ESSAYS ON BAYESIAN INFERENCE OF TIME-SERIES AND ORDERED PANEL DATA MODELS

BY JEEHYUN PARK

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 approved by

New Brunswick, New Jersey

October, 2012

ABSTRACT OF THE DISSERTATION

Essays on Bayesian inference of time-series and ordered panel data models

by Jeehyun Park

Dissertation Director: Professor Hiroki Tsurumi

At the heart of my dissertation is the study of Markov chain Monte Carlo algorithms and their applications.My dissertation consists of three essays as follow.

The first chapter is on MCMC algorithms for the dynamic ordered probit model with random effects. I have tried to estimate the model with four representative MCMC algorithms: two algorithms by Albert and Chib (1993) and Albert and Chib (2001), Liu and Sabatti (2000), and Chen and Dey (2000). I have found that the autocorrelations still remain high in the cutoffs compared to other parameters even though the levels of autocorrelation are reduced in the algorithms by Liu and Sabatti (2000), and Chen and Dey (2000).

In the second chapter, I have developed the dynamic ordered probit model studied in the first chapter. It is natural for panel data to have missing data problem because there is no guarantee that subjects will stay over the study periods. This chapter provides Bayesian statistical methods that permits non-ignorable missing data in panel datasets. In order to incorporate non-random missing data in the model, I jointly model observed and non-ignorable missing ordinal data with selection model approach. In the empirical section, I have used the model to examine determinants of self-rated

health of old people in the Health and Retirement Study. I have concluded that in this elderly American population, the longest occupation that respondents have held over their careers is strongly associated with self-rated health.

In the third chapter of my dissertation, I analyze financial time-series data before and after the Wall Street meltdown in 2008. In this chapter, I develop MCMC algorithms for the CKLS model and examine (1) time-series characteristics of the credit default swap index, stock index and federal funds rate from January 2007 to September 2009, the highly volatile period. (2) The lead-lag relationship between the credit default swap and stock markets are examined using the CKLS model employing multivariate analysis.

Acknowledgements

I always think that I am a lucky person. The luckiest moment in my graduate study was to be able to meet two great mentors: my advisor, Professor Hiroki Tsurumi, and my dissertation committee, Professor Louise B. Russell.

First, I would like to thank Professor Tsurumi for his numerous invaluable suggestions helping me get around impediments and his supportive encouragement at all times. The strategies he suggested at every step during my graduate study, regarding research, teaching and other academic endeavors, were infallible, to say the least. All my thanks to him for being on my side in every difficult moment.

It would be an injustice to and a gross understatement of my profound respect for Professor Russell if I just state that I wish to express my sincerest thanks to her. Of course, I really thank her for her precious comments on my drafts. In every moment of my distress, she was there with me by providing her sympathetic ear. As she showed me, I would like to follow her: being sincere to research and being warm.

I also would like to thank my dissertation committee, Professor John Landon-Lane , for his valuable comments on my dissertation.

Dedication

To my family

Table of Contents

Abstract ii

Acknowledgements iv

Dedication v

List of Tables ix

List of Figures xi

1. Introduction 1

2. Bayesian inference of a dynamic ordered probit model with random effects 3

2.1. Introduction 3

2.2. Dynamic Ordered Probit with Random Effects 6

2.2.1. Model 6

2.2.2. Bayesian Inferences 7

2.2.3. Identification 8

2.3. Algorithms for the Estimation of the Bayesian Dynamic Panel-Ordered Probit Model 9

2.3.1. Algorithm1: Albert and Chib (1993) 10

2.3.2. Algorithm2: Liu and Sabatti (2000) 11

2.3.3. Algorithm3: Albert and Chib (2001) 11

2.3.4. Algorithm4: Chen and Dey (2000) 12

2.4. Results with Simulated Data 13

2.4.1. Data Generating Process 13

2.4.2. MCMC Results 13

2.5. Conclusion 14

3. Bayesian inference of an ordered probit model with non-ignorable missing data: The determinants of self-rated health using the Health and Retirement Study 16

3.1. Introduction 16

3.2. Missing data mechanism 18

3.3. Models 22

3.3.1. Complete-data model 22

3.3.2. Incomplete-data model under MAR 23

3.3.3. Selection model for non-ignorable missing data mechanisms . . . 25

3.4. Bayesian inference 27

3.4.1. Bayesian inference for complete-data model 27

3.4.2. Bayesian inference under ignorable missing data mechanisms . . 27

Numerical example with simulated data 28

3.4.3. Bayesian inference under non-ignorable missing data mechanisms 31

Numerical example with simulated data (continued) 32

3.5. Empirical application: Self-rated health in the Health and Retirement Study 34

3.5.1. Self-rated health 34

3.5.2. Model and variables 39

Data 39

Self-rated health 40

Explanatory variables 40

1. Socioeconomic variables: 40

2. Health risk factors: 41

3. Other controls: 41

3.5.3. Results 42

Analysis1: Explanation of self-rated health using only baseline covariates 43

Analysis2: Explanation of self-rated health using time-varying covariates 45

Analysis3: Explanation of self-rated health using additional covariates 46

3.6. Conclusion 48

4. Time-series Characteristics and Lead-lag Relations of Credit Default Swap, S&P500, and Overnight Federal Funds Rate during the Great Recession 63

4.1. Introduction 63

4.2. Time-series analysis 67

4.2.1. CKLS-ARMA-GARCH-EPD model 67

4.2.2. Empirical Results 69

4.3. Lead-Lag Relationship 75

4.3.1. Model 78

4.3.2. Empirical Results 78

4.4. Conclusions 83

4.A Appendix 92

4.A.1 MCMC Algorithms 92

4.A.2 Credit Default Swap 95

4.A.3 Credit Default Swap Index (CDX) 98

4.A.4 History of Regulation and Deregulation of the U.S. Financial Markets 101

Bibliography 106

Vita 112

List of Tables

2.1. Posterior results (Simulated data) 15

3.1. Inference for ignorable MDM with MNAR data 31

3.2. Inference for nonignorable MDM with MNAR data 33

3.3. Analysis1: Average partial effects on excellent self-rated health (including only baseline covariates) 43

3.4. Analysis2: Average partial effects on excellent self-rated health (using time-varying covariates) 45

3.5. Analysis3: Average partial effects on excellent self-rated health (using additional covariates) 47

3.6. Posterior summary statustics for missing data mechanisms, ψ 48

3.7. Missing data 52

3.8. Summary statistics 53

3.9. Summary statistics by self-rated health (All waves) 54

3.10. Summary statistics by self-rated health (Wave1) 55

3.11. Summary statistics by self-rated health (Wave7) 56

3.12. Analysis1 (Wave2-7): Average partial effects on poor and excellent self-rated health 57

3.13. Analysis1 (Baseline): Average partial effects on poor and excellent self-rated health 58

3.14. Analysis2 (Wave2-7): Average partial effects on poor and excellent self-rated health 59

3.15. Analysis2 (Baseline): Average partial effects on poor and excellent self-rated health 60

3.16. Analysis3 (Wave2-7): Average partial effects on poor and excellent self-rated health 61
3.17. Analysis3 (Baseline): Average partial effects on poor and excellent self-rated health 62
4.1. CKLS-ARMA-GARCH-EPD Models of CDX, SP500, and Federal Funds rates 88
4.2. Adjusted CKLS model with two variables 89
4.3. VAR Models with one lag 90
4.4. VAR Model with 3 lags 91

List of Figures

3.1. Self-rated health by wave 51

4.1. Posterior pdf's of α_0 for S&P500 and CDX 86

4.2. Posterior pdf's of $\alpha\beta$ 87

4.3. Semi-annual data for CDS notional amounts outstanding (Source: ISDA 2009) 96

Chapter 1

Introduction

Bayesian statistics began with a posthumous publication in 1763 by Thomas Bayes, a non-conformist minister from the small English town of Tunbridge Wells. His work was formalized as Bayesian theorem which, when expressed mathematically, is a simple and uncontroversial result in probability theory. However, specific uses of the theorem have been the subject of continued controversy for over a century, giving rise to a steady stream of polemical arguments in a number of disciplines (Spiegelhalter et al. (2004)).

The basic idea of Bayesian analysis is reasonably straightforward. Using information from earlier studies or expert opinion, it begins by defining a prior probability distribution for the parameter of interest. Or the prior can be one of several distributions that are conventionally used to represent no previous information. The prior distribution is then updated by combining it with data, which is represented by a likelihood function. The result is the posterior probability distribution, which combines the earlier information and the new data into a single probability distribution that describes the possible values of the parameter and the probability of each value based on the data.

Hence, the objective of Bayesian is to obtain the posterior probability distribution of each model parameter. To make inference on the parameters, Bayesian inference heavily relies on high-dimensional integration over the posterior distribution of other model parameters, which is hard to calculate analytically. Markov chain Monte Carlo (MCMC) algorithm, a simulation-based integration using Markov chain, could provide a solution to the problem. MCMC algorithms enable us to draw samples from probability distributions by constructing a Markov chain that has the target distribution as its equilibrium distribution. Since late 1980s, MCMC algorithms have been widely used in many fields from natural sciences to social sciences.

The introduction of MCMC algorithms has made Bayesian inference more approachable and flexible than frequentist sample theory inference. The advantages of Bayesian inference employing MCMC algorithms are: 1) Bayesian incorporates prior knowledge formally into data analysis; (2) important parameters for policy decisions such as elasticity and marginal effect can be easily obtained by MCMC algorithms; and (3) using predictive densities, economic business forecasts can be obtained not only as point forecasts but also as forecast densities.

In my dissertation, I attempt to devise the MCMC algorithms. Then, I apply my MCMC algorithms to real data and explain the implications.

Part II of the dissertation consists of three essays on the application of Bayesian inference using MCMC to topics in economics. The first chapter is on MCMC algorithms for the ordered probit model to analyze panel data using Bayesian approach. In the second chapter, I develop the ordered probit model studied in the first chapter by incorporating non-random missing data analysis in the model. Then, I study the determinants of self-rated health in Health and Retirement study from 1992 to 2004. In the third chapter, I analyze financial time-series data before and after the Wall Street meltdown in 2008 using the CKLS model in Bayesian.

Chapter 2

Bayesian inference of a dynamic ordered probit model with random effects

2.1 Introduction

According to Green and Hensher (2010), the ordered response model has been developed through three discernible steps in the literature. In step 1, Aitchison and Silvey (1957) treatment of stages in the life cycle of a certain insect; in step 2, Snell (1964)'s analysis of ordered outcomes (without a regression interpretation); and in step 3, McKelvey and Zavoina (1975)'s proposal of the modern form of the *ordered probit* regression model. From the frequentist viewpoint, the ordered probit model can be estimated by using the maximum likelihood (ML) method. A Bayesian approach of the ordered probit model builds on the estimation for the binary probit model, which is pioneered by Tanner and Wong (1987) and Albert and Chib (1993). This paper is based on the inferential framework of Albert and Chib (1993) by using the latent variable representation in MCMC algorithm.

The fundamental advantage of a panel dataset over a cross-section is that it will allow us greater flexibility in modeling differences in behavior across individuals. Hence, the major motivation for using a panel dataset is its ability to control for individual heterogeneity. The literatures on panel data place emphasis on the fixed and random effects models in order to specify the unobserved individual heterogeneity. Consider a standard linear panel regression model:

$$y_{it} = x'_{it}\beta + b_i + \varepsilon_{it}$$

where x_{it} includes k covariates, but a constant term; and b_i is the heterogeneity, or individual effects. The individual effects (b_i) is specified as random effects, but it could be specified as the fixed effects, to be estimated together with β. An important consideration to determine fixed effects or random effects model is whether the covariates (x_{it}) are correlated with the individual effects (b_i).

If the unobserved individual effects b_i is correlated with covariates x_{it}, then it is specified with fixed effects, i.e., the fixed effects model allows the unobserved individual effects to be correlated with the covariates. If this model, in which b_i is unobserved, but correlated with x_{it}, is estimated by the least squares, then the least squares estimator of β is biased and inconsistent. Adding a dummy variable for each individual will solve the problem, but the least square dummy variable approach (LSDV) may be prohibitive if there are a large number of cross-section observations. When the unobserved individual effects are uncorrelated, the model leads to the random effects model Greene (2002). In sum, the crucial distinction between these two cases is whether the unobserved individual effects embodies elements that are correlated with the covariates in the model.

My work using the panel dataset of the Health and Retirement Study (HRS) has to deal with the existence of unobserved heterogeneity and with the need to use nonlinear models to employ ordered discrete dependent variables, i.e., the ordered probit model with panel data. As indicated above, two panel models are possible: the ordered probit model with fixed effects or random effects. When the frequentist methods are employed, there are two problems that this ordered probit model with fixed effects shares with other non-linear fixed effects models. First, regardless of how estimation and analysis are approached, time-invariant variables are not identified. Since I am interested in the effects of demographic variables such as gender, education level, etc. that are time invariant, this is likely to be a significant obstacle. Second, there is no sufficient statistic available to condition the fixed effects out of the model. That would imply that in order to estimate the model, one must maximize the full log likelihood. If the sample is small enough, one may simply insert the individual group dummy variables and treat the entire pooled sample as a cross-section. I am interested in the longitudinal dataset in which this would not be feasible due to a large sample size.

The larger methodological problem with fixed effects model approach would be the incidental parameters problem[1]. This is even more severe when estimating dynamic models as mine, the dynamic ordered probit model. The incidental parameters problem is reflected in the inconsistency of standard estimates like maximum likelihood estimator (MLE) when the number of individuals N goes to infinity while T is fixed. Even when T goes to infinity, if it does at a smaller or the same rate as N, the asymptotic normal distribution is not centered at zero due to the bias coming from the incidental parameters. Moreover, this problem results in large finite sample biases of the MLE when using panels where T is not very large. Recent proposals for the bias reduction methods can be grouped in three approaches: (1) to construct an analytical or numerical bias correction of a fixed effects estimator, (2) to correct the bias in moment equations, and (3) to correct the objective function.

A random effects approach also has the drawbacks of imposing a strong assumption of independence between the unobserved heterogeneity and other covariates: consistency requires that the effects be uncorrelated with the included variables. It also has the drawback of having to deal with the so-called initial condition problem when estimating the dynamic ordered probit model. Taking a fixed effects approach relaxes the independent assumptions between individual effects and the covariates, and also allows that there is no initial condition problem. Despite these advantages, there have been only few applications in health economics of nonlinear panel models with the fixed effects, as in Jones (2000). This is due to the difficulty of solving the incidental parameters problem.

McCulloch and Rossi (1994) state that in a Bayesian point of view, there is no distinction between fixed and random effects models, only between hierarchical and

[1]Because of the individual-specific fixed effects, the total number of parameters in the models with fixed effects equals the number of individuals plus the dimension of the common parameters. When the number of individuals (N) goes to the intinity and the time-series dimension (T) is fixed, the maximum likelihood estimator (MLE) typically results in inconsistent estimates of the common parameter of interest (Neyman and Scott (1948)).

non-hierarchical models. This paper estimates the dynamic ordered probit model with random effects by employing four Bayesian MCMC algorithms as in Hasegawa (2009). The first algorithm developed by Albert and Chib (1993) presents Bayesian implementations of the ordered probit model using the Gibbs sampler applying the data augmentation method of Tanner and Wong (1987).

Hoever, Albert and Chib (1993)'s method of data augmentation has the problem of the high autocorrelation in the estimated cutoff points, c_j in equation (**??**). Along with Albert and Chib (1993)'s algorithm, this paper employs additional three algorithms that present the ways to mitigate the high autocorrelation. First, Liu and Sabatti (2000) develop the generalized Gibbs sampling approach (Algorithm 2); Albert and Chib (2001) transforms the cutoff points and employs the MH-within-Gibbs (Algorithm3); and Chen and Dey (2000) present one of the approaches to reduce high autocorrelation between cutoff points, which is "reparameterization" with the different identification restriction (Algorithm4).

2.2 Dynamic Ordered Probit with Random Effects

2.2.1 Model

The dynamic ordered probit model can be used to model a discrete dependent variable that takes ordered multinomial outcomes, e.g., $y_{it} = 1, ..., J$. Let y^*_{it} $(-\infty < y^*_{it} < \infty)$ be the underlying latent variable for a respondent i at time t for $t = 0, 1, ..., T$ and $i = 1, ..., N$. The model can be expressed as

$$y_{it} = j if c_j \leq y^*_{it} < c_{j+1}, j = 1, ..., J \tag{2.1}$$

where c_j is a cutoff point of ordinal responses and is specified as

$$-\infty < c_1 = 0 < c_2 < \ ... \ < c_{J-1} < \infty$$

Define y^*_{it} as the underlying latent variable as follows

$$\begin{aligned} y^*_{it} &= \phi y^*_{it-1} + v_{it}\beta_1 + x_i\beta_2 + b_i + \varepsilon_{it}, t = 2, ..., T \qquad (2.2) \\ y^*_{i0} &= v_{i0}\beta_{10} + x_i\beta_{20} + b_i + \varepsilon_{i0}, t = 1 \end{aligned}$$

where $v_{it} = (v_{it1}, ..., v_{itk})$ is the time-variant covariates, $x_i = (x_{i1}, ..., x_{ip})$ is the cross-sectional time-invariant covariates, b_i is an individual-specific and time-invariant random effects, $\beta_1 = (\beta_{11}, ..., \beta_{1k})'$, $\beta_2 = (\beta_{21}, ..., \beta_{2p})'$, $\beta_{10} = (\beta_{101}, ..., \beta_{10k})'$, and $\beta_{20} = (\beta_{201}, ..., \beta_{20p})'$. ε_{it} is a time and individual-specific error term which is assumed to be normally distributed and uncorrelated across individuals and times and uncorrelated with b_i. b_i is distributed with a mean and constant variance, σ^2, and independent of ε_{it} for all t. ε_{it} is assumed to be strictly exogenous, that is, the ε_{it} are uncorrelated with ε_{is} for all t and s (Hasegawa (2009)).

Conditioned on the individual effects b_i, the observations on y_{it}, $t = 1, ..., T$, are assumed to be independent. Then, the contribution to the likelihood for individual i, conditional on the covariates and the individual effects, would be the joint probability

$$\begin{aligned} & p\left(y_{i1} = j_{i1}, ..., y_{iT} = j_{iT} | v_{it}, x_i, b_i\right) \\ = & \prod_{t=2}^{T}[\Phi\left(c_j - \phi y_{it-1}^* - v_{it}\beta_1 - x_i\beta_2 - b_i\right) - \Phi\left(c_{j-1} - \phi y_{it-1}^* - v_{it}\beta_1 - x_i\beta_2 - b_i\right)] \\ & \times\left[\Phi\left(c_j - v_{i1}\beta_{10} - x_i\beta_{20} - b_i\right) - \Phi\left(c_{j-1} - v_{i1}\beta_{10} - x_i\beta_{20} - b_i\right)\right] \end{aligned}$$

where $\Phi\left(.\right)$ is the normal distribution function.

