationship between $t$ , $h$ and $N$  | 44 |
| 2.2 Posterior pdf for $\theta_2$ and $\beta$ from Empirical Analysis  | 62 |
| 3.1 Assessing Convergence for Parameters in the CKLS with Stochastic Volatility Model   | 72 |
| 3.2 Weekly 3-month Treasury-Bill Rate (01/08/1982-10/06/2006)   | 77 |
| 3.3 Predictive Densities: One-and two-period ahead  | 81 |
| 3.4. Cumulative Density Functions of Absolute Deviation of the Predicted Values from the Actual Value   | 82 |

## Part I

### INTRODUCTION

Bayesian statistical tools have been widely used for analyses in social and natural science fields. In this context, several textbooks on Bayesian econometrics have been recently published in economics, long after the classic on Bayesian econometrics by Zellner (1971). These textbooks emphasize the application of Bayesian inference to issues in micro and macro economics, and cover various topics ranging from models in time series (Koop(2003), Bauwens et al.(2000)), Geweke(2005), and Koop et al. (2007)) to topics in micro econometrics including analysis of limited dependent variables or panel data (Koop(2003) and Lancaster(2004)). In addition to the development in theoretical aspects of analysis, progress in the technical side, i.e., improvement in hard- and software in computational tools, contributed to the rise in the popularity of Bayesian inference in the applied fields as Bayesian inference often requires computer intensive analysis. Specific software has been developed for Bayesian analysis – such as *WinBugs*<sup>1</sup> and the Bayesian module for SAS 9.1.3.<sup>2</sup>, – so that more researchers obtained access to tools of Bayesian inference.

---

<sup>1</sup> The *WinBugs* was developed through the BUGS project (<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>). The BUGS (Bayesian inference Using Gibbs Sampling) project started in the MRC Biostatistics Unit, Cambridge University, and developed jointly with the Imperial College School of Medicine at St Mary's London and the University of Helsinki. WinBugs uses R language.

<sup>2</sup> Three procedures for Bayesian inference, in generalized linear models, accelerated life failure models, Cox regression models, and piecewise constant baseline hazard models, are available as experimental versions as of March 2008.

Bayesian statistical inference is made about a parameter  $\theta$ , or unobserved data  $\tilde{y}$  in the case of predictive inference, in terms of probability terms,  $p(\theta | y)$  or  $p(\tilde{y} | y)$  given the realized data  $y$  using Bayes' rule. For parameter inference, given *the prior* distribution  $p(\theta)$  and the data distribution  $p(y | \theta)$ , the joint probability density for  $\theta$  and  $y$  is written as:

$$p(\theta, y) = p(y | \theta)p(\theta).$$

The posterior density  $p(\theta | y)$  is therefore

$$\begin{aligned} p(\theta | y) &= \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y | \theta)}{p(y)} \\ &\propto p(\theta)p(y | \theta). \end{aligned}$$

The data distribution  $p(y | \theta)$  is also called the likelihood function, which is a function of  $\theta$  for fixed (realized value)  $y$ . The posterior density  $p(\theta | y)$  shows the revision of our initial probability on  $p(\theta)$  using new information in our data  $y$  with Bayes' rule.

A similar argument can be made for predictive inference. The distribution of unobserved  $\tilde{y}$ , given the realized value  $y$ , is obtained by the following equation.

$$\begin{aligned} p(\tilde{y} | y) &= \int p(\tilde{y}, \theta | y) d\theta \\ &= \int p(\tilde{y} | \theta, y) p(\theta | y) d\theta \\ &= \int p(\tilde{y} | \theta) p(\theta | y) d\theta. \end{aligned}$$

, where the last equality holds because  $y$  and  $\tilde{y}$  are independent conditioned on  $\theta$ . The distribution of  $\tilde{y}$  given the realized value  $y$  is called the *posterior predictive distribution*.

The object of Bayesian is to obtain the posterior distribution for each model parameter given data, which requires integrating over the posterior distribution of models parameters to find the proportional to constant. Therefore, Bayesian inference potentially requires high-dimensional integration to make inference on model parameters or predictions, which is often analytically difficult or impossible. Markov chain Monte Carlo (MCMC) methodology, a simulation-based technique, provides a solution to the problem. The MCMC enables us to draw samples from probability distributions by constructing a Markov chain that has the target distribution as its equilibrium distribution. The two most well-known ways of constructing a Markov chain that has such a property are the Metropolis-Hastings algorithm (Metropolis et al. (1953) and Hasting (1970)) and the Gibbs sampling (Geman and Geman (1984)), which is the special case of the MH algorithm<sup>3</sup>. MCMC is an essential tool in recent research in economics and econometrics using Bayesian inference.

Part II of the dissertation consists of three essays on the application of Bayesian inference using MCMC to topics in economics. Chapter 1 deals with an application in modeling in health and medicine, and develops a cost-effectiveness analysis using Bayesian approach. Cost-effectiveness analysis based on the Bayesian approach is a form of probabilistic cost-effectiveness analysis. Uncertainty in inference is well-captured by the probabilistic cost-effectiveness analysis, which is one of the reasons why this type of analysis has been increasing.

---

<sup>3</sup> Casella and George(1992) and Chib and Greenberg(1995) provide an intensive summary about the Gibbs Sampling and the Metropolis-Hastings algorithm.

A Bayesian approach is sometimes preferred because of its computational ease. For example, simulation-based inference is powerful enough to handle models with latent observations or variables. Chapter 2 and 3 discuss such cases.

Chapter 2 consists of a Bayesian analysis of a discrete-time asset pricing model. It is widely known that a bias occurs when one estimates a discrete-time asset pricing model, if the approximation from the original continuous-time model to the one in discrete-time is naive. One way of avoiding the bias was proposed by Eraker(2002), who suggested augmenting data between the observables to make the approximation closer to the original using MCMC.

Chapter 3 analyzes another type of model in finance, the stochastic volatility model. The stochastic volatility model includes the vector of volatility state, which is latent in the model. The existence of the latent variables makes estimation of the model cumbersome. With the help of MCMC, the estimation can be relatively easily handled.

In Chapter 1 and Chapter 3, we use the concept of Bayesian predictive densities. Predictive densities can show uncertainty accompanied with forecasted values, which is important for decision making based on the forecast. MCMC methodology is also useful to obtain predictive densities when the analytical predictive density is unavailable.

## Part II

### CHAPTER 1: A BAYESIAN APPROACH TO MODELING DEPENDENCE IN COST-EFFECTIVENESS ANALYSIS

#### 1.1 Introduction

Cost-effectiveness analyses have been widely used not only in the field of health economics but also in other health and medical fields. The purpose of cost-effectiveness analysis is to compare the differences, or increments in cost and effectiveness between two treatments. Since the two variables are often related to each other, it is important to model cost and effectiveness simultaneously.

O'Hagan et al.(2001) assume that cost  $C$  and effectiveness  $E$  are distributed as;

$$f(e, c) \sim N(\alpha, \Sigma) \quad (1.1)$$

In their model, it is supposed that the measures of cost and effectiveness follow the bivariate normal distribution with the mean vector  $\alpha$  and the variance-covariance matrix  $\Sigma$ . Although the normal distribution is easy to handle, the model based on the normal distribution has some limitations. First, the normality assumption is not always appropriate for cost data that is often skewed; second, the model also does not allow a non-normal distribution in effectiveness measure, where survival function is often used in analysis.

An alternative model assumes a joint density function  $f(e, c)$  where the variable  $E$  is categorical (O'Hagan and Stevens (2001); Heitjan and Li (2001).) For



example, O'Hagan and Stevens model the joint pdf of cost and effectiveness using the following equation:

$$f(e, c | \phi, \theta_0, \theta_1) = \begin{cases} (1 - \phi)f(c | \theta_0) & \text{if } e = 0 \\ \phi f(c | \theta_1) & \text{if } e = 1 \end{cases}$$

, where  $\phi$  is the probability of  $e = 1$  and  $\theta_0$  is a vector of parameters of function  $f(c)$  when  $e = 0$  and  $\theta_1$  is a corresponding vector when  $e = 1$ . This approach is appropriate when the effectiveness is a categorical variable.

The aim of this paper is to propose an alternative approach to model the joint distribution of  $f(e, c)$  using copulas. According to Nelson (2006), "Copulas are functions that join or "couple" multivariate distribution functions to their one dimensional marginal distribution functions." The study of copulas started in probability and statistics, and applications of copulas to finance and survival analysis have also been increasing (Cherubini et al. (2004); Shemyakin and Youn (2006); Rodeo et al.(2006).)

One of the useful features of joint distributions based on copulas lies in the flexibility of the assumption one can make on the marginals. Equation (1.1) assumes that the marginal distributions of  $E$  and  $C$  are both normal, but one can choose any distributions as marginals in copula models. In the example presented in this paper, we construct the distribution function  $f(e, c)$  using a Weibull distribution for  $f(e)$  and a log-normal distribution for  $f(c)$ .

Introduction of copulas into cost-effectiveness analysis was first made by Quinn (2005). Our model is different from his in three aspects. First, we conduct a Bayesian analysis while Quinn conducts his analysis from frequentists' point of view. The Bayesian framework has recently become popular in cost-effectiveness analyses (Briggs

(1999); Heitjan et al.(1999); Al and Van Hout (2000); O'Hagan and Stevens (2001); O'Hagan et al. (2001); Vanness and Kim (2002); Vázquez Polo et al.(2005)). Bayesian methods allow us to incorporate prior information or ideas into the analysis through prior distributions.

Second, we face censored data in the effectiveness measure while Quinn uses non-censored data. Our model assumes survival time as effectiveness, and available data for survival is often censored. Willan et al.(2005) developed a model for censored data with a joint normal assumption.

Third, Quinn assumes normal distribution for effectiveness equation while this study assumes Weibull distribution.

The paper is organized as follows. Section 1.2 introduces the model of a joint function of cost and effectiveness using three kinds of copulas. In section 1.3, Bayesian estimation of the copula models is developed. Section 1.4 explains the cost-effectiveness analysis. Section 1.5 illustrates our model with a sample of adults aged between 25 and 74 years from the NHANES I Epidemiologic Follow-up Study (NHEFS). A simulation of cost-effectiveness analysis is conducted using the estimated model. Section 1.6 concludes the chapter.

## **1.2 The Model**

Our model of the joint distribution of cost and effectiveness consists of two parts: (1) modeling the marginals for cost and effectiveness, and (2) choice of copulas. It is supposed that we have patient-level data from a clinical trial or a survey with observations on subject  $i$  ( $i = 1, \dots, n$ ). Our approach is regression-based analysis, and for

each  $i$ , the data includes measurements of effectiveness  $e_i$ , cost  $c_i$  and a set of covariates  $z_i = (1, x_i, y_{i1}, y_{i2}, \dots, y_{ik})$ . Here  $x_i$  is the intervention dummy, which takes 1 if the patient received the intervention, and 0 if she did not.  $y_{i1}, y_{i2}, \dots, y_{ik}$  are other covariates with  $k$  the total number of the control variables. The control variables can include characteristics of patients such as age, sex and measures of health condition. It is of interest to estimate the cost and effectiveness equations using the Data  $D = \{e_i, c_i, z_i\}$  for  $i = 1, \dots, n$ . Here, we assume that the covariates used in the two equations are the same, but more flexible assumptions would be possible.

In this study, the duration of survival is used as the effectiveness measure  $E$  with  $E > 0$ . The survival function is written as  $S_E(e) = \Pr(E > e) = 1 - F_E(e)$ , where  $E$  is a random variable denoting time to death and  $e$  is an arbitrary number, and  $F_E(e)$  is the cumulative distribution function of  $E$ . The cost variable  $C$  is assumed to be a continuous random variable that has the cumulative distribution function  $F_C(c)$ .

The next step is to introduce the dependence between  $E$  and  $C$  using copulas. Copulas are multivariate distribution functions where the marginal of each variable is a uniform distribution. This study deals with two-dimensional copulas to construct joint distributions of variables  $E$  and  $C$ , and formally defines them as follows:

A bivariate copula is a function  $C : [0,1]^2 \rightarrow [0,1]$  with the following properties:

1. For every  $u$  and  $v \in [0,1]$ ,

$$C(u,0) = 0 = C(0,v)$$

and

$$C(u,1) = u \text{ and } C(1,v) = v;$$

2. For every  $u_1, u_2, v_1, v_2 \in [0,1]$  such that  $u_1 \leq u_2$  and  $v_1 \leq v_2$ ,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

The most well-known theorem on copulas is Sklar's theorem that states a copula recovers the joint distribution function of two random variables from their marginal distributions. For example, one can obtain the joint distribution function

$F_{XY}(x, y) = C(F_X(x), F_Y(y))$  of the cumulative distribution functions (cdf) of two random variables  $F_X(x)$  and  $F_Y(y)$  using the associated copula  $C$  with  $u \equiv F_X(x)$  and  $v \equiv F_Y(y)$ .

Another example of copulas is survival copulas. A joint survival function of two random variables,  $X$  and  $Y$ , is  $S_{XY}(x, y)$  and is defined by  $\hat{C}(S_X(x), S_Y(y))$  where  $\hat{C}$  shows a survival copula and  $S_X(x)$  is a survival function of random variable  $X$  and  $S_Y(y)$  is the corresponding function of random variable  $Y$ . Applications of copulas in survival analysis are found in Romeo et al. (2006) and Shemyakin and Youn (2006).

In our case, effectiveness variable  $E$  is more closely related to the survival function  $S_E(e)$  rather than  $F_E(e)$  in its interpretation, and therefore it is natural to build the joint function of  $E$  and  $C$  based on  $S_E(e)$  and  $F_C(c)$ . Assuming  $u \equiv S_E(e)$  and  $v \equiv F_C(c)$  given a dependence parameter, the joint function of the survival function and the distribution function for the cost equation based on a copula  $C$  is defined by

$$H(e, c) = C(S_E(e), F_C(c); \alpha) = \Pr(E > e, C \leq c) \quad (1.2)$$

It is important to notice that the function expressed by equation (1.2) is not a joint distribution function or joint survival function. Indeed,  $H(e, -\infty) = 0$  but

$$H(-\infty, c) \neq 0 \text{ and } H(\infty, \infty) \neq 1.$$

There are many classes of copulas, and an important class is known as Archimedean Copulas. This class of copulas is generated using the following function:

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)), \quad u, v \in [0, 1] \quad (1.3)$$

, where  $\varphi$  is a continuous, strictly decreasing, convex function  $[0, 1] \rightarrow [0, \infty]$  such that

$\varphi(0) = \infty$  and the inverse function  $\varphi^{-1}$  satisfies  $\varphi^{-1} : [0, \infty] \rightarrow [0, 1]$ ,  $\varphi^{-1}(0) = 1$

and  $\varphi^{-1}(\infty) = 0$ . One of the important features of Archimedean copulas is their symmetry;

$C(u, v) = C(v, u)$  for all  $u, v \in [0, 1]$ .

Archimedean copulas are applied in wide range of studies mainly because of their ease of construction (Nelson 2006). Among a number of Archimedean copulas, the following three well-known families are considered: Clayton, Gumbel-Hougaard and Frank families. The three families are different in dependence structure between variables, and each copula is explained as follows.

### 1. Clayton family

A Clayton copula is generated by equation (1.3) with

$$\varphi(x) = \frac{1}{\alpha}(x^{-\alpha} - 1).$$

From this equation, we obtain  $\varphi(S_E(e)) = \frac{1}{\alpha}(S_E(e)^{-\alpha} - 1)$  and  $\varphi(F_C(c)) = \frac{1}{\alpha}(F_C(c)^{-\alpha} - 1)$ .

The joint function equation (1.2) based on Clayton copula is given by

$\varphi^{-1}(\varphi(S_E(e)) + \varphi(F_C(c)))^{-1}$ , which is

$$H(e, c) = [\max(S_E(e)^{-\alpha} + F_C(c)^{-\alpha} - 1, 0)]^{-1/\alpha}, \quad \alpha \in [-1, \infty) \setminus \{0\}. \quad (1.4)$$

The original form in Clayton(1978) allows only positive dependence, but this general version allows positive or negative dependence. As  $\alpha \rightarrow 0$ , we have

$$H(e, c) = S_E(e)F_C(c).$$

## 2. Gumbel-Hougaard family

The generating function for Gumbel-Hougaard family is given by

$$\varphi(x) = (-\ln x)^{-\alpha}.$$

In a similar way as Clayton family, this function together with (1.3) yields the following equation:

$$H(e, c) = \exp\left\{-\left[(-\ln(S_E(e)))^\alpha + (-\ln(F_C(c)))^\alpha\right]^{1/\alpha}\right\}, \quad \alpha \in [1, \infty) \quad (1.5)$$

As  $\alpha \rightarrow 1$ , we have  $H(e, c) = S_E(e)F_C(c)$ . This family permits only positive dependence.

## 3. Frank family

The generating function for Frank family is given by

$$\varphi(x) = -\ln\left(\frac{e^{-\alpha x} - 1}{e^{-\alpha} - 1}\right).$$

The joint function given by equation (2) is therefore:

$$H(e, c) = \left(-\frac{1}{\alpha}\right) \ln\left(1 + \frac{(\exp(-\alpha S_E(e)) - 1)(\exp(-\alpha F_C(c)) - 1)}{\exp(-\alpha) - 1}\right) \quad \alpha \in (-\infty, \infty) \setminus \{0\} \quad (1.6)$$

This copula allows both negative and positive dependence. As  $\alpha \rightarrow 0$ , we

have  $H(e, c) = S_E(e)F_C(c)$ .

### 1.3 Bayesian Estimation

A Bayesian estimation of models with copulas is made in Romeo et al.(2006), and we closely follow their method. The estimation consists of two steps: in the first step, parameters in the survival and cost equations are estimated assuming their independence; and in the second step, the parameter for dependence ( $\alpha$ ) is estimated by the pseudo-likelihood function with the estimated parameters from the first step integrated.

The probability density function (pdf) of the survival function  $S_E(e)$  is denoted as  $f_E(e)$ , and the pdf of  $F_C(c)$  is written as  $f_C(c)$ . The effectiveness measure, survival time, is usually censored and estimations are made with the observed (censored) variable  $w_i = \min\{e_i, t_i\}$ , where  $e_i$  is a time to death and  $t_i$  is a censoring time. The binary for censoring is denoted as  $\delta_i = I[w_i = e_i]$ .

We define  $\phi_E$  and  $\phi_C$  as the vectors of parameters in the survival function and cost equation respectively. The likelihood function for the survival function with censored data is expressed as:

$$L(\phi_E | w, \delta) = \prod_{i=1}^n f_E(w_i)^{\delta_i} S_E(w_i)^{1-\delta_i}$$

The likelihood function for the cost equation is straightforward and given by

$$L(\phi_C | c) = \prod_{i=1}^n f_C(c_i)$$

Together with the prior distributions for parameters,  $\pi^{\phi_E}(\phi_E)$  and  $\pi^{\phi_C}(\phi_C)$ , the posterior pdfs of the parameters are given by

$$\pi(\phi_E | w, \delta) \propto L(\phi_E | w, \delta) \pi^{\phi_E}(\phi_E) \tag{1.7}$$

and

$$\pi(\phi_C | c) \propto L(\phi_C | c) \pi^{\phi_C}(\phi_C) \quad (1.8)$$

In the second step, the pseudo-likelihood function for the dependence parameter  $\alpha$  is formed, given the posterior means of parameters in the two equations,  $\hat{\phi}_E$  and  $\hat{\phi}_C$ .

The derivation of the pseudo likelihood function is referred to Appendix A.

$$L(\alpha | w, \delta, c, \phi_E, \phi_C) = \prod_{i=1}^n f(w_i, c_i; \alpha)^{\delta_i} \times \frac{\partial H(w_i, c_i; \alpha)^{1-\delta_i}}{\partial c_i}$$

Given the likelihood function and the prior distribution of  $\alpha$ ,  $\pi^\alpha(\alpha)$ , the posterior pdf of the dependence parameter is given by

$$\pi(\alpha | w, \delta, c, \hat{\phi}_E, \hat{\phi}_C) \propto L(\alpha | w, \delta, c, \phi_E, \phi_C) \pi^\alpha(\alpha) \quad (1.9)$$

Three equations (1.7)(1.8) and (1.9) are obtained by the Markov Chain Monte Carlo method once the likelihood functions are formed.

#### 1.4 Cost-Effectiveness Analysis

From the estimated joint distribution function of  $E$  and  $C$ , a cost-effectiveness analysis is conducted. The cost-effectiveness of two treatments is often measured by the incremental cost-effectiveness ratio (ICER)

$$ICER \equiv \frac{\overline{C_2} - \overline{C_1}}{\overline{E_2} - \overline{E_1}} \quad (1.10)$$

where  $\overline{C_j}$  shows average costs for treatment  $j$  and  $\overline{E_j}$  shows average effects received from treatment  $j$  ( $j=1,2$ ). The ICER shows the ratio of increment in cost of Treatment 2 over Treatment 1 to increment in effectiveness of Treatment 2 over Treatment 1; therefore a negative ICER suggests either Treatment 2 is more costly and



less effective, or less costly and more effective. If the ICER is a positive number, either Treatment 2 exceeds Treatment 1 in both cost and effectiveness, or vice versa. Costs in cost-effectiveness analysis include direct costs of intervention, and effectiveness is often measured by life year saved. In the regression model in cost-effectiveness analysis, the different treatments are captured by a binary  $x$  to show which of the treatments people received. If an individual  $i$  received Treatment 1  $x_i$  takes 0; if she received Treatment 2,  $x_i$  is 1.

One can narrow down the information conveyed by the ICER by plotting the incremental cost and effectiveness,  $\Delta\bar{C} \equiv \bar{C}_2 - \bar{C}_1$  and  $\Delta\bar{E} \equiv \bar{E}_2 - \bar{E}_1$ , in the x-y plane. In the plane, called cost-effectiveness plane, the x-axis represents the difference in effects and y-axis represents the difference in costs between two treatments. The south-east quadrant of the plane shows Treatment 2 is more effective and less costly than Treatment 1. The north-west quadrant shows the dominance of Treatment 1. The north-east quadrant shows Treatment 2 is more effective but more costly; and the south-west quadrant means Treatment 1 is more effective but more costly. The plot of distributions of  $\Delta\bar{C}$  and  $\Delta\bar{E}$  reveals the probability of each possibility in the plane.

### 1.5 An Example

We apply the model developed in earlier sections to a sample of adults between 25 and 74 years old from the NHANES I and its Epidemiologic Follow-up Study (NHEFS). The study examines the baseline health of a cross-section of individuals between 1971 and 75 and then follows their history over the next 20 years. Unfortunately the survey does not include an intervention, and we suppose a hypothetical scenario for a

cost-effectiveness analysis. The scenario assumes that a hypothetical intervention would lower systolic blood pressure below 140. Treatment 1 is "do- nothing" to treat high blood pressure, and Treatment 2 immediately cures high blood pressure as a result of the hypothetical intervention. In this setting, the binary of having high blood pressure is considered as the intervention variable in the model.