2.2.2 Bayesian Inferences

Let $\Theta = \{\phi, \beta_1, \beta_2, \beta_{10}, \beta_{20}, b, \mu, \tau, c\}$. The prior distribution of the random effects b is assumed to have a hierarchical structure. Then, the prior distributions are specified as follows:

$$p\left(\Theta\right) = p\left(\phi\right) p\left(\beta_1\right) p\left(\beta_2\right) p\left(\beta_{10}\right) p\left(\beta_{20}\right) p\left(b|\mu, \tau\right) p\left(\mu\right) p\left(\tau\right) p\left(c\right)$$

The joint posterior distribution for Θ is

$$\begin{aligned} p\left(\Theta, y^*, y_0^* | y, y_0\right) & \propto p\left(\Theta, y, y_0\right) p\left(y, y_0 | \Theta, y^*, y_0^*\right) \\ & = p\left(\Theta\right) \prod_{i=1}^{N} \{p\left(y_i^* | \Theta_{-\beta_0}, y_{i0}^*\right) \cdot p\left(y_i | \Theta_{-\beta_0}, y_{i0}^*\right) \\ & \quad \cdot p\left(y_{i0}^* | \Theta_{-\phi,\beta}\right) \cdot p\left(y_{i0} | \Theta_{-\phi,\beta}, y_{i0}^*\right) \end{aligned}$$

(2.4)

where $\Theta_{-\beta_0} = \{\phi, \beta_1, \beta_2, b, \mu, \tau, c\}$ except β_{10} and β_{20}, and $\Theta_{-\phi,\beta} = \{\beta_{10}, \beta_{20}, b, \mu, \tau, c\}$.

2.2.3 Identification

In equation (2.2), we assume that the error terms, ε_{it}, has a mean zero and a standard normal distribution with variance one. The latent variable y_{it}^* is the unobservable, and only the ordered responses, y_{it} $(y_{it} = 1, 2, ..., J)$, are observed. Therefore, equation 2.2 links the observed ordered responses with the latent variable. Two restrictions are necessary in order to uniquely identify the parameters of the model. First, suppose $c_j^* = c_j + a$ and $\phi^* y_{it-1}^* + v_{it}\beta_1^* + x_i\beta_2^* + b_i^* = \phi y_{it-1}^* + v_{it}\beta_1 + x_i\beta_2 + b_i + a$ for some constant a. Then, because $c_j^* - (\phi^* y_{it-1}^* + v_{it}\beta_1^* + x_i\beta_2^* + b_i^*) = c_j - \left(\phi y_{it-1}^* + v_{it}\beta_1 + x_i\beta_2 + b_i\right)$, it is straightforward to verify that $p\left(y_{it} = j|\phi, \beta_1, \beta_2, b_i\right) = p\left(y_{it} = j|\phi^*, \beta_1^*, \beta_2^*, b_i^*\right)$. This identification problem is usually corrected by fixing a cutoff point (in addition to $c_0 = -\infty$ and $c_J = \infty$), in particular, letting $c_1 = 0$ removes the possibility for shifting the distribution without changing the probability of observing y_{it}.

Second, suppose the variance of ε_{it} is scaled by an unrestricted parameter σ_ε^2. The latent regression 2.2 will be $y_{it}^* = \phi y_{it-1}^* + v_{it}\beta_1 + x_i\beta_2 + b_i + \sigma_\varepsilon\varepsilon_{it}$. However, $(y_{it}^*/\sigma_\varepsilon) = \phi\left(y_{it-1}^*/\sigma_\varepsilon\right) + v_{it}\left(\beta_1/\sigma_\varepsilon\right) + x_i\left(\beta_2/\sigma_\varepsilon\right) + (b_i/\sigma_\varepsilon) + \varepsilon_{it}$ is the same model with the same data: the observed data will be unchanged. This means that there is no information about σ_ε in the data so that it cannot be estimated. This usual approach to achieve identification in this case is to fix the variance of ε, σ_ε. For example, in the case of the probit model, ε_{it} is assumed to be normally distributed with the variance of 1. The algorithms of Albert and Chib (1993), Liu and Sabatti (2000), and Albert and Chib (2001) follow this standard approach. Instead of fixing the variance of ε_{it} to be 1, it is also possible to restrict one of the cutoff points in addition to $c_1 = 0$. For example, Chen and Dey (2000) restrict the cutoff point c_2 to be 1. This restriction precludes the simultaneous rescaling of the numerator and denominator in $\Phi\left(\frac{c_j - \left(\phi y_{it-1}^* + v_{it}\beta_1 + x_i\beta_2 + b_i\right)}{\sigma_\varepsilon}\right)$ because it would violate $c_2 = 1$.

2.3 Algorithms for the Estimation of the Bayesian Dynamic Panel-Ordered Probit Model

In theory, one could directly apply standard computational tools such as the Gibbs sampler coupled with a few Metropolis-within-Gibbs steps to fit the model. However, it has been shown that use of the standard Gibbs sampler in the models with ordered responses (Albert and Chib (1993)) suffers from slow mixing due to a high correlation between the simulated cutoff points and latent variables. Cowles (1996) noted the possibility that sampling of the cutoff points conditional on the latent variables can lead to small changes in the cutoff points between successive iterations, especially as more data become available. Li and Tobias (2006) observed only very small local movements from iteration to iteration when the chain mixes slowly. As a result, it may take a very long time for the simulator to traverse the entire parameter space. When the lagged autocorrelation between the simulated parameters are very high, estimates of posterior features may be quite inaccurate, and numerical standard errors associated with those estimates will be unacceptably large.

To mitigate this slow mixing problem and move closer to a simulation where one can obtain i.i.d samples from the posterior distribution, Cowles (1996) suggests sampling the latent data, y^*, and the cutoff points, c, jointly by drawing from $c \sim p(c|y, \beta)$ marginalized over the latent variable and, subsequently, sampling $y^* \sim p(y^*|y, \beta, c)$. Nandram and Chen (1996) improve Cowles (1996) that the cutoff points should be sampled jointly, not one-at-a time, and that the particular Metropolis-Hasting (MH) proposal density suggested in Cowles (1996) may be difficult to tune. Nandram and Chen (1996) suggest a reparameterization of the model and present a sampler that allows joint sampling of the reparameterized cut points in a single block and also marginally of the latent variable using a Dirichlet proposal density that depends on the previous cutoff points, but does not depend on the other parameters or the latent variable. Chen and Dey (2000) point out that the Dirichlet density will generally work well when the cell (category) counts are balanced, but may fail to serve as a good proposal density when the category counts are unbalanced.

Subsequent works, including Chen and Dey (2000) and Albert and Chib (2001), are built upon these ideas. They show that the cutoffs c can easily be sampled jointly in a single block by well-tailored independent chains, marginally of y^*, to improve the efficiency of MCMC algorithms. Maintaining the identification restriction that the variance of the error terms equals 1, Albert and Chib (2001) simplify the sampling of the cutoffs c by transforming them so as to remove the ordering constraint by the one-to-one map.

In order to reduce high autocorrelation for the cutoff points, Liu and Sabatti (2000) also propose the generalized Gibbs sampler that provides a framework encompassing methods of the parameter expansion and reparameterizing such as Albert and Chib (2001) and Chen and Dey (2000). The generalized version of the Gibbs sampler is based on conditional moves along the traces of groups of transformations in the sample space. The details of the four algorithms employed are as follows. The details of the four algorithms employed are as follow.

2.3.1 Algorithm1: Albert and Chib (1993)

Applying the data augmentation idea of Tanner and Wong (1987), Albert and Chib (1993) treat the unknown latent variable y^* values as additional parameters to be simulated in the Gibbs sampler. Once values are obtained for y^*, the problem of estimating β in the ordered probit model simplifies to that of doing so in a standard normal linear model. In order that the domain of the y^* may be the entire real line, the cutoff points, c_0 and c_J must be fixed at $-\infty$ and ∞, respectively. Albert and Chib (1993) also note that in order to make the parameters of the model identifiable, one additional cutpoint must be fixed: without loss of generality, c_1 is fixed at 0. The algorithm of drawing the parameters in the dynamic ordered probit model based on the method of Albert and Chib (1993) is as follows.

1. Sample $\phi,\ \beta,\ \beta_0,\ b,\ \mu,\ \tau$ from their full conditional distributions.

2. Sample $y^*,\ y_0^*$ from their full conditional distributions.

3. Sample c from the full conditional distribution.

2.3.2 Algorithm2: Liu and Sabatti (2000)

Liu and Sabatti (2000)'s generalized version of the Gibbs sampler is based on conditional moves along the traces of groups of transformations in the sample space. Liu and Sabatti (2000) indicate that the Gibbs sampler employed Albert and Chib (1993) suffers from high autocorrelation: conditional on the values of the latent variable, y^*_{it}, and the other cutoff points c_{-j}, the cutoff point c_j has very little room to move.

The algorithm of the generalized Gibbs samplers in the dynamic panel ordered probit model employing the method of Liu and Sabatti (2000) is as follows:

1. Sample $\phi,\ \beta,\ \beta_0,\ b,\ \mu,\ \tau$ from the full conditional distributions.
2. Sample $y^*,\ y^*_0$ from their full conditional distributions.
3. Sample c from the full conditional distribution.
4. ImplementLiu and Sabatti (2000)'s generalized Gibbs sampler.

2.3.3 Algorithm3: Albert and Chib (2001)

Albert and Chib (2001) proposed the algorithm for drawing the cutoff points by transforming them as follows:

$$\xi_c = \log(c_j - c_{j-1}),\ j = 2, ..., J-1$$

where $\xi = [\xi_2, ..., \xi_{C-1}]'$ is unrestricted. The algorithm of drawing the parameters in the dynamic ordered probit model is as follows.

1. Sample $\phi,\ \beta,\ \beta_0,\ b,\ \mu,\ \tau$ from the full conditional distributions.
2. Sample $y^*,\ y^*_0$ from their full conditional distributions.
3. Sample ξ from the Metropolis-Hastings (MH) algorithm.
4. Calculate $c_j = \sum_{j=1}^{J} \exp(\xi_j),\ j = 2,\ ...\ ,\ J-1$

2.3.4 Algorithm4: Chen and Dey (2000)

Chen and Dey (2000) consider other transformations of the cutoff points and introduce alternative identification restrictions instead of the traditional one (i.e., variance of error terms is 1 so that error terms are normally distributed with $N(0,1)$). In particular, Chen and Dey (2000) leave the variance of error terms as σ^2 as an unrestricted parameter to be estimated, but instead fix another cutoff in addition to having $c_0 = -\infty,\ c_1 = 0,$ and $c_J = +\infty$ in order to determine the scale of the latent data. Li and Tobias (2008)indicate that there are several advantages of working with this reparameterization. First, the rescaling helps to mitigate correlation between the simulated cutoff points and latent variable and thus improves the performance of the posterior simulator. Second, the reparameterization effectively eliminates one cutoff point from each equation in the model. However, the main drawback to working with the reparameterized model is that it requires us to place priors on the transformed parameters Li and Tobias (2006). Under the identification, the model is modified as

$$
\begin{aligned}
y^*_{it}\ |. &\sim N(\phi y^*_{it-1} + v_{it}\beta_1 + x_i\beta_2 + b,\ \sigma^2), t = 2, ..., T \\
y^*_{i0}|. &\sim N(v_{i0}\beta_{10} + x_i\beta_{20} + b_i,\ \sigma^2), t = 1
\end{aligned}
$$

for $i = 1\ ,\ ...\ ,\ N$. Further, the cutoff points are transformed as

$$\xi_j = \log(\frac{c_j - c_{j-1}}{1 - c_j}),\ j = 2, ..., J-2$$

where $\xi = [\xi_2,\ ...\ , \xi_{J-1}]'$ is unrestricted. The prior distributions are specified as follows

$$p(\Theta) = p(\phi)\, p(\beta_1)\, p(\beta_2)\, p(\beta_{10})\, p(\beta_{20})\, p(b|\mu,\tau)\, p(\mu)\, p(\tau)\, p\left(\sigma^2\right) p(\xi)$$

where $\Theta = [\phi, \beta_1, \beta_2, \beta_{10}, \beta_{20}, b, \mu, \tau, \sigma^2, \xi]$. The prior distribution for σ^2 and ξ are

$$
\begin{aligned}
\sigma^2 &\sim InvGam(\tilde{c}, \tilde{d}) \\
\xi &\sim N\left(\tilde{\xi}, \tilde{G}\right)
\end{aligned}
$$

The algorithm of drawing the parameters in the dynamic ordered probit model based on the algorithm of Nandram and Chen (1996) and Chen and Dey (2000) is as follows:

1. Sample $\phi,\ \beta,\ \beta_0,\ b,\ \mu,\ \tau$ from the full conditional distributions.

2. Sample $y^*,\ y_0^*$ from their full conditional distributions.

3. Sample ξ from the Metropolis-Hastings (MH) algorithm.

4. Calculate $c_j = \frac{c_{j-1}+\exp(\xi_j)}{1+\exp(\xi_j)},\ j = 2,\ \dots\ ,\ J-2$

2.4 Results with Simulated Data

2.4.1 Data Generating Process

This section provides a numerical example with simulated data for checking four MCMC algorithms. I simulate data following the previous simulation study by Hasegawa (2009). I set up $N = 200$ and $T = 10$, and the ordered response variables y_{it} $(i = 1, ..., 200,$ $t = 1, ..., 10)$ take 4 values, i.e., $y_{it} = 1,\ 2,\ 3,$ or 4. The latent variable, y_{it}^*, is distributed as follows:

$$y_{it}^*|.\ \sim\ N\left(\phi y_{it-1}^* + v_{it}\beta_1 + x_i\beta_2 + b_i,\ 1\right), t = 2, ..., 10$$

$$y_{i1}^*|.\ \sim\ N\left(v_{i1}\beta_{10} + x_i\beta_{20} + b_i, 1\right), t = 1$$

for $i = 1, ..., 200$. The parameters and the variables are set up as

$$\begin{aligned}
\phi &= 0.5;\ \beta_1 = \beta_{10} = 2;\ \beta_2 = \beta_{20} = 1.5; \\
b_i &\sim N(\mu,\ \tau)\, where \mu = 1 and \tau = 1 \\
c_1 &= 0;\ c_2 = 5;\ c_3 = 10 \\
v_{i0} &\sim N(1,3)\,; v_{it} = 0.3v_{it-1} + u_{it}, where u_{it}\ \sim\ N(0,1) \\
x_i &\sim N(2,4)
\end{aligned}$$

2.4.2 MCMC Results

Table 2.1 Here.

For the analysis, the MCMC algorithms run for $24,000$ iterations, keeping 20th draws after the first $4,000$ draws are burned. Table 2.1 shows posterior summary statistics for the parameters.

The posterior means of the key variables, ϕ, β, β_o, and c, are very close to the true values in all four algorithms. However, those of individual random effects, μ and τ, vary across the algorithms. Especially, the posterior mean of standard deviation (τ) of random effects from Chen and Dey (2000)'s algorithm is estimated as 0.37, far from the true value, 4. Generally, the posterior results show that random effects are hard to draw to be close to the true values. It might be due to relatively small sample sizes in my simulation study ($N = 200$ and $T = 10$).

The main issue is to compare autocorrelation of posterior draws of cutoff points from four algorithms. In the first algorithm employing Albert and Chib (1993), autocorrelation for c_2 and c_3 are 0.98 and 0.97; in the second algorithm (Liu and Sabatti (2000)), those are 0.33 and 0.27; in the third algorithm (Albert and Chib (2001)), 0.89 and 0.93; and in the fourth algorithm (Chen and Dey (2000)), 0.18 and 0.17. In terms of autocorrelation, the best algorithm to reduce autocorrelation of cutoff points is Chen and Dey (2000), using generalized Gibbs sampler, and the second one is Liu and Sabatti (2000), employing Metropolis-Hasting algorithm. As expected, MCMC draws from Gibbs samplers by Albert and Chib (1993) show the highest autocorrelation.

2.5 Conclusion

In this paper, I have examined four representative MCMC algorithms for estimating the dynamic ordered probit model with random effects. For the initial conditions problem, I employed the approach proposed by Hasegawa (2009). The result of simulated data suggested that the algorithm using Chen and Dey (2000) is the best one in terms of reducing autocorrelation. Even though the levels of autocorrelation are reduced in Chen and Dey (2000) and Liu and Sabatti (2000), I have found that autocorrelation still high in the cutoffs compared to other parameters. Hence, another MCMC algorithm should be devised to correct the problem. In the future study, I would like to develop a new MCMC algorithm using a probability integral transformation method for the cutoff points.

	True Value	Mean	Std.	AR(1)
Algorithm1: Albert & Chib (1993)				
ϕ	0.5	0.491014	0.016534	0.521104
β_1	2	2.042886	0.09483	0.622086
β_2	1.5	1.495506	0.09557	0.873232
β_{10}	2	1.82096	0.106477	0.487809
β_{20}	1.5	1.528753	0.094678	0.692229
c_2	5	4.946154	0.209125	0.980274
c_3	10	10.109342	0.408196	0.971385
μ	-3	-2.689466	0.261125	0.675941
τ	4	3.2116	0.472037	0.441375
Algorithm2: Liu & Sabatti (2000)				
ϕ	0.5	0.478799	0.085205	0.360457
β_1	2	1.907619	0.194558	0.191619
β_2	1.5	1.469513	0.219317	0.502431
β_{10}	2	1.695263	0.19826	0.230064
β_{20}	1.5	1.479918	0.211768	0.364171
c_2	5	4.489116	0.466112	0.325568
c_3	10	9.360203	0.943748	0.269957
μ	-3	-2.742552	0.501644	0.410578
τ	4	3.209466	0.911796	0.390433
Algorithm3: Albert & Chib (2001)				
ϕ	0.5	0.491323	0.016492	0.556116
β_1	2	2.19801	0.115296	0.758481
β_2	1.5	1.623513	0.109723	0.904076
β_{10}	2	1.976785	0.122717	0.586622
β_{20}	1.5	1.652241	0.112536	0.765145
c_2	5	5.266243	0.275124	0.885248
c_3	10	10.937321	0.523085	0.929045
μ	-3	-2.946039	0.278986	0.686912
τ	4	3.84305	0.617692	0.845353
Algorithm4: Chen & Dey (2000)				
ϕ	0.5	0.488499	0.017588	0.583684
β_1	2	2.183088	0.123529	0.155371
β_2	1.5	1.618349	0.132593	0.57606
β_{10}	2	1.967679	0.134613	0.198425
β_{20}	1.5	1.644063	0.132182	0.433074
c_2	5	5.225161	0.301872	0.183204
c_3	10	10.832213	0.568302	0.170061
μ	-3	-2.937164	0.32182	0.591519
τ	4	0.373715	0.055038	0.271458

Table 2.1: Posterior results (Simulated data)

Chapter 3

Bayesian inference of an ordered probit model with non-ignorable missing data: The determinants of self-rated health using the Health and Retirement Study

3.1 Introduction

Missing data is an inherent problem in panel surveys or longitudinal studies due to the characteristic of panel studies: the subjects of the study are chosen at baseline and samples are taken from the subjects over time. In other words, there is no guarantee that subjects will stay over the study periods. My study provides statistical methods to address such missing data problems that panel studies face. Especially, in health surveys, missing data should not be overlooked because non-response to the survey tends to be highly correlated with health status of the subjects. If missing data are related to the respondents' health status or the explanatory variables on health, ignoring missing data may lead to an imprecise or incorrect analysis. Furthermore, it is well understandable that the probability of responding is likely to be related to health status of older people who are subjects in this study.

Despite missing data problems, the fundamental advantage of a panel dataset over a cross-section is that it will allow us greater flexibility in modeling differences in behavior across individuals. Accordingly, the major motivation for using panel data is its ability to control individual heterogeneity. In order to specify the unobserved individual heterogeneity, I employ the random effects model. In addition, I suspect that missing data are likely to depend on unobserved data, therefore, I jointly model observed and missing data through selection model approaches.

Selection models first appeared in econometrics, with Heckman (1979)'s work on sample selection bias. Later, Diggle and Kenward (1994) proposed selection models for continuous longitudinal data subject to non-ignorable dropout. Selection model approach in missing data is the most straightforward way to handle non-ignorable missing data mechanism since selection models allow a direct estimation of the parameters of interest from the marginal distribution of responses. Many models have been proposed that link the response and missing values. Recently, Gad (2011) proposes a model for continuous longitudinal data with not only non-ignorable dropout, but intermittent missing data as well, by specifying missing data mechanism with a multinomial logit model.

In this paper, I extend the model by Gad (2011) to ordinal responses since Gad (2011) models continuous longitudinal responses with a multivariate normal distribution. It becomes possible since an estimation of ordered probit model by Bayesian methods employs the data augmentation approach proposed by Albert and Chib (1993). Albert and Chib (1993) treat the unknown latent variable y^* values as additional parameters to be simulated within Markov chain Monte Carlo (MCMC) algorithms. Once values are obtained for y^*, the problem of estimating the model parameters in ordered probit model simplifies to that of doing so in a standard normal linear model.

In addition, Gad (2011) uses a stochastic expectation-maximization[1] (EM) algorithm, which adds an imputation step for missing data to a step of estimation through maximizing a likelihood function. I also impute latent variables and missing data from their conditional distributions, but estimate the model parameters by MCMC methods. Although using selection models has an important advantage that we can specify familiar econometric models for response data and missing data, it also has disadvantages that the estimates from selection model are very sensitive both to misspecification of

[1] The expectation-maximization (EM) iterative algorithm is a broadly applicable statistical technique for maximizing complex likelihoods and handling the incomplete data problem. At each iteration step of the algorithm, two steps are performed: (1) E-step consisting of projecting an appropriate functional containing the augmented data on the space of the original, incomplete data, and (2) M-step consisting of maximizing the function.

the complete distribution of the response and to the assumed shape of the dependence between the dropout process and response process (Kenward (1998)).

This paper provides a model to analyze panel data on ordinal dependent variable with missing observations by employing a selection model with Bayesian methods of inferences. The model is applied to seven waves $(1992 - 2004)$ of the Health and Retirement Study (HRS) in order to examine the determinants of self-rated health in older Americans. Since one can never observe missing data, a certain assumption needs to be made on the missing data mechanism. Rubin (1976) developed a framework for inference for incomplete data. First, I explain missing data mechanisms following Rubin (1976) in section 2. In section 3, I develop a model to analyze ordinal responses under the assumptions of missing data, and introduce Bayesian inferences for the developed models in section 4. Using Bayesian inferences, the developed model is applied to the Health and Retirement Study to explicate the determinants of self-rated health in section 5. Then, I conclude the paper.

3.2 Missing data mechanism

Missing data exhibit two types of patterns: intermittent missingness and monotone missingness. Subjects may withdraw from the study prematurely resulting in a monotone missing pattern (a dropout) or they may miss some occasions resulting in an intermittent missing pattern.

Following Rubin (1976), missing data are generally classified into three groups: MCAR (*missing completely at random*), MAR (*missing at random*), and MNAR (*missing not at random*). A missing process is MCAR if a missing observation is independent of both unobserved and observed data, and MAR if, conditional on the observed data, the missing observation is independent of the unobserved responses. A process that depends on the unobserved responses is MNAR.

First, let us discuss missing data, following Little and Rubin (2002). Let $y = (y_{i1}, ..., y_{iT})'$ denote the full data response vector for $i = 1, ..., N$ individuals over time

$t = 1, ..., T$. y can be partitioned into observed, y_{obs}, and missing, y_{mis}, values, i.e., $y = (y_{obs}, y_{mis})$. Now define m_{it} to be a binary indicator variable such that

$$m_{it} = \begin{cases} 0 & \text{if } y_{it} \text{ is observed} \\ 1 & \text{if } y_{it} \text{ is missing} \end{cases}$$

The missing data mechanism (MDM) can then be defined by the conditional distribution of m given y, i.e., $p(m|y, \psi)$, where ψ denotes the unknown parameters of the missing function. And, $y = \{y_{obs}, y_{mis}\}$ is denoted as the complete data, $\{y_{obs}, y_{mis,}m\}$ as the full data, and $\{y_{obs}, m\}$ as the observed data.

We can describe MCAR by

$$p(m|y_{obs}, y_{mis}, \psi) = p(m|\psi)$$

Note that m depends on ψ, but not on the values of any variables in y. For MAR data, the conditional distribution becomes

$$p(m|y_{obs}, y_{mis}, \psi) = p(m|y_{obs}, \psi) \tag{3.1}$$

and for MNAR, there is no simple way of presenting the conditional distribution of m.

The implications of these different missing data mechanisms (MDM) can be understood by considering the joint distribution of y and m, i.e.,

$$p(y, m|\beta, \psi) = p(y_{obs}, y_{mis}, m|\beta, \psi)$$

where β denotes the unknown parameters of the model of interest. The marginal distribution of the observed data can be obtained by integrating out the missing data

$$p(y_{obs}, m|\beta, \psi) = \int p(y_{obs}, y_{mis}, m|\beta, \psi)\, dy_{mis} \tag{3.2}$$

The integrand may be specified as the product of the distribution of y and the conditional distribution of m given y, i.e.,

$$p(y_{obs}, y_{mis}, m|\beta, \psi) = p(y_{obs}, y_{mis}|\beta, \psi)\, p(m|y_{obs}, y_{mis}, \beta, \psi)$$

and this can be simplified to

$$p(y_{obs}, y_{mis}, m|\beta, \psi) = p(y_{obs}, y_{mis}|\beta)\, p(m|y_{obs}, y_{mis}, \psi)\,, \tag{3.3}$$

if we assume that $m|y, \psi$ is conditionally independent of β, and $y|\beta$ is conditionally independent of ψ, which is usually reasonable in practice. This form of the joint distribution is known as selection model.

If the data are MAR, then (3.1) and (3.3) imply that (3.2) can be rewritten as

$$\begin{aligned} p\left(y_{obs}, m|\beta, \psi\right) &= p\left(m|y_{obs}, \psi\right) \int p\left(y_{obs}, y_{mis}|\beta\right) dy_{mis} \\ &= p\left(m|y_{obs}, \psi\right) p\left(y_{obs}|\beta\right) \end{aligned}$$

In this case, the joint distribution is factored into two terms, one involving the observed data and the parameters β, and the other involving the missing indicator, m, and parameters ψ.

The missing data mechanism (MDM) is termed *ignorable* for likelihood inference about β if the missing data are MAR. In addition, the full parameters can be decomposed as the parameters of the response model, β, and the missingness mechanism, ψ, i.e., β and ψ are distinct[2] (Little and Rubin (2002)). For ignorability to hold in Bayesian inference, in addition the priors for β and ψ need to be independent (Daniels and Hogan (2008)).

When the MDM is not *ignorable* (i.e., in the case of non-ignorable MDM), information on the observed data should be combined with assumptions about the MDM by building a joint model. Data in MDM is usually denoted by $y = \{y_{obs}, y_{mis}\}$ as the complete data, $\{y_{obs}, y_{mis}, m\}$ as the full data, and $\{y_{obs}, m\}$ as the observed data.

There are three models according to factorization of outcome and missingness: 1) *selection* model (outcome-dependent factorization), 2) *pattern-mixture* model (pattern dependent factorization), and 3) *shared parameter* model (parameter dependent factorization).

First, *selection model* approach factorizes the joint distribution for the complete data and the missing indicator (m), i.e., the full data, into the marginal distribution

[2]The parameter ψ is distinct from β if there are no a *priori* ties, via parameter space restrictions or prior distributions, between ψ and β (Rubin (1976)).

for the complete data and the conditional distribution for the missing indicator given the complete data as follows:

$$p\left(y_{obs}, y_{mis}, m|\beta, \psi\right) = p\left(y_{obs}, y_{mis}|\beta\right) p\left(m|y_{obs}, y_{mis}, \psi\right) \tag{3.4}$$

One of the advantages of the selection model is that it specifies the complete data distribution, $p\left(y_{obs}, y_{mis}|\beta\right)$, directly. As a result, the primary parameters of interest are explicit in the model, and the nature of dependence between missingness and responses has a transparent representation. The main disadvantage of the selection model is the inability to partition the full data response vector into identified and unidentified components (y_{obs} and y_{mis}).

Pattern-mixture models classify individuals by their missingness and allow different model structures for each pattern of missing data as follows:

$$p\left(y_{obs}, y_{mis}, m|\beta, \psi\right) = p\left(y_{obs}, y_{mis}|m, \beta\right) p\left(m|\psi\right)$$

with the assumptions of distinctness of parameters. The complete data model is a mixture of these patterns:

$$p\left(y_{obs}, y_{mis}|\beta, \psi\right) = \sum_{m \in M} p\left(y_{obs}, y_{mis}|m, \beta\right) p\left(m|\psi\right) \tag{3.5}$$

Since mixture models treat dropout or missingness as a source of variation in the full data distribution, specifying a different distribution of each dropout time or missing data pattern seems cumbersome. However, it frequently has the advantage of making explicit the parameters that cannot be identified by observed data. The ability to partition the parameters makes a sensitivity analysis possible.

Shared parameter models are specified by adding random coefficients to the models in order to allow for individual specific latent effects. Incorporating random coefficients to selection models yields random coefficient selection models and to pattern-mixture models yields random-coefficient pattern-mixture models (Little (1995)). These models are also known as shared parameter models. Let b denote a set of random effects and φ associated unknown parameters. Then, the joint distribution, $p\left(y_{obs}, y_{mis}, m, b|\beta, \psi, \varphi\right)$, can be factorized as a selection model:

$$p\left(y_{obs}, y_{mis}, m, b|\beta, \psi, \varphi\right) = p\left(y_{obs}, y_{mis}|b, \beta\right) f\left(m|y_{obs}, y_{mis}, b, \psi\right) p\left(b|\varphi\right)$$

or a pattern mixture model:

$$p\left(y_{obs}, y_{mis}, m, b|\beta, \psi, \varphi\right) = p\left(y_{obs}, y_{mis}|m, b, \beta\right) p\left(m|b, \psi\right) p\left(b|\varphi\right)$$

3.3 Models

In developing a model for the ordinal panel data with both dropout and intermittent missing values, a number of issues need to be addressed:

1. Subjects are followed over time; that is, the observations are correlated across time.

2. The missing data mechanism of the response variables might be non-ignorable.

3. Both intermittent and monotone missing data patterns are found in the panel study.

4. There are also missing values in time-varying covariates.

I handle issue 1 by using a dynamic probit model to capture the correlation across time within a subject and by using random effects to model the correlation across subjects. For issue 2, I use selection model assuming that missing data depend on observed responses and missed responses as well. For issue 3, I adopt selection model proposed by Gad (2011), incorporating both intermittent missingness and dropouts into the model. Finally, issue 4 is handled by using a joint multivariate distribution for the response and the time-varying covariates.

3.3.1 Complete-data model

Let y_{it} the discrete dependent variable that takes ordered categorical outcomes, i.e., $y_{it} = 1, ..., J$, and y^*_{it} $(-\infty < y^*_{it} < \infty)$ be the underlying latent variable for an individual

i at time t for $t = 1, ..., T$ and $i = 1, ..., N$. The model can be expressed as

$$y_{it} = j \quad if \;\; c_j \leq y_{it}^* < c_{j+1} \tag{3.6}$$

where c_j is a cutoff point of ordinal responses and is specified as

$$-\infty < c_1 = 0 < c_2 < ... < c_{J-1} < \infty$$

The latent variable, y_{it}^*, is modelled as (Hasegawa (2009))

$$\begin{aligned} y_{it}^* &= \phi y_{it-1}^* + v_{it}\beta_1 + x_i\beta_2 + b_i + \varepsilon_{it}, t = 2, ..., T \\ y_{i0}^* &= v_{i1}\beta_{10} + x_i\beta_{20} + b_i + \varepsilon_{it}, t = 1 \end{aligned} \tag{3.7}$$

where v_{it} is the set of the time-variant covariates, x_i is the set of the cross-sectional time-invariant covariates, and b_i is an individual specific and time-invariant random effects. ε_{it} is a time- and individual-specific error term that is assumed to be normally distributed and uncorrelated across individuals and times. In addition, ε_{it} is uncorrelated with random effects of the individual, b_i. b_i is distributed a mean, μ, and constant variance, τ, and independent of ε_{it} for all t. ε_{it} is assumed to be strictly exogenous, that is, the v_{it} is uncorrelated with ε_{is} for all t and s.

Assume that the observations on y_{it}, $t = 1, ..., T$, are independent conditional on the individual effects b_i. Then, the contribution to the likelihood for individual i, conditional on the covariates and the individual effects would be the joint probability:

$$\begin{aligned} & p\left(y_{i1} = j_{i1}, ..., y_{iT} = j_{iT} | v_{it}, x_i, b_i\right) \\ = & \prod_{t=2}^{T} \left[\Phi\left(c_j - \phi y_{it-1}^* - v_{it}\beta_1 - x_i\beta_2 - b_i\right) - \Phi\left(c_{j-1} - \phi y_{it-1}^* - v_{it}\beta_1 - x_i\beta_2 - b_i\right)\right] \\ & \times \left[\Phi\left(c_j - v_{i1}\beta_{10} - x_i\beta_{20} - b_i\right) - \Phi\left(c_{j-1} - v_{i1}\beta_{10} - x_i\beta_{20} - b_i\right)\right] \end{aligned} \tag{3.8}$$

where $\Phi(.)$ is the normal distribution function.

3.3.2 Incomplete-data model under MAR

Typically, when missing data occur in a response, it is likely that time-varying covariates are also missing. When missing values occur in both the response and time-varying covariates, we need to expand the model (3.8) to jointly model the response and the time-varying covariates. The conditional joint probability model for y and v is as follows:

$$\begin{aligned} & p(y, v|x, \beta, b, \alpha, a) \\ = & \prod_{t=1}^{T_i} p(y_t|y_{t-1}, x, v_t, \beta, b)\, p(v_t|v_{t-1}, x, \alpha, a) \\ = & \prod_{t=1}^{T_i}\prod_{i=1}^{N} p(y_{it}|u_{it}, \beta, b)\, p(v_{it}|w_{it}, \alpha, a_i) \end{aligned} \tag{3.9}$$

where $u_{it} \equiv (y_{it-1}, x_i, v_{it})$ and $w_{it} \equiv (v_{it-1}, x_i)$. In addition, b_i ˜ $N(\mu_y, \tau_y)$, and a_i ˜ $N(\mu_v, \tau_v)$ are the individual random effects in the model for y and v, respectively. In (3.9), $p(y_{it}|u_{it}, \beta, b_i)$ is the same as in given in (3.8).

Assume the time-variant covariate, v_{it}, is a continuos variable and v_{it} is normally distributed with mean $w_{it} \cdot \alpha + a_i$ and variance σ^2. a_i is an individual random effects, normally distributed with mean μ_v and τ_v.

Then,

$$\begin{aligned} & p(y, v|x, \beta, b, \alpha, a) \\ = & \prod_{t=1}^{T_i}\prod_{i=1}^{N} p(y_{it}|u_{it}, \beta, b_i)\, p(v_{it}|w_{it}, \alpha, a_i) \end{aligned}$$

Using parallel notations, I develop the model $p(v_{it}|w_{it}, \alpha, a_i)$ for time-varying covariate, v_{it}, when it is a binary variable. Specifically, let $\theta_{it} = p(v_{it} = 1|w_{it}, \alpha, a_i)$ and

$$logit\ (\theta_{it}) = \alpha_0 + w_{it}\alpha^* + a_i$$

where $\alpha = \{\alpha_0,\ \alpha^*\}$. Random effects a_i is introduced to account for individual heterogeneity and are assumed to be normally distributed with mean μ_a and variance τ_a. The probability density function of v is

$$p(v_{it}|\theta_{it}) = \theta_{it}^{v_{it}} (1 - \theta_{it})^{1-v_{it}}$$

The joint distribution y^* and v conditional on random effects, b and a respectively, is given by

$$
\begin{aligned}
p\left(y^{*}, v|b, a, x, \beta, \alpha\right) &= \prod_{t=1}^{T_i} p\left(y_t^{*}|y_{t-1}^{*}, v_t, x, \beta, b\right) \cdot p\left(v_t|v_{t-1}, x, \alpha, a\right) \\
&= \prod_{i=1}^{N}\prod_{t=1}^{T} p\left(y_{it}^{*}|u_{it}, \beta, b_i\right) \cdot p\left(v_{it}|w_{it}, \alpha, a_i\right)
\end{aligned}
\tag{3.10}
$$

where

$$
p\left(v_{it}|w_{it}, \alpha, a_i\right) = \frac{1}{1+\exp\left(-\left(\alpha_0 + w_{it}\alpha^{*} + a_i\right)\right)}
$$

3.3.3 Selection model for non-ignorable missing data mechanisms

Non-ignorable missing data mechanisms imply that the distributions of the response variables for the respondents and nonrespondents are systematically different, even after controlling for all known covariates. In such situations, the inferences based on the likelihood function of the observed data while ignoring the MDM would not be valid. There are two broad approaches for incorporating non-ignorable MDM: selection and pattern-mixture models. Here, I use a selection model approach for partitioning the joint distribution of observables and the missing data.

First, I consider a model where the MDM for time-varying covariates are ignorable. The dependent variable y is partitioned into observed, y_{obs}, and missing, y_{mis}, values, i.e., $y = (y_{obs}, y_{mis})$. The likelihood function is factorized following the selection model approach under MNAR

$$
p\left(y_{obs}, y_{mis}, m|\beta, \psi\right) = p\left(y_{obs}, y_{mis}|\beta\right) \cdot p\left(m|y_{obs}, y_{mis}, \psi\right) \tag{3.11}
$$

where ψ is a vector of the parameters for the missing data mechanism and is distinct from the parameter β.