The costs of Treatment 1 are the costs of medical care for the untreated high blood pressure. The costs of Treatment 2 are the costs of hypothetical intervention, which we assume to be zero for simplicity, and the costs of medical care that occur despite the treatment. The cost of medical care in each treatment is measured by hospital days per year. The monetary cost can be obtained by multiplying the hospital days per year by the cost per inpatient day, but the multiplication affects only the unit of cost and does not change the entire argument regarding the model itself. Hence we show the results based on hospital days per year as the cost variable  $C$ .

To estimate the survival time  $E$ , the variable of duration at risk is used, which is the length of time between the first interview date of the survey and either the last interview date or the date of death, whichever occurred first.

We assume the Weibull distribution for survival function, whose pdf is shown by the following equation:

$$f_E(e | z, \phi_E) = \exp(\omega - \exp(\omega))$$

,where  $\omega = p(\ln e - \beta_E' z)$  and  $z$  is a vector of constant term and a set of variables. A vector of parameters is defined as  $\phi_E \equiv \{\beta_E, p\}$ . The cost variable  $C$ , conditioned on the covariates is assumed to follow log-normal distribution

$$f_C(c | z, \phi_2) = \frac{1}{cv\sqrt{2\pi}} \exp\left(-\frac{(\ln c - \beta_C' z)^2}{2v^2}\right)$$

,where  $\phi_C \equiv \{\beta_C, v\}$ .

The covariates include only three variables and the constant term: (1) binary for having systolic blood pressure of 140 or higher as the "intervention" variable, (2) age, and (3) the number of chronic diseases from the following nine conditions: asthma/pleurisy/emphysema, diabetes, heart attack/failure, stroke, malignant tumor, spine/hip/wrist fracture, arthritis/gout, ulcer, and tuberculosis.

### 1.5.1 Data

Among the 11,419 respondents in the survey, we only use respondents who reported more than one hospital day during the sample period (n=7866). This is because our simple model assumes Log-Normal distribution in the cost equation, which does not allow 0 values. Table 1.1 shows descriptive statistics of the sample. The duration at risk, in the first row of the table, shows the period between the first interview and date of death, or date of the last interview when the respondent is alive at the time of the last interview.

The mean of hospital days per year is 3.48. As shown in Figure 1.1, the distribution of the variable is highly skewed and the median value is 2.25. It shows that 31% of the entire respondents in the sample belong to the range above 0 and less than or equal to 0.5.

**Table 1.1: Descriptive statistics: sample from NHEFS**

| (n=7866)                              | Mean | Std.D | Max    | Min  |
|---------------------------------------|------|-------|--------|------|
| Duration at risk*, in years           | 16.4 | 4.81  | 22.06  | 0.02 |
| Hospital days per year                | 3.48 | 16.63 | 219.15 | 0.04 |
| Age, in years                         | 50.7 | 15.28 | 76     | 24   |
| High blood pressure** (0 or 1)        | 0.39 | 0.49  | 1      | 0    |
| Number of chronic disease (9 maximum) | 0.91 | 1.03  | 7      | 0    |

\* Duration at risk is the length of time between the first interview date and either the last interview date or date of death whichever occurred first

\*\* The variable takes 1 if the systolic blood pressure is 140 mm Hg or higher; 0 otherwise.

As for the variables for covariates, 39% of the people have a systolic blood pressure of 140 or more in the sample and the mean age of the sample is 50.7. On average, people have one chronic disease where the maximum number of chronic diseases is 7 and the minimum is 0.

The second panel of Figure 1.1 shows the plot of the duration at risk, from which the survival time is estimated, and the cost. A slight negative association between the variables is observed, and we attempt to capture the association between the two variables using copulas.

### 1.5.2 Estimation

The Bayesian estimation of Weibull survival and log-normal cost was done by the experimental version of SAS Bayesian module downloadable at <http://support.sas.com/rnd/app/da/bayesproc.html>. The module uses Gibbs-sampling to obtain the posterior density of the parameters. The Weibull regression, performed by the BLIFEREG procedure in the module, uses the algorithm proposed in Ibrahim et al.(2001). There after the full conditional distribution of each parameter is analytically

obtained, they sample from the full conditional using a rejection algorithm. For the log-normal regression, the BGENMOD procedure was used. To make data smooth, the natural logarithm is taken in age.

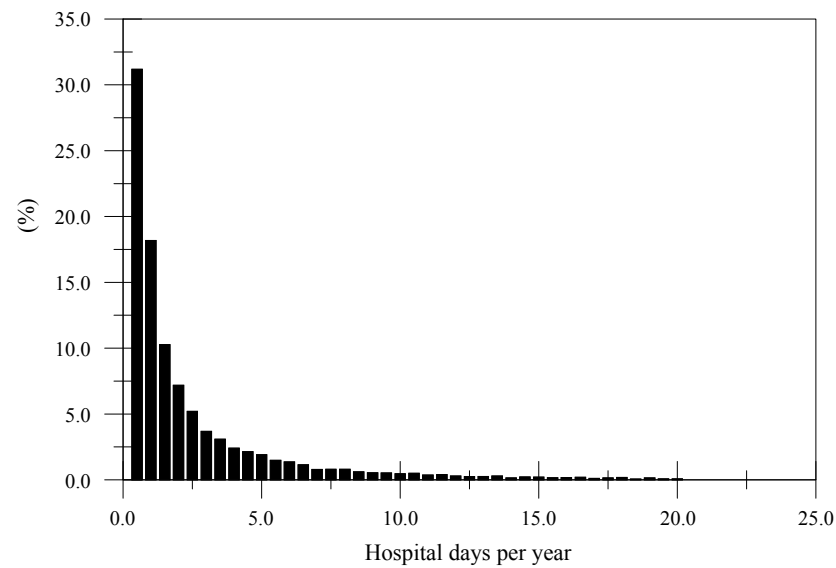
For the estimation of the dependence parameter  $\alpha$ , we use an efficient jump proposed by Gelman et al. (2004). The efficient jump method is one of the Metropolis-Hastings algorithms, and useful in case we do not find a candidate theoretically as the proposal density. The estimates were obtained by 8000 iterations after a burn-in of 2000 values. To obtain the posterior statistics, every 3rd draw is kept. The priors of all the parameters are assumed to be non-informative and flat in this study. Choices of priors is an issue, and some discussions on the effect of prior distributions in cost-effectiveness analysis is found in O'Hagan and et al.(2001).

### 1.5.3 Estimation Results

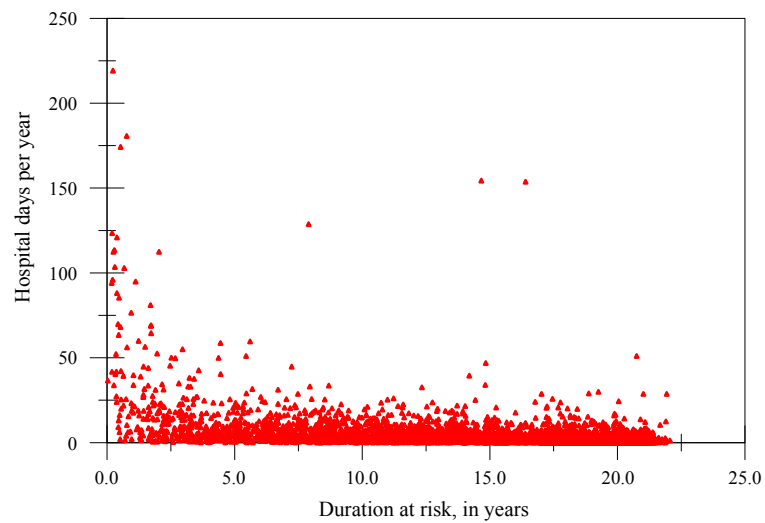
Table 1.2 shows the posterior means and standard errors of the parameters in the model. Having high blood pressure, older age, and more chronic diseases are negatively related to survival, and positively related to hospital days per year. The scale parameter in the log-normal distribution for cost is 1.25, showing the high skewness of the cost variable. The posterior mean for the dependence parameter is 1.12 with Clayton, 1.48 with Gumble-Hougaard, and 4.24 with Frank copulas. All the values indicate positive dependence between the two functions  $S_E(e)$  and  $F_C(c)$ . This implies that the variables  $E$  and  $C$  are negatively associated because function  $S_E$  is a decreasing function of  $E$  and  $F_C(c)$  is an increasing function of  $C$ .

**Figure 1.1: Plot of variables from NHEFS**

(a) Distribution of hospital days per year



(b) Duration at risk and hospital days per year \*



\* Duration at risk is the length of time between the first interview date and either the last interview date or date of death whichever occurred first.

The plot of posterior samples and posterior density for  $\alpha$  is also shown for each copula in Figure 1.2. The convergence of Markov Chains is visually confirmed in the top panels of the figure. The dependence parameters for different copulas cannot be compared directly with each other as the models are different; however, the range of the posterior densities stays away from 0 for Clayton, 1 for Gumble-Hougaard, and 0 for Frank, which indicates some dependence between the effectiveness and the cost. From the dependence parameter in copulas, one can easily obtain measures for dependence such as Kendall's tau and Spearman's rho. These measures of dependence are rank correlation coefficients, and have the following properties. If the agreement between the two rankings is perfect, the coefficient takes 1; if one ranking is the reverse of the other, the coefficient takes -1; and the rankings are independent, the coefficient takes 0 value. Unlike measures of linear dependence between random variables, such as Pearson's correlation coefficient, Kendall's tau is a scale-invariant measure of dependence.

From the estimated dependence parameter in each copula function, we obtained the following estimates for Kendall's tau. Tau is 0.36 for Clayton, 0.32 for Gumbel-Hougaard, and 0.41 for Frank. They show moderate association between the functions.

To have a better idea of the shape of the density function with each copula, Figure 1.3 shows the surface and contour of the pdf  $f(e, c)$  for each copula. It is important to note that this is not the joint posterior density of  $E$  and  $C$ , but a simple draw of the joint pdf of the two variables when the parameters are fixed. The function is the joint pdf conditioned on the parameters  $\phi_E, \phi_C, \alpha$  and covariates  $z$ . In the figure, the parameter values are fixed at the posterior means, and the following assumption is used for

covariates  $z$ : binary for high blood pressure is assumed to be 1 (Treatment 2), age is 60, and the number of chronic diseases is 0. In each family of copula, the negative association between  $E$  and  $C$  is observed with the estimated value for the given parameters: the larger number for hospital days per year is associated with the shorter survival time.

#### 1.5.4 A simulation

A small simulation is made using the estimated model to show an example of cost-effectiveness. This is a counter-factual experiment, and assumes that persons who have high blood pressure reduce their blood pressure to safe level. We assume a specific type of patient, who has high blood pressure, is 60 years old and has no other chronic conditions than high blood pressure, and examine the effect on costs and effects of reduction of blood pressure under 140. We can think of this as two different treatments for high blood pressure. The original treatment (Treatment 1) does not treat high blood pressure actively and leave the condition.

First, we obtained the cost-effectiveness plane. The mean costs and effects are defined by the expected values of  $E$  and  $C$ , and denoted as  $E(E)$  and  $E(C)$ . Using the

expected values of  $E$  and  $C$ , the equation can be rewritten as  $ICER = \frac{\gamma_2 - \gamma_1}{\mu_2 - \mu_1} = \frac{\Delta\gamma}{\Delta\mu}$

, where  $\gamma_j = E(C_j)$  and  $\mu_j = E(E_j)$  with  $j = 1, 2$ . For copula models, the expected values were calculated using Simpson's rule, a method of numerical integration. When assuming the independence between the two equations, the expected values were obtained using the standard formulas for expected values in Weibull and log-normal distributions.



Figure 1.4 shows the cost-effectiveness plane. Copulas affect only the distribution of  $E$  and  $C$ , and the means are invariant among the models. Therefore, we do not find much differences in the distribution of conditional expectations of  $E$  and  $C$  among different models. This is also shown in summary statistics shown in Table 1.3.

Although the cost-effectiveness analysis is usually conducted by using the mean cost and effectiveness for a particular group, it would be worth while seeing the distribution of cost and effectiveness for a particular person or patient at a patient level. The conditional distribution of mean cost and effectiveness shown in Table 1.3 and Figure 1.4 shows the information for a group of persons who have certain characteristics (persons aged 60 with no chronic condition in this example) while the distribution of individual cost and effectiveness shows the information for an individual patient who has certain characteristics. The distribution of cost and effectiveness for an individual can be obtained using Bayesian predictive density. Using predictive densities, we would be able to obtain predicted distributions of  $E$  and  $C$  with a set of patient characteristics given. When a set of patient characteristics is denoted as  $\tilde{z}$ , and the predictive values of  $E$  and  $C$  associated with the person's characteristics  $\tilde{z}$  are denoted as  $\tilde{e}$  and  $\tilde{c}$ , the joint posterior density for  $\tilde{e}$ ,  $\tilde{c}$  and parameter  $\alpha$ , is written as

$$h(\tilde{e}, \tilde{c}, \alpha \mid w, \delta, c, z, \tilde{z}, \hat{\phi}_C, \hat{\phi}_E) = f(\tilde{e}, \tilde{c} \mid \alpha, \tilde{z})\pi(\alpha \mid w, \delta, c, z, \hat{\phi}_C, \hat{\phi}_E)$$

The predictive density is obtained by integrating the parameter  $\alpha$  out:

$$h(\tilde{e}, \tilde{c} \mid w, \delta, c, z, \tilde{z}) = \int f(\tilde{e}, \tilde{c} \mid \alpha, \tilde{z})\pi(\alpha \mid w, \delta, c, z)d\alpha \quad (1.11)$$

**Table 1.2: Posterior statistics of parameters**

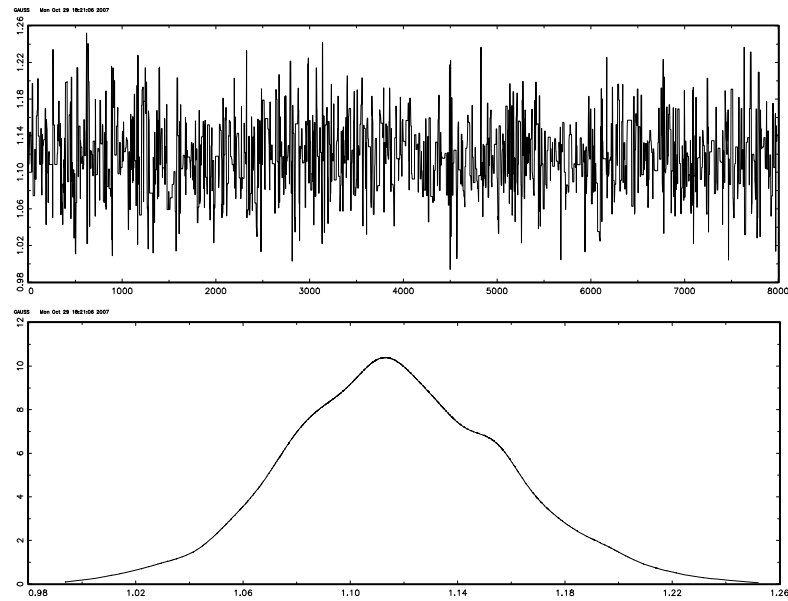
|                                  | Mean  | Std.D | A.R.* |
|----------------------------------|-------|-------|-------|
| Weibull regressions for survival |       |       |       |
| Intercept                        | 18.1  | 0.16  | N.A.  |
| Binary for high blood pressure   | -0.12 | 0.02  | N.A.  |
| Log of age                       | -2.17 | 0.04  | N.A.  |
| Number of chronic diseases       | -0.06 | 0.01  | N.A.  |
| Scale(1/p)                       | 0.52  | 0.01  | N.A.  |
| Log-normal regressions for cost  |       |       |       |
| Intercept                        | -5.81 | 0.22  | N.A.  |
| Binary for high blood pressure   | 0.26  | 0.03  | N.A.  |
| Log of age                       | 1.44  | 0.06  | N.A.  |
| Number of chronic diseases       | 0.24  | 0.02  | N.A.  |
| Scale(v)                         | 1.25  | 0.01  | N.A.  |
| Dependence Parameter             |       |       |       |
| Clayton                          | 1.12  | 0.04  | 0.16  |
| Gumbel-Hougaard                  | 1.48  | 0.02  | 0.07  |
| Frank                            | 4.24  | 0.11  | 0.39  |

\* A.R.=Acceptance rates for the Metropolis-Hastings algorithm

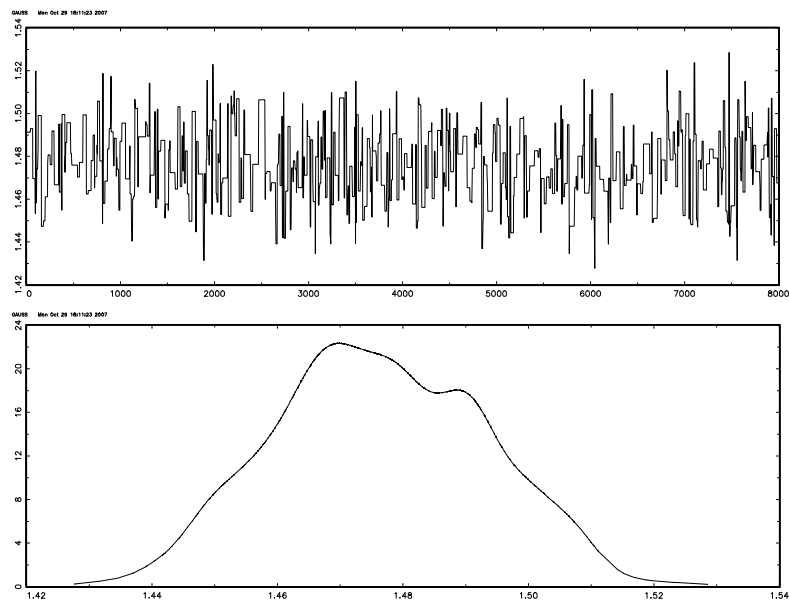
N.A.=Not applicable

**Figure 1.2: Plot of posterior draws and posterior pdf for the dependence parameter**

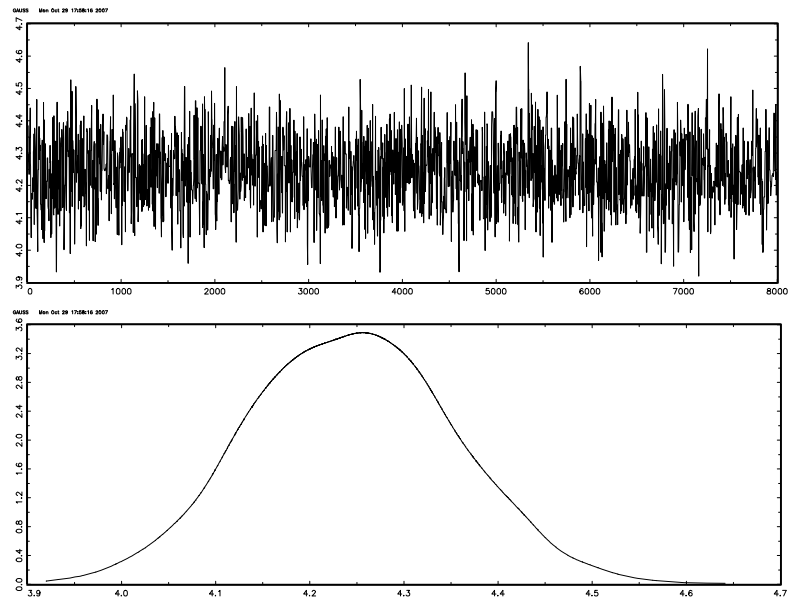
(a) Clayton family



(b) Gumbel-Hougaard family



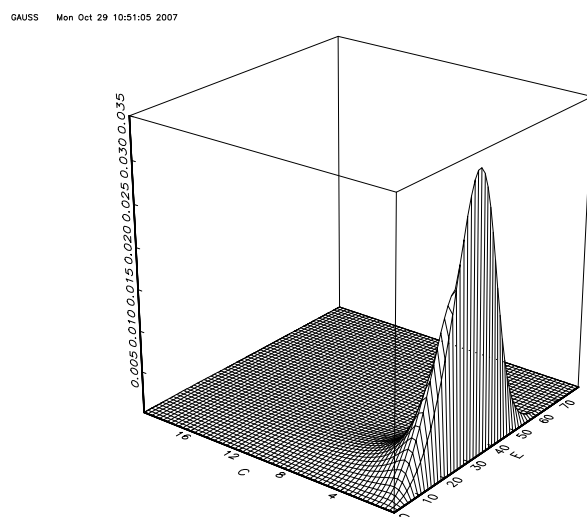
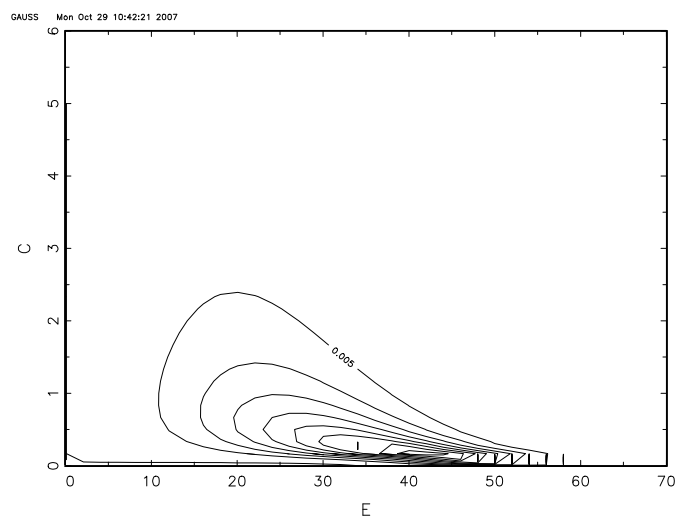
(c) Frank family



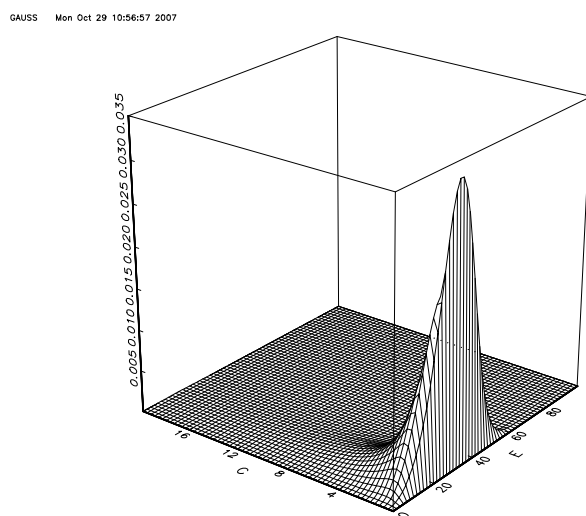
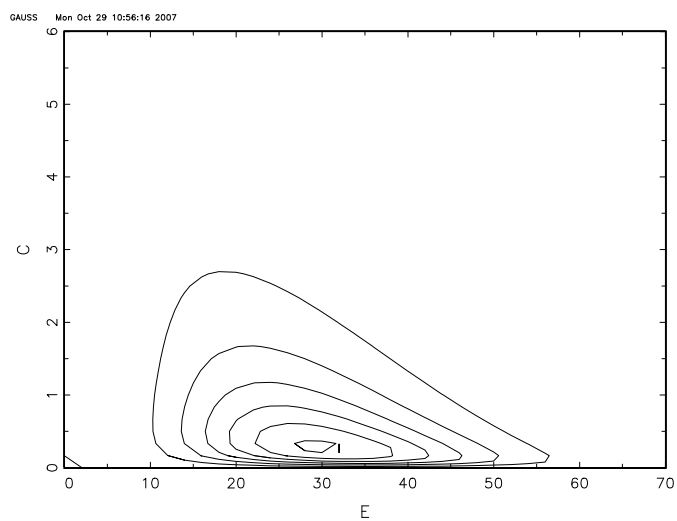
The efficient jump method was used to obtain the posterior densities for the dependence parameter. The number of Markov chains is 10,000 and the first 2,000 was burned.

**Figure 1.3: Contour and surface of joint pdf of effectiveness and cost with various copulas**

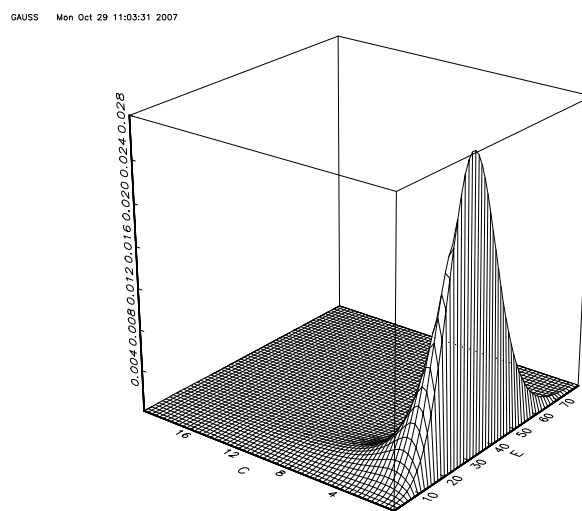
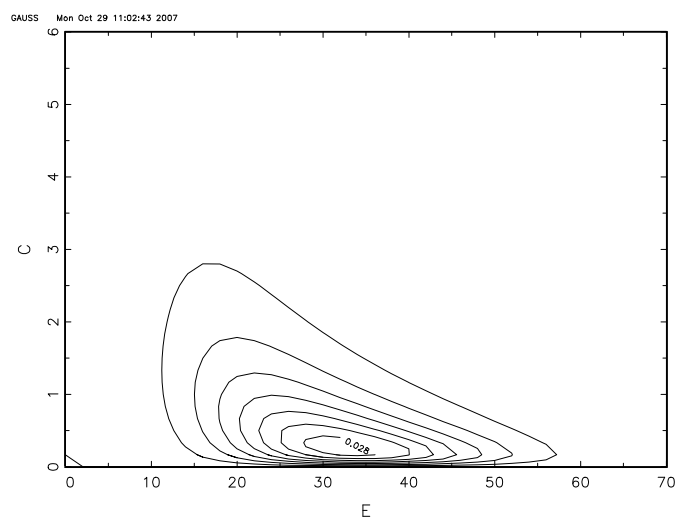
(a) Clayton family



## (b) Gumbel-Hougaard family



## (c ) Frank Family



The parameters are fixed at the posterior mean. The covariates are set assuming the following characteristics: aged 60, with no chronic conditions and high blood pressure (systolic blood pressure 140 mm Hg or more.)

**Table 1.3: Means and standard deviations for cost, effectiveness and incremental cost and effectiveness for a group of persons aged 60 and no chronic condition (Low-risk persons)**

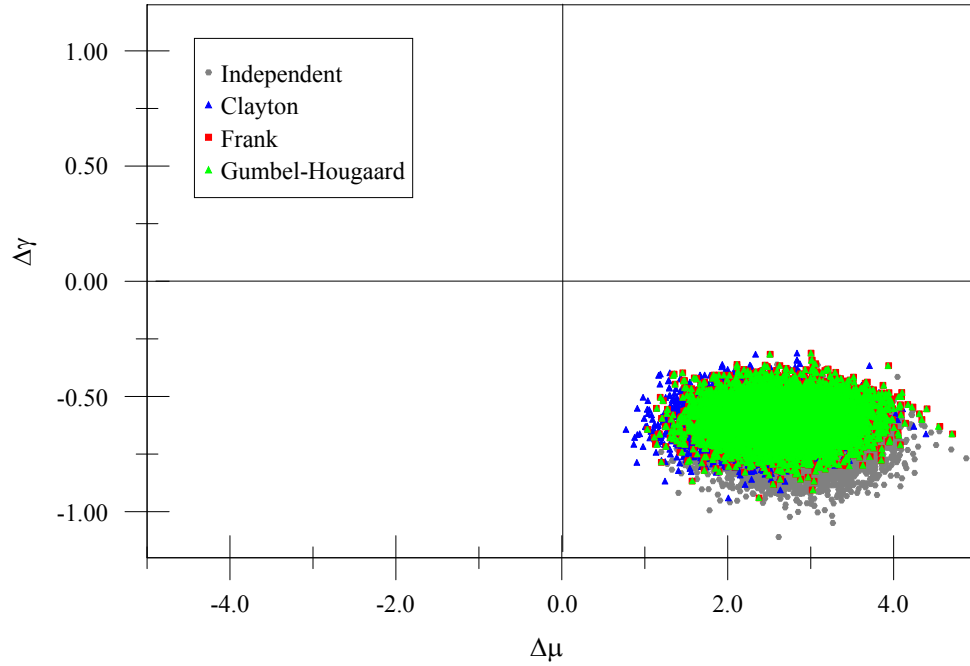
|                  |       | Treatment 1   |          | Treatment 2   |          | Difference of two treatments |          |
|------------------|-------|---------------|----------|---------------|----------|------------------------------|----------|
|                  |       | Hospital days | Survival | Hospital days | Survival | Hospital days                | Survival |
| Independent      |       |               |          |               |          |                              |          |
|                  | Mean  | 3.04          | 21.25    | 2.35          | 24.08    | -0.69                        | 2.83     |
|                  | Std.D | 0.09          | 0.38     | 0.08          | 0.51     | 0.09                         | 0.49     |
| Clayton          |       |               |          |               |          |                              |          |
|                  | Mean  | 2.89          | 20.85    | 2.3           | 23.27    | -0.6                         | 2.41     |
|                  | Std.D | 0.08          | 0.37     | 0.07          | 0.49     | 0.08                         | 0.48     |
| Gumbel-Hauggaard |       |               |          |               |          |                              |          |
|                  | Mean  | 2.9           | 21.19    | 2.3           | 23.87    | -0.6                         | 2.68     |
|                  | Std.D | 0.08          | 0.38     | 0.07          | 0.5      | 0.08                         | 0.49     |
| Frank            |       |               |          |               |          |                              |          |
|                  | Mean  | 2.89          | 21.23    | 2.3           | 23.9     | -0.59                        | 2.68     |
|                  | Std.D | 0.08          | 0.38     | 0.07          | 0.51     | 0.08                         | 0.49     |

Treatment 1 is no active treatment for high blood pressure conducted, and Treatment 2 is treatment for high blood pressure was conducted and patients' blood pressure was reduced to safe level.

We obtain the joint predictive density of  $\tilde{e}$  and  $\tilde{c}$ , for a fixed value for the vector of covariates  $\tilde{z}$  using Equation (1.11). Since the binary for high blood pressure is used as the intervention variable, it is set to be 1 for Treatment 1 and 0 for Treatment 2. As a result, each treatment arm has a different predictive density for variables  $E$  and  $C$ . Under treatment  $i$  the effectiveness is  $E_i$  and the cost is  $C_i$  for  $i = 1, 2$ . We can choose the values for other variables in  $z$  freely. We consider two sets of characteristics of a patient for  $\tilde{z}$ . The first type is the same as the previous case, and it is a person aged 60 with no chronic condition (low-risk person). In addition to this, we also consider another type of person: a person aged 60 who has 4 chronic conditions (high-risk person).



**Figure 1.4: Cost-effectiveness plane for persons aged 60 with no chronic conditions (low-risk persons)**



$\Delta\mu \equiv \mu_2 - \mu_1 = E(E_2) - E(E_1)$ , and  $\Delta\gamma \equiv \gamma_2 - \gamma_1 = E(C_2) - E(C_1)$ .  $E_1$  and  $C_1$  show effectiveness and cost under Treatment 1 (No active treatment for high blood pressure was conducted) and  $E_2$  and  $C_2$  show effectiveness and under Treatment 2 (treatment for high blood pressure was conducted). The expected values of cost and effectiveness are conditioned on the parameters.

*(1) Person aged 60 with no chronic conditions (Low-risk person)*

The plot of predictive densities of cost and effectiveness under Treatment 1 is presented in Figure 1.5. Each panel shows the result from a different type of copula. For a comparison, the plot of predictive densities from two independent marginal densities (Weibull and log-normal distributions) is also shown in the first panel of the figure. From the model assuming two independent marginals, the realized  $\tilde{e}$  and  $\tilde{c}$  are randomly scattered; however, we observe a negative dependence - longer life time is associated with less cost, and shorter life time is associated with higher cost- in the result from the

three models with copulas. The figure shows that the predictive density from the copula model successfully incorporates the negative dependence between the effectiveness and cost.

From the two sequences draws  $(E_1, C_1)$  and  $(E_2, C_2)$ , the sequence of the new variable  $(\Delta E, \Delta C)$  were calculated where  $\Delta E \equiv E_2 - E_1$  and  $\Delta C \equiv C_2 - C_1$ . The plot of  $\Delta E$  against  $\Delta C$ , the cost-effectiveness plane, is presented in Figure 1.6. The draws are most widely spread in the South-East quadrants of the plane among the four, indicating the probability that Treatment 2 dominates Treatment 1 is the highest. We observe the negative dependence between  $E$  and  $C$  here too.

The top panel of Table 1.4 shows the summary statistics. The mean of survival time under Treatment 1 ranges from 22.02 to 22.28 years while the mean survival time under Treatment 2 from 25.16 to 25.51. The highest mean cost is 3.10 and the lowest is 2.34 in Treatment 1 and the corresponding costs are 2.31 and 1.83 days respectively in Treatment 2. From the results, we conclude that the means of the incremental cost are negative and those of the incremental effectiveness are positive values.

Pearson's correlation coefficients between  $E_i$  and  $C_i$  as well as  $\Delta E$  and  $\Delta C$  are calculated. The model with copulas show moderate correlation between the two variables where the correlation coefficients range between -0.22 (Clayton family) and -0.29 (Gumbel-Hougaard and Frank families).

## *(2) Person aged 60 with 4 chronic conditions (High-risk person)*

A similar exercise was performed with the number of chronic conditions set to be a different value. The number of chronic conditions is set to be 4 instead of 0. Table

1.4(b) shows the summary statistics. The means of  $C_1$  and  $C_2$  are more than twice as large as the low-risk individual case. The mean of the simulated survival time reduces from 22 to 17 years for  $E_1$  and from 25 to 19 years. The means of  $\Delta C$  range between -1.53 and -1.30, and those of  $\Delta E$  range between 1.91 and 2.15.  $E_1$ ,  $C_1$ ,  $\Delta C$  and  $\Delta E$  are plotted in Figure 1.7. The correlation among the variables remains unchanged.

**Table 1.4: Means and standard deviations for cost, effectiveness, and incremental cost and effectiveness for two types of an individual**

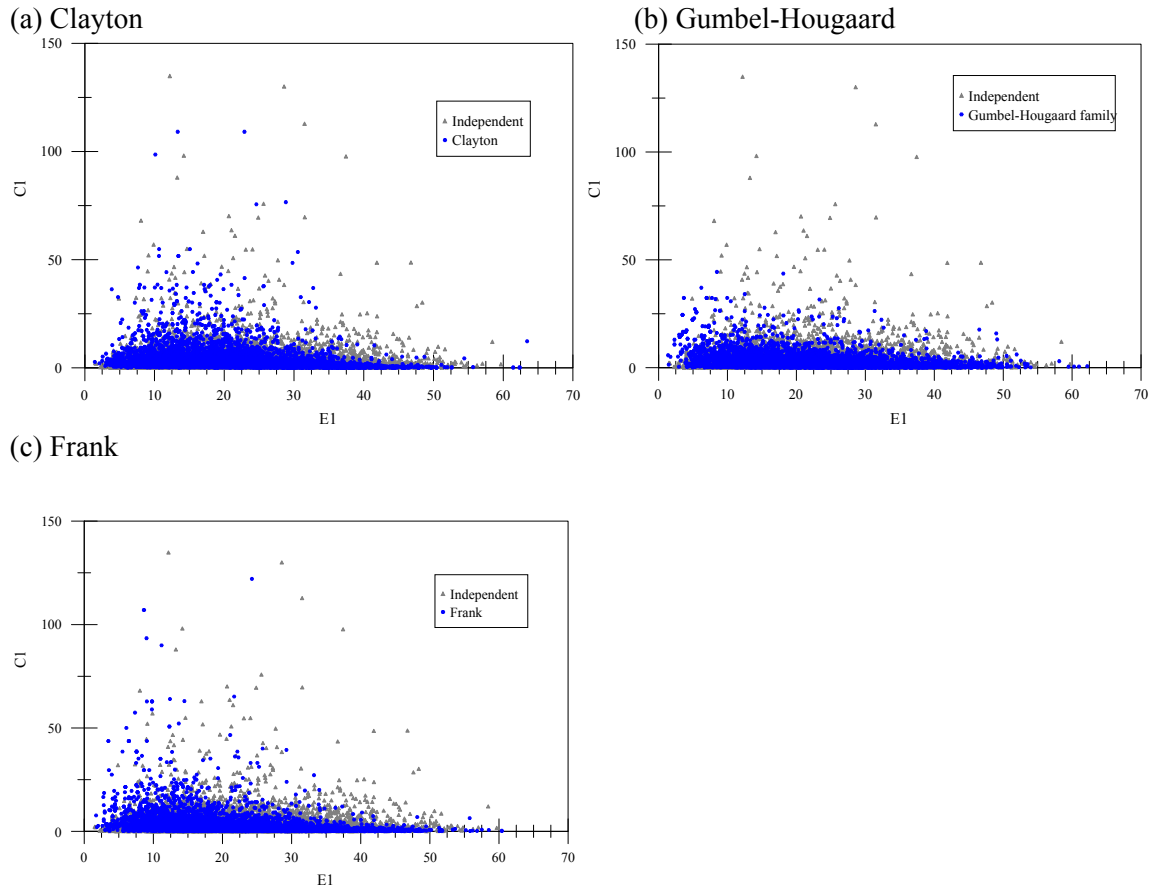
(a) Person with age 60 and no chronic condition (Low-risk person)

|                 |       | Treatment 1   |          | Treatment 2   |          | Difference of two treatments |          |
|-----------------|-------|---------------|----------|---------------|----------|------------------------------|----------|
|                 |       | Hospital days | Survival | Hospital days | Survival | Hospital days                | Survival |
| Independent     | Mean  | 3.1           | 22.16    | 2.31          | 25.16    | -0.79                        | 3        |
|                 | Std.D | 10.75         | 9.28     | 3.84          | 10.42    | 11.41                        | 13.95    |
| Clayton         | Mean  | 2.88          | 22.28    | 2.17          | 25.51    | -0.7                         | 3.22     |
|                 | Std.D | 5.18          | 9.49     | 3.88          | 10.6     | 6.48                         | 14.1     |
| Gumbel-Haugaard | Mean  | 2.34          | 22.16    | 1.83          | 25.29    | -0.51                        | 3.13     |
|                 | Std.D | 3.37          | 9.35     | 2.65          | 10.53    | 4.28                         | 14.08    |
| Frank           | Mean  | 2.76          | 22.02    | 2.11          | 25.22    | -0.65                        | 3.2      |
|                 | Std.D | 5.38          | 9.32     | 4.28          | 10.37    | 6.94                         | 14.14    |
| Independent     | Corr  | 0.01          |          | 0.02          |          | 0                            |          |
| Clayton         | Corr  | -0.23         |          | -0.23         |          | -0.22                        |          |
| Gumbel-Haugaard | Corr  | -0.29         |          | -0.29         |          | -0.29                        |          |
| Frank           | Corr  | -0.29         |          | -0.28         |          | -0.29                        |          |

(b) Person with age 60 and 4 chronic conditions (High-risk person)

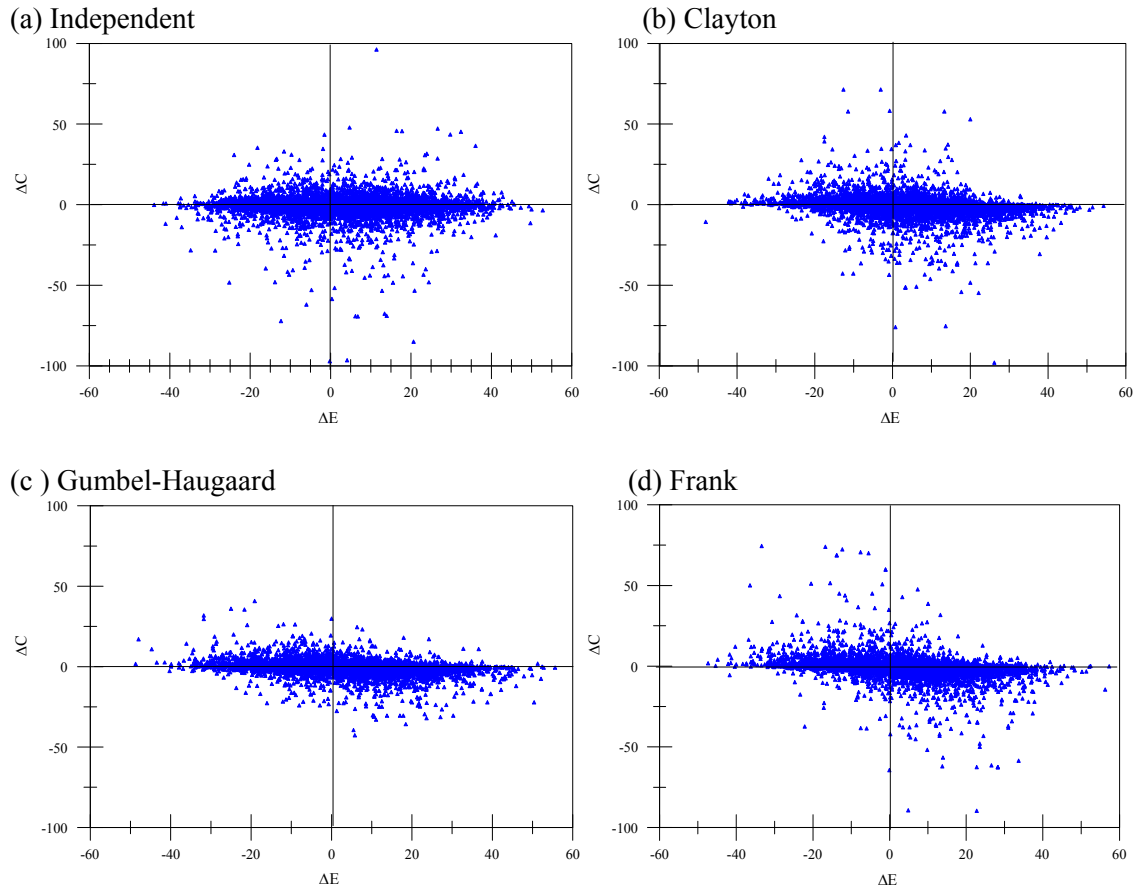
|                 |       | Treatment 1   |          | Treatment 2   |          | Difference of two treatments |          |
|-----------------|-------|---------------|----------|---------------|----------|------------------------------|----------|
|                 |       | Hospital days | Survival | Hospital days | Survival | Hospital days                | Survival |
| Independent     | Mean  | 7.85          | 17.21    | 6.32          | 19.35    | -1.53                        | 2.14     |
|                 | Std.D | 13.91         | 7.18     | 12.47         | 8.04     | 18.66                        | 10.75    |
| Clayton         | Mean  | 7.58          | 17.2     | 6.11          | 19.11    | -1.47                        | 1.91     |
|                 | Std.D | 14.54         | 7.11     | 11.56         | 7.87     | 18.51                        | 10.46    |
| Gumbel-Haugaard | Mean  | 6.19          | 17.14    | 4.89          | 19.29    | -1.3                         | 2.15     |
|                 | Std.D | 8.58          | 7.17     | 6.78          | 8.1      | 10.89                        | 10.77    |
| Frank           | Mean  | 6.76          | 17.35    | 5.25          | 19.49    | -1.51                        | 2.14     |
|                 | Std.D | 14.7          | 7.25     | 9.52          | 8.14     | 17.53                        | 10.77    |
| Independent     | Corr  | 0             |          | 0.01          |          | 0                            |          |
| Clayton         | Corr  | -0.21         |          | -0.2          |          | -0.21                        |          |
| Gumbel-Haugaard | Corr  | -0.27         |          | -0.29         |          | -0.27                        |          |
| Frank           | Corr  | -0.24         |          | -0.28         |          | -0.25                        |          |

**Figure 1.5: Posterior predictive densities of effectiveness and cost for a person aged 60 with no chronic conditions (low-risk person) under Treatment 1**



$E_1$  and  $C_1$  show effectiveness and cost under Treatment 1 (no active treatment for high blood pressure was conducted).

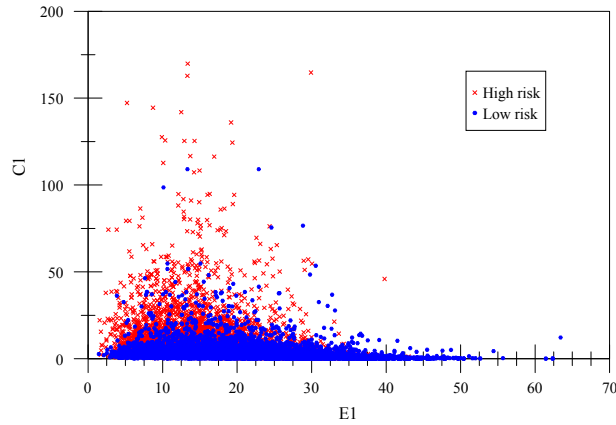
**Figure 1.6: Posterior predictive densities of incremental effectiveness and cost for a person aged 60 with no chronic conditions (low-risk person)\***



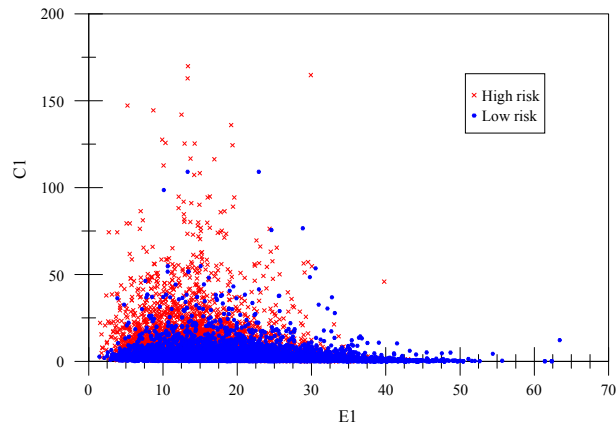
$E_1$  and  $C_1$  show effectiveness and cost under Treatment 1 (no active treatment for high blood pressure was conducted) and  $E_2$  and  $C_2$  show effectiveness and cost under Treatment 2 (treatment for high blood pressure was conducted.)  $\Delta E \equiv E_2 - E_1$  and  $\Delta C \equiv C_2 - C_1$ . The y-axis of each figure is truncated at -100 and 100 to focus on the center of the figure.