The first term, $p\left(y_{obs}, y_{mis}|\beta\right)$, in equation (3.11) is modified as $p\left(y_{i,obs}, y_{i,mis}|\beta, y_{i,obs}^{*}, y_{i,mis}^{*}\right)$ and same as given in (3.7).

For the missing data mechanism (MDM), $p\left(m|y_{obs}, y_{mis}, \psi\right)$ is adjusted as

$p\left(m_i|y^*_{i,obs}, y^*_{i,mis}, \psi\right)$ in the model. Let m_{it} be a missing value indicator that takes three values as

$$m_{it} = \begin{cases} 0, & if y_{it} is observed \\ 1, & if y_{it} is intermittent missing \\ 2, & if y_{it} is dropped out \end{cases} \tag{3.12}$$

The MDM is assumed to conditionally depend on the past and current values of the dependent variable, y^*_{it-1} and y^*_{it}. Let, for simplicity,

$$\begin{aligned} \eta_{it,1} &\equiv \psi_{11} y^*_{it} + \psi_{12} y^*_{it-1} \\ \eta_{it,2} &\equiv \psi_{21} y^*_{it} + \psi_{22} y^*_{it-1} \end{aligned}$$

The MDM is modeled as a multinomial logit with three states as

$$p_{m,it} \equiv p\left(m_{it} = m | y^*_{it-1}, m_{it-1} \neq 2; \psi\right) = \begin{cases} \frac{1}{1+\sum_{m=1}^{2} \exp(\eta_{it,m})}, & m = 0 \\ \frac{\exp(\eta_{it,m})}{1+\sum_{m=1}^{2} \exp(\eta_{it,m})}, & m = 1, 2 \end{cases}$$

The parameters ψ_1 relate the intermittent missing and ψ_2 relate the dropout process with the response process. The MDM is non-ignorable when these two parameters takes non-zero values. It is assumed that there are no missing values at the baseline, i.e., always $m_{i1} = 0$. Note that when $m_{it} = 2$, it is an absorbing state, i.e.,

$$p\left(m_{it} = 2 | m_{it-1} = 2\right) = 1$$

For an individual i, the likelihood function of $p\left(m_i|y^*_{i,obs}, y^*_{i,mis}, \psi\right)$ is as follows:

$$\prod_{t=1}^{t_i} \left(1 - p_{1,it} - p_{2,it}\right)^{1(m_{it}=0)} p_{1,it}^{1(m_{it}=1)} p_{2,it}^{1(m_{it}=2)}$$

When there are missing time-varying covariates, and an ignorable MDM for missing covariates cannot be assumed, another selection model for covariates with different parameters is necessary. The parameters of the missingness for the response (y) model are identified in the same way as described above as the case where the MDM for missing covariates is ignorable. Since the parameters of the model for response, y, are distinct from the parameters of the model for time-varying covariates, the framework developed for selection models for y, can be extended for covariates in the same way. If a time-varying covariate is binary or ordinal, the same selection model as in (3.11) is used. If a time-varying covariate is continuous, $p\left(y_{obs}, y_{mis}|\beta\right)$, equation (3.11) is specified with a linear regression model with normally distributed error terms.

3.4 Bayesian inference

3.4.1 Bayesian inference for complete-data model

Let $\Theta = \{\phi, \beta_1, \beta_2, \beta_{10}, \beta_{20}, b, \mu, \tau, c\}$. The prior distribution of the random effects b is assumed to have a hierarchical structure. Then, the prior distributions are specified as follows:

$$p\left(\Theta\right) = p\left(\phi\right) p\left(\beta_1\right) p\left(\beta_2\right) p\left(\beta_{10}\right) p\left(\beta_{20}\right) p\left(b|\mu, \tau\right) p\left(\mu\right) p\left(\tau\right) p\left(c\right)$$

The joint posterior distribution for Θ is

$$\begin{aligned} p\left(\Theta, y^*, y_0^*|y, y_0\right) &\propto p\left(\Theta, y, y_0\right) p\left(y, y_0|\Theta, y^*, y_0^*\right) \\ &= p\left(\Theta\right) \prod_{i=1}^{N} \{p\left(y_i^*|\Theta_{-\beta_0}, y_{i0}^*\right) \cdot p\left(y_i|\Theta_{-\beta_0}, y_{i0}^*\right) \\ &\quad \cdot p\left(y_{i0}^*|\Theta_{-\phi,\beta}\right) \cdot p\left(y_{i0}|\Theta_{-\phi,\beta}, y_{i0}^*\right)\} \end{aligned}$$

where $\Theta_{-\beta_0} = \{\phi, \beta_1, \beta_2, b, \mu, \tau, c\}$ except β_{10} and β_{20}, and $\Theta_{-\phi,\beta} = \{\beta_{10}, \beta_{20}, b, \mu, \tau, c\}$.

3.4.2 Bayesian inference under ignorable missing data mechanisms

Let $\xi = [\beta, \beta_0, b, \mu_y, \tau_y, \alpha, a, \mu_v, \tau_v, c]$. Bayesian inference under ignorable MDM is modified based on the complete-data model is

$$\begin{aligned} p\left(\xi, y^*, y_0^*, v^*|y, y_0, v\right) &\propto p\left(\xi, y^*, y_0^*, v^*\right) \cdot p\left(y, y_0, v|\xi, y^*, y_0^*, v^*\right) \qquad (3.13) \\ &= p\left(\xi\right) \prod_{i=1}^{N} \{p\left(y_i^*|\xi_{-\beta_0}, y_{i0}^*, v_i^*\right) p\left(y_i|\xi_{-\beta_0}, y_i^*, y_{i0}^*, v_i^*\right) \\ &\quad \cdot p\left(y_{i0}^*|\beta_0, b_i, \mu_b, \tau_b, c\right) p\left(y_{i0}|\beta_0, b_i, \mu_b, \tau_b, c, y_{i0}^*\right) \\ &\quad \cdot p\left(v_i^*|\alpha, a_i, \mu_a, \tau_a\right) p\left(v_i|\alpha, a_i, \mu_a, \tau_a\right)\} \end{aligned}$$

As before, I use proper prior distributions for β, μ_b, τ_b, α, μ_a, and τ_a with β and α having a diffuse normal prior, μ_b and μ_a having a diffuse normal prior, and τ_b and τ_a having an inverse gamma distribution.

When there are missing values in y and/or v, and if the missing data mechanism is ignorable, the Gibbs sampling for the complete-data model can be easily modified. Especially, the Gibbs sampling for the model already include a data augmentation step for y^*. Hence, we only need to include the missing values in v in the Gibbs sampling steps by drawing values from its conditional predictive distribution, given the observed values and the current draws of the parameters.

The Gibbs sampling for ignorable missing data mechanism involves two steps: imputation (I) and posterior (P) steps. In imputation step, each missing value is replaced by a draw from its conditional distribution given the observed data and the current values of the parameters. And in posterior step, these drawn values of the missing data are treated as if they were the actual observed values of the data, and one draw of the parameters is made from the complete data posterior distribution. First, I fill in the missing values of v_{it} with the current values of parameters. With imputed v_{it}, y^*_{it} is drawn according to the data augmentation method.

Note that I only impute all intermittent missing data and the first dropout values for the purpose of computation because the dropout probability at T_i only depends on the current and previous values. Imputing more dropout values is not necessary for computation and does not provide additional information since all information actually comes from the available data and the assumptions we based on: the models remain unchanged over time.

Numerical example with simulated data

This section provides a numerical example with simulated data for checking MCMC algorithms of an ignorable MDM. I simulate data based on an assumption of *missing not at random* (MNAR); therefore, non-ignorable MDM should be considered to analyze data. Then, the MNAR dataset is analyzed as though the missing values are ignorable (so that I use an ignorable MDM). Hence, we can check the performance of the algorithm of ignorable MDM when a MDM is misspecified.

I set up $N = 200$ and $T = 10$, and the ordered response variables y_{it} ($i = 1, ..., 200$, $t = 1, ..., 10$) take 3 values, i.e., $y_{it} = 1$, 2, or 3. The latent variable, y^*_{it}, is distributed as follows:

$$y^*_{it}|. \sim N\left(\phi y^*_{it-1} + v_{it}\beta_1 + x_i\beta_2 + b_i,\ 1\right), t = 2, ..., 10$$

$$y^*_{i1}|. \sim N\left(v_{i1}\beta_{10} + x_i\beta_{20} + b_i, 1\right), t = 1$$

for $i = 1, ..., 200$. The parameters and the variables are set up as

$$\begin{aligned} \phi &= 0.5;\ \beta_1 = \beta_{10} = 2;\ \beta_2 = \beta_{20} = 1.5;\ c_1 = 0;\ c_2 = 5; \\ b_i &\sim N(\mu,\ \tau)\, where \mu = 1 and \tau = 1 \\ v_{i0} &\sim N(1, 3); v_{it} = 0.3v_{it-1} + u_{it}, where u_{it} \sim N(0, 1) \\ x_i &\sim N(2, 4) \end{aligned}$$

Missing values are generated under an assumption of MNAR. By definition, MNAR depends on the values of both observed and missed dependent variables. Since the dependent variable y_{it} is ordered in our model, the latent variable of the dependent variable, y^*_{it}, is used to create MNAR. The first response at $t = 1$, Y_{i1}, is assumed to be observed for every subject in the study, and it is assumed that there are no missing values in the covariates.

Missing data mechanism follows a multinomial logit regression model as follows:

$$\begin{aligned} \eta_{it,1} &= \psi_{11}y^*_{it} + \psi_{12}y^*_{it-1} \\ \eta_{it,2} &= \psi_{21}y^*_{it} + \psi_{22}y^*_{it-1} \end{aligned}$$

where

$$\begin{aligned} \psi_{11} &= -1;\ \psi_{12} = 0 \\ \psi_{12} &= -2;\ \psi_{22} = 0 \end{aligned}$$

By fixing $\psi_{12} = 0$ and $\psi_{22} = 0$, I make the missing values only depend on the current value of y^*_{it} for simplicity.

First, the probability of the intermittent missing value is calculated

$$p\left(m_{it} = 1|y^*_{it-1}, m_{it-1} \neq 2; \psi\right) = \frac{\exp\left(\eta_{it,1}\right)}{1 + \sum_{m=1}^{2} \exp\left(\eta_{it,1}\right)}$$

and the probability of the dropout is

$$p\left(m_{it} = 2|y^*_{it-1}, m_{it} \neq 2; \psi\right) = \frac{\exp\left(\eta_{it,2}\right)}{1 + \sum_{m=1}^{2} \exp\left(\eta_{it,2}\right)}$$

The probability of a response is, therefore,

$$p\left(m_{it} = 0|y^*_{it-1}, m_{it} \neq 2; \psi\right) = 1 - p_1\left(m_{it} = 1|.\right) - p_2\left(m_{it} = 2|.\right)$$

Once an observation is dropped out from the study, then the observation must not be measured in the future. Hence, when m_{it} reaches 2, there are no follow-up observations, i.e., an absorbing state.

Although a number of missing values depend on the data generating process of y^*_{it}, the parameter values that I choose create $1.75 \sim 2.2\%$ intermittent missing values and $13.1 \sim 21.1\%$ dropouts. The remaining 76.7% of the observations respond.

The MCMC algorithms run for $22,000$ iterations, keeping every10th draws after the first $2,000$ draws are burned. When there are missing values in the dependent variable, and if the MDM is ignorable, the Gibbs sampling for the complete-data model can be easily modified. I fill in the missing values in y^* in the Gibbs sampling steps by drawing values from their conditional predictive distribution, given the observed values and the current draw of the parameters.

1. Sample ϕ, β, β_0 from their full conditional distributions
2. Sample $[b|\mu, \tau]$, μ, τ from their full conditional distributions

   The random effects, b_i, have a hierarchical structure. b_i are drawn from the full conditional distribution (FCD) with a mean, μ, and a standard deviation, τ, both μ and τ has their own prior parameters.

3. Sample c following Albert and Chib (1993)

4. Sample ψ_{11}, ψ_{12}, ψ_{21}, ψ_{22} from the MH algorithm

5. Since we do not have a full conditional distribution of y^* and y_0^*, the MH algorithm is employed for the latent variables. y^* and y_0^* are drawn from a proposal density, normal linear regression model, given current draws of the parameters, and then the accept-reject algorithm is used through the posterior distribution (3.13).

The simulated data have 2.2% of intermittent missing values, and 16.9% of dropouts. There is 80.9% of the responses. The results are as follow:

	TRUE	mean	std	AR(1)	95%HPDI	
ϕ	0.5	0.464041	0.038634	0.748277	0.389917	0.538269
β_1	2	1.648129	0.129983	0.801044	1.402893	1.905542
β_2	1.5	1.266007	0.128838	0.773085	1.013718	1.51616
c_2	5	5.691052	0.287419	0.994797	5.153038	6.159729
μ	1	1.872833	0.30236	0.891786	1.281703	2.500762
τ	1	0.754516	0.188127	0.591337	0.409215	1.118818
β_{10}	2	1.865389	0.16241	0.715006	1.537643	2.17642
β_{20}	1.5	1.428823	0.14304	0.788267	1.157142	1.709506

Table 3.1: Inference for ignorable MDM with MNAR data

From the results, the posterior means of key parameters (ϕ, β_1, β_2, β_{10}, β_{20}) are underestimated. For β_1 and μ, the 95% HPDIs do not include the true values. This example will be compared with one from correctly specifying non-ignorable MDM.

3.4.3 Bayesian inference under non-ignorable missing data mechanisms

Let $\beta = \{\phi, \beta_1, \beta_2\}'$ and $\Theta = \{\beta, b, \mu, \tau, c, \psi\}$. The posterior distribution under non-ignorable MDM is

$$
\begin{aligned}
p\left(\Theta, y^*_{obs}, y^*_{mis} | y_{obs}, y^*_{mis}, m\right) &\propto p\left(\Theta, y^*_{obs}, y^*_{mis}\right) \cdot p\left(y_{obs}, y_{mis}, m | \Theta, y^*_{obs}, y^*_{mis}\right) \quad (3.14) \\
&= p\left(\Theta\right)_i p\left(y^*_{i,obs}, y^*_{i,mis} | \Theta\right)
\end{aligned}
$$

$$
\begin{aligned}
& \cdot p\left(y_{i,obs}, y_{i,mis}, m_i|\Theta, y^*_{i,obs}, y^*_{i,mis}\right) \\
= \quad & p\left(\Theta\right)_i p\left(y^*_{i,obs}|\Theta_{-\psi}, y^*_{i,mis}\right) \cdot p\left(y^*_{i,mis}|\Theta_{-\psi}\right) \\
& \cdot p\left(m_i|y^*_{i,obs}, y^*_{i,mis}, \psi\right) \cdot p\left(y_{i,obs}, y_{i,mis}|\beta, y^*_{i,obs}, y^*_{i,mis}\right)
\end{aligned}
$$

where y^*_{obs} and y^*_{mis} are the latent variables as in (3.7) and

$$
\begin{aligned}
y_i &\equiv (y_{i1}, ..., y_{iT})' \\
y &\equiv (y'_1, ..., y'_N)'
\end{aligned}
$$

Numerical example with simulated data (continued)

For this numerical example for non-ignorable MDM, the same data generating process is used as in section (3.4.2). After augmenting the values of underlying latent variable and missing values in y^*, model parameters are sampled by following the algorithms:

1. Sample ϕ, β, β_0 from the MH algorithm

2. Sample $[b|\mu, \tau]$, μ, τ from their full conditional distributions

   The random effects, b_i, are drawn from the full conditional distribution (FCD).

3. Sample c following Albert and Chib (1993)

4. Sample ψ_{11}, ψ_{12}, ψ_{21}, ψ_{22} from the MH algorithm

5. Since the full conditional distribution of y^* and y^*_0 does not have a closed form, the MH algorithm is employed for the latent variables. y^* and y^*_0 are drawn from a proposal density, normal linear regression model, given current draws of the parameters, and then the accept-reject algorithm is used through the posterior distribution (3.14).

The data generating process simulates a dataset with 1.75% of intermittent missing values, 13.1% of dropout values, and therefore, 85.15% of responses.

So far, I have examined two numerical examples: in the examples, missing data are generated based on an assumption of *missing not at random*, i.e., missingness depends

	TRUE	mean	std	AR(1)	95%HPDI	
ϕ	0.5	0.471161	0.033515	0.689461	0.405424	0.536378
β_1	2	1.900816	0.147563	0.811174	1.62278	2.19259
β_2	1.5	1.71629	0.189391	0.894441	1.372818	2.071071
c_2	5	5.001917	0.283764	0.992227	4.556836	5.576122
μ	1	0.834036	0.395551	0.903113	0.035744	1.641484
τ	1	1.540151	0.406483	0.644241	0.832063	2.37792
β_{10}	2	2.107924	0.173894	0.697289	1.778993	2.46646
β_{20}	1.5	1.517123	0.177661	0.888077	1.183802	1.865441
ψ_{11}	-1	-0.71194	0.146569	0.725916	-0.99162	-0.43838
ψ_{12}	0	-0.15773	0.091341	0.304052	-0.33061	0.022277
ψ_{21}	-2	-1.59185	0.278491	0.722029	-2.12838	-1.06164
ψ_{22}	0	0.06644	0.115394	0.936178	-0.18944	0.290959

Table 3.2: Inference for nonignorable MDM with MNAR data

on the dependent variable. The data with missing data under *missing not at random* are specified with the model under ignorable missing data mechanisms in section 3.4.2 and under non-ignorable missing data mechanisms in section 3.4.3. In the example for non-ignorable MDM (3.4.3), the response data and missing data are jointly modeled; therefore, the parameters for missing data mechanism (ψ_{11}, ψ_{12}, ψ_{21}, ψ_{22}) are drawn, different from the ignorable case in section 3.4.2. Based on the results, we can observe that the key parameters, $\beta' s$, are underestimated in section 3.4.2. However, in section 3.4.3, the posterior means of every parameter are closer to true values than in section 3.4.2. Only one 95% HPDIs of ψ_{11} exclude the true value, but the difference is not much because the true value is -1 and the upper bound of the HPDIs is -0.99. This example implies that if missingness might not be at random, we should consider non-ignorable missing data mechanism to prevent biased estimations.

In sum, these examples imply that if missing observations depend on unobserved data (i.e., MNAR), ignoring the missing observations may lead to an imprecise or wrong data analysis. Hence, we need to consider non-ignorable MDM if missing data are suspected to be MNAR. For analyzing such missing observations, I have developed Bayesian approach for the ordered probit model with non-ignorable missing data mechanism through selection models. In next section, I apply the model to the Health and Retirement Study to examine self-rated health and its determinants.