**Figure 1.7: Comparison between low-risk (number of chronic condition 0) and high-risk persons (number of chronic conditions 4)**

(a) Predictive densities of effectiveness and cost under Treatment 1



(b) Predictive densities of incremental cost and effectiveness



$E_1$  and  $C_1$  show effectiveness and cost under Treatment 1 (no active treatment for high blood pressure was conducted) and  $E_2$  and  $C_2$  show effectiveness and cost under Treatment 2 (treatment for high blood pressure was conducted.)  $\Delta E \equiv E_2 - E_1$  and  $\Delta C \equiv C_2 - C_1$ . The y-axis of figure (b) is truncated at -100 and 100 to focus on the center of the figure.

## 1.6. Discussion

This study developed a Bayesian framework of cost-effectiveness analysis based on copulas. Copulas allow us to have both flexibility of assumptions in distributions for underlying cost and effectiveness, and to incorporate dependence between the variables into the model at the same time. We used a Weibull regression for the survival function as effectiveness and a log-Normal distribution for cost, which are common distributions for survival and cost analyses.

The application to a sample from the NHANES I and NHEFS showed the existence of positive dependence between the survival and cost functions, which in turn indicated negative dependence between measures of effectiveness and cost. Two types of analysis were conducted. One is the estimation of the expected value of the differences in cost and effectiveness, and this type of analysis would be useful for policy makers. The other type of analysis is the prediction for an individual patient. By simulation using Bayesian predictive densities, it was shown that the negative dependence between cost and effectiveness affects the distribution of incremental cost and effectiveness for an individual patient. Enormous variability was found at the level of the patient.

The following extension of this paper would be considered: The first point is with respect to model selection. It is essential to compare good-of-fit among different types of copulas. For the model to be used for decision making, it is important to examine which type of copula would best fit the data. Second, it is important to discuss under what circumstance the copula method is superior over other methods such as the model using joint normal assumption.



Third, in Bayesian estimation, assumptions on prior distributions of parameters are important. Our simplest model assumed the flat prior, but more realistic assumptions should be imposed regarding priors. Finally, a more practical setting in cost-effectiveness analysis could be introduced. For example, we need to assume positive direct costs for intervention. Also, costs in cost-effectiveness analyses are usually indicated in a monetary unit. Converting the unit of cost from hospital days per year to the monetary cost of hospital stays would give us more intuitive understanding of the results.

## CHAPTER 2: BAYESIAN ESTIMATION OF DISCRETE TIME ASSET PRICING MODELS

### 2.1 Introduction

Diffusion models are widely used for pricing models in finance as well as in economics. In econometrics of finance, the main concern is in estimating the parameters of the model. Although continuous time processes are useful tools for analysis of these models, estimating the parameters in diffusion models is often difficult because data are usually available only at discrete times. To estimate parameters in diffusion models, therefore, a number of researchers have suggested various methods.

One way of estimation is to use discrete approximations of the continuous model. Chan, Karolyi, Longstaff, and Sanders (1992) (CKLS hereafter) estimates short-term risk-free interest rate models using a discrete-time specification. In their paper, the continuous-time model is given by a stochastic differential equation that implies the conditional mean and variance of changes in the short-term interest rate depend on the level of interest rates. The model is often referred as CEV (constant elasticity of variance) models, and widely nests many short-term interest rate models. The discrete-time specification is constructed so that it allows the variance of changes to depend on the level of the interest rate in a consistent manner with the continuous-time model. Brenner, Harjes and Kroner(1996) propose a discrete type short-term interest rate model, as an extension of CKLS model, that could have GARCH effects in the conditional variance of interest rate changes.

However, as Eraker(2001) pointed out, naïve discretization might bias estimates towards incorrect ones if sampling times are infrequent. The problem that arises from

using discrete-time data to estimate continuous-time models is called discretization bias, and estimates with discrete-approximation specification may be subject to discretization bias.

A key point about discretization bias is how we set the value of time change. In general, researchers set the value to be equal to one when estimating the equation, which is naive discretization. However the value should be small enough so that the discrete-time model could well approximate the continuous-time model. Therefore, Eraker suggested that the value should be an arbitrary number chosen by econometricians so that the discrete-time model could fully approximate the continuous-time model. To make it possible for us to choose a correct value, it is important to know how the choice of the value affects the estimates. In this paper, we investigate how the value of time change affects the estimates in various settings using simulated data. It is virtually impossible to create continuous-time data by simulation, and hence we will use the simulated data generated by the discrete-time model.

Before starting to analyze a discrete-time asset pricing model, I will also consider another approach to estimate continuous-time models proposed by Aït-Sahalia(1999, 2002b). He approximates transition probability functions by Hermite polynomial around the normal density and obtains an approximate likelihood functions. His approach is different from the discrete-time approach, but it is important to realize that his approach also depends on the value of time change.

As for estimation method, Bayesian methods will be used to estimate models all through this paper. A number of other econometric methods have been developed; some of them are based on the Generalized Method of Moments (CKLS, Hansen and

Scheinkman(1995)) while the Efficient Method of Moments are discussed by Gallant and Tauchen(1996), Andersen and Lund(1997). Nonparametric methods are proposed by Aït-Sahalia(1996a, b). Aït-Sahalia(1999,2002b) develops the approximation method of the maximum likelihood as mentioned earlier. From a Bayesian point of view, Eraker(2001), Durham and Gallant(2002), Durham(2002) and Jones(2003) conducted studies on estimation of continuous-time interest rate models. Jensen and Poulsen(1999) compare some of those methods in terms of multiple criteria. As widely known, MCMC methods have been proven very useful for handling diffusion models that involve the computation of a high-dimensional integral such as state-space models and the stochastic volatility models (Eraker 2001). In this sense, MCMC methods would be a good tool to handle diffusion models although our analysis in this paper is limited to a one-factor model. Lastly, a standard non-informative prior is used in all estimations in this paper.

As for the model, I will use a one-factor asset pricing model proposed by CKLS in most of my analysis in this paper. However, the Vasicek(1977) model is considered in section 2 for the purpose of analysis because the latter model has a known transition function.

The remaining of this paper consists of the following: In section 2, the model we use in this paper is introduced and two approaches for estimation are discussed. In section 3, one of the approaches, the Aït-Sahalia(1999, 2002b) approach, is discussed in terms of the importance of the value of time change through some numerical examples. In section 4, numerical examples with the discrete-time model are shown in various settings: knowledge of the data generating process, different frequency of data, and use of averaged data. In section 4.1, we discuss when the data generation process is known to

us. In section 4.2, a more realistic case, when we do not know the data generation process, is dealt with. In section 4.3, Eraker's data augmentation algorithm is introduced and we see how it works. In section 4.4, we discuss estimation with averaged data. In section 5, we show some empirical results with daily data of U.S. effective Federal Funds rate. Section 6 contains concluding remarks.

## 2.2 Continuous Asset Pricing Models

There are mathematics of finance and econometrics of finance. In the field of mathematical finance a stochastic differential equation such as

$$dY_t = \mu(Y_t; \theta)dt + \sigma(Y_t; \theta)dW_t \quad (2.1)$$

is specified to model the dynamics of an asset price,  $Y_t$ . In equation (2.1),  $W_t$  is a Wiener or Brownian motion, and the drift  $\mu : R^d \times \Theta \rightarrow R^d$  and diffusion  $\sigma : R^d \times \Theta \rightarrow R^{d \times d}$  are known functions. The problem is then to solve for  $Y_t$  under some market conditions such as arbitrage relations.

In the field of econometrics of finance, on the other hand, the problem is to estimate unknown parameters  $\theta$  in equation (2.1), using data that are taken at discrete time  $t$ . The discrete time data are assumed to be generated according to equation (2.1).

It is obvious that equation (2.1) cannot be used to form a likelihood function even if we know the distribution of  $W_t$  since it is impossible to obtain data on  $dY_t$ . How can we reconcile this dilemma? There are two ways to solve it:

1. Given the interval  $h$  at which the discrete data is collected, we form an approximate likelihood function using (2.1).
2. We use a discrete-time model.

The first approach is followed by Aït-Sahalia(1999, 2002b). His approach is based on an idea of expanding a density function of  $Y_t$  by using Hermite polynomials. The idea of Hermite expansion goes back to Gramer (1925), but Aït-Sahalia(2002a, b) introduced it to the area of finance. Since a continuous-time diffusion is a Markov process, the log-likelihood function is expressed by the following equation:

$$l_n(\theta) \equiv n^{-1} \sum_{i=1}^n \ln \{p_Y(h, Y_{ih} | Y_{(i-1)h}; \theta)\} . \quad (2.2)$$

Conditional on  $h$ , his method gives us the closed-form expression of  $p_Y$ , and therefore the likelihood function  $l_n$ .

The second approach proceeds as follows. The discrete-time model version of equation (2.1) is

$$\Delta Y_n = \mu(Y_n; \theta)h + \sigma(Y_n; \theta)\Delta W_n \quad (2.3).$$

$\Delta W_n$  is a  $d$ -dimensional iid  $N(0, hI_d)$  random vector, where  $I_d$  denotes a  $d$ -dimensional identity matrix, and  $h$  is the unit of time at which data is recorded. The subscript  $n$  runs from 0,1 to  $N$ . Let us introduce another index  $T$  and  $m$  such that  $N = T/h = T \cdot m$ . As  $h \rightarrow 0$ , equation (2.3) approaches equation (2.1) under some regularity conditions<sup>4</sup>.

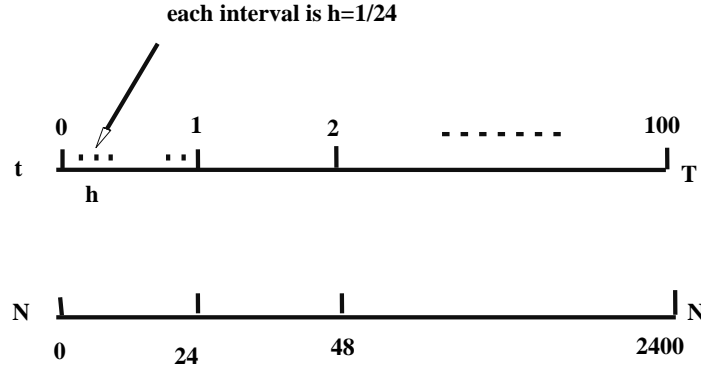
The discrete time model approach has been commonly used in econometrics of finance. As long as we know  $h = 1/m$  for  $h$  to be small enough, and use the data from  $n = 1, \dots, N$ , there is no problem of *discretization bias*<sup>5</sup>. Discretization bias arises if we

---

<sup>4</sup> We should remind ourselves that in general the solutions of difference equation (2.3) and that of the differential equation(2.1) are different, since differentiability in time is lost in the discrete time approach, e.g. see Kariya and Liu (2003, p4).

<sup>5</sup> Generally the phrase *discretization bias* refers to the bias that occurs approximating  $dY_t$  in equation(2.1) by  $\Delta Y_t$  in equation (2.3). However, unless  $\Delta \rightarrow 0$ , the bias will not disappear. In the discrete-time model of (2.3), a bias occurs if we do not know  $m$ . In this paper, we call this the discretization bias.

have data at  $t = 1, \dots, T$  and we do not know  $h$ . The relationship between  $h, T$  and  $N$  is explained in Figure 2.1.



**Figure 2.1: Relationship between  $t$ ,  $h$ , and  $N$**

In Figure 2.1, suppose that  $t$  is a day (i.e. 24 hours or  $24 \times 60$  minutes, and we have a sample size of 100 ( $T = 100$ ) i.e. 100 daily data. If the data is generated every hour, then  $h = 1/24$ , and if it is generated every minute, then  $h = 1/(24 \times 60)$ .

As for the specification of functions  $\mu(Y_n; \theta)$  and  $\sigma(Y_n; \theta)$ , there has been a number of research on the specification of these functions. The most general specification of one-factor models was proposed by CKLS (1992):

$$dY_t = (\theta_1 + \theta_2 Y_t)dt + \sigma |Y_t|^\beta dW_t \quad (2.4)$$

This model becomes the Vasicek(1977) model with  $\beta = 0$ . If  $\beta = 1$ , it is the Brennan and Schwartz(1980) model. With  $\beta = 0.5$ , it is the short-term interest rate model studied by Cox, Ingersoll and Ross (1985). A summary of these one-factor models is presented in CKLS (1992). Here  $\beta$  measures the elasticity of volatility with respect to the level of the interest rate, and  $\theta_2$  captures the mean-reversion effect.

The discrete time specification of the CKLS model (2.4) is given by the following equation:

$$\Delta Y_t = (\theta_1 + \theta_2 Y_{t-1})h + \sigma |Y_{t-1}|^\beta \sqrt{h} u_t, u_t \sim N(0,1) \quad (2.5).$$

### 2.3 Numerical Examples with the Transition Function

In this section, some numerical examples with Aït-Sahalia (1999, 2002b)'s method are demonstrated. The motivation is to reveal the relationship between  $h$  and the estimates of coefficients. His method is applied for several models of short-term interest rates in his paper (1999), but I only use the Vasicek(1977) model because it has a known transition density of  $p_Y$ . This makes it possible to compare the approximated transition function of  $p_Y$  to the exact  $p_Y$ .

The model we consider in this section is the Ornstein-Uhlenbeck specification by Vasicek(1977) for the short-term interest rate

$$dY_t = \kappa(\alpha - Y_t)dt + \sigma dW_t. \quad (2.6)$$

This model is a special case of the CKLS model shown by equation (2.4).

The Gaussian transition density of  $y$  is

$$p_Y(h, y | y_0; \theta) = (\pi\gamma^2 / \kappa)^{(-1/2)} \exp\left\{-\frac{(y - \alpha - (y_0 - \alpha)e^{-\kappa h})^2 \kappa}{\gamma^2}\right\} \quad (2.7)$$

where  $\theta \equiv (\alpha, \kappa, \sigma)$  and  $\gamma^2 \equiv \sigma^2(1 - e^{-2\kappa h})$ .  $h$  is a given small number that shows the sampling interval. The detail of the derivation can be found in Appendix B.

In this simplest case we form the likelihood function using the closed-form expression (2.7), given the initial value  $y_0$  and  $\theta$ . Our concern is how to set the value of  $h$  in the equation. Since the transition density function is a function of  $h$ , the estimates



depend on the value of  $h$ . In the study,  $h$  is set to be  $1/12$  for a monthly sampling frequency and  $1/52$  for a weekly sampling frequency, normalizing a year to be 1.

To show the role of  $h$  in estimation, we conduct the following analysis: first, we generate the data of  $Y$  with equation (2.7) with  $h$  set to be a particular value,  $\alpha = 1$ ,  $\kappa = .3$ , and  $\sigma = 0.7791$ <sup>6</sup>. Here,  $h$  is set to be  $1/24$  so that the interpretation of  $h$  should be consistent with other analyses in this paper. Second, we estimate the coefficients  $\kappa$ ,  $\alpha$  and  $\sigma$  with  $h = 1$  (wrong value) as well as  $h = 1/24$  (true value), and compare between the results. To estimate the coefficients, Markov Chain Monte Carlo is used. In particular, Gibbs-Sampling method can be used because we know the correct distributions of the parameters. The detail of Gibbs-Sampling method with the transition density is explained in Appendix C.

The estimation results are presented in Table 2.1. There are a couple of remarks. First, there is no influence of  $h$  in the estimation of  $\alpha$ . This is simply because the coefficient  $\alpha$  is not a function of  $h$  as can be seen in Appendix C. Secondly, the estimates of  $\kappa$  and  $\sigma$  are influenced by how you set the value of  $h$ . That is, if we know the correct value of  $h$ , we obtain the almost correct results. However, if we miss-specify the value of  $h$ , the estimates are far from the true values. The reason is the same as the earlier case, and we need a correct value of  $h$  to calculate that parameters from estimates. Aït-Sahalia's approach is explained in Appdenix D.

---

<sup>6</sup> The value of  $\sigma$  is taken from the empirical study of short-term interest rate in Aït-Sahalia (1999).  $\kappa$  and  $\alpha$  are given arbitrary numbers by the author because the values in his study are closed to zero and it is difficult to see clear results.

**Table 2.1: Posterior statistics for parameters with different values of  $h$  with Vasicek model: Using the transition function**

| Posterior Statistics |          | True h<br>h=1/24 | Wrong h<br>h=1 |
|----------------------|----------|------------------|----------------|
| Alpha                | mean     | 0.7254           | 0.7254         |
|                      | variance | 0.858            | 0.858          |
| Kappa                | mean     | 0.4              | 0.0167         |
|                      | variance | 0.1166           | 0.0049         |
| Sigma                | mean     | 0.7536           | 0.1538         |
|                      | variance | 0.0018           | 0.0004         |

The true value for parameters are;  $\alpha = 1$ ,  $\kappa = 0.3$ , and  $\sigma = 0.7791$ . Data is generated with  $h = 1/24$ . The same value is used for true  $h$  to estimate the parameters, and 1 is used for wrong  $h$ . For all parameters, Gibbs sampling based on the transition function is used.

## 2.4 Numerical Examples Using the CKLS Model

### 2.4.1 When $m$ Is Known

In the previous section, we saw that setting  $h = 1$  causes problems in estimation and we needed to know a correct value for  $h$ .

In this section, we will show some numerical examples with the discrete-time CKLS model<sup>7</sup>. As mentioned earlier, our concern is how to set  $h$  when estimating equation (2.5). We obtain the posterior probability distributions of the parameters in equation (2.5) in various settings.

Our approach is as follows: First, We will generate data by equation (2.5), estimate parameters with the data, and see how the simulation works. To avoid discretization bias, it is important to choose an appropriate value for  $h$ . Therefore, we will start our analysis assuming that  $m$ , which is defined by  $1/h$ , is *known* in subsection 4.1 as a benchmark case, and deal with more realistic case, where  $m$  is *unknown*, in the next subsection. Next, we will introduce Eraker's data augmentation method. Lastly, we will use averaged data instead of point-counted data.

#### 2.4.4.1 Using all data

We generate data from equation (2.5) with various values of  $T$  and  $h = 1/m$ .

These generated data sets will approach to continuous data as  $m$  gets large, or in this

---

<sup>7</sup> A numerical example in Bayesian inferences is equivalent to a Monte Carlo sampling experiment in frequentists' inference. In Bayesian inference, a sample is generated given the parameters and the error terms. Then given the sample, posterior probability densities (pdf's) of the parameters are drawn by using Markov Chain Monte Carlo algorithms. In frequentists' sampling experiments, a sample is drawn given the parameters and error terms, and point estimates of the parameters are obtained. Then another sample is drawn to obtain other point estimates. Repeating sample draws, we obtain the frequencies (empirical distributions) of the point estimates.

case equivalently  $h$  gets small. For example, when  $m$  is 24 and  $T$  is 100, the data stands for hourly data, and we have all observations which consists of  $N = 24 \times 100$ .

We generate data using (2.5) with

$$\theta_1 = 0.007393, \theta_2 = -0.0876, \sigma = 0.7791 \text{ and } \beta = 1.48$$

and  $T$  is given by 100 and 500. The values of the coefficients are from the empirical study of short-term interest rates in Aït-Sahalia (1999). Since we use all the data, and  $m$  is known, the number of observations we have is  $N = T \times m$ . Two values of  $m$  are chosen: 1 and 24. The values correspond to the cases data is sampled daily and hourly respectively. We will try to obtain the posterior means of parameters using the generated data so that we could see whether the simulation works well. We will assume that the value of  $h$  (or equivalently  $m$  here) is known all through this subsection.

We use Markov Chain Monte Carlo (MCMC) algorithms to draw the posterior pdfs of the parameters. The MCMC algorithms are given in appendix E. In MCMC, we obtain 6000 MCMC draws and discard the first 1000 draws. To save space, rather than presenting the graphs of the posterior pdf's, we present the posterior summary statistics in Table 2.2. Panel A) reports the result with Metropolis-Hastings algorithm and panel B) with Gibbs-sampling for  $\theta$ , which is a vector of coefficients.

In Table 2.2, the posterior probability densities (pdf) of the parameters are, in general, well centered around the true values. These well-behaved results come from the fact that all data generated are available and know  $h$  (or equivalently  $m$ , in this case). If we have all data in our hand and know the value of  $h$ , we can obtain posterior distributions of the parameters that are centered around the true parameter values. The value of  $h$  does not change the result. *So long as we know  $h$  and we use all the*

**Table 2.2: Posterior statistics for parameters with various values for  $m$  : Using discrete-time approximation**

A) Theta with Metropolis-Hastings algorithm

|                      |          | T=100  |        | T=500  |        |
|----------------------|----------|--------|--------|--------|--------|
| Posterior statistics |          | m=1    | m=24   | m=1    | m=24   |
| Sample size          |          | 100    | 2400   | 100    | 2400   |
| Theta1               | mean     | 0.010  | 0.009  | 0.007  | 0.009  |
|                      | variance | 0.005  | 0.005  | 19.438 | 0.002  |
| Theta2               | mean     | -0.100 | -0.094 | -0.074 | -0.090 |
|                      | variance | 0.063  | 0.080  | 0.009  | 0.036  |
| Sigma                | mean     | 0.228  | 0.762  | 0.843  | 0.785  |
|                      | variance | 0.068  | 0.045  | 0.096  | 0.024  |
| Beta                 | mean     | 1.009  | 1.477  | 1.511  | 1.485  |
|                      | variance | 0.112  | 0.024  | 0.041  | 0.012  |
| Acceptance rate      | Theta    | 0.151  | 0.075  | 0.161  | 0.043  |
|                      | Beta     | 0.152  | 0.077  | 0.171  | 0.040  |

B) Theta with Gibbs-sampling algorithm

|                      |          | T=100  |        | T=500  |        |
|----------------------|----------|--------|--------|--------|--------|
| Posterior statistics |          | m=1    | m=24   | m=1    | m=24   |
| Sample size          |          | 100    | 2400   | 100    | 2400   |
| Theta1               | mean     | 0.007  | 0.012  | 0.007  | 0.009  |
|                      | variance | 0.003  | 0.082  | 0.003  | 0.034  |
| Theta2               | mean     | -0.075 | -0.110 | -0.075 | -0.089 |
|                      | variance | 0.010  | 0.026  | 0.010  | 0.011  |
| Sigma                | mean     | 0.768  | 0.767  | 0.768  | 0.775  |
|                      | variance | 0.025  | 0.011  | 0.025  | 0.005  |
| Beta                 | mean     | 1.487  | 1.478  | 1.487  | 1.479  |
|                      | variance | 0.015  | 0.008  | 0.015  | 0.003  |
| Acceptance rate      | Theta    | 1.000  | 1.000  | 1.000  | 1.000  |
|                      | Beta     | 0.012  | 0.004  | 0.012  | 0.002  |

The true values for parameters are;  $\theta_1 = 0.007393$ ,  $\theta_2 = -0.0876$ ,  $\sigma = 0.7791$  and  $\beta = 1.48$ .  $T$  is the number of observations.  $m$  shows how many data are in between each  $T$ . Data frequency is assumed to be daily when  $m = 1$ , hourly when  $m = 24$ . The last two rows report the acceptance rates of Metropolis-Hastings algorithm for  $\theta$  and  $\beta$ . In panel A), Metropolis-Hastings algorithm for  $\theta$ , and in panel B), Gibbs-sampling is used. As for  $\sigma$ , Gibbs-sampling is used in both cases, and therefore acceptance rate is 1.