3.5 Empirical application: Self-rated health in the Health and Retirement Study

3.5.1 Self-rated health

This section examines self-rated health and its determinants in the Health and Retirement Study (HRS) by estimating the model developed in the previous sections. Self-rated health is a widely used indicator of general health. This single item global rating is used in health surveys all over the world, in many languages, to serve as health indicators for the population and to track trends over time. Usually, health surveys ask respondents to rate their health with a single question; for example, the Health and Retirement Study asks respondents ”would you say your health is (1) excellent, (2) very good, (3) good, (4) fair, or (5) poor?”.

Self-rated health has long been a focus of interdisciplinary research on social and psychological factors in health. The the first study on self-rated health and mortality appeared in 1982 (Mossey and Shapiro (1982)). Since Mossey and Shapiro (1982), many other studies across areas have found that self-rated health has an independent effect even beyond objective clinical measures of health and other risk factors. As a measure of health, the reliability and validity of self-rated health have been well-established, that is, self-rated health provides a valid assessment of overall health (Idler and Benyamini (1997)) and is a strong predictor of mortality (Mossey and Shapiro (1982)), functional limitation (Idler et al. (2000)), health-related behavior (Cott et al. (1999)), and health care utilization (Pinquart (2001)). It is, therefore, natural to attempt to gain better understanding of what underlies self-rated health.

In this section, I investigate the determinants of self-rated health among older Americans using seven waves of the Health and Retirement Study (HRS). Previous researchers, in economics and health science, focus on explaining differences of health status with socioeconomic inequalities. For example, Smith and Kington (1998) provide evidence to support that socioeconomic status plays a role in explaining racial and ethnic differences in health outcomes of older Americans using the HRS. Recently, Berry

(2007) examines the effect of household financial resources on health over six panels of the HRS (1992 – 2002). Based on results estimated from fixed-effects models, Berry (2007) concludes that there is a significant influence of long-term income on health, but not short-term income, in the elderly. Frijters and Ulker (2008) examine robustness of the common determinants of health to explain six health measures, including self-rated health, in the HRS (1992 – 2002). In case of self-rated health, all of the key variables (income, drinking, smoking, and exercise) are significant to explain self-rated health when using the pooled sample. However, only two of them, exercise and smoking, remain significant after controlling for fixed effects. Kim (2011) investigates socioeconomic inequalities in self-rated health among middle-aged and older adults in the HRS (1992 – 2006). The findings show more income, assets, and education, and having private health insurance predict better self-rated health.

Although the previous studies use panel data from the HRS, which suffer from dropouts of respondents over time, most of them do not consider missing data in their analyses by simply dropping missing observations. My analysis starts from this point: if panel attrition is related to the respondent's health status or the determinants of health, then this attrition might have implications for explaining the relationship between health and its determinants. Banks et al. (2010) show that wealth appears to predict attrition in the group aged 55 – 64 using three waves (2002 – 2006) of the HRS. Kapteyn et al. (2006) categorize the respondents in the HRS, by using six waves (1992 – 2002), into four groups: "always-in" (who provide interview in all six waves), "ever-out" (who have ever dropped out, but come back into the survey), "died", and "permanent attritors" (who drop out permanently because of other reasons than death). Estimating from a multinomial logit, Kapteyn et al. (2006) show that the characteristics of respondents who drop out over time are quite different from those in the retention samples. Particularly, those who attrit, but are recruited back into the survey are very different from those who permanently drop out from the HRS. The differences are mainly on race and ethnicity, education, health and household income. Kapteyn et al. (2006) conclude that it is likely that *missing at random* (MAR) assumption for the

HRS is violated. This conclusion provides justification to assume missing data in the HRS as *missing not at random*.

Statistical methods to address such attrition in panel data have been actively studied. This empirical study employs one of the statistical methods under the assumption of *missing not at random*. Since missing data on self-rated health are supposed to be not at random, my model adapts non-ignorable missing data mechanism through selection model approaches. It is common that time-varying covariates are also missing when missing data occur in responses. Hence, missing covariates are also considered in my model. Another significant feature of my model is its capability to analyze intermittent missing values (caused by respondents who drop out, but return to the survey) as well as dropouts (due to permanent attritors). Using seven waves of the HRS $(1992-2004)$ and correcting missing data problems in the analysis, this paper investigates the determinants of self-rated health of old people. Gueorguieva et al. (2009) examine occupational differences in self-rated health after accounting for demographics, health behaviors, economic attributes, and employment characteristics over seven waves $(1992-2004)$ in the HRS, using hierarchical linear models. By using a different way to account for missing data and a different model, this paper examines if the same conclusions as Gueorguieva et al. (2009) are obtained.

Since the Whitehall study, established in 1985 as a panel survey in Britain, studies have found that occupation has a significant impact on health, with a marked social gradient between British civil service grades and a variety of health outcome (Bosma et al. (1997), Ferrie et al. (2002), Marmot et al. (1997a), Marmot et al. (1991), Marmot et al. (1997b)). The effects of occupation on health are particularly important for older people due to the cumulative effects of an individual's occupational commitment over time and the decline in health that occurs as one ages.

This paper provides three sets of estimation results from a dynamic ordered probit model with non-ignorable missing data mechanism through selection model approaches. Main differences of my paper, across all sets of results, from Gueorguieva et al. (2009) are (1) the models: self-rated health is specified with hierarchical linear models in

Gueorguieva et al. (2009), but with a dynamic ordered probit model with random effects in my analyses, (2) the estimation methods: although Gueorguieva et al. (2009) use a frequentist method, my analysis is based on Bayesian inferences, using MCMC algorithms, and (3) the methods to handle missing data: Gueorguieva et al. (2009) assume missing data as random and include a dummy variable for respondents who drop out, but I jointly model responses and missing data based on the assumption of *missing not at random.*

In the first analysis, I employ the same covariates and periods as Gueorguieva et al. (2009) do. All covariates are measured at the study baseline (1992) and, therefore, there are no missing values in covariates over time. Only self-rated health, the dependent variable, changes over time and has missing values. In this analysis, only differences are that I use continuous variables for wealth and income rather than using binary variables for separating the ranges into five categories as in Gueorguieva et al. (2009). The number of subjects is $9,557$ in my analysis rather than $9,586$ in Gueorguieva et al. (2009). The difference might be due to my assumption that there are no missing data at baseline. From this part of analyses, we could observe how the results are affected solely by the different statistical model and the different way to handle missing data.

In the second analysis, I employ the same covariates and periods as in Gueorguieva et al. (2009), but let the time-varying covariates change over time. As indicated as one of their limitations, Gueorguieva et al. (2009) only consider baseline characteristics in covariates to examine changes in self-rated health. However, I suspect that changes in status or health habits might have significant effects on explaining self-rated health over time. For example, the analysis uses subjects who were $50 - 64$ years of age or older in 1992 (baseline), and these subjects were $62 - 76$ years old in 2004 (the year of the last wave in the analysis). Most subjects, then, became eligible to receive Medicare, government-provided health insurance. The dataset from the HRS shows that only 12% of the subjects in my dataset was covered by government health insurance program at baseline, but it increased to 52% in 2004. The HRS also shows significant changes in employment status: 20% of subjects at baseline were retired, but it increased to 50% in

2004. Hence, in the second part of the analyses, we can see how the results change if we consider changes of respondents' status and characteristics over time in the analysis.

In the third analysis, I include additional control variables to explain self-rated health: regular exercise, objective health problems (doctor-diagnosed chronic diseases, body mass index (BMI)), and respondents' mental health, represented by depression. There has been a long-standing interest in social inequalities in health and survival, and in the behavioral, psychosocial, and environmental mechanisms that may account for these disparities. However, social scientists have only recently begun to examine the underlying biological pathways linking social status to mental and physical well-being (Goldman et al. (2011)). Hence, it will be interesting to see how much self-rated health is explained by objective health status in terms of a number of the chronic diseases that are diagnosed by doctors and BMI.

It has been studied that even physical activity from non-leisure activities, e.g., walking, household chores, and job-related activity, as well as leisure-time physical activity are associated with a substantial reduction in all-cause mortality (Arrieta and Russell (2008)). In the HRS, vigorous physical activity is defined as sports, heavy housework, or a job that involves physical labor. Frijters and Ulker (2008) create a binary indicator for regular exercise showing whether the respondent participates in vigorous physical activity at least 3 times a week, and examine the effect of regular exercise on self-rated health in the HRS. The results support the significant association between regular exercise and self-rated health. Following Frijters and Ulker (2008), I recode a variable for regular exercise to examine the effects of regular exercise on self-rated health.

Finally, I explore the relationship between mental health and self-rated health in the analysis. I focus on depression, the most prevalent mental health condition in older population and a leading cause of disability. According to Leon et al. (2003), the prevalence in patients aged more than 65 years can be as high as 30% in outpatient setting and 40% in hospitalized patients. Chang-Quan et al. (2010) conduct a meta-analysis of eleven longitudinal studies, examining the relationship between self-rated health and depression for the elderly. They conclude that poor self-rated health is very closely associated with depression. Hence, I hypothesize that depression is a significant explanatory variable that should not be overlooked on an explanation of self-rated health in my study.

3.5.2 Model and variables

The empirical analyses employ the model (3.11) developed in section 3.3.3 and Bayesian inference and MCMC algorithms explained in section 3.4.3. For the second and third analyses, I consider missing time-varying covariates and assume that missing covariates are *missing at random*. Although I assume the missing covariates as MAR, this assumption can be easily relaxed to *missing not at random* by employing another selection model for missing covariates with different parameters in section 3.3.3.

Data

I use seven waves $(1992 - 2004)$ from the Health and Retirement Study (HRS)[3] as Gueorguieva et al. (2009). The RAND Center for the Study of Aging provide publicly available dataset from the HRS. This study uses the most recent RAND HRS data version K that includes information on people's health, socioeconomic status, and health care uses. The HRS is a longitudinal survey of individuals aged $51 - 61$ in 1992 in the U.S. Data were collected every two years and cover a wide range of aspects of the life of the population over 50 years old. In 1992, $12,652$ interviews were conducted for a random sample of individuals born between 1931 and 1941. Spouses of these individuals were included irrespective of their age. I exclude respondents who have missing values in self-rated health and covariates at baseline, and include ones who were $50-64$ years of

[3]The original cohort entering the HRS study in 1992 was composed of individuals born between 1931 and 1941 and their spouses, irrespective of their age. The next year, a much older cohort was interviewed, the Study of Assets and Health Dynamics among the Oldest Old (AHEAD) cohort, which was born before 1923. Both these cohorts have been followed every two years up through 2006 (the AHEAD cohort is interviewed in 1993 and 1995, and then merged with the main study for 1998 onwards). In 1998, two new cohorts were added and blended into the original sample, the so-called Children of the Depression Age Cohort (CODA), born between 1924 and 1930, and the War Babies cohort, born between 1942 and 1947.

age or older at baseline (1992), following Gueorguieva et al. (2009). The number of the respondents included in my analysis is reduced to $9,557$[4], therefore, total observations including missing values are $66,899$ over seven waves. Table (3.7) in the appendix shows the proportions of respondents with missing observations due to dropouts and intermittently missingness in self-rated health and covariates. Since there are people who were intermittently missing for a while, but came back, then dropped out, a sum of the probabilities might be greater than 1.

Self-rated health

Self-rated health is used as a dependent variable. The respondents are asked to rate their current health as 1 (excellent), 2 (very good), 3 (good), 4 (fair), and 5 (poor), having discrete ordinal values with five categories. Figure 3.1 describes the distribution of self-rated health across all seven waves and total waves. The distribution shows that the trend of self-rated health became worse over time. For example, the percentage of respondents who reported their health as excellent declined from 22% in the first wave to 11% in the seventh wave.

Explanatory variables

1. Socioeconomic variables:

Following Gueorguieva et al. (2009), eight dummy variables categorizing occupation are used in this analysis. These eight categories are (1) professional and technical support (reference category), (2) managerial, (3) clerical and administrative support, (4) sales, (5) mechanical, construction, and precision production, (6) service (including private household services, protective services, food preparation, health services, and personal services), (7) operators, fabricators, and laborers, and (8) farming, forestry, and fishing.

[4] At baseline, $12,652$ subjects were interviewed and included spouses. All of those subjects reported self-rated health, but $1,417$ subjects did not answer their longest occupation at wave 1. I eliminated the subjects with missing occupation and those who are less than 50 ($n = 1218$) or more than 64 ($n = 460$).

As Gueorguieva et al. (2009), occupation corresponds to occupational group for the job with the longest reported tenure at baseline.

Other socioeconomic variables used in the study are years of education, household wealth, and household income, which are all continuous variables. I use the consumer price index (CPI) as a deflator for constructing real variables for household wealth and income (1992 as a common year).

2. Health risk factors:

For chronic disease, I use count variables indicating how many chronic diseases among eight are diagnosed by doctors: cancer, heart condition, lung disease, a stroke, high blood pressure, diabetes, arthritis, and psychiatric problems.

For depression, respondents are asked about eight common symptoms of depression, taken from the Center for Epidemiologic Studies Depression (CESD) instrument. In validation studies against the full CESD battery, the presence of four out of the eight symptoms is associated with clinically significant depression (Karp (2007)). Depression variable used in this study has nine ordered values according to level of depressive symptoms: 0 indicates no depression and 8 does the most depression.

3. Other controls:

Other control variables are age, gender, marital status, race/ethnicity (black or Hispanic), smoking, types of health insurance (employer sponsored, government sponsored, other private insurance), employment status, body mass index (BMI), and regular exercise. Following Frijters and Ulker (2008), I recode regular exercise variable as a binary indicator showing whether the respondent participates in vigorous physical activity at least 3 times a week.

Summary statistics across all waves and all categories of self-rated health are presented in appendix table (3.8) and summary statistics by self-rated health across all waves are shown in appendix table (3.9). For comparisons of characteristics over time, I present summary statistics of covariates by self-rated health at wave 1 in table (3.10) and at wave 7 in table (3.11).

3.5.3 Results

The scaling of the ordered probit coefficients is arbitrary. To provide an indication of the magnitude of the associations between self-rated health and the explanatory variables, I present the average partial effects of the explanatory variables on the response probabilities. The partial effects give the impact on the specific probability per unit change in the covariates. For continuous covariates, such as household wealth and income, these are obtained by taking the derivative of the ordered probit probabilities with respect to the variable in question given random effects. For discrete covariates, such as binary variables for dummy variables for the longest occupation, they are obtained by taking differences given random effects.

For example, suppose x is a binary variable in the model and β is the coefficient on x. I measure the effect of a change in x from 0 to 1 with all other variables held at the values of interest using

$$\Delta \Pr(y = j) = \Pr(y = j|x = 1) - \Pr(y = j|x = 0)$$

where j is the ordered value in the dependent variable. It is possible to compute partial effects for each of five categories of self-rated health in the ordered probit model. For simplicity, I present average partial effects on probability of reporting excellent and poor self-rated health in the appendix. Hence, partial effects with a positive coefficient on probability of reporting excellent self-rated health imply a positive effect on self-rated health, whereas partial effects with a positive coefficient on probability of reporting poor self-rated health mean a negative effect on self-rated health.

In this section, I present the results based on average partial effects from three sets of analyses. In the first analysis, the posterior results from my model are compared with Gueorguieva et al. (2009) by using same baseline covariates to analyze self-rated health. In the second analysis, the results are presented when time-varying covariates are considered to explain the self-rated health over time. Finally, in the third analysis, the results from employing additional control variables are presented.

Analysis1: Explanation of self-rated health using only baseline covariates

	wave2-7				Baseline				
Occupation	means	95%HPDI		Pr (PE ¡0)*	means	95%HPDI		Pr (PE ¡0)	Differences*
Managerial	-0.028	-0.053	-0.002	0.981	-0.035	-0.068	0.000	0.979	0.006
Sales	-0.028	-0.056	0.001	0.967	-0.065	-0.102	-0.030	1.000	0.037
Clerical and administrative	-0.015	-0.040	0.013	0.878	-0.046	-0.078	-0.012	0.995	0.031
Service	-0.102	-0.129	-0.075	1.000	-0.128	-0.160	-0.096	1.000	0.026
Farming and fishing	-0.107	-0.153	-0.058	1.000	-0.152	-0.206	-0.092	1.000	0.045
Mechanical	-0.082	-0.111	-0.053	1.000	-0.123	-0.159	-0.086	1.000	0.042
Operator	-0.120	-0.145	-0.095	1.000	-0.165	-0.195	-0.133	1.000	0.044

*Pr(PE¡0): probability of the partial effects to be less than zero

*Differences: the differences of posterior means of the partial effects between wave2-7 and baseline

Table 3.3: Analysis1: Average partial effects on excellent self-rated health (including only baseline covariates)

This first analysis can be compared with Gueorguieva et al. (2009) since this analysis only employs baseline covariates as Gueorguieva et al. (2009). The results from the analysis examine whether their findings are robust when the analysis employs an ordered probit model with random effects specifying missing observations with non-ignorable missing data mechanism. For the analyses, the MCMC algorithms run for $100,000$ iterations, keeping every 15th draws after the first $40,000$ draws are burned. Table 3.3 shows posterior summary statistics for the average partial effects of occupation on excellent self-rated health and complete posterior results are presented in tables 3.12 for wave $2-7$ and 3.13 for baseline in the appendix.

Measuring occupation as the longest occupation that respondents have held over their careers, Gueorguieva et al. (2009) find substantial variations in self-rated health across longest occupation at baseline of the HRS. However, they do not find any significant impact of occupation on rate changes in health over time. That is, occupational differentials in self-rated health are persistent as individuals age, but such differentials neither widen nor narrow.

First, regarding the directions of coefficients in table 3.3, all posterior means of the partial effects show same negative directions at baseline and over time (wave $2-7$). In other words, all occupational groups are negatively associated with excellent self-rated health relative to the reference group, professional, based on cross-sectional and longitudinal evidence.

In general, occupational groups could be separate into two groups in terms of 95% HPDIs of the partial effects: those of three occupational groups, managerial, sales, and clerical and administrative, overlap each other, and those of the remaining four do so. It shows that the managerial, sales, and clerical and administrative groups have similar effects on explaining excellent self-rated health, either service, farming and fishing, mechanical, and operator groups do. This pattern is observed not only at baseline but also over time from wave 2 to 7. It implies that there are variations in self-rated health across longest occupation even though it fall into two. Hence, the results provide evidence to support Gueorguieva et al. (2009) that there are substantial variations in self-rated health across longest occupation at baseline and they persist over time.

Another empirical question is whether such occupational differentials in self-rated health either widen or narrow at older ages. Gueorguieva et al. (2009) find that gaps remain while the health disparities persist. My results in table 3.3 indicate that absolute values of the partial effects are reduced from baseline to waves 2-7. That is, the impact on health narrow in old age.