*observations that are generated and sampled at every  $h$ -th interval we have no problem of discretization bias.*

#### **2.4.4.2 Using every $r$ -th observation**

The second case is where we do not use all data that are generated. Here, we remove the first assumption that we have all the observations at hand, and see how the result will be affected.

Let us assume that every  $r$ -th observation is available. That means that we could use limited frequency of data out of all generated data. For example, data is generated every hour but we use the subset that consists of observations at every two hours to estimate parameters.

The way to construct the new data set is explained as follows; first, we generate data in the same way as we do in the previous subsection.  $m$  is set to 24, and the number of observations,  $T$ , is set to 100. That is, data are *sampled* hourly. This data set is called the original data set, and it has  $N = T \times m$  data points in it. Next, we sample every  $r$ -th draw from the original data set we have obtained. The sample size now becomes  $(T \times m) / r$ . How can we interpret this new data set? Let us take an example of  $r = 2$ . As mentioned earlier, the original data set is assume to be hourly, so we are supposed to have data every two hours in our hand when  $r = 2$ . In the same way, when  $r = 24$ , we would think that only daily data are available.

We have to distinguish between *generated* and *sampled* data. The data is always generated at every  $1/m$  interval but it is assumed to be sampled at  $r$ -th interval. We always know what  $r$  is since it shows the data frequency once  $m$  is fixed.

To obtain the posterior statistics for parameters, we need to specify the value of  $h$  in equation (2.5). When we *know* that the data is generated at every  $1/m$  interval (*i.e.* we know  $m$ ) and we only have data that are sampled at every  $r$ -th interval, an appropriate  $h$  is equal to  $(1/m) \times r$ . For example, we know that the data are generated at every hour (*i.e.*  $1/24$ ) and the data we have at hand is sampled every two hours ( $r=2$ ). In this case, the value for  $h$  should be  $(1/24) \times 2$ .

Table 2.3 shows the results. The choices of the number of  $r$  are 1, 2, 4, 8, 10 and 24. Clearly, the first column ( $r = 1$ ) of Table 2.3 corresponds to the second column of panel (A) of Table 2.2, which is the case when all data are available. The MCMC algorithms used for  $\sigma$  and  $\beta$  are same as before, and Metropolis-Hastings algorithm is used for  $\theta_1$  and  $\theta_2$ .

When  $r$  is small, we obtain almost exactly the same posterior means for all parameters as those when all data are available. However, the posterior means for parameters are less close to those when all data are available when  $r$  is large. This is especially true for parameter  $\beta$  and  $\sigma$ , and we see that the posterior means is 1.207 when  $r = 24$  for  $\beta$  although the true value is 1.48. For  $\sigma$ , the posterior mean is 0.404 when  $r = 24$  while the true value is 0.7791. This could be explained in terms of the error of approximation of equation (2.5) from equation (2.4). Equation (2.5) will be a good approximation of equation (2.4) only when  $h$  is small. But when  $r$  is a large number,  $h$  is not small enough because  $h$  is defined as  $(1/m) \times r$ . For example, let us think about the case of  $r = 24$ . Then the value of  $h$  we used is one, which comes from  $(1/24) \times 24$ , while the data is generated with  $h = 1/24$ . In this case, the results would be biased even if we use an appropriate value for  $h$ .

**Table 2.3: Posterior statistics for parameters after sampling every  $r$ -th interval when  $m$  is known**

|                 |          | Posterior Statistics | r=1    | r=2    | r=4    | r=8    | r=10   | r=24   |
|-----------------|----------|----------------------|--------|--------|--------|--------|--------|--------|
| Sample size     |          |                      | 2400   | 1200   | 600    | 300    | 240    | 100    |
| Theta1          | mean     |                      | 0.009  | 0.009  | 0.010  | 0.010  | 0.009  | 0.010  |
|                 | variance |                      | 0.005  | 0.005  | 0.005  | 0.005  | 0.005  | 0.005  |
| Theta2          | mean     |                      | -0.094 | -0.093 | -0.108 | -0.105 | -0.095 | -0.107 |
|                 | variance |                      | 0.080  | 0.082  | 0.082  | 0.086  | 0.072  | 0.081  |
| Sigma           | mean     |                      | 0.762  | 0.691  | 0.701  | 0.613  | 0.506  | 0.404  |
|                 | variance |                      | 0.045  | 0.067  | 0.077  | 0.090  | 0.091  | 0.110  |
| Beta            | mean     |                      | 1.477  | 1.432  | 1.438  | 1.382  | 1.318  | 1.207  |
|                 | variance |                      | 0.024  | 0.036  | 0.042  | 0.056  | 0.067  | 0.095  |
| Acceptance rate | theta    |                      | 0.075  | 0.118  | 0.156  | 0.185  | 0.188  | 0.221  |
|                 | beta     |                      | 0.077  | 0.121  | 0.154  | 0.184  | 0.189  | 0.215  |

The true values for parameters are;  $\theta_1 = 0.007393$ ,  $\theta_2 = -0.0876$ ,  $\sigma = 0.7791$  and  $\beta = 1.48$ . The statistics are obtained with  $m = 24$  and  $T = 100$ .  $r$  is the number of interval of draw.  $(1/m) \times r$  is used as  $h$ . The last two rows report the acceptance rates of Metropolis-Hastings algorithm for  $\theta$  and  $\beta$ . For  $\sigma$ , Gibbs-sampling is used, and acceptance rate is 1.

## 2.4.2 When $m$ Is Unknown

As we saw in the previous subsection, data discretization bias can be avoided to some extent by using appropriate values for  $h$  as long as we have knowledge about how the data is generated. However, this assumption does not often hold for the data we usually use. We usually do not know at what interval the data is generated. The only thing we know is that the data we have are *sampled (not generated)* every  $r$ . For example, let us assume that we have data that is *sampled* every 2 hours but we do not know the fact that the data is *generated* every hour. In such a case, in the literature, it is the common practice to set  $h = 1$ . This is called the *naïve discretization*.



**Table 2.4: Posterior statistics for parameters after sampling every  $r$ -th interval when  $m$  is unknown**

|                 |          | Posterior | r=1    | r=2    | r=4    | r=8    | r=10   | r=24   |
|-----------------|----------|-----------|--------|--------|--------|--------|--------|--------|
| Statistics      |          |           |        |        |        |        |        |        |
| Sample size     |          |           | 2400   | 1200   | 600    | 300    | 240    | 100    |
| Theta1          | mean     |           | 0.000  | 0.001  | 0.002  | 0.003  | 0.004  | 0.010  |
|                 | variance |           | 0.000  | 0.000  | 0.001  | 0.002  | 0.002  | 0.005  |
| Theta2          | mean     |           | -0.005 | -0.009 | -0.017 | -0.035 | -0.038 | -0.107 |
|                 | variance |           | 0.004  | 0.007  | 0.013  | 0.028  | 0.032  | 0.081  |
| Sigma           | mean     |           | 0.148  | 0.194  | 0.292  | 0.367  | 0.306  | 0.404  |
|                 | variance |           | 0.008  | 0.014  | 0.033  | 0.055  | 0.066  | 0.110  |
| Beta            | mean     |           | 1.457  | 1.422  | 1.445  | 1.396  | 1.290  | 1.207  |
|                 | variance |           | 0.021  | 0.029  | 0.044  | 0.059  | 0.087  | 0.095  |
| Acceptance rate | theta    |           | 0.036  | 0.063  | 0.105  | 0.151  | 0.148  | 0.221  |
|                 | beta     |           | 0.039  | 0.066  | 0.102  | 0.152  | 0.145  | 0.215  |

The true value for parameters are;  $\theta_1 = 0.007393$ ,  $\theta_2 = -0.0876$ ,  $\sigma = 0.7791$  and  $\beta = 1.48$ . The statistics are obtained with  $m = 24$  and  $T = 100$ .  $r$  is the number of interval of draw. Since  $m$  is unknown, 1 is used as  $h$ . The last two rows report the acceptance rates of Metropolis-Hastings algorithm for  $\theta$  and  $\beta$ . For  $\sigma$ , Gibbs-sampling is used, and acceptance rate is 1.

The result is shown in Table 2.4. The posterior pdfs are not well-centered around the true value any more. This is because we do not use a proper value for  $h$ . This means, if we do not know  $m$ , the results would be biased.

From the argument in sections 2.4.1 and 2.4.2, I would conclude that we need two things to obtain the right estimates for the parameters; (1) the right knowledge on both  $m$  and  $r$ , and (2) small value for  $h$ . The results may not be correct if you do not have both (1) and (2).

### 2.4.3 Eraker's Data Augmentation

In section 2.4.1, we generated data and estimated parameters assuming that we know what  $m$  is. We found that as long as we know  $m$ , and all data are used for estimation (in our case, it also guarantees the small value of  $h$ ), then we could obtain correct results. In section 2.4.2, we assumed we do not know about  $m$ . In this case, we could not obtain good results anymore. In this section, we will introduce Eraker's data augmentation algorithm, and augment the latent data that we sampled in section 2.4.1 and 2.4.2, and see how it works.

To obtain posterior probability density functions for parameters, again we need to specify  $h$  in equation (2.5). In naïve discretization case, we usually set  $h = 1$  and we would have data discretization bias. Now we will augment the missing data which is sampled every  $r$ -th draw, using Eraker's data augmentation algorithm. The number of augmentation is fixed to 2, and the number of  $r$  is changed with  $r$  set to be 2, 4, 10 and 24. When  $r = 2$ , for example, we can recover the same sample size of data set as the original one, because we sample every two draw, and augment the same number of data. Eraker's data augmentation algorithm we use is explained in Appendix F.

The posterior summary statistics are presented in Table 2.5(a). The estimation results become better than those before augmentation in the sense that it is closer to the true values except for  $\beta$ . This is because we have increased the number between data. The sample size is now double, so it gets closer by the same amount. However, you can find that they are still far from the true values. If we have no idea on  $m$ , it is not possible to reach the true values even we augment data.

So far we have used  $h = 1$ . If we know  $m$ , then we may use a correct value for  $h$ . The correct value for  $h$  is  $(1/m) \times r$  in section 4.3.2, assuming that we know  $m$ . Now we augment one data between the observations, so the right value for  $h$  should be  $(1/m) \times r \times (1/2)$  here. Let us think about the case of  $r = 2$ . In this case, we have a new data set which consists of exactly the same amount of data. The correct value for  $h$  should be  $(1/m) \times 2 \times (1/2) = 1/m$  in this case.

The results are shown in Table 2.5(b). Here, we can see that the posterior means for each parameter never attain the true values even if we use the correct value for  $h$  and modify the data discretization bias.

**Table 2.5 (a) : Posterior statistics for parameters with Eraker's data augmentation : naïve discretization**

| Posterior Statistics |          | r=2    | r=4    | r=10   | r=24   |
|----------------------|----------|--------|--------|--------|--------|
| Sample size          |          | 2400   | 1200   | 480    | 200    |
| Theta1               | mean     | 0.059  | 0.054  | 0.061  | 0.064  |
|                      | variance | 0.001  | 0.016  | 0.002  | 0.004  |
| Theta2               | mean     | -1.000 | -0.906 | -1.000 | -1.000 |
|                      | variance | 0.000  | 0.257  | 0.000  | 0.000  |
| Sigma                | mean     | 0.664  | 2.893  | 0.676  | 0.682  |
|                      | variance | 0.029  | 35.522 | 0.073  | 0.129  |
| Beta                 | mean     | -1.181 | -0.751 | -1.175 | -1.161 |
|                      | variance | 0.014  | 0.994  | 0.031  | 0.046  |

The true value for parameters are;  $\theta_1 = 0.007393$ ,  $\theta_2 = -0.0876$ ,  $\sigma = 0.7791$  and  $\beta = 1.48$ . The statistics are obtained with  $m = 24$  and  $T = 100$ .  $r$  is the number of interval of draw. 1 is used for. The data is augmented by Eraker's data augmentation algorithm, which is explained in Appendix F.

**Table 2.5(b) : Posterior statistics for parameters with Eraker's data augmentation : exact value for  $h$**

| Posterior Statistics |          | r=2    | r=4    | r=10   | r=24   |
|----------------------|----------|--------|--------|--------|--------|
| Sample size          |          | 2400   | 1200   | 480    | 200    |
| Theta1               | mean     | 0.004  | 0.158  | 0.133  | 0.090  |
|                      | variance | 0.002  | 0.087  | 0.005  | 0.006  |
| Theta2               | mean     | -0.047 | -1.681 | -2.191 | -1.414 |
|                      | variance | 0.025  | 0.924  | 0.000  | 0.000  |
| Sigma                | mean     | 0.845  | 0.667  | 0.667  | 0.685  |
|                      | variance | 0.043  | 0.327  | 0.069  | 0.138  |
| Beta                 | mean     | 1.395  | 0.531  | -1.196 | -1.173 |
|                      | variance | 0.068  | 0.479  | 0.034  | 0.051  |

The true value for parameters are;  $\theta_1 = 0.007393$ ,  $\theta_2 = -0.0876$ ,  $\sigma = 0.7791$  and  $\beta = 1.48$ . The statistics are obtained with  $m = 24$  and  $T = 100$ .  $r$  is the number of interval of draw.  $(1/m) \times r \times (1/2)$  is used for  $h$ . The data is augmented by Eraker's data augmentation algorithm, which is explained in Appendix F.

#### 2.4.4 When Averaged Data is Used

In this subsection, we look at a different type of data: data are not sampled at a point in time, but are averaged over some interval. Data are often given by the averaged value in some interval, say averaged daily, weekly or monthly. In such a case, the data are far from continuous any more, and therefore we expect that  $h$  plays totally a different role in equation (2.5).

To investigate this point, we now draw a data set which consists of the averaged data. The new data set is constructed by the following: First, data is generated with  $m = 24$  and  $T = 100$ . Second, the data is averaged every  $m$ -th, which means the averaged data is assumed to be daily averaged data. For example, when  $m = 24$ , the data is generated every hour. Now, a new data set is averaged every 24-th, and it becomes daily averaged data.

Using the new data set, we obtained the posterior means for the parameters with the same MCMC algorithms as before.

Table 2.6 shows the results. An interesting finding is observed by comparing the result in Table 2.6 with the last column of Table 2.3. We can interpret that the former result is with the average price in a day, and the latter is with the opening of a day. In the latter case, the last column of Table 2.3 shows that we could obtain good performance in estimation at least for  $\theta_1$  and  $\theta_2$  by setting  $h = 1$  assuming that we know  $m$ . When we use averaged data, however, this is not the case any more. The posterior means are  $\theta_1 = 0.004$  and  $\theta_2 = -0.037$ , and both are not well centered around the true value. Also sigma is not close to the true value, which is a very intuitive result.

**Table 2.6: Posterior statistics for parameters when observations are averaged**

|                 |          |        |
|-----------------|----------|--------|
| Theta 1         | mean     | 0.004  |
|                 | variance | 0.004  |
| Theta 2         | mean     | -0.037 |
|                 | variance | 0.063  |
| Sigma           | mean     | 0.488  |
|                 | variance | 0.129  |
| Beta            | mean     | 1.391  |
|                 | variance | 0.1    |
| Acceptance rate | Theta    | 0.22   |
|                 | Beta     | 0.229  |

The true values for parameters are  $\theta_1 = 0.007393$ ,  $\theta_2 = -0.0876$ ,  $\sigma = 0.7791$  and  $\beta = 1.48$ . The statistics are obtained with  $m = 24$  and  $T = 100$ . 1 is used as  $h$ . The last two rows report the acceptance rates of Metropolis-Hastings algorithm for  $\theta$  and  $\beta$ . For  $\sigma$ , Gibbs-sampling is used, and acceptance rate is 1.

## 2.5 Empirical Evidence with Short-term Interest Rates

In the previous section, we saw the numerical example with the CKLS model. We found that  $h$  affected the posterior means of the parameters, and therefore we should be careful in using equation (2.5) as a discrete-time approximation of the continuous-time model shown by equation (2.4).

In this section, we estimate the model with short-term interest rates. We use the daily effective federal funds rate from 1/1/1960 to 7/31/2004. The sample size is 16284. We obtained the data from the website of the Federal Reserve Bank of St.Louis.<sup>8</sup> The daily effective Federal Funds rate is a volume-weighted average of the rates, and it is calculated by the Federal Reserve Bank. The use of averaged data might lose the continuity of data as we discussed in 3.4. However, the errors in estimates of  $\theta_1$  and  $\theta_2$

<sup>8</sup> <sup>8</sup> <http://research.stlouisfed.org/fred2/>

are small when the coefficients take small values. As can be seen later, our case is one of those cases. Table 2.7 shows the descriptive statistics.

We use the same MCMC algorithms as those we have used to obtain posterior pdfs...: Metropolis-Hasting Algorithms for  $\theta_1$ ,  $\theta_2$ , and  $\beta$ , and Gibbs-sampling for  $\sigma$ .

Here, MCMC draws are 25000 times, and the first 5000 draws are burned.  $h$  is set to be one.

Table 2.8 shows the posterior statistics for parameters. The pdfs  $\theta_2$  and  $\beta$  are also presented in Figure 2.2. The mean-reversion coefficient  $\theta_2$  shows a small value, -0.007, and this implies that there is no mean-reversion in daily effective Federal Funds rate. These results are similar to earlier studies on short-term interest rates (Brenner, Harjes and Kroner(1996)). The heteroscedasticity coefficient,  $\beta$ , takes 0.158 as the posterior mean. The value is smaller than those of 0.7608 in Eraker(2001), 0.67 in Andersen and Lund(1997), and 1.5 in CKLS(1992). The parameter  $\beta$  allows the volatility of interest rate to depend on the level of the interest rate, and therefore the higher value of  $\beta$  means that the interest rate volatilities are more sensitive to interest rate levels. Our result shows that the volatility of interest rate depends less on the interest rate level than the earlier studies have found.

**Table 2.7: Descriptive statistics of Daily Effective Federal Funds Rate (1/1/1960-7/31/2004)**

|                 | Mean     | Median   | Maximum  | Minimum           |       |
|-----------------|----------|----------|----------|-------------------|-------|
| y               | 6.22     | 5.53     | 22.36    | 0.13              |       |
| Difference in y | 0.00     | 0.00     | 7.79     | -7.89             |       |
|                 | Std.Dev. | Skewness | Kurtosis | # of observations |       |
| y               | 3.37     | 1.19     | 5.12     | 16283             |       |
| Difference in y | 0.38     | 0.67     | 52.56    | 16282             |       |
| lag             | 1        | 2        | 3        | 4                 | 5     |
| y               | 0.99     | 0.99     | 0.99     | 0.99              | 0.99  |
| Difference in y | -0.20    | -0.12    | -0.08    | -0.02             | -0.04 |

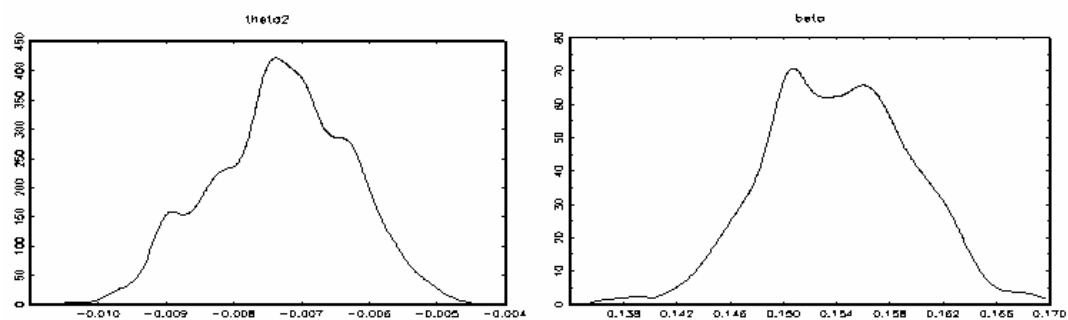
**Table 2.8: Empirical results**

|                 | Posterior statistics | Effective Federal Funds Rate (Daily) |
|-----------------|----------------------|--------------------------------------|
| Theta 1         | mean                 | 0.047                                |
|                 | variance             | 0.008                                |
| Theta 2         | mean                 | -0.007                               |
|                 | variance             | 0.001                                |
| Sigma           | mean                 | 0.289                                |
|                 | variance             | 0.013                                |
| Beta            | mean                 | 0.158                                |
|                 | variance             | 0.035                                |
| Acceptance rate | theta                | 0.031                                |
|                 | beta                 | 0.031                                |

The data used is daily effective Federal Funds rate, covering 1/1/1960 to 7/31/2004. The last two rows report the acceptance rates of Metropolis-Hastings algorithm for  $\theta$  and  $\beta$ . For  $\sigma$ , Gibbs-sampling is used, and acceptance rate is 1.



Figure 2.2: Posterior pdf for  $\theta_2$  and  $\beta$  from Empirical Analysis



## 2.6 Concluding Remarks

This paper was concerned with the Bayesian estimation of a discrete-time asset pricing model with Markov-Chain Monte Carlo (MCMC) methods. We have mainly dealt with a one-factor asset pricing model proposed by CKLS (1992).

To estimate stochastic differential equations of the asset pricing model, it is common use the discrete-time approximation of the continuous-time equation. When we estimate the discrete-time model, in equation (2.5), the specification of  $h$  is important. To set  $h = 1$  is the most common way to specify  $h$ , and it is called naïve discretization. In this paper, we have examined how naïve discretization could cause problems, using simulated data. Here are some findings we have obtained through our simulation-based analysis.