In sum, I examine if the ordered probit model with non-ignorable missing data mechanism obtains same results as Gueorguieva et al. (2009) by using same baseline covariates as Gueorguieva et al. (2009). Based on evidence from 95% HPDIs of the partial effects, my results indicate that there are substantial variations in self-rated health across longest occupation at baseline and they remain over time, not widen, but narrow. My first analysis with baseline covariates confirm the findings of Gueorguieva et al. (2009) that self-rated health is explained by longest occupation differently, but narrowed gaps do not provide evidence to support Gueorguieva et al. (2009) that the health disparities persist over time, not widening nor narrowing.

Analysis2: Explanation of self-rated health using time-varying covariates

	wave 2-7				Baseline				
Occupation	means	95%HPDI		Pr (PE ¡0)*	means	95%HPDI		Pr (PE ¡0)	Differences*
Managerial	-0.025	-0.052	0.000	0.975	-0.031	-0.064	0.002	0.968	0.005
Sales	-0.066	-0.094	-0.036	1.000	-0.094	-0.132	-0.056	1.000	0.027
Clerical and administrative	-0.037	-0.064	-0.012	0.996	-0.064	-0.097	-0.028	1.000	0.026
Service	-0.155	-0.183	-0.127	1.000	-0.165	-0.200	-0.130	1.000	0.011
Farming and fishing	-0.143	-0.189	-0.095	1.000	-0.175	-0.227	-0.123	1.000	0.032
Mechanical	-0.118	-0.148	-0.090	1.000	-0.151	-0.187	-0.117	1.000	0.033
Operator	-0.167	-0.196	-0.137	1.000	-0.201	-0.234	-0.168	1.000	0.033

*Pr(PE¡0): probability of the partial effects to be less than zero

*Differences: the differences of posterior means of the partial effects between wave2-7 and baseline

Table 3.4: Analysis2: Average partial effects on excellent self-rated health (using time-varying covariates)

In the second analysis, I examine if the same results as the first analysis are obtained when covariates change over time rather than being fixed at baseline. Still the same covariates used by Gueorguieva et al. (2009), but the covariates allowed to change over time in this analysis. Missing covariates are assumed as MAR, and missing observations in covariates are augmented in the algorithms described in the previous section. By allowing covariates to be time-varying, this analysis controls for changes in status and health habits of a respondent on self-rated health over seven waves of the HRS: changes in health behaviors, health insurance, employment status, household wealth and income, and importantly, longest occupation for respondents. Complete sets of average partial effects on poor and excellent self-rated health are presented in tables 3.14 and 3.15 in the appendix.

The results in the second analysis are same as ones from the first analysis in terms of signs on coefficients. That is, compared with the professional workers, all other occupation have negative effects on reporting excellent self-rated health. However, the absolute values of the partial effects are greater in the second analysis. Hence, when changes in covariates over time are considered, the results imply stronger negative associations between occupational groups and excellent self-rated health, relative to the reference, professionals.

Regarding 95% HPDIs, the results in the second analysis show the same pattern as in the first analysis: 95% HPDIs overlap for managerial, sales, and clerical and administrative, and for service, farming and fishing, mechanical, and operator, respectively. Decreases in the absolute values of the partial effects over time indicate that occupational differentials in self-rated health narrow. In sum, the second analysis also show the same results as the first analysis that there are substantial and continuous differentials of longest occupation in self-rated health over time, but the degree of the differentials narrow as ones age.

It is interesting to see the differences between partial effects between wave 2-7 and baseline in the first (table 3.3) and second (table 3.4) analyses. Decreasing absolute values of the partial effects from baseline to wave 2-7 for each analysis provides us significant evidence to support that the occupation-related differences narrow at older ages. In the second analysis, the differences between partial effects over time are reduced when compared with the first analysis. Although both analyses show the health disparities narrow over time, the magnitude is reduced in the case of considering changes in individuals' status and health habits.

My analysis allowing covariates to vary over time shows that there are significant different effects of longest occupation on explaining self-rated health not only at baseline, but over time as well. Such health disparities narrow over the study period. However, the degrees of narrowed gaps are smaller when compared with the first analysis considering only baseline characteristics. Controlling for changes in socioeconomic status and health habits of respondents might be important to explain self-rated health over time.

Analysis3: Explanation of self-rated health using additional covariates

The third analysis includes additional control variables (regular exercise, doctor-diagnosed chronic diseases, BMI, depression) to the second analysis to explain self-rated health. Tables 3.16 and 3.17 in the appendix show complete results about average partial effects of covariates on poor and excellent self-rated health.

Since the analysis includes additional control variables, average partial effects of

Occupation	wave2-7 means	95%HPDI		Pr (PE ¡0)	Baseline means	95%HPDI		Pr (PE ¡0)
Managerial	-0.011	-0.033	0.011	0.863	-0.020	-0.050	0.009	0.909
Sales	-0.037	-0.061	-0.011	0.999	-0.058	-0.092	-0.022	1.000
Clerical and administrative	-0.018	-0.040	0.004	0.937	-0.038	-0.068	-0.008	0.994
Service	-0.087	-0.111	-0.064	1.000	-0.090	-0.121	-0.059	1.000
Farming and fishing	-0.111	-0.147	-0.074	1.000	-0.136	-0.186	-0.084	1.000
Mechanical	-0.083	-0.107	-0.058	1.000	-0.108	-0.140	-0.077	1.000
Operator	-0.121	-0.142	-0.099	1.000	-0.149	-0.179	-0.120	1.000
BMI	-0.004	-0.005	-0.003	1.000	-0.006	-0.007	-0.004	1.000
Exercise	0.073	0.066	0.080	0.000	0.077	0.059	0.095	0.000
CESD	-0.044	-0.046	-0.042	1.000	-0.048	-0.052	-0.044	1.000
Chronic diseases	-0.127	-0.132	-0.123	1.000	-0.191	-0.199	-0.184	1.000

Table 3.5: Analysis3: Average partial effects on excellent self-rated health (using additional covariates)

longest occupation on excellent self-rated health are reduced from the second analysis. And, 95% HPDI of some occupational groups include zero: managerial at baseline and wave2-7, and clerical and administrative for wave 2-7. By using one of the advantages in Bayesian inferences, I calculate posterior probabilities that average partial effects are less than zero. At baseline, the probability for managerial is over 90%; however, that for wave 2-7 are lower than 90%, but the probability of clerical and administrative is over 90%. The analysis coincide with the previous findings that there are substantial and persistent variations in self-rated health, but the differentials narrow over time.

In addition, I investigate several other determinants of self-rated health in this analysis. First, the partial effects of BMI are negative; i.e., a unit increase in BMI has negative effects on reporting excellent self-rated health at baseline and wave 2-7. A number of chronic diseases also show significant relationship with excellent self-rated health, i.e., a number of chronic diseases is negatively associated with excellent self-rated health at baseline and over time. Particularly, average partial effects of a number of chronic diseases are the second highest after ”age” among all covariates. This implies that chronic diseases have important meaning to explain self-rated health in older people.

It has been studied that depression has a negative and significant effect on self-rated health. Also, poor self-rated health is viewed as a concomitant phenomenon of depression. My results also indicate that depressive symptoms are negatively associated with self-rated health. This negative effects remain significantly different from zero over time in table 3.5. Hence, depression should be considered as one of the determinants when we examine self-rated health over time in the elderly. In terms of regular exercise, my results support the findings of Frijters and Ulker (2008) about significantly positive association between regular exercise and excellent self-rated health. It implies that the respondents in the HRS with vigorous physical activity including sports, heavy housework, or a job involving physical activity rated their health status as healthier.

	mean	std	AR(1)	95%HPDI	
ψ_1	-3.19349	0.037557	0.628769	-3.26904	-3.12062
	-0.07359	0.022708	0.98161	-0.12057	-0.02775
	0.090063	0.018862	0.967074	0.051738	0.128077
ψ_2	-3.18834	0.035635	0.711621	-3.25924	-3.12113
	-0.17576	0.028674	0.982354	-0.22923	-0.1205
	0.365284	0.022708	0.992981	0.322141	0.409721

Table 3.6: Posterior summary statustics for missing data mechanisms, ψ

Lastly, in table 3.6, the posterior means for ψ_1 on the dependent variable are $\{-0.07,\ 0.09\}$ and for ψ_2 are $\{-0.18,\ 0.37\}$. Note that the parameters in ψ_1 relate the intermittent missing data and ψ_2 relate the dropout process. Any of the parameters in ψ_1 and ψ_2 do not include zero in their 95% HPDIs. The parameters different from zero imply that missing data might not be at random and be highly related to unobserved data. Hence, in order to analyze self-rated health over time, non-ignorable missing data mechanism should be considered to avoid incorrect analyses.

3.6 Conclusion

In this paper, I have developed Bayesian ordered probit model with random effects to analyze ordinal panel responses subject to non-random missing data. I used Bayesian MCMC algorithms to estimate the model. The strength of the Bayesian method in this paper lies in its ability to incorporate all of available information of randomness and uncertainty in inference, including those in missing data mechanism.

This paper provides a statistical method to correct missing data not only in the dependent variable, but also in time-varying covariates by jointly modeling responses and missing data. Also, the model is able to analyze missing data due to respondents who not only dropout, but also those who missed some waves of the survey, but come back to the survey later.

In order to examine the determinants of self-rated health, the model was applied to the Health and Retirement Study using seven waves. The empirical analyses were conducted by employing three different sets of covariates. The first analysis used only baseline characteristics to explain self-rated health over time, following Gueorguieva et al. (2009). Using same dataset as Gueorguieva et al. (2009), but different models specifying missing data with non-ignorable missing data mechanism, my results confirm the findings of Gueorguieva et al. (2009) that there are substantial differences in self-rated health across longest occupation at baseline and the differences persist over time, but narrow at older ages.

In the second analysis, I employed the same covariates and periods as Gueorguieva et al. (2009), but allow the covariates to be changed over time. As indicated as one of their limitations, Gueorguieva et al. (2009) only consider baseline characteristics in covariates in examining changes in self-rated health. By allowing covariates to be time-varying, my second analysis controls for changes in status and health habits of a respondent on self-rated health over time. Using time-varying covariates, the results show that there are important variations in self-rated health across occupation not only at baseline, but over time as well. The health disparities narrow same as in the first analysis, but the degree of narrowed disparities became smaller.

Finally, in the third analysis, I included additional control variables to explain self-rated health to the second analysis to explain self-rated health. A number of chronic diseases turned out to be key variables to explain self-rated health in the elderly. Also, degree of depressive symptoms has high impact on explaining older people's self-rated health.

I have developed selection models to examine the determinants of self-rated health for older people based on the assumption that missing data might be highly related to older people's health status. However, it is true that we can never observe true missing data mechanism, so, in further analysis, the sensitivity of the posterior inferences should be compared against other missing data models.

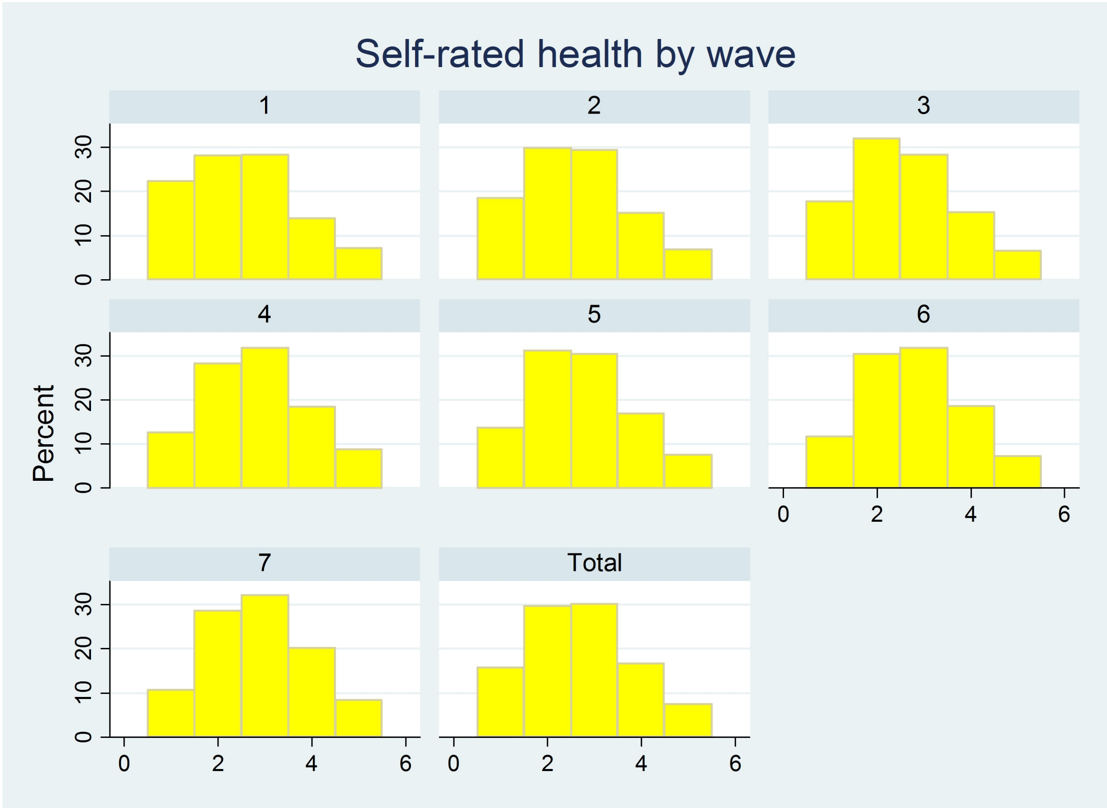

Figure 3.1: Self-rated health by wave

	Missing status (N=9,557)					
Variables	Completed (N)	(%)	Dropped out (N)	%	Missed intermittently (N)	(%)
Self-rated health	5830	61.00	2882	30.00	1206	12.62
Marital status	5805	60.74	2884	30.18	1238	12.95
Smoking	5639	59.00	2883	30.17	1444	15.11
Drinking	5639	59.00	2883	30.17	1444	15.11
Wealth	5837	61.08	2879	30.17	1199	12.55
Income	5837	61.08	2879	30.17	1199	12.55
BMI	5639	59.00	2954	30.91	1354	14.17
Exercise	5631	58.92	2882	30.16	1446	15.13
CESD	5125	53.63	3384	35.41	1504	15.73
Chronic diseases	5837	61.08	2879	30.12	1199	12.55

Table 3.7: Missing data

Variable	Obs	Mean	Std. Dev.
Self-rated health	55134	2.70	1.14
Age (years)	55151	61.15	5.31
Female (%)	66899	48.74	0.50
Not married/partnered (%)	55094	25.15	0.43
Black (%)	66899	16.89	0.37
Hispanic (%)	66899	8.32	0.28
Education (years)	66899	12.17	3.14
Drinking	66899	32.81	0.47
Smoking	54865	20.59	0.40
Health insurance (%)			
Government sponsored	66899	30.38	0.46
Employer sponsored	66899	36.52	0.48
Other private	66899	12.76	0.33
Occupation (%)			
Professional*	66899	12.23	0.33
Managerial	66899	11.69	0.32
Sales	66899	7.40	0.26
Clerical and administrative	66899	12.86	0.33
Service	66899	11.69	0.32
Farming and fishing	66899	2.45	0.15
Mechanical	66899	9.44	0.29
Operator	66899	13.42	0.34
Employment status (%)			
Employed full-time*	66899	31.43	0.46
Employed part-time	66899	6.26	0.24
Not employed/not retired	66899	1.04	0.10
Retired	66899	36.10	0.48
Household wealth ($)	55151	318099	1041136
Household income ($)	55151	54678	85406
BMI	54634	27.50	5.13
Exercise (%)	54854	36.62	0.48
CESD	51806	1.49	1.93
Number of chronic diseases	55151	1.51	1.32

Table 3.8: Summary statistics

self-rated health	1 (Excellent)		2 (Very good)		3 (Good)		4 (Fair)		5 (Poor)	
Nobs	8696	(15.77%)	16417	(29.78%)	16675	(30.24%)	9219	(16.72%)	4127	(7.49%)
	mean	std	mean	std	mean	std	mean	std	mean	std
Age (years)	59.80	5.20	61.10	5.30	61.42	5.29	61.78	5.24	61.72	5.27
Female (%)	49.29	0.50	51.31	0.50	48.97	0.50	50.91	0.50	49.33	0.50
Not married/partnered (%)	20.75	0.41	20.99	0.41	24.83	0.43	31.42	46.42	0.38	0.49
Black (%)	8.59	0.28	11.57	0.32	18.01	0.38	24.00	42.71	0.24	0.43
Hispanic (%)	5.44	0.23	4.42	0.21	8.26	0.28	14.67	35.38	0.13	0.33
Education (years)	13.39	2.79	13.03	2.64	12.15	2.94	10.96	3.30	10.15	3.64
Drinking (%)	22.01	0.41	20.62	0.40	18.55	0.39	14.67	0.35	10.86	0.31
Smoking (%)	15.53	0.36	17.60	0.38	21.62	0.41	25.06	0.43	29.07	0.45
Health insurance (%)										
Government sponsored	24.33	0.43	30.65	0.46	35.74	0.48	48.77	0.50	65.71	0.47
Employer sponsored	52.10	0.50	50.42	0.50	45.65	0.50	33.88	0.47	21.32	0.41
Other private	16.35	0.37	16.55	0.37	15.90	0.37	13.86	0.35	11.34	0.32
Occupation (%)										
Professional*	21.73	0.41	18.49	0.39	13.26	0.34	8.29	0.28	6.71	0.25
Managerial	19.28	0.39	17.72	0.38	13.09	0.34	8.22	0.27	7.10	0.26
Sales	10.44	0.31	9.25	0.29	8.63	0.28	8.02	0.27	8.31	0.28
Clerical and administrative	16.62	0.37	18.12	0.39	15.82	0.36	12.34	0.33	9.79	0.30
Service	8.22	0.27	11.02	0.31	14.97	0.36	20.77	0.41	21.37	0.41
Farming and fishing	2.37	0.15	2.25	0.15	3.04	0.17	3.98	0.20	4.56	0.21
Mechanical	8.99	0.29	10.11	0.30	12.52	0.33	12.78	0.33	14.56	0.35
Operator	9.73	0.30	11.63	0.32	17.40	0.38	24.04	0.43	26.70	0.44
Employment status (%)										
Employed full-time*	51.68	0.50	43.67	0.50	39.46	0.49	25.99	0.44	9.30	0.29
Employed part-time	9.13	0.29	8.13	0.27	7.96	0.27	6.55	0.25	3.10	0.17
Not employed/not retired	1.38	0.12	1.01	0.10	1.45	0.12	1.37	0.12	1.09	0.10
Retired	32.88	0.47	41.54	0.49	43.89	0.50	51.39	0.50	58.32	0.49
Household Wealth ($)	463238	1007710	406846	1476230	275436	801433	193317	604290	109959	323327
Household Income ($)	76265	100347	64235	91053	51430	85710	36487	65263	24845	32238
BMI	25.91	4.09	26.94	4.39	28.02	5.08	28.70	5.89	28.37	6.82
Exercise	50.66	0.50	43.31	0.50	35.75	0.48	25.48	0.44	13.02	0.34
CESD	0.77	1.30	0.94	1.45	1.40	1.76	2.43	2.22	3.83	2.35
Number of chronic diseases	0.60	0.79	1.10	0.99	1.58	1.15	2.29	1.34	3.05	1.54

Table 3.9: Summary statistics by self-rated health (All waves)

self-rated health	1 (Excellent)		2 (Very good)		3 (Good)		4 (Fair)		5 (Poor)	
Nobs	2133	(22.32%)	2694	(28.19%)	2717	(28.43%)	1325	(13.86%)	688	(7.20%)
	mean	std	mean	std	mean	std	mean	std	mean	std
Age (years)	55.38	3.58	55.73	3.66	56.04	3.70	56.16	3.63	56.44	3.54
Female (%)	48.57	0.50	50.00	0.50	47.18	0.50	51.32	0.50	45.49	0.50
Not married/partnered (%)	18.24	0.39	18.71	0.39	20.54	0.40	29.21	0.45	36.34	0.48
Black (%)	8.91	0.28	13.29	0.34	19.58	0.40	27.17	0.45	25.29	0.43
Hispanic (%)	6.00	0.24	4.94	0.22	9.35	0.29	15.17	0.36	11.48	0.32
Education (years)	13.24	2.82	12.90	2.64	11.82	3.06	10.81	3.25	10.00	3.68
Drinking (%)	18.85	0.39	17.41	0.38	15.90	0.37	14.79	0.36	11.48	0.32
Smoking (%)	20.30	0.40	24.20	0.43	29.41	0.46	33.51	0.47	38.66	0.49
Health insurance (%)										
Government sponsored	6.84	0.25	5.98	0.24	8.65	0.28	20.75	0.41	45.78	0.50
Employer sponsored	58.65	0.49	56.98	0.50	53.63	0.50	39.02	0.49	24.56	0.43
Other private	13.22	0.34	13.40	0.34	12.81	0.33	12.38	0.33	9.88	0.30
Occupation (%)										
Professional*	20.72	0.41	16.70	0.37	12.18	0.33	6.64	0.25	6.25	0.24
Managerial	19.74	0.40	16.67	0.37	12.44	0.33	7.40	0.26	6.69	0.25
Sales	9.61	0.29	9.73	0.30	8.50	0.28	8.68	0.28	7.27	0.26
Clerical and administrative	16.36	0.37	17.71	0.38	14.72	0.35	12.15	0.33	8.87	0.28
Service	8.67	0.28	11.73	0.32	15.46	0.36	22.19	0.42	19.62	0.40
Farming and fishing	1.97	0.14	2.49	0.16	3.05	0.17	4.30	0.20	5.38	0.23
Mechanical	9.70	0.30	11.02	0.31	12.81	0.33	12.91	0.34	15.99	0.37
Operator	10.45	0.31	12.69	0.33	19.62	0.40	24.45	0.43	28.92	0.45
Employment status (%)										
Employed full-time*	69.10	0.46	65.18	0.48	61.10	0.49	42.57	0.49	17.88	0.38
Employed part-time	10.97	0.31	11.92	0.32	10.75	0.31	11.70	0.32	3.63	0.19
Not employed/not retired	1.92	0.14	2.15	0.15	3.86	0.19	2.57	0.16	2.76	0.16
Retired	13.36	0.34	15.29	0.36	17.56	0.38	27.92	0.45	47.82	0.50
Household Wealth ($)	310944	531896	262893	553766	168495	315877	123549	291189	74975	185769
Household Income ($)	64255	64898	53876	50061	43741	40887	30336	30230	21544	22785
BMI	25.69	3.94	26.72	4.40	27.72	4.93	28.64	6.19	28.20	6.50
Exercise (%)	27.33	0.45	19.97	0.40	16.97	0.38	14.42	0.35	9.93	0.29
CESD	1.34	1.51	1.77	1.71	2.11	1.84	3.08	2.08	4.26	2.07
Number of chronic diseases	0.41	0.65	0.75	0.81	1.15	0.98	1.83	1.18	2.53	1.40

Table 3.10: Summary statistics by self-rated health (Wave1)

self-rated health	1 (Excellent)		2 (Very good)		3 (Good)		4 (Fair)		5 (Poor)	
Nobs	712	(10.67%)	1914	(28.67%)	2147	(32.16%)	1345	(20.15%)	557	(8.34%)
	mean	std	mean	std	mean	std	mean	std	mean	std
Age (years)	66.97	3.46	67.38	3.61	67.45	3.59	67.47	3.64	67.87	3.77
Female (%)	52.11	0.50	52.77	0.50	49.65	0.50	51.97	0.50	49.91	0.50
Not married/partnered (%)	24.72	0.43	25.04	0.43	28.31	0.45	36.16	0.48	40.57	0.49
Black (%)	7.16	0.26	11.13	0.31	17.28	0.38	22.08	0.41	19.21	0.39
Hispanic (%)	5.48	0.23	4.34	0.20	6.94	0.25	15.61	0.36	14.00	0.35
Education (years)	13.58	2.76	13.13	2.64	12.35	2.82	11.17	3.38	10.44	3.66
Drinking (%)	26.54	0.44	20.27	0.40	18.40	0.39	14.72	0.35	9.69	0.30
Smoking (%)	10.41	0.31	11.18	0.32	14.21	0.35	19.57	0.40	21.58	0.41
Health insurance (%)										
Government sponsored	69.10	0.46	72.62	0.45	76.53	0.42	82.30	0.38	89.95	0.30
Employer sponsored	38.20	0.49	38.09	0.49	33.35	0.47	24.83	0.43	18.85	0.39
Other private	21.63	0.41	21.47	0.41	19.93	0.40	17.77	0.38	12.21	0.33
Occupation (%)										
Professional*	23.46	0.42	19.44	0.40	14.67	0.35	9.59	0.29	6.28	0.24
Managerial	18.68	0.39	18.08	0.38	13.51	0.34	8.33	0.28	8.98	0.29
Sales	11.24	0.32	9.67	0.30	8.38	0.28	6.91	0.25	9.87	0.30
Clerical and administrative	16.99	0.38	19.02	0.39	16.67	0.37	13.31	0.34	11.13	0.31
Service	7.30	0.26	10.34	0.30	14.90	0.36	19.33	0.40	20.47	0.40
Farming and fishing	2.67	0.16	2.25	0.15	2.61	0.16	4.31	0.20	4.31	0.20
Mechanical	8.15	0.27	9.09	0.29	12.62	0.33	12.79	0.33	12.03	0.33
Operator	8.85	0.28	10.76	0.31	15.09	0.36	23.94	0.43	25.13	0.43
Employment status (%)										
Employed full-time*	23.31	0.42	20.22	0.40	15.84	0.37	11.38	0.32	4.13	0.20
Employed part-time	8.29	0.28	4.96	0.22	4.42	0.21	3.42	0.18	0.90	0.09
Not employed/not retired	0.56	0.07	0.37	0.06	0.14	0.04	0.37	0.06	0.36	0.06
Retired	63.90	0.48	69.80	0.46	74.10	0.44	75.84	0.43	76.66	0.42
Household Wealth ($)	809831	1876759	618473	2334577	418154	1118255	244392	585994	157294	324548
Household Income ($)	82636	101200	69638	132754	58949	115464	37302	48111	28308	29224
BMI	26.00	4.27	27.09	4.51	28.52	5.31	28.91	6.13	28.70	6.78
Exercise (%)	50.98	0.50	35.86	0.48	26.47	0.44	16.65	0.37	7.55	0.26
CESD	0.47	1.00	0.70	1.29	1.13	1.64	2.15	2.20	3.62	2.31
Number of chronic diseases	0.96	0.96	1.57	1.12	2.12	1.18	2.78	1.42	3.62	1.53

Table 3.11: Summary statistics by self-rated health (Wave7)

(Wave2-7)	Poor self-rated health mean	std	95%HPDI		Excellent self-rated health mean	std	95%HPDI	
lag	0.046	0.002	0.043	0.049	-0.048	0.002	-0.051	-0.045
Age (years)	0.729	0.028	0.673	0.781	-0.758	0.028	-0.811	-0.699
Female	-0.006	0.008	-0.023	0.010	0.006	0.009	-0.011	0.024
Not married/partnered	0.036	0.010	0.017	0.055	-0.037	0.010	-0.057	-0.018
Black	0.076	0.010	0.055	0.095	-0.079	0.011	-0.099	-0.057
Hispanic	0.066	0.013	0.042	0.091	-0.069	0.013	-0.095	-0.043
Education (years)	-0.015	0.001	-0.017	-0.012	0.015	0.001	0.013	0.018
Drinking	-0.023	0.009	-0.041	-0.004	0.023	0.010	0.004	0.042
Smoking	0.098	0.007	0.083	0.112	-0.103	0.008	-0.117	-0.087
Health insurance								
Government sponsored	0.154	0.012	0.129	0.178	-0.155	0.012	-0.178	-0.130
Employer sponsored	-0.049	0.008	-0.065	-0.033	0.051	0.008	0.035	0.067
Other private	-0.024	0.012	-0.048	-0.002	0.025	0.012	0.003	0.049
Occupation								
Managerial	0.027	0.013	0.002	0.051	-0.028	0.013	-0.053	-0.002
Sales	0.027	0.014	0.000	0.054	-0.028	0.015	-0.056	0.001
Clerical and administrative	0.014	0.012	-0.012	0.039	-0.015	0.013	-0.040	0.013
Service	0.098	0.013	0.072	0.124	-0.102	0.014	-0.129	-0.075
Farming and fishing	0.103	0.023	0.056	0.147	-0.107	0.024	-0.153	-0.058
Mechanical	0.078	0.014	0.051	0.107	-0.082	0.015	-0.111	-0.053
Operator	0.115	0.012	0.091	0.138	-0.120	0.013	-0.145	-0.095
Employment status								
Employed part-time	-0.031	0.011	-0.053	-0.009	0.032	0.012	0.009	0.055
Not employed/not retired	0.027	0.021	-0.012	0.068	-0.028	0.021	-0.070	0.013
Retired	0.058	0.011	0.037	0.080	-0.060	0.011	-0.082	-0.038
Household wealth	-0.008	0.002	-0.011	-0.004	0.008	0.002	0.004	0.012
Household income	-0.059	0.005	-0.069	-0.050	0.062	0.005	0.052	0.071

Table 3.12: Analysis1 (Wave2-7): Average partial effects on poor and excellent self-rated health

(Baseline)	Poor self-rated health mean	std	95%HPDI		Excellent self-rated health mean	std	95%HPDI	
Age (years)	0.757	0.033	0.692	0.821	-0.800	0.033	-0.864	-0.735
Female	-0.002	0.010	-0.022	0.018	0.002	0.011	-0.019	0.023
Not married/partnered	0.045	0.012	0.022	0.067	-0.046	0.012	-0.070	-0.023
Black	0.119	0.014	0.093	0.146	-0.120	0.013	-0.146	-0.095
Hispanic	0.066	0.018	0.030	0.100	-0.067	0.018	-0.101	-0.032
Education (years)	-0.021	0.002	-0.024	-0.018	0.022	0.002	0.019	0.025
Drinking	-0.056	0.011	-0.078	-0.033	0.061	0.013	0.035	0.086
Smoking	0.081	0.010	0.061	0.101	-0.084	0.010	-0.104	-0.064
Health insurance								
Government sponsored	0.268	0.015	0.238	0.299	-0.250	0.012	-0.273	-0.225
Employer sponsored	-0.065	0.010	-0.083	-0.045	0.069	0.010	0.048	0.089
Other private	-0.016	0.014	-0.043	0.011	0.017	0.015	-0.012	0.046
Occupation								
Managerial	0.033	0.017	0.000	0.066	-0.035	0.017	-0.068	0.000
Sales	0.064	0.019	0.029	0.101	-0.065	0.019	-0.102	-0.030
Clerical and administrative	0.044	0.017	0.011	0.076	-0.046	0.017	-0.078	-0.012
Service	0.127	0.017	0.093	0.160	-0.128	0.017	-0.160	-0.096
Farming and fishing	0.157	0.032	0.090	0.218	-0.152	0.029	-0.206	-0.092
Mechanical	0.122	0.019	0.084	0.160	-0.123	0.018	-0.159	-0.086
Operator	0.165	0.017	0.132	0.198	-0.165	0.016	-0.195	-0.133
Employment status								
Employed part-time	-0.028	0.014	-0.055	-0.001	0.031	0.015	0.000	0.060
Not employed/not retired	0.006	0.025	-0.043	0.057	-0.006	0.027	-0.059	0.045
Retired	0.144	0.013	0.119	0.170	-0.146	0.012	-0.171	-0.121
Household wealth	-0.009	0.002	-0.013	-0.005	0.009	0.002	0.005	0.014
Household income	-0.058	0.005	-0.068	-0.048	0.061	0.005	0.051	0.071

Table 3.13: Analysis1 (Baseline): Average partial effects on poor and excellent self-rated health

(Wave2-7)	Poor self-rated health mean	std	95%HPDI		Excellent self-rated health mean	std	95%HPDI	
lag	-0.020	0.001	-0.023	-0.017	-0.045	0.002	-0.049	-0.041
Age (years)	0.044	0.002	0.040	0.047	-0.784	0.023	-0.830	-0.739
Female	-0.001	0.008	-0.017	0.015	0.001	0.009	-0.015	0.018
Not married/partnered	0.016	0.006	0.004	0.027	-0.016	0.006	-0.028	-0.004
Black	0.109	0.010	0.089	0.127	-0.113	0.010	-0.132	-0.092
Hispanic	0.082	0.013	0.056	0.107	-0.085	0.014	-0.111	-0.058
Education (years)	-0.020	0.001	-0.023	-0.017	0.021	0.001	0.018	0.024
Drinking	-0.045	0.006	-0.056	-0.034	0.047	0.006	0.035	0.058
Smoking	0.016	0.006	0.004	0.028	-0.016	0.006	-0.029	-0.004
Health insurance								
Government sponsored	0.015	0.004	0.007	0.023	-0.015	0.004	-0.023	-0.007
Employer sponsored	-0.012	0.004	-0.021	-0.004	0.013	0.004	0.004	0.022
Other private	-0.018	0.005	-0.027	-0.009	0.019	0.005	0.009	0.028
Occupation								
Managerial	0.025	0.013	0.000	0.051	-0.025	0.013	-0.052	0.000
Sales	0.064	0.014	0.036	0.092	-0.066	0.015	-0.094	-0.036
Clerical and administrative	0.036	0.013	0.012	0.063	-0.037	0.013	-0.064	-0.012
Service	0.150	0.014	0.123	0.177	-0.155	0.014	-0.183	-0.127
Farming and fishing	0.139	0.023	0.092	0.184	-0.143	0.024	-0.189	-0.095
Mechanical	0.115	0.015	0.088	0.144	-0.118	0.015	-0.148	-0.090
Operator	0.161	0.014	0.131	0.188	-0.167	0.015	-0.196	-0.137
Employment status								
Employed part-time	-0.003	0.007	-0.017	0.011	0.003	0.007	-0.011	0.017
Not employed/not retired	0.001	0.015	-0.029	0.031	-0.001	0.015	-0.031	0.030
Retired	0.017	0.004	0.009	0.025	-0.017	0.004	-0.026	-0.009
Household wealth	-0.002	0.001	-0.003	-0.001	0.002	0.001	0.001	0.003
Household income	-0.004	0.001	-0.007	-0.002	0.004	0.001	0.002	0.007

Table 3.14: Analysis2 (Wave2-7): Average partial effects on poor and excellent self-rated health

(Baseline)	Poor self-rated health mean	std	95%HPDI		Excellent self-rated health mean	std	95%HPDI	
Age (years)	0.867	0.032	0.804	0.931	-0.913	0.031	-0.972	-0.852
Female	0.005	0.010	-0.016	0.026	-0.006	0.011	-0.027	0.017
Not married/partnered	0.024	0.010	0.004	0.043	-0.025	0.010	-0.045	-0.004
Black	0.146	0.013	0.121	0.172	-0.147	0.012	-0.171	-0.123
Hispanic	0.075	0.017	0.043	0.109	-0.077	0.017	-0.110	-0.045
Education (years)	-0.024	0.002	-0.028	-0.021	0.026	0.002	0.022	0.030
Drinking	-0.055	0.009	-0.072	-0.038	0.059	0.010	0.040	0.078
Smoking	0.006	0.008	-0.010	0.022	-0.006	0.009	-0.023	0.011
Health insurance								
Government sponsored	0.155	0.012	0.131	0.179	-0.154	0.011	-0.176	-0.133
Employer sponsored	-0.029	0.008	-0.045	-0.014	0.031	0.008	0.015	0.047
Other private	-0.001	0.010	-0.021	0.018	0.002	0.011	-0.019	0.022
Occupation								
Managerial	0.030	0.017	-0.002	0.062	-0.031	0.017	-0.064	0.002
Sales	0.093	0.020	0.055	0.132	-0.094	0.019	-0.132	-0.056
Clerical and administrative	0.062	0.018	0.027	0.096	-0.064	0.018	-0.097	-0.028
Service	0.166	0.019	0.129	0.203	-0.165	0.018	-0.200	-0.130
Farming and fishing	0.182	0.031	0.122	0.241	-0.175	0.027	-0.227	-0.123
Mechanical	0.151	0.019	0.115	0.189	-0.151	0.018	-0.187	-0.117
Operator	0.202	0.018	0.167	0.238	-0.201	0.017	-0.234	-0.168
Employment status								
Employed part-time	-0.003	0.011	-0.024	0.019	0.003	0.012	-0.020	0.026
Not employed/not retired	-0.014	0.021	-0.055	0.025	0.015	0.022	-0.027	0.060
Retired	0.099	0.010	0.080	0.117	-0.102	0.010	-0.120	-0.083
Household wealth	-0.007	0.002	-0.011	-0.002	0.007	0.003	0.002	0.012
Household income	-0.024	0.005	-0.033	-0.015	0.025	0.005	0.016	0.035

Table 3.15: Analysis2 (Baseline): Average partial effects on poor and excellent self-rated health

(Wave2-7)	Poor self-rated health mean	std	95%HPDI		Excellent self-rated health mean	std	95%HPDI	
lag	0.0430	0.0016	0.0402	0.0459	-0.0459	0.0016	-0.0489	-0.0429
Age (years)	0.2091	0.0325	0.1483	0.2722	-0.2229	0.0343	-0.2894	-0.1588
Female	-0.0247	0.0068	-0.0387	-0.0117	0.0263	0.0073	0.0124	0.0414
Not married/partnered	-0.0097	0.0052	-0.0199	0.0003	0.0104	0.0055	-0.0003	0.0212
Black	0.0702	0.0078	0.0552	0.0855	-0.0756	0.0084	-0.0925	-0.0593
Hispanic	0.0976	0.0103	0.0773	0.1173	-0.1054	0.0112	-0.1268	-0.0830
Education (years)	-0.0132	0.0012	-0.0154	-0.0109	0.0140	0.0013	0.0117	0.0164
Drinking	-0.0216	0.0050	-0.0316	-0.0119	0.0231	0.0053	0.0127	0.0338
Smoking	0.0401	0.0054	0.0293	0.0505	-0.0429	0.0058	-0.0540	-0.0313
Health insurance								
Government sponsored	-0.0096	0.0043	-0.0180	-0.0012	0.0102	0.0046	0.0013	0.0191
Employer sponsored	-0.0157	0.0043	-0.0241	-0.0073	0.0168	0.0045	0.0079	0.0256
Other private	-0.0178	0.0047	-0.0270	-0.0086	0.0190	0.0050	0.0091	0.0287
Occupation								
Managerial	0.0108	0.0101	-0.0099	0.0310	-0.0115	0.0108	-0.0330	0.0107
Sales	0.0341	0.0119	0.0105	0.0564	-0.0365	0.0128	-0.0606	-0.0111
Clerical and administrative	0.0169	0.0108	-0.0042	0.0374	-0.0180	0.0115	-0.0402	0.0045
Service	0.0807	0.0114	0.0594	0.1030	-0.0866	0.0122	-0.1105	-0.0637
Farming and fishing	0.1028	0.0172	0.0686	0.1353	-0.1109	0.0187	-0.1467	-0.0738
Mechanical	0.0772	0.0117	0.0544	0.0997	-0.0829	0.0126	-0.1071	-0.0582
Operator	0.1116	0.0104	0.0914	0.1315	-0.1205	0.0113	-0.1422	-0.0986
Employment status								
Employed part-time	-0.0029	0.0067	-0.0161	0.0100	0.0030	0.0071	-0.0109	0.0172
Not employed/not retired	-0.0163	0.0154	-0.0469	0.0137	0.0173	0.0163	-0.0144	0.0497
Retired	0.0051	0.0041	-0.0028	0.0132	-0.0054	0.0044	-0.0142	0.0030
Household wealth	-0.0018	0.0005	-0.0028	-0.0008	0.0019	0.0005	0.0009	0.0030
Household income	-0.0040	0.0012	-0.0063	-0.0018	0.0043	0.0013	0.0019	0.0067
BMI	0.0038	0.0005	0.0028	0.0047	-0.0040	0.0005	-0.0050	-0.0030
Exercise	-0.0678	0.0034	-0.0743	-0.0613	0.0725	0.0036	0.0657	0.0796
CESD	0.0412	0.0010	0.0394	0.0431	-0.0441	0.0011	-0.0462	-0.0421
Chronic diseases	0.1181	0.0022	0.1139	0.1223	-0.1274	0.0025	-0.1322	-0.1226

Table 3.16: Analysis3 (Wave2-7): Average partial effects on poor and excellent self-rated health

(Baseline)	Poor self-rated health mean	std	95%HPDI		Excellent self-rated health mean	std	95%HPDI	
Age (years)	0.1700	0.0407	0.0930	0.2490	-0.1906	0.0453	-0.2781	-0.1060
Female	-0.0371	0.0091	-0.0551	-0.0197	0.0416	0.0101	0.0222	0.0618
Not married/partnered	-0.0075	0.0089	-0.0245	0.0100	0.0084	0.0100	-0.0112	0.0276
Black	0.1015	0.0104	0.0807	0.1212	-0.1113	0.0110	-0.1323	-0.0892
Hispanic	0.1185	0.0137	0.0906	0.1446	-0.1299	0.0145	-0.1579	-0.1006
Education (years)	-0.0190	0.0015	-0.0218	-0.0161	0.0214	0.0017	0.0181	0.0247
Drinking	-0.0415	0.0090	-0.0590	-0.0238	0.0470	0.0102	0.0266	0.0672
Smoking	0.0467	0.0078	0.0316	0.0622	-0.0522	0.0087	-0.0694	-0.0354
Health insurance								
Government sponsored	0.0895	0.0112	0.0676	0.1119	-0.0971	0.0118	-0.1201	-0.0745
Employer sponsored	-0.0278	0.0074	-0.0421	-0.0135	0.0312	0.0083	0.0150	0.0473
Other private	-0.0039	0.0097	-0.0232	0.0154	0.0044	0.0109	-0.0174	0.0262
Occupation								
Managerial	0.0181	0.0135	-0.0083	0.0452	-0.0201	0.0150	-0.0501	0.0092
Sales	0.0524	0.0164	0.0196	0.0835	-0.0580	0.0178	-0.0918	-0.0218
Clerical and administrative	0.0344	0.0141	0.0073	0.0614	-0.0383	0.0155	-0.0682	-0.0083
Service	0.0812	0.0148	0.0529	0.1101	-0.0896	0.0160	-0.1207	-0.0589
Farming and fishing	0.1260	0.0247	0.0761	0.1742	-0.1358	0.0255	-0.1859	-0.0837
Mechanical	0.0979	0.0155	0.0694	0.1285	-0.1077	0.0166	-0.1400	-0.0767
Operator	0.1357	0.0143	0.1086	0.1643	-0.1493	0.0151	-0.1788	-0.1199
Employment status								
Employed part-time	0.0043	0.0106	-0.0170	0.0256	-0.0048	0.0119	-0.0285	0.0191
Not employed/not retired	-0.0051	0.0197	-0.0444	0.0329	0.0058	0.0221	-0.0372	0.0503
Retired	0.0624	0.0090	0.0454	0.0803	-0.0690	0.0098	-0.0883	-0.0503
Household wealth	-0.0012	0.0022	-0.0055	0.0031	0.0014	0.0025	-0.0034	0.0062
Household income	-0.0233	0.0041	-0.0315	-0.0154	0.0262	0.0046	0.0173	0.0352
BMI	0.0052	0.0007	0.0038	0.0065	-0.0058	0.0008	-0.0073	-0.0043
Exercise	-0.0671	0.0079	-0.0830	-0.0518	0.0767	0.0092	0.0591	0.0951
CESD	0.0432	0.0019	0.0395	0.0468	-0.0478	0.0020	-0.0516	-0.0439
Chronic diseases	0.1814	0.0034	0.1748	0.1879	-0.1914	0.0037	-0.1985	-0.1840

Table 3.17: Analysis3 (Baseline): Average partial effects on poor and excellent self-rated health

Chapter 4

Time-series Characteristics and Lead-lag Relations of Credit Default Swap, S&P500, and Overnight Federal Funds Rate during the Great Recession

4.1 Introduction

In this paper, I analyze, first, the time-series characteristics of the credit default swap (CDS), stock, and federal funds rate during the Great Recession, $2007 - 2009$. Second, the lead-lag relationship between the CDS and stock markets are examined.

In 2008, we observed the collapse of the financial institutions: Bear sterns was acquired for $2 a share by J.P.Morgan Chase; Merrill Lynch was sold to Bank of America; Lehman Brothers declared bankruptcy with $690 billion in assets; the leading insurance company AIG, the leading bank Citigroup, and the two largest mortgage companies were bailed out by the government. Economists refer such financial market turmoils as the worst financial crisis since the Great Depression of the 1930s. The main cause of the crisis was claimed as the collapse of the United States housing bubble that peaked in 2006. Moreover, credit default swap (CDS) exacerbated the credit crisis by hastening the demise of the financial companies.

While CDS has existed in the past, the beginning of the modern form was created by a team of J.P.Morgan in 1997. Although CDS notional outstanding was about $632 million in 2001, it expanded to $62 trillion at the end of 2007 (International Swap and Derivatives Association (ISDA)). The expansion of the CDS market was in line with that of structured credit products such as mortgage-backed security (MBS), asset-backed security (ABS), and collateralized debt obligation (CDO). The expansion was based on a trend of the low interest rate caused by large inflows of foreign funds and the lax interest rate policy by the Federal Reserve. The low interest rate environment created easy credit conditions for years prior to the crisis, and moreover easy credit conditions promoted a housing construction, debt-financed consumption, and easily achievable mortgage loans.

At the same time, the banking system underwent an important transformation such that banks pooled and tranched various types of loans (e.g., mortgages, corporate bonds, credit card loans, etc), and resold loans via securitization instead of holding loans on banks' balance sheets until maturity. Through securitization, banks could sheft risk to those who wish to bear it, and it allowed banks to relax the credit requirement. Individual low-rated loans were transformed into investment-grade structured products via securitization. However, the complicated process of securitization has a drawback that investors cannot obtain any information about the reference entities on these structured credit products except credit grades. The drawback seemed to be relieved by insuring against the default of the structured products or a particular bond by purchasing CDSs. That is one reason that a growth of CDS has been accompanied with that of the structured products.

Although there has been criticism about CDS that exacerbated the credit crunch of 2008, it is also true that CDS brings investors enormous benefits through reducing their exposure on default of the underlying entities. This benefit has carried a great weight on making investors hard to disregard the CDS market, and the CDS market still live and active today. For example, the overall amount of insurance on Greek debt hit $85 billion in February 2010, up from $38 billion a year ago, against Greece default on debt through CDS according to the Depository Trust and Clearing Corporation.

In spite of significance of the CDS market, previous studies have been addressed the time-series properties of CDS only in a cursory manner. My study focuses carefully on 1) time-series characteristics of financial markets including CDS, stock, and federal funds markets in the U.S., and 2) intertemporal relations between the CDS and stock markets in the U.S.

This empirical study has the following significant implications. First, with the index-level data rather than firm-specific data, I examine the time-series characteristics of each market and information flow between the CDS and stock markets driven by market-wide risk. Movements in the individual CDS and stock prices can be analyzed with two layers, the systematic[1] and idiosyncratic risk. Hence, price changes in individual CDS or stock (i.e., the firm level) can be explained as a response of the individual security to changes in the market-wide systematic risk and/or the nonsystematic shocks such as individual corporate events. By using index-level data, I can smooth the disturbances attributed to firm-specific risk, and it allows us to observe time-series characteristics and intertemporal relationship only driven by the market-wide risk. In addition, each market index can be regarded as a representative of investors' prediction to the aggregate market risk since indices represent a standardized portfolio of single CDSs or stocks. Hence, this study about indices would provide information on the portfolio level. Since modelling and forecasting portfolios are crucial for risk management, the study using the indices has significance on risk management.

Especially using the CDX index, this study has some advantages compared to studies using single-name CDSs. Firstly, empirical studies of single-name CDSs tend to be distorted by the liquidity problem on each firm-specific level. However, this study would be free of the bias problem induced by liquidity since market liquidity tends to be more concentrated on the CDX indices, which are traded in higher volumes than the individual CDS. In addition, more liquid CDX index is more appropriate to reflect the arrival of credit information than the individual CDS. Secondly, Acharya and Johnson (2007) present evidence of insider trading in the CDS market. Since most of the major players are insiders in the CDS market, it should not be overlooked that asymmetric information and insider trading problems cause studies about the single-name CDS to be biased. By using the CDX index, this study needs not consider idiosyncrasies arising from insider trading; therefore, I might reduce the bias problem induced by

[1]In finance, systematic risk, also called market risk, or undiversifiable risk, is the risk associated with the overall aggregate market return. Systematic risk is a risk of security that cannot be reduced through diversification. It should not be confused with systemic risk, which is the risk that the entire financial system will collapse as a result of some catastrophic event, not to any individual's entity.

insider trading.

Second important implication of this study is a model specification. In order to examine the time-series characteristics of the CDS, stock, and federal funds markets, the CKLS model (Chan et al. (1992)) is employed. In original CKLS model, the conditional variance of changes in the interest rate depends on the level of the interest rate. Since the conditional variance is constant and depends on the interest rate, the original CKLS model is considered to be restrictive. Rather than following the original CKLS model, I use the CKLS model allowing both GARCH and level effects by employing CKLS-ARMA-GARCH-EPD model, which is developed by Li (2009). By letting the error terms to follow an exponential power distribution, the CKLS model in this study allows us to estimate excess kurtosis of the residuals. Using the CKLS-ARMA-GARCH-EPD model allows us to compare the markets with respect to time-series properties such as volatility, mean reversion, and kurtosis.

In addition, in order to examine intertemporal relations between the markets, the CKLS-ARMA-GARCH-EPD model is modified to employ multivariables as a vector autoregression model (VAR). The modified CKLS model is different from the VAR model in three aspects: (1) the error term follows ARMA-GARCH processes, (2) the error term is specified by $CDS_{t-1}^{c}u_t$, not by u_t, and (3) the error term, u_t, follows the exponential power distribution. In general, our specification become more general than the original CKLS and VAR models by specifying the heteroscedastic conditional variance and level effects by employing GARCH and ARMA processes and allowing excess kurtosis of the residuals by allowing the error term to follow the exponential power distribution.

Third implication is about intertemporal relationship between the CDS and stock market during the sample period of this study during the credit crunch in 2008. Empirical studies have provided mixed results showing a leading role between the CDS and stock markets. With the sophisticated model specification rather than the VAR model and the time-series samples for highly volatile periods from 2007 to 2009, I could observe that the CDS market leads the stock market in the U.S. The result is comparable with a result from Fung et al. (2008). Fung et al. (2008) examine the market-wide relations between the CDS and stock markets in the U.S. during the period of 2001 to 2007, which does not cover the period of the credit crunch. By employing two different CDS indices, investment grade CDX (CDX.NA.IG) and high yield CDX (CDX.NA.HY), they distinguish respective information flows according to credit qualities. They find that the stock market leads less risky CDX.NA.IG market, whereas more risky CDX.NA.HY market leads the stock market. Interestingly, I find the same result by employing investment grade CDX index during the credit crunch as Fung et al. (2008), which employs high yield CDX index during the period before the credit crunch. The relationship is that credit information arrives first in the CDS market and the changes in the CDS index affect the stock market next. This can be a distinct characteristic of lead-lag relations between the markets during the financial crisis.

Organization of this paper is as follows. In Section 2, the CKLS-ARMA-GARCH-EPD model is introduced and the empirical results are explained. In Section 3, the modified CKLS model is introduced and the following results are described. In Section 4, I discuss further study in the future and conclude the paper. In the appendix, MCMC algorithms employed in this study are presented, and a brief history of regulations of the U.S. financial market is examined.

4.2 Time-series analysis

In this section, I use the CKLS-ARMA-GARCH-EPD model and MCMC algorithms to make Bayesian inference on CDX.NA.IG index, S&P500 index, and federal funds rate. MCMC algorithms employed are explained in the appendix.

4.2.1 CKLS-ARMA-GARCH-EPD model

Given the CKLS model

$$y_t = a + by_{t-1} + y_{t-1}^c u_t \tag{4.1}$$

we specify the error term u_t to follow an ARMA(p,q) process

$$u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j} \tag{4.2}$$

and $\{e_t\}$ is given by an exponential power distribution (EPD) with a GARCH(r,s) process

$$\begin{aligned}
e_t &= \sigma_t \varepsilon_t &(4.3)\\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{r} \alpha_i e_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 &(4.4)\\
\alpha_0 &> 0, \quad \alpha_i > 0, \quad i = 1, ...r\\
\beta_i &\geq 0, \quad i = 1, ..., s\\
1 &\geq \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i)
\end{aligned}$$

where ε_t follows the exponential power distribution (EPD) with variance normalized to be unity. The probability density function of ε_t is given by

$$f(\varepsilon_t | Y, X) = \frac{1}{\lambda 2^{1+\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)} \exp\left[-\frac{1}{2}\left|\frac{\varepsilon_t}{\lambda}\right|^{\alpha}\right] \tag{4.5}$$

$$\lambda = \sqrt{\frac{2^{-\frac{2}{\alpha}} \Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)}} \tag{4.6}$$

λ is a constant to make the variance of ε_t as unity.

Let

$$\Theta = \left[a, b, c, \{\phi_i\}_{i=1}^{p}, \{\theta_i\}_{i=1}^{q}, \{\alpha_i\}_{i=0}^{r}, \{\beta_i\}_{i=1}^{s}, \alpha\right].$$

The posterior distribution for Θ is

$$p(\Theta | Y, X) \propto p(\Theta) \prod_{t=1}^{T} \frac{y_{t-1}^{-c}\ \sigma_t^{-1}}{\lambda 2^{1+\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right)} \exp\left[-\frac{1}{2}\left|\frac{\varepsilon_t}{\lambda}\right|^{\alpha}\right]\ p(\Theta) \tag{4.7}$$

where

$$\begin{aligned}
\varepsilon_t &= \frac{e_t}{\sigma_t} &(4.8)\\
e_t &= \frac{y_t - (a + b y_{t-1})}{y_{t-1}^{c}} - \sum_{j=1}^{p} \phi_j u_{t-j} - \sum_{j=1}^{q} \theta_j e_{t-j}
\end{aligned}$$

and the prior is given by

$$
\begin{aligned}
p(\Theta) &= p\left[a, b, c, \{\phi_i\}_{i=1}^{p}, \{\theta_i\}_{i=1}^{q}, \{\alpha_i\}_{i=0}^{r}, \{\beta_i\}_{i=1}^{s}, \alpha\right] \qquad (4.9) \\
&= p(a)p(b)p(c)\prod_{i=1}^{p} p\{\phi_i\} \prod_{i=1}^{q} p\{\theta_i\} \prod_{i=0}^{r} p\{\alpha_i\} \prod_{i=1}^{s} p\{\beta_i\}\, p(\alpha) \\
&= N_a(\mu_a, \Sigma_a)\, N_b(\mu_b, \Sigma_b)\, N_c(\mu_c, \Sigma_c) \prod_{i=1}^{p} N_{\phi_i}(\mu_{\phi_i}, \Sigma_{\phi_i}) \prod_{i=1}^{q} N_{\theta_i}(\mu_{\theta_i}, \Sigma_{\theta_i}) \\
&\quad \prod_{i=0}^{r} N_{\alpha_i}(\mu_{\alpha_i}, \Sigma_{\alpha_i}) \prod_{i=1}^{s} N_{\beta_i}(\mu_{\beta_i}, \Sigma_{\beta_i})\; N_{\alpha}(\mu_{\alpha}, \Sigma_{\alpha})
\end{aligned}
$$

4.2.2 Empirical Results

Using the daily sample from January 2007 to September 2009, I estimate the CKLS-ARMA-GARCH-EPD model for each of the CDX.NA.IG index, S&P500 index, and federal funds rate using MCMC algorithms that are explained in the appendix.

I ran $50,000$ MCMC iterations, discarded the first $10,000$ draws and kept every 20th draw. The acceptance rates are higher than 0.2006 for the S&P500 index; 0.4625 for the CDX index; and 0.2545 for federal funds rates. The convergence of the MCMC