The first problem caused by setting  $h = 1$  arises when we do not know the value of  $m$ . This case is obvious because the reason we use  $h = 1$  is we do not know what  $m$  is. The second case happens the value of  $h$  is not small enough. Equation (2.5) is a discrete-time approximation of the continuous-time model, and therefore, equation (2.5) works as a good approximation when  $h$  is small. Our estimation with sampled data shows that we sample data at lower frequency, which means  $h$  takes larger value, the posterior means of the parameters are not close to their true values. When we sample at every 24th, for example, the posterior pdf for  $\beta$  is not well-centered around the true value. However, we also found that the proper value for  $h$  also depends on how we sample data. One possible way to solve the data discretization problem is to augment the data and increase the number of observations artificially. We used the MCMC-based data augmentation algorithm proposed by Eraker (2001), and investigated how it works. Our results show

that the results with the augmented data are not well-behaved both when we know the value of  $m$  and when do not it. Finally, when we use averaged data, the true value for  $h$  does not work well anymore. When we use the averaged data of a certain period, we could not obtain good estimates even we use a proper value for  $h$ .

This problem with respect to how we set  $h$  is not unique in the discrete-time approach. One such example was shown using Aït-Sahalia(1999,2002b) method. His approach that uses an approximate likelihood function is subject to a similar problem. That is, if we mis-specify the value of  $h$ , we do not obtain good performance in estimation.

We have also applied the estimation with daily data of U.S. effective Federal Funds rate between 1/1/1960 to 7/31/2004. Our result shows that there is no evidence of mean reversion, which is constant with the earlier studies. Also, we have found the relatively small value for the coefficient showing the volatility of interest rate to the level, to the earlier studies.

Finally, there are a couple of possible extensions to be mentioned. First, in terms of Bayesian method of estimation, how to set prior functions is a big issue. In this paper, the simplest prior function was considered, but they need to be extended to incorporate different priors. In particular, the Litterman priors might be one of the candidates in the context of VAR approach. Second, the model can be extended. One possible extension could be a multi-factor model such as the Stochastic Volatility model. Another would be a model where the error term follows an ARMA-GARCH process.

## CHAPTER 3: A BAYESIAN COMPARISON IN SHORT-TERM INTEREST RATE MODELS

### 3.1 Introduction

Short-term riskless interest rates are one of the most important macroeconomics variables, and significant efforts have been devoted to develop models describing the dynamics of the variables. One of the remarkable studies for estimation of short-term interest rate models was conducted by Chan, Karolyi, Longstaff and Sanders(1992) (CKLS hereafter), where they proposed an econometric model for short-term interest rates based on the diffusion processes for continuous-time asset pricing models. Their model allows the variance of interest rate to change depending on the level of the interest rate in a consistent way with the continuous time model, and encompasses other important specifications such as Merton(1973), Vasicek(1977) ,Cox, Ingersoll and Ross(1985) and Brennan and Schwartz(1979).

One of the extensions of the CKLS model is to allow non-Gaussian error terms. Brenner, Harjes and Kroner (1996) assume a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) specification in the error term. This specification allows that the model to capture the phenomenon that the sensitivity of volatility to levels is higher when shocks are larger in absolute value than when they are stable. Other studies of the short-term interest model with GARCH class errors were conducted by Longstaff and Schwartz (1995), Andersen and Lund(1997), Koedijk, Nissen, Schotman, and Wolff(1997), and Bali(2000).

Another natural extension of the CKLS model would allow errors to have stochastic volatility. Stochastic volatility models have been often used to describe time-varying volatility of financial time-series data. Such models are particularly helpful to capture sudden changes in the magnitude of variation of the observed values since it allows the conditional mean and variance to follow separate stochastic processes. In the estimation of stochastic volatility models, simulation based inference has been widely used (Jacquier, Polson and Rossi(1994 and 2004) (JPR hereafter); Kim, Shephard and Chib(1998); and Watanabe (2000)) because of the intractability of the likelihood function (Mahieu and Schotman(1998)).

In this chapter, we develop a model for short-term interest rates, based on the CKLS model with stochastic volatility. After developing MCMC algorithms for the model, the model will be applied with the 3 month Treasury bill rate. A comparison of our model with the CKLS with GARCH (1, 1) error term will be performed by using predictive densities to show a potential usefulness of the CKLS with stochastic volatility.

The remaining of the paper goes as follows: In section 3.2, the model and MCMC algorithms are explained. Section 3.3 demonstrates some numerical examples to show how the proposed algorithms work. Empirical results with the 3 month Treasury bill rate are shown in Section 4. Section 5 contains a comparison of our model with the model with CKLS-GARCH(1,1) model using predictive densities. Section 5 summarizes the findings of the paper.

### 3.2 The Model and MCMC algorithms

Our model is the extension of the short-term interest rate model by CKLS, where unobservable error terms are assumed to evolve stochastically. The model specification is given by the following equation:

$$\begin{aligned}\Delta r_t &= \theta_1 + \theta_2 r_{t-1} + \sigma_t r_{t-1}^\gamma \varepsilon_t & \varepsilon_t &\sim N(0,1) \\ \ln \sigma_t^2 &= \alpha_1 + \alpha_2 \ln \sigma_{t-1}^2 + v_t & v_t &\sim N(0,1) \\ t &= 1, \dots, n\end{aligned}\tag{3.1}$$

, where  $\Delta r_t$  is defined as  $r_t - r_{t-1}$ , and  $\varepsilon_t$  and  $v_t$  are independently and identically distributed across time, and  $\varepsilon_t$  and  $v_t$  are assumed to be independently distributed. To perform Bayesian analyses of the CKLS model with stochastic volatility, we need the posterior density function of the model. The likelihood function of the model is given by

$$\begin{aligned}p(\Theta_1, \Theta_2, V | Y) &\propto p(Y | V, \Theta_1, \Theta_2) p(V, \Theta_2 | \Theta_1) p(\Theta_1) \\ &\propto p(Y | V, \Theta_1, \Theta_2) p(V | \Theta_1) p(\Theta_1) p(\Theta_2)\end{aligned}$$

where  $\Theta_1 \equiv \{\alpha_1, \alpha_2, \sigma_v^2\}$ ,  $\Theta_2 \equiv \{\theta_1, \theta_2, \gamma\}$ ,  $V \equiv \{\sigma_t^2\}_{t=1}^n$ ,  $Y \equiv \{r_t\}_{t=1}^n$ .

To draw  $p(\Theta_1 | \Theta_2, V, Y)$  and  $p(\Theta_2 | \Theta_1, V, Y)$ , the Metropolis-Hasting algorithm is applied with equation (3.1). The likelihood function of the model is

$$l(Y | V, \Theta_1, \Theta_2) = \prod_{t=1}^n \left[ \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_{t,t}^2}{2\sigma_t^2}\right) \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{v_t^2}{2\sigma_v^2}\right) \right] \tag{3.2}$$

where

$$\begin{aligned}\varepsilon_t &= \varepsilon_0 & (t=0) \\ &= \frac{\Delta r_t}{r_{t-1}^\gamma} - \left( \frac{\theta_1 + \theta_2 r_{t-1}}{r_{t-1}^\gamma} \right) & (t=1, \dots, n)\end{aligned}$$

and

$$\begin{aligned} v_t &= v_0 & (t=0) \\ &= \sigma_t^2 - (\alpha_1 + \alpha_2 \sigma_{t-1}^2) & (t=1, \dots, n) \end{aligned}$$

As a prior, we assume the flat prior.

The proposal distribution is Normal distribution for the coefficients of the models, and inverse gamma distribution for  $\sigma_v^2$ . Since there is no obvious proposal density for the parameter  $\gamma$ , the efficient jumping rule is applied (German, Carlin, Stern and Rubin (2002)).

Draw of  $p(V | \Theta_1, \Theta_2, Y) = p(V | \Theta_1, Y)$  follows Jacquier, Polson, and Rossi (1994 and 2004), and we extend their algorithm to our modified CKLS model. The difficulty of creating Markov chains remains in drawing the vector of volatility  $V = \{\sigma_t^2\}_{t=1}^n$ , that is latent in this model. We have the marginal density for the volatility state  $p(V | \Theta_1, Y)$ , which is given by the following equation:

$$\begin{aligned} p(V | \Theta_1, Y) &= \sum_{t=1}^n p(\sigma_t^2 | \sigma_{t-1}^2, \sigma_{t+1}^2, \alpha_1, \alpha_2, \sigma_v, y_t) \\ &\propto \sum_{t=1}^n p(y_t | \sigma_t^2) p(\sigma_t^2 | \sigma_{t-1}^2) p(\sigma_{t+1}^2 | \sigma_t^2) \\ &\propto \sum_{t=1}^n \frac{1}{\sigma_t} \exp \frac{-y_t^2}{2\sigma_t^2} \times \frac{1}{\sigma_t^2} \exp \frac{-(\ln \sigma_t^2 - \mu_t)^2}{2\sigma^2} \end{aligned}$$

where

$$\mu_t = (\alpha_1(1 - \alpha_2) + \alpha_2(\ln \sigma_{t+1}^2 + \ln \sigma_{t-1}^2)) / (1 + \alpha_2^2)$$

and

$$\sigma^2 = \sigma_v^2 / (1 + \alpha_2^2).$$

The derivation of  $p(V | \Theta_1, Y)$  is found in JPR (1994 and 2004).

Let us denote  $p(h_t | \cdot)$ .

$$p(\sigma_t^2 | \cdot) = \frac{1}{\sigma_t} \exp \frac{-y_t^2}{2\sigma_t^2} \times \frac{1}{\sigma_t^2} \exp \frac{-(\ln \sigma_t^2 - \mu_t)^2}{2\sigma^2} \quad (3.3)$$

Equation (3.3) is not recognized as a density and direct draws from the equation are not feasible. JPR and Johannes and Polson (2002) suggest finding a proposal density by approximating equation (3.3). In their studies, they provide the proposal density based on the fact that the first term of equation (3.3) is the density function for an inverse gamma distribution. The second term of equation (3.3), which is log-normal kernel was approximated by an inverse gamma with same mean and variance, and was multiplied by the inverse gamma kernel shown by the first term of equation (3.3). This procedure yields an inverse gamma distribution:

$$q(\sigma_t^2 | \cdot) \propto \sigma_t^{-2(\phi+1)} e^{\frac{-\lambda}{\sigma_t^2}}$$

where

$$\phi = 0.5 + (1 - 2 \exp \sigma^2) / (1 - \exp \sigma^2)$$

and

$$\lambda_t = 0.5 y_t^2 + (\phi - 1) \exp(\mu_t + 0.5 \sigma^2)$$

In addition to the proposal density for  $V = \{\sigma_t^2\}_{t=1}^n$  by JPR, I propose two new proposal densities. One of the proposal densities is the approximation of equation (3.3) to inverted gamma distribution as JPR suggested, but the mean and variance are obtained numerically instead of analytically. On each iteration, we calculate the numerical mean and variance of equation (3.3) using quadratic formula such as Simpson's rule, and set



the parameter values of the invertse gamma distribution obtained from the mean and variance. We use the inverse gamma distribution as the proposal density. The same process is conducted using inverse Gaussian distribution and used as the third proposal density.

Given all of the arguments above, the MCMC algorithm works as follows:

- (i) Set initial values for  $\Theta_1^0, \Theta_2^0$ , and  $\{\sigma_t^2\}_{t=1}^n$ .
- (ii) Draw  $\{\sigma_t^2\}_{t=1}^n$  by the MH algorithm with the proposal density (JPR, inverse gamma with numerical mean and variance, or inverse Gaussian with numerical mean and variance.)
- (iii) Given (ii), draw  $\gamma^{(i)}$  by the efficient jump rule.
- (iv) Given (ii), draw  $\Theta_1^{(i)}$  by the MH algorithm.
- (v) Given all of above, draw  $\theta_1^{(i)}$  and  $\theta_2^{(i)}$  by the MH algorithm.
- (vi) Repeat (ii)-(v) until the sequences become stable.

### 3.3 Numerical Examples

This section shows the numerical examples of the algorithms developed for the CKLS stochastic volatility model. In each experiment, 20,000 MCMC samples were drawn, the first 10,000 were discarded, and every 10th iterate was saved. The data were generated where parameters values are given, and Bayesian estimations on the model parameters were made using the proposed algorithms. The parameters were set to be:

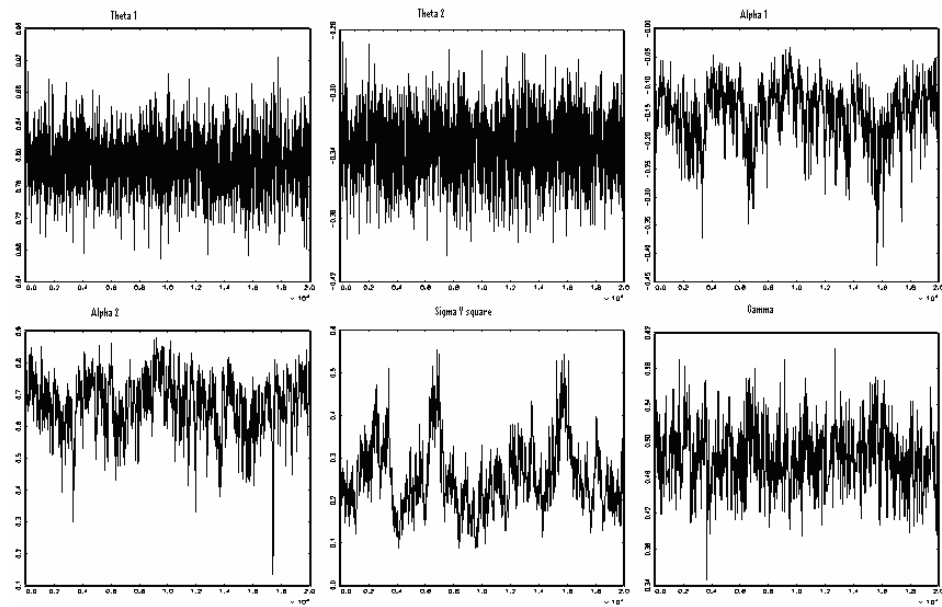
$$\theta_1=0.7; \theta_2=-0.3; \gamma=0.5; \alpha_1=-0.1; \alpha_2=0.8 \text{ and } \sigma_v=0.5.$$

Figure 3.1 plots the MCMC draws using three different algorithms. Panel (A) shows the figure generated with the proposal density by JPR, and panel (B) and panel (C) are experiments with the proposal density with numerical mean and variance. Inverted gamma and the inverse Gaussian are used in panel (B) and panel (C) respectively as proposal densities. This figure gives us an idea of the stability of Markov chain, but formal convergence tests would be necessary.

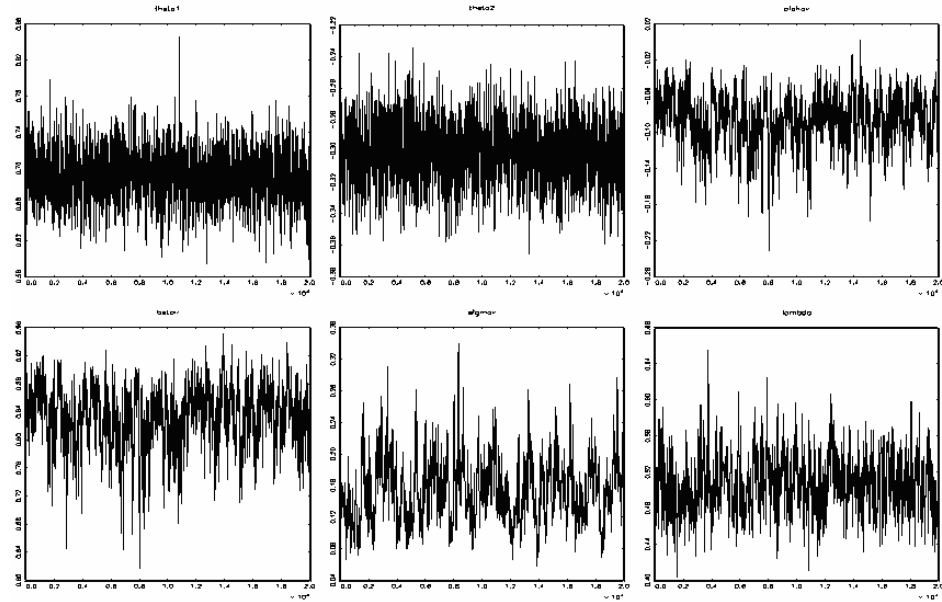
The mean, standard deviation of posterior distributions and autocorrelation of MCMC draws are shown in Table 3.1. Overall, the posterior mean is well-centered on the true value. The acceptance rate is 1 except for one parameter, the parameter for volatility persistence. High to moderate autocorrelation are observed in the parameters in the volatility evolution equation.

**Figure 3.1: Assessing convergence for parameters in the CKLS with stochastic volatility model**

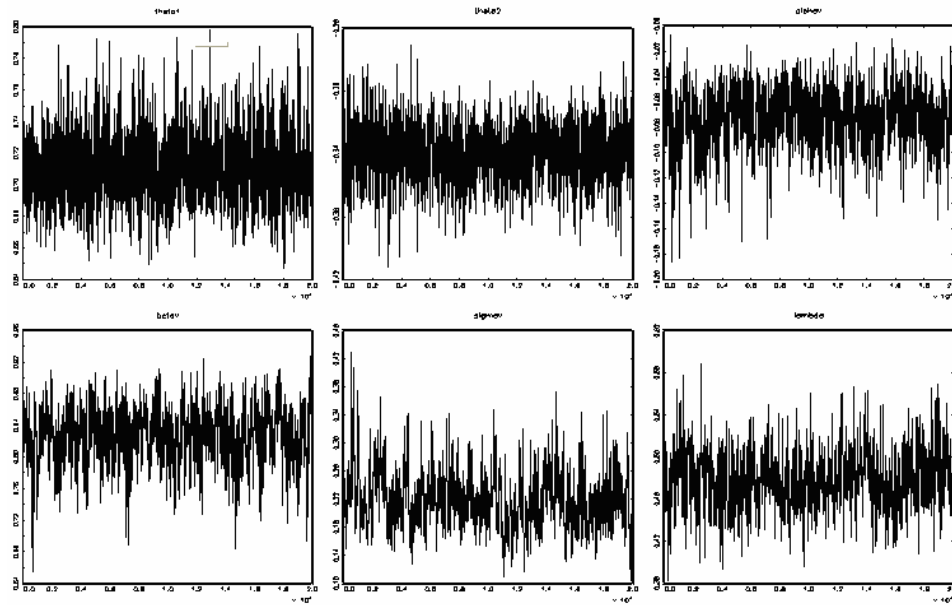
(a) Jacquire, Polson and Rossi



(b) Inverted Gamma with numerical mean and variance



(c ) Inverse Gaussian with numerical mean and variance



This table plots the MCMC samples of each parameter of a simulated numerical example for the CKLS with stochastic volatility model. 20,000 Markov chains are generated, the first 10,000 are discarded and every 10 draw is kept.

**Table 3.1: Numerical examples for the CKLS model with stochastic volatility**

|   | Theta 1 | Theta 2 | Alpha 1 | Alpha 2 | Sigma v square | Gamma |
|---|---------|---------|---------|---------|----------------|-------|
| <i>True Value</i>                                     | -0.3    | 0.7     | -0.1    | 0.8     | 0.2            | 0.5   |
| <i>Posterior Statistics</i>                           |         |         |         |         |                |       |
| (A) JPR   |         |         |         |         |                |       |
| Mean  | -0.301  | 0.691   | -0.087  | 0.831   | 0.154          | 0.506 |
| Std.D   | 0.021   | 0.032   | 0.03    | 0.046   | 0.042          | 0.033 |
| AC*   | 0.084   | 0.06    | 0.607   | 0.708   | 0.854          | 0.645 |
| AR**  | 1       | 1       | 1       | 1       | 1              | 0.098 |
| (B) Inverted gamma with numerical mean and variance   |         |         |         |         |                |       |
| Mean  | -0.335  | 0.791   | -0.153  | 0.671   | 0.252          | 0.483 |
| Std.D   | 0.021   | 0.037   | 0.053   | 0.091   | 0.082          | 0.033 |
| AC*   | 0.038   | 0.031   | 0.766   | 0.807   | 0.908          | 0.637 |
| AR**  | 1       | 1       | 1       | 1       | 1              | 0.096 |
| (C) Inverse Guassian with numerical mean and variance |         |         |         |         |                |       |
| Mean  | -0.34   | 0.711   | -0.073  | 0.827   | 0.219          | 0.478 |
| Std.D   | 0.019   | 0.023   | 0.024   | 0.036   | 0.042          | 0.029 |
| AC*   | 0.106   | 0.059   | 0.402   | 0.568   | 0.768          | 0.584 |
| AR**  | 1       | 1       | 1       | 1       | 1              | 0.095 |

This table provides means, standard deviations, and autocorrelations as well as acceptance rates for the Metropolis-Hastings algorithms for parameters of a simulated numerical example with sample size 1,000. 20,000 Markov chains are generated, the first 10,000 are discarded and every 10 draw is kept.

\* Autocorrelation

\*\* Acceptance rate for the MH algorithm

### 3.4 Empirical Results, Predictive Densities and Comparison with GARCH model

This section shows an application of the model with a short-term interest rate. The data consist of the 3-month Treasury constant maturity rates for the period 01/08/82 to 10/06/06, and were obtained from the Fred® by the Federal Reserve St. Louis. The data are weekly data, which is the average of business days in a week. The total number of the observations is 1292. Panel (A) of Figure 3.2 plots the observations of the interest rate while panel (B) presents the changes in the interest rate. In the beginning of the sample period, high interest rates higher than ten percent were recorded. This phenomenon is called Volcker disinflation, where the Fed conducted the monetary targeting experiment.

Table 3.2 shows summary statistics for the data. The original sequence  $r_t$  is slightly right-skewed, which shows the asymmetry of the data. On the other hand,  $\Delta r_t$  shows the skew the left, which indicates increases in the interest rate occurred less often than decreases in the sample period.

The stochastic volatility model is applied with the weekly interest rate. For a comparison purpose, the CKLS model with GARCH(1,1) error, is also estimated:

$$\begin{aligned}\Delta r_t &= \phi_1 + \phi_2 r_{t-1} + \sigma_t r_{t-1}^\tau \varepsilon_t \quad \varepsilon_t \sim N(0,1) \\ \sigma_t^2 &= \beta_1 + \beta_2 u_{t-1}^2 + \beta_3 \sigma_{t-1}^2 \\ t &= 1, \dots, n\end{aligned}\tag{3.4}$$

, where  $u_t \equiv \sigma_t \varepsilon_t$  and  $\varepsilon_t$  is independently and identically distributed. The definition of  $\Delta r_t$  follows the earlier model. This model allows the variance of error term to follow GARCH (1, 1) process, and was originally developed by Bollerslvev(1986).

The algorithm for the model with GARCH (1, 1) effect closely follows the algorithm for regression models with ARMA-GARCH errors by Nakatsuma(2000). In his study, the MH algorithm was employed to draw the Monte Carlo sample from the joint posterior distribution.

Our empirical results from the CKLS models with the stochastic volatility and with the model with GARCH (1, 1) error term are shown in Table 3.3. In stochastic volatility models, the volatility evolves with high autocorrelation, ranging from 0.90 to 0.95.

One way of comparing models in Bayesian context can be done by Bayesian predictive densities. The draws of predicted values  $r_{t+j}$  are fairly easily obtained using MCMC. The joint pdf of the predicted values  $\tilde{r} = \{r_{n+j}\}_{j=1}^m$  and the parameter vector  $\Theta \equiv \{\Theta_1, \Theta_2\}$  is given by

$$h(\tilde{r}, \Theta | Y) = f(\tilde{r} | \Theta, Y) p(\Theta | Y)$$

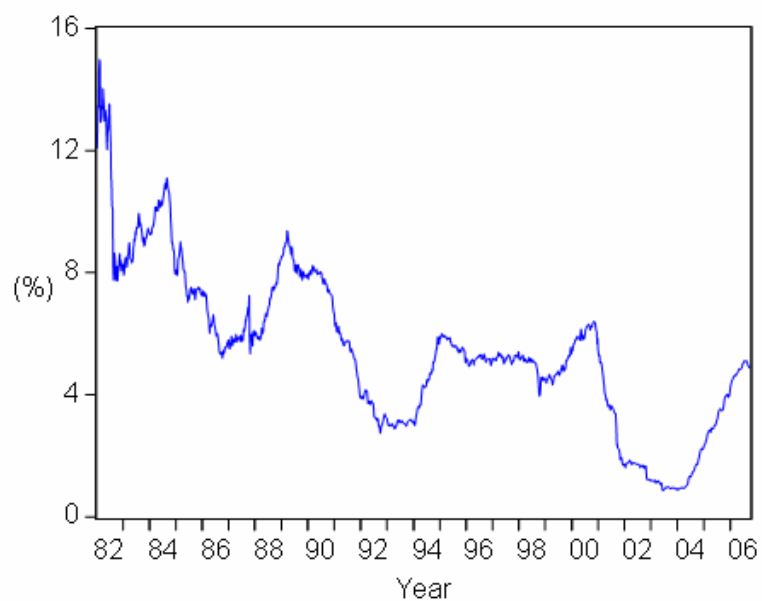
,where  $p(\Theta | Y)$  is the posterior pdf of  $\Theta$ , and  $f(\tilde{r} | \Theta, Y)$  is the density of  $\tilde{r}$ . Integrating this over the parameters, the following equation is obtained:

$$f(\tilde{r} | Y) = \int f(\tilde{r} | \Theta, Y) p(\Theta | Y) d\Theta$$

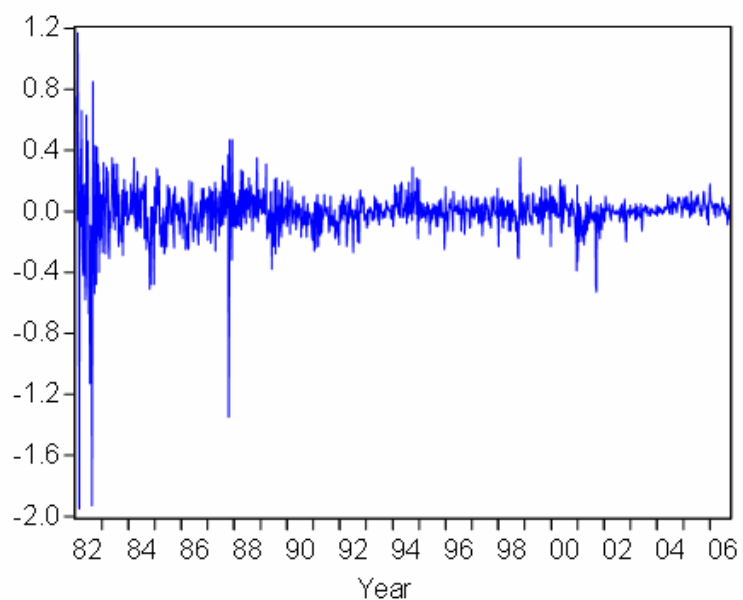
Figure 3.3 illustrates one and two period ahead predictive densities of the 3-month Treasury bill rates based on the two models: CKLS model with stochastic volatility and with GARCH (1, 1) error terms. The JPR algorithm was employed for the stochastic volatility model. The model with GARCH (1, 1) error terms has by far smaller variance in both one and two period ahead predictions.

**Figure 3.2: Weekly 3-month Treasury bill rate (01/08/1982-10/06/2006)**

(a) Original sequence



(b) First difference



The data consist of the observations of the 3-month Treasury constant maturity rate for the period 01/08/82 to 10/06/06, and are weekly data that are the average of business days in a week.

Source: Fred® by the Federal Reserve St. Louis



**Table 3.2: Summary statistics: weekly 3-month Treasury bill rate (01/08/1982-10/06/2006)**

|             | r     | Delta r |
|-------------|-------|---------|
| Sample size | 1292  | 1291    |
| Mean        | 5.466 | -0.006  |
| Median      | 5.27  | 0       |
| Maximum     | 14.97 | 1.17    |
| Minimum     | 0.86  | -1.95   |
| Std.D       | 2.616 | 0.161   |
| Skewness    | 0.468 | -3.091  |
| Kurtosis    | 3.422 | 44.524  |

The data consist of the observations of the 3-month Treasury constant maturity rate for the period 01/08/82 to 10/06/06, and are weekly data that are the average of business days in a week.  
Source: Fred® by the Federal Reserve St. Louis

Figure 3.4 shows the cumulative density functions (CDFs) of absolute deviation between the predicted and actual values. The actual values, for example, are the interest rate dated on 10/13/06 for one period ahead, which is 5.03, and the one dated on 10/20/06 for two period ahead, which is 5.09. Since the GARCH (1, 1) model produces the tight prediction (Mean: 4.93, Std. D.: 0.009 for one-period ahead prediction; and Mean: 4.94: Std. D: 0.017 for two-period ahead prediction), the actual values do not fall in the predictive densities. The probabilities of making the prediction that is exactly equal to the realized value model is 0 from the GARCH, which is shown by the starting points of the CDFs at positive values in Figure 3.4. However since the GARCH model produces predictions with tighter variance, the CDFs attain 1 more rapidly than the SV models. The SV model produces predictions that allow wider uncertainty (Mean: 4.93, Std. D: 0.36 for one period ahead prediction; and Mean: 4.93, Std.D: 0.68 for two period ahead prediction). As a result, in one-period ahead prediction, the probability of making prediction which is close to the true value is higher than the probability from GARCH(1,1) model until the absolute value of error reaches approximately 0.1. The

CDFs increase at a slower rate, and it reaches 1 around the point where the absolute error is approximately 0.6. A similar argument can be made for two-period ahead prediction.

**Table 3.3: Estimation results of the CKLS with stochastic volatility model and CKLS with GARCH (1, 1) model with 3 month Treasury bill rate**

| <i>1. CKLS with Stochastic Volatility</i>             |         |         |         |         |                |       |
|---|---------|---------|---------|---------|----------------|-------|
|   | Theta 1 | Theta 2 | Alpha 1 | Alpha 2 | Sigma v Square | Gamma |
| (A) JPR   |         |         |         |         |                |       |
| Mean  | 0.005   | -0.001  | -0.517  | 0.915   | 0.281          | 0.268 |
| Std.D   | 0.002   | 0       | 0.278   | 0.037   | 0.078          | 0.224 |
| AC*   | 0.184   | 0.105   | 0.91    | 0.865   | 0.909          | 0.998 |
| AR**  | 1       | 1       | 1       | 1       | 1              | 0.064 |
| (B) Inverted gamma with numerical mean and variance   |         |         |         |         |                |       |
| Mean  | 0.012   | -0.003  | -0.11   | 0.95    | 0.012          | 0.001 |
| Std.D   | 0.021   | 0.004   | 0.028   | 0.013   | 0.001          | 0.001 |
| AC*   | -0.005  | 0.002   | 0.414   | 0.418   | 0.607          | 0.978 |
| AR**  | 1       | 1       | 1       | 1       | 1              | 0.004 |
| (C) Inverse Gaussian with numerical mean and variance |         |         |         |         |                |       |
| Mean  | 0.012   | -0.003  | -0.217  | 0.904   | 0.018          | 0.001 |
| Std.D   | 0.021   | 0.004   | 0.063   | 0.028   | 0.002          | 0.001 |
| AC*   | -0.015  | 0.001   | 0.748   | 0.749   | 0.753          | 0.973 |
| AR**  | 1       | 1       | 1       | 1       | 1              | 0.003 |
| <i>2. CKLS with GARCH(1,1)</i>                        |         |         |         |         |                |       |
|   | Theta 1 | Theta 2 | Beta 1  | Beta 2  | Beta 3         | Tau   |
| Mean  | 0.005   | -0.001  | 0.315   | 0.489   | 0.817          | 0.817 |
| Std.D   | 0.002   | 0       | 0.037   | 0.04    | 0.048          | 0.048 |
| AC*   | 0.227   | 0.117   | 0.437   | 0.633   | 0.704          | 0.704 |
| AR**  | 0.579   | 0.579   | 0.375   | 0.375   | 0.511          | 0.456 |

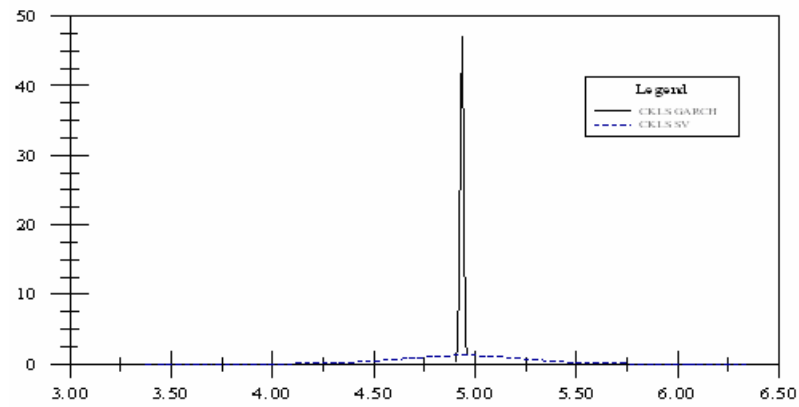
This table provides means, standard deviations, and autocorrelations as well as acceptance rates for the Metropolis-Hastings algorithm for parameters of the CKLS SV model and CKLS GARCH(1,1) model. 20,000 Markov chains are generated, the first 10,000 are discarded and every 10 draw is kept.

\* Autocorrelation

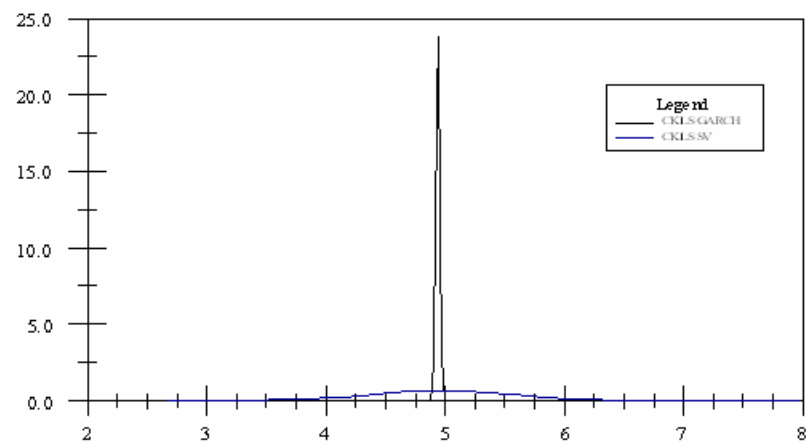
\*\* Acceptance rate for the MH algorithm

**Figure 3.3: Predictive densities: one and two period ahead**

(a) Predictive density at  $t+1$



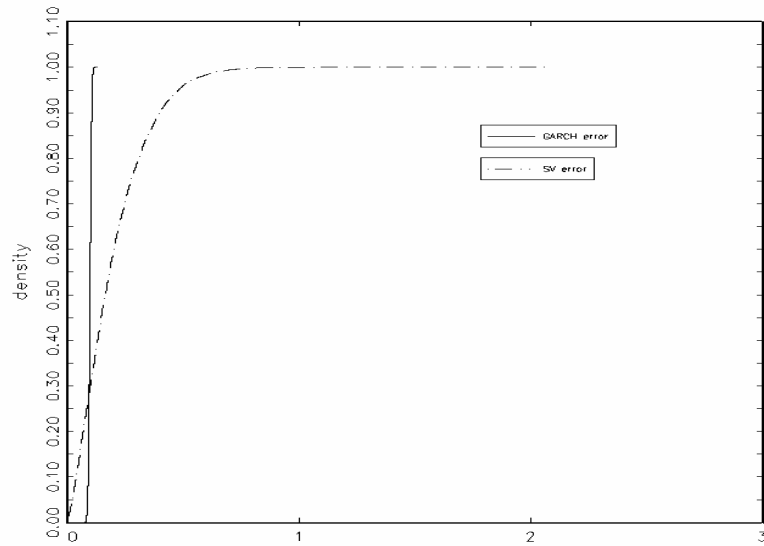
(b) Predictive density at  $t+2$



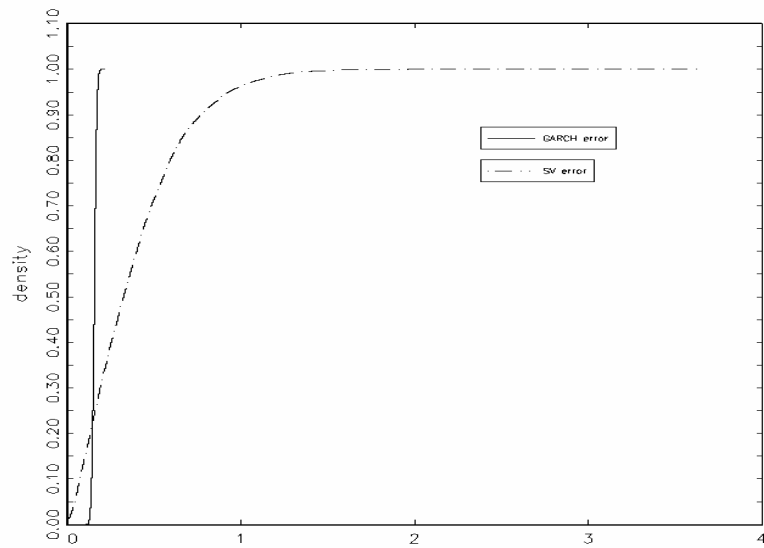
The figure provides the predictive densities of 3 month Treasury bill rate in one and two period ahead by the CKLS SV and CKLS GARCH (1, 1) model.

**Figure 3.4: Cumulative density functions of absolute deviation of the predicted values from the actual value**

(a) Prediction at  $t+1$



(b) Prediction at  $t+2$



The X axis shows the absolute deviation from the realized value.

### 3.5. Concluding Remarks

In this chapter, I made a Bayesian inference of a short-term interest rate model with stochastic volatility. The model was developed based on the CKLS model, and MCMC algorithms suitable for the model were constructed. In addition to the algorithm that is a simple extension of the algorithm proposed by JPR, two additional algorithms were considered. After confirming that the algorithms worked well using numerical examples, the model was applied with the 3-month Treasury constant maturity rate for the period 01/08/82 to 10/06/06. The empirical results suggested that there was high autocorrelation in volatility of the error terms.

Finally, Bayesian predictive densities were used to compare the developed model was compared with the model with GARCH (1, 1) error. The predictive densities obtained by CKLS with stochastic volatility have wider variance than the ones from CKLS with GARCH in one- and two-period ahead predictions. However, the realized value did not fall in the support of the predictive values in CKLS with GARCH model because of the tighter variance in predictive density; therefore the probability that the absolute deviation from the realization is smaller was higher in the CKLS with stochastic volatility in a certain range of deviation.

## Appendix A. Likelihood function of the joint function of survival and distribution functions

Let  $F_E(e)$  and  $F_C(c)$  be cumulative density functions (cdfs) of  $E$  and  $C$  respectively, and  $F(e, c)$  be the joint cdf of  $E$  and  $C$ . Using a copula  $C$ , the joint function of  $S_E(e) \equiv 1 - F_E(e)$  and  $F_C(c)$  is written as follows;

$$H(e, c) = C(S_E(e), F_C(c))$$

In practice, available data to estimate the survival function is right censored. Bayesian estimation of the function  $H(e, c)$  is based on the following likelihood function:

$$L(\alpha | e, c, \delta) = \prod_{i=1}^n \Pr(E = e, C = c)^{\delta_i} \times \Pr(E > e, C = c)^{1-\delta_i}$$

Although function  $H(e, c)$  itself is not a distribution function, it is possible to define the two probability terms in the likelihood function using  $H(e, c)$ .

First, the following equation holds:

$$\begin{aligned} H(e, c) &= \Pr(E > e, C \leq c) \\ &= \Pr(C \leq c) - \Pr(E \leq e, C \leq c) \\ &= F_C(c) - F(e, c) \end{aligned}$$

Using this, the probability density function (pdf) of  $F(e, c)$  denoted as  $f(e, c)$  is written as the following:

$$\begin{aligned} f(e, c) &= \Pr(E = e, C = c) \\ &= \frac{\partial F(e, c)}{\partial e \partial c} \\ &= \frac{\partial (F_C(c) - H(e, c))}{\partial e \partial c} \end{aligned}$$

$$\begin{aligned}
&= \frac{\partial(F_C(c) - C(S_E(e), F_C(c)))}{\partial e \partial c} \\
&= -c(S_E(e), F_C(c))(-f_E(e))(f_C(c)) \\
&= c(S_E(e), F_C(c))f_E(e)f_C(c)
\end{aligned}$$

, where  $c(S_E(e), F_C(c)) \equiv \partial C(S_E(e), F_C(c)) / \partial S_E \partial F_C$ , which is the density function associated with copula  $C$ .

$$\begin{aligned}
\Pr(E > e, C = c) &= \Pr(E > e \mid C = c) \cdot \Pr(C = c) \\
&= \int_e^\infty f_{E|C}(e \mid c) de \cdot f_C(c) \\
&= \left(1 - \int_{-\infty}^e f_{E|C}(e \mid c) de\right) \cdot f_C(c) \\
&= (1 - F_{E|C}(e \mid c))f_C(c) \\
&= \left(1 - \frac{1}{f_C(c)} \frac{\partial F(e, c)}{\partial c}\right) \cdot f_C(c) \\
&= f_C(c) - \frac{\partial F(e, c)}{\partial c} \\
&= f_C(c) - \frac{\partial(F_C(c) - H(e, c))}{\partial c} \\
&= \partial H(e, c) / \partial c
\end{aligned}$$



## Appendix B. Derivation of the transition function of Vasicek model

The model is

$$dX(t) = \kappa(\alpha - X(t))dt + \sigma dB(t). \quad (B1)$$

Multiplying equation (B1) by the integrating factor  $e^{\kappa t}$ , we have

$$e^{\kappa t} dX(t) = -\kappa e^{\kappa t} X(t)dt + \kappa \alpha e^{\kappa t} dt + \sigma e^{\kappa t} dB(t) \quad (B2)$$

With  $\alpha = 0$ , equation (B1) becomes

$$dX(t) = -\kappa X(t)dt + \sigma dB(t). \quad (B3)$$

Applying the same procedure, we have

$$e^{\kappa t} dX(t) = -\kappa e^{\kappa t} X(t)dt + \sigma e^{\kappa t} dB(t) \quad (B4)$$

The total differentiation of  $e^{\kappa t} dX(t)$  is

$$d[e^{\kappa t} X(t)] = \kappa e^{\kappa t} X(t)dt + e^{\kappa t} dX(t),$$

or

$$e^{\kappa t} dX(t) = -\kappa e^{\kappa t} X(t)dt + d[e^{\kappa t} X(t)]. \quad (B5)$$

Equating equation (B2) and (B5), we have

$$d[e^{\kappa t} X(t)] = \kappa \alpha e^{\kappa t} dt + \sigma e^{\kappa t} dB(t). \quad (B6)$$

Integrating out over  $t_0$  to  $t$ , we obtain

$$\begin{aligned} & e^{\kappa t} X(t) - e^{\kappa t_0} X(t_0) \\ &= \kappa \alpha \int_{t_0}^t e^{\kappa s} ds + \sigma \int_{t_0}^t e^{\kappa s} dB(s) \\ &= \alpha(e^{\kappa t} - e^{\kappa t_0}) + \sigma \int_{t_0}^t e^{\kappa s} dB(s) \end{aligned}$$

, or

$$X(t) = X(t_0)e^{-\kappa(t-t_0)} + \alpha(1 - e^{-\kappa(t-t_0)}) + \sigma e^{-\kappa t} \int_{t_0}^t e^{\kappa s} dB(s). \quad (B7)$$

If we define  $Y(t) = \int_{t_0}^t e^{\kappa s} dB(s)$ , which is the Wiener integration with  $f(s) = e^{\kappa s}$ ,  $Y(t)$  is a

Gaussian process with mean 0 and variance,  $Var(Y_t)$ , given by

$$\begin{aligned}
 Var(Y_t) &= \int_{t_0}^t e^{2\kappa s} dV_X(s) \\
 &= \int_{t_0}^t e^{2\kappa s} ds \\
 &= \left[ \frac{1}{2\kappa} e^{2\kappa s} \right]_{t_0}^t = \frac{1}{2\kappa} (e^{2\kappa t} - e^{2\kappa t_0})
 \end{aligned} \tag{B8}$$

Equation (B6) in the text is, therefore, the transition probability from previous  $h$  to next  $h$  by setting the previous  $h$  to 0 each time.

### Appendix C. MCMC algorithm for obtaining the posterior p.d.f. of parameters of Vasicek model using the transition function

A Gibbs sampling MCMC goes as follows.

1. We can think of the following regression model using the transition probability function (2.6).

$$y_h = \alpha^* + \beta^* y_0 + \varepsilon \quad \varepsilon \sim N(0, \tau^2) \quad (C1)$$

, where  $y_0$  is the previously drawn  $y_h$ .

Here,  $\alpha^* = \alpha(1 - \beta)$ ,  $\beta^* = e^{-\kappa h}$ , and  $\tau^2 = \frac{\gamma^2}{2\kappa}$ .

Draw  $\alpha^*, \beta^*$  and  $\tau^2$ .

2. Transform  $\alpha^*, \beta^*$  and  $\tau^2$  to  $\alpha, \kappa, \gamma$  and  $\sigma$ .

The relationships between them are:

$$\begin{aligned} \kappa &= -\frac{\ln \beta^*}{h} \\ \alpha &= \frac{\alpha^*}{1 - \beta^*} \\ \gamma &= \tau \sqrt{2\kappa} \\ \sigma &= \frac{\gamma}{\sqrt{1 - \beta^{*2}}} \end{aligned} \quad (C2)$$

## Appendix D. Ait-Sahalia's approximation method of the transition functions

Suppose that the model is expressed by the following equation.

$$dY_t = \mu(Y_t; \theta)dt + \sigma(Y_t; \theta)dW_t \quad (D1)$$

where  $W_t$  is a standard Brownian motion and the drift  $\mu$  and diffusion  $\sigma^2$  are known functions. His method produces approximations in closed form to the (usually) unknown transition function  $p_Y(h, y | y_0; \theta)$ . The procedure consists of the following three steps<sup>9</sup>:

1. Standardize the diffusion function of  $Y$  into another diffusion  $X$

Diffusion  $X$  is defined as

$$X_t \equiv \gamma(Y_t; \theta) = \int^{Y_t} du / \sigma(u; \theta) \quad (D2)$$

where any primitive of the function  $1/\sigma$  may be selected. In finance the domain of  $Y$  and  $X$  would be either whole real line  $(-\infty, +\infty)$  or the half line  $(0, +\infty)$  in most cases.

By applying Ito's Lemma,  $X$  has unit diffusion:

$$dX_t = \mu_X(X_t; \theta)dt + dW_t \quad (D3)$$

where

$$\mu_X(x; \theta) = \frac{\mu(\gamma^{-1}(x; \theta); \theta)}{\sigma(\gamma^{-1}(x; \theta); \theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial y}(\gamma^{-1}(x; \theta); \theta). \quad (D4)$$

Once we have the transition density of  $p_X$ , we can easily obtain  $p_Y$  from

$p_X$  through the Jacobian formula

$$p_Y(h, y | y_0; \theta) = \frac{p_X(h, \gamma(y; \theta) | \gamma(y_0; \theta); \theta)}{\sigma(\gamma(y; \theta); \theta)}. \quad (D5)$$

---

<sup>9</sup> The detail of the explanations can be found in Ait-Sahalia (1999, 2002b).

## 2. Explicit Expression for the Approximation

One can derive an explicit expansion for the transition density of the variable  $X$  based on a Hermite expansion of its density  $x \mapsto p_X(h, x | x_0; \theta)$  around a Normal density function. The expansion of  $p_X$  up to order  $K$  is given by

$$\tilde{p}_X^{(K)}(h, x | x_0; \theta) = h^{-1/2} \phi\left(\frac{x - x_0}{h^{1/2}}\right) \exp\left(\int_{x_0}^x \mu_X(w; \theta) dw\right) \sum_{k=0}^K c_k(x | x_0; \theta) \frac{h^k}{k!} \quad (\text{D6})$$

where  $\phi(z) \equiv e^{-z^2/2} / \sqrt{2\pi}$  is the  $N(0,1)$  density function,  $c_0(x | x_0; \theta) = 1$ , and for all  $j \geq 1$ ,

$$\begin{aligned} c_j(x | x_0; \theta) &= j(x - x_0)^{-j} \int_{x_0}^x (w - x_0)^{j-1} \\ &\times \left\{ \lambda_X(w) c_{j-1}(w | x_0; \theta) + (\partial^2 c_{j-1}(w | y_0; \theta) / \partial w^2) / 2 \right\} dw \end{aligned} \quad (\text{D7})$$

where  $\lambda_X(x; \theta) \equiv -(\mu_X^2(x; \theta) + \partial \mu_X(x; \theta) / \partial x) / 2$ .

## 3. Transform $\tilde{p}_X^{(K)}$ into $\tilde{p}_Y^{(K)}$ using equation (D5).

### Appendix E. MCMC algorithms for obtaining posterior p.d.f. of parameters

To obtain posterior probability density functions of parameters, the following MCMC algorithms are used. The model is expressed by

$$\Delta Y_t = (\theta_1 + \theta_2 Y_{t-1})h + \sigma |Y_{t-1}|^\beta \sqrt{h} u_t, u_t \sim N(0,1).$$

The posterior probability density functions of parameters are given by

$$\pi(Y_t, \theta) \propto \prod_{t=1}^T p(Y_t | \theta) p(\theta) \quad (\text{E1})$$

where the likelihood function is give by

$$p(Y_t | \theta) = \frac{1}{\sigma |Y_{t-1}|^\beta \sqrt{h}} \exp \left\{ -\frac{1}{2} \left( \frac{(\Delta Y_t - (\theta_1 + \theta_2 Y_{t-1})h)^2}{\sigma^2 Y_{t-1}^{2\beta} h} \right) \right\} \quad (\text{E2})$$

and I assume the standard non-informative prior for  $p(\theta)$ . Once we define

$y \equiv (Y_t - Y_{t-1})/(Y_{t-1}^\beta \sqrt{h})$  and  $X \equiv [\sqrt{h} Y_{t-1}^{-\beta} \quad \sqrt{h} Y_{t-1}^{1-\beta}]$  for all  $t$ , equation(E2) is written as

the likelihood function of a linear regression model of  $y$  on  $X$ , and therefore

$$(\theta_1, \theta_2) | \sigma, Y \sim N(\bar{\theta}, \sigma^2 (X'X)^{-1}) \quad (\text{E3})$$

and

$$\sigma^{-2} | Y \sim IG(n-2, \bar{s}^2) \quad (\text{E4})$$

where  $\bar{\theta} = (X'X)^{-1}(X'y)$  and  $\bar{s}^2 = 1/n \sum_t (y_t - X_t \bar{\theta})^2$ . Eraker uses Gibbs sampling

method to draw two  $\theta$ s, but here we use Metropolis-Hastings algorithm with proposal

density set equation (9), as well as Gibbs sampling. Metropolis-Hastings algorithm for  $\theta$ s

works as follows: we generate  $\theta_1$  and  $\theta_2$  by equation (9). We accept  $\theta^{(j)}$  with

probability

$$\theta_0 = \min \left\{ \frac{p(\theta^{(j)}, \sigma^{(j-1)}, \beta^{(j-1)} | Y)}{p(\theta^{(j-1)}, \sigma^{(j-1)}, \beta^{(j-1)} | Y)}, 1 \right\}.$$

Otherwise set  $\theta^{(j)} = \theta^{(j-1)}$ .

To estimate  $\beta$ , we use an efficient jump proposed by German, Carlin, Stern and Rubin(1995). The effective jump rule is conducted as follows;

$$\beta^{(j)} | Y \sim N(\beta^{(j-1)}, c), c = .045$$

If  $0 < \beta < 1$ , we accept  $\beta^{(j)}$  with probability

$$\beta_0 = \min \left\{ \frac{p(\theta^{(j)}, \sigma^{(j)}, \beta^{(j)} | Y)}{p(\theta^{(j)}, \sigma^{(j)}, \beta^{(j-1)} | Y)}, 1 \right\}.$$

Otherwise set  $\beta^{(j)} = \beta^{(j-1)}$ .

Having all of these in your hand, MCMC works as follows;

1. Set initial values. I use OLS estimators for  $\theta_1$  and  $\theta_2$ ,  $\sigma = 0.7791$  and  $\beta = 1.46$ .  
Set  $j=1$ .
2. Draw  $(\theta_1, \theta_2)^{(j)}$  by either Metropolis-Hastings algorithm with proposal density as equation (D3), or Gibbs-sampling with equation (E3).
3. Draw  $(\sigma^{(j)})$  by Gibbs-sampling with equation (E4).
4. Draw  $(\beta^{(j)})$  by an efficient jump method.
5. Increase  $j$  and return to step 2.

## Appendix F: MCMC algorithm for Data Augmentation proposed by Eraker

To augment the data, we use the algorithms proposed by Eraker(2001). In short, he introduces a set of latent points in between each observation based on MCMC approach. He formulates the joint density for parameters together with observed and unobserved data, and then integrates the unobserved data out of the joint density. This method is referred to as “data augmentation”, originally formalized by Tanner and Wong (1987).

The equation I work with is given by equation (2), and it is

$$\Delta Y_t = \mu(Y_t; \theta) \Delta t + \sigma(Y_t; \theta) \Delta W_t.$$

As denoted in section 2,  $\mu$  is a drift function and  $\sigma$  is a diffusion function.  $Y_t$  includes all data which consists of both observable and non-observable data. Now what I would like to do is to substitute the missing data,  $Y_t$  with simulations  $\hat{Y}_t$ . Let  $\hat{Y}$  be the  $1 \times N$  matrix, where all missing data are replaced by simulated data. Let  $\hat{Y}_i$  denote the  $i$ -th column of  $\hat{Y}$ . Conditioning on the first observation, the joint posterior density is given by

$$\pi(\hat{Y}, \theta) \propto \prod_{i=1}^n p(\hat{Y}_i | \theta) p(\theta) \quad (\text{F1})$$

where  $p(\theta)$  is the prior density for the parameter and the likelihood function  $p(\hat{Y}_i | \theta)$  is given by

$$p(\hat{Y}_i | \theta) = |\sigma_{i-1}^2|^{1/2} \exp \left\{ -\frac{1}{2} \left\| (\Delta \hat{Y}_i - \mu_{i-1} \Delta t) \sigma_{i-1}^{-1} (\Delta t)^{-1/2} \right\|^2 \right\} \quad (\text{F2})$$



where  $\|\cdot\|$  denotes the Euclidean norm. For simplicity of notation, the following are used:  $\mu_i = \mu(\hat{Y}_i; \theta)$ ,  $\sigma_i = \sigma(\hat{Y}_i; \theta)$  and  $\sigma_{i-1}^{-2} = (\sigma_{i-1} \sigma'_{i-1})^{-1}$ .

To draw  $\hat{Y}$ , we use the same blocking schemes in the Gibbs sampling as Eraker suggested, which is the Gibbs sampling updates one column vector of  $\hat{Y}$  at a time. The density for  $\hat{Y}_i$  is defined by the following relationship

$$\pi(\hat{Y}_i | \hat{Y}_{\setminus i}, \theta) \propto p(\hat{Y}_i | \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta),$$

where  $Y_{\setminus i}$  denotes all elements of  $\hat{Y}$  except the  $i$ th column, and

$$p(\hat{Y}_i | \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta) = |\sigma_{i-1}^{-2}|^{1/2} |\sigma_i^{-2}|^{1/2} \times \exp \left\{ -\frac{1}{2} \left\| (\Delta \hat{Y}_i - \mu_{i-1} \Delta t) \sigma_{i-1}^{-1} (\Delta t)^{-1/2} \right\|^2 - \frac{1}{2} \left\| (\Delta \hat{Y}_{i+1} - \mu_i \Delta t) \sigma_i^{-1/2} (\Delta t)^{-1} \right\|^2 \right\}.^{10} \quad (\text{F3})$$

Therefore, at the  $h$ th iteration of Gibbs sampling, I draw

$$\hat{Y}_i^{(h)} \sim \pi(\hat{Y}_i | \hat{Y}_{i-1}^{(h)}, \hat{Y}_{i+1}^{(h-1)}; \theta)$$

for all  $i = 0, 1, \dots, n$ . Eraker showed the following; For a scalar process,  $Y_t \in R$  with constant drift and diffusion functions,

$$\hat{Y}_i | \hat{Y}_{i-1}, \hat{Y}_{i+1}, \theta \sim N \left( \frac{1}{2} (\hat{Y}_{i-1} + \hat{Y}_{i+1}), \frac{1}{2} \sigma^2 \Delta t \right) \quad (\text{F4})$$

where given parameter vector  $\theta = \{\mu, \sigma\}$ .

Eraker has showed that if drift and diffusion functions satisfying Assumptions A1-A4 in Eraker, the following holds:

$$\Delta t^{-1/2} \left( \hat{Y}_i - \frac{1}{2} (\hat{Y}_{i-1} + \hat{Y}_{i+1}) \right) \Rightarrow N \left( 0, \frac{1}{2} \sigma_{i-1}^2 \right)$$

---

<sup>10</sup> See Eraker(2001) for the detail of the derivation.

as  $\Delta t \rightarrow 0$ . Therefore, we have approximately

$$\hat{Y}_i | \hat{Y}_{i-1}, \hat{Y}_{i+1} \sim N\left(\frac{1}{2}(\hat{Y}_{i-1} + \hat{Y}_{i+1}), \frac{1}{2}\sigma_{i-1}^2\Delta t\right),$$

and (F4) can be used to draw latent variables.

## BIBLIOGRAPHY

- Al, M. and Van Hout B.A.(2000), A Bayesian Approach to Economic Analyses of Clinical Trials: The Case of Stenting Versus Balloon Angioplasty, *Health Economics*, 9(7):599-609.
- Aït-Sahalia, Y.(1996a), Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies*, 9(2):385-426.
- Aït-Sahalia, Y.(1996b), Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica*, 64(3):527-560.
- Aït-Sahalia, Y. (2002a), Telling from Discrete Data Whether the Underlying Continuous-Time Model Is a Diffusion, *The Journal of Finance*, 57(5):2075-2112.
- Aït-Sahalia, Y. (2002b), Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach, *Econometrica*, 70(1):223-262.
- Andersen, T.G. and J. Lund (1997), Estimating Continuous-Time Stochastic Volatility Models of the Short-term Interest Rate, *Journal of Econometrics*, 77(2):343-377.
- Bali, T.G. (2000), Modeling the Conditional Mean and Variance of the Short Rate Using Diffusion, GARCH, and Moving Average Models, *Journal of Futures Markets*, 20(8):717-751.
- Bauwens, L., M. Lubrano, and J-F Richard (2000), *Bayesian Inference in Dynamic*, Oxford University Press.
- Black,F. and M.S. Scholes(1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81(3):637-654.
- Box, G.E.P. and G.C. Tiao (1992), *Bayesian Inference in Statistical Analysis*, Wiley-Interscience.
- Brenner, R.J., R.H. Harjes and K.F. Kroner(1996), Another Look at Models of the Short-Term Interest Rate, *Journal of Financial Quantitative Analysis*, 31(1):85-107.
- Brennan, M.J. and E.S. Schwartz (1979), A Continuous Time Approach to The Pricing of Bonds, *Journal of Banking and Finance*, 3(2):133-155.
- Brennan, M.J. and E.S. Schwartz (1980), Analyzing Convertible Bonds, *The Journal of Financial and Quantitative Analysis*, 15(4):907-929.
- Briggs A.H. (1999), A Bayesian approach to stochastic cost-effectiveness analysis, *Health Economics*,8(3):257-261.
- Briggs A.H., C.Z. Mooney, D.E. Wonderling (1999), Constructing Confidence Intervals for Cost- Effectiveness Ratios: An Evaluation of Parametric and Non-parametric Techniques, *Statisticsin Medicine*, 18(23):3245-3262.
- Casella, G. and E.I. George (1992), Explaining the Gibbs Sampling, *The American Statistician*, 46(3):167-174.
- Chan, K.C., A. Karolyi, F. Longstaff and A.B. Sanders(1992), An Empirical Comparison of Alternative Models of Short-Term Interest Rates, *Journal of Finance*, 47(3):1209-1227.
- Cherubini, U., E. Luciano and W. Vecchiato (2004), *Copula Methods in Finance*, Wiley.
- Chib, S. and E. Greenberg (1995), Understanding the Metropolis-Hastings Algorithms, *The American Statistician*, 49(4):327-335.
- Cox, J., J. Ingersoll and S. Ross (1981), A Re-examination of Traditional Hypotheses

- about the Term Structure of Interest Rates, *The Journal of Finance*, 48(2):769-799.
- Dothan, L.U. (1978), On the term structure of interest rates, *Journal of Financial Economics*, 6(1):59-69.
- Durham, G.B. (2002), Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate, *Journal of Financial Econometrics*, 70(3):463-487.
- Durham, G.B. and A.R. Gallant (2002), Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes, *Journal of Business and Economic Statistics*, 20(3): 297-316.
- Eraker, B. (2001) MCMC Analysis of Diffusion Models with Application to Finance, *Journal of Business and Economic Statistics*, 19(2):177-190.
- Gallant, A.R. and G. Tauchen, (1996), Which Moments to Match? *Econometric Theory*, 12(4):657- 681.
- Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin (2003), *Bayesian Data Analysis*, Chapman and Hall.
- Geman, S. and D. Geman (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721-741.
- Genest C., and L-P. Rivest (1993), Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of the American Statistical Association*, 88(423):1034-1043.
- Gelfand, A.E. and D.K. Dey (1994), Bayesian Model Choice: Asymptotic and Exact Calculations, *Journal of Royal Statistical Society*, 56(3):501-514.
- Gelfand, A.E. and S.K. Ghosh (1998), Model choice: A Minimum Posterior Predictive Loss Approach, *Biometrika*, 85(1): 1-11.
- Geweke J (2005), *Contemporary Bayesian Econometrics and Statistics*, Wiley-Interscience.
- Greene, W.H. (2000), *Econometric Analysis* (Fourth Edition), Prentice Hall.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.
- Hastings, W.K. (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 57(1):97-109.
- Heitjan D.F., H. Li (2004), Bayesian Estimation of Cost-effectiveness: An Importance-sampling Approach, *Health Economics*, 13(2):191-198.
- Heitjan D.F., A.J. Moskowitz, and W. Whang (1999), Bayesian Estimation of Cost-effectiveness Ratios from Clinical Trials, *Health Economics*, 8(3):191-201.
- Hoch J.S., A.H. Briggs, and A.R. Willan (2002), Something Old, Something New, Something Borrowed, Something Blue: A Framework for The Marriage of Health Econometrics and Cost-Effectiveness Analysis, *Health Economics*, 11(5):415-430.
- Ibrahim J.G., M. Chen, and D. Sinha (2001), *Bayesian Survival Analysis*, Springer.
- Jacquier, E., N.G. Polson, and P.E. Rossi (1994), Bayesian Analysis of Stochastic Volatility Models, *Journal of Business and Economic Statistics*, 12(4):371-89.
- Jacquier, E., N.G. Polson, and P.E. Rossi (2004), Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors, *Journal of Econometrics*, 122(10):185-212.

- Jensen, B. and R. Poulsen (1999), A Comparison of Approximation Techniques for Transition Densities of Diffusion Process, Working Paper Series No.38, Center for Analytical Finance, University of Aarhus.
- Jones, C.S. (2003), Nonlinear Mean Reversion in the Short-Term Interest Rate, *Review of Financial Studies*, 16(3):793-843.
- Johannes and Polson(2002), MCMC methods for Financial Econometrics, forthcoming in *the Handbook of Financial Econometrics*, edited by Yacine Ait-Sahalie and Lars Hansen.
- Kariya,T. and R.Y Liu (2002), *Asset Pricing-Discrete Time Approach*, Kluwer Academic Publishers.
- Merton, R.C. (1973), Theory of rational option pricing, *Bell Journal of Economics and Management Science*, 4 (1):141-183, available in Merton eds.(1990) *Continuous-time Finance*, Blackwell.
- Kim, S., N. Shephard and S. Chib (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65(4):406-410.
- Koedijk, K.G., F.G.J.A. Nissen, P.C. Schotman and C.C.P. Wolff (1997), The Dynamics of Short-Term Interest Rate Volatility Reconsidered, *European Finance Review*, 1(1): 105-130.
- Koop, G.(2003), *Bayesian Econometrics*, Wiley-Interscience.
- Koop, G. D.J. Poirier, and J.L. Tobias (2007), *Bayesian Econometric Methods*, Cambridge University Press.
- Lancaster, T.(2004), *Introduction to Modern Bayesian Econometrics*, Blackwell.
- Longstaff, F.A. and E. S. Schwartz (1995), A Simple Approach to Valuing Risky Fixed and Floating Rate Debt, *The Journal of Finance*, 1(3):789-819.
- Mahieu, R.J. and P.C. Schotman (1998), An Empirical Application of Stochastic Volatility Models, *Journal of Applied Econometrics*, 13:333-360.
- Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H.Teller, and E.Teller (1953), Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, 21(6):1087-1092.
- Nakatsuma, T. (2000) Bayesian analysis of ARMA--GARCH models: A Markov Chain Sampling Approach, *Journal of Econometrics*, 95(1):57-69.
- Nelson R.B.(2006), An Introduction to Copulas. Springer.
- O'Hagan, A. and J.W. Stevens (2001), A Framework for Cost-effectiveness Analysis from Clinical Trial data, *Health Economics*, 10(4):303-315.
- O'Hagan, A., J.W. Stevens, and J. Montmartin (2001), Bayesian Cost-effectiveness Analysis from Clinical Trial Data, *Statistics in Medicine*, 20(5):733-753.
- Quinn, C.(2005), Generalisable Regression Methods for Cost-effectiveness Using Copulas. HEDG Working Paper 05/13. Health, Econometrics and Data Group, The University of York.
- Robert, C.P. and G. Casella(1999), *Monte Carlo Statistical Methods*, Springer.
- Romeo, J.S., N.I. Tanaka, and A.C. Pedroso-de-Lima (2006), Bivariate Survival Modeling: A Bayesian Approach Based on Copulas, *Lifetime Data Analysis*, 12(2):205-222.
- Shemyakin, A.E., and H. Youn (2006), Copula Models of Joint Last Survivor Analysis, *Applied Stochastic Models in Business and Industry*, 22(2):211-224.
- Tanner, M.A. and W.H. Wong (1987), The Calculation of Posterior Distributions by Data

- Augmentation, *Journal of the American Statistical Association*, 82(398):528-540.
- Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics*, 22(4):1701-1728.
- Vanness, D.J., and W.R. Kim (2002), Bayesian Estimation, Simulation and Uncertainty Analysis: the Cost-effectiveness of Ganciclovir Prophylaxis in Liver Transplantation, *Health Economics*, 11(6):551-566.
- Vasicek, O.(1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5(2):177-188.
- Vazquez-Polo, F.J., M.A. Negrin Hernandez, and B.G. Lopez-Valcarcel (2005), Using Covariates to Reduce Uncertainty in the Economic evaluation of Clinical Trial Data, *Health Economics*, 14(6): 545-557.
- Vazquez-Polo, F.J M., Negrin, X. Badia, and M. Roset (2005), Bayesian Regression Models for Cost-effectiveness Analysis, *European Journal of Health Economics*, 6(1):45-52.
- Willan A.R., A.H. Briggs, and J.S. Hoch (2004), Regression Methods for Covariate Adjustment and Subgroup Analysis for Mon-censored Cost-effectiveness Data. *Health Economics*, 13(5):461-475.
- Watatnabe, T. (2000), Bayesian Analysis of Dynamic Bivariate Mixture Models: Can They Explain the Behavior of Returns and Trading Volume? *Journal of Business and Economic Statistics*, 18(2):199-210.
- Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley & Sons.

## CURRICULUM VITA

Yoko Ibuka

|           |   |
|-----------|---|
| 2004-2008 | Ph.D. in Economics, Rutgers University  |
| 2002-2004 | M.A. in Economics, Rutgers University   |
| 2000-2002 | M.A. in Economics, Keio University, Tokyo, Japan  |
| 1996-2000 | B.A. in Economics, Keio University, Tokyo, Japan  |
| 2005-2008 | Graduate Assistant, Institute for Health, Health Care Policy and Aging Research, Rutgers University |
| 2006      | Consultant, World Bank  |
| 2004-2005 | Teaching Assistant, Department of Economics, Rutgers University                                     |
| 2002      | Research Assistant, Faculty of Economics, Keio University, Tokyo, Japan                             |

Russell L.B., Y. Ibuka, and K.G. Abraham, Health-Related Activities in the American Time Use Survey, *Medical Care* 2007; 45(7): 680-685.

Yoshino N., S.Kaji and Y. Ibuka, The Stabilization of an Open Economy with Capital Controls: An Analysis Using Malaysian Data, *Asian Economic Papers* 2003; 2(3): 63-83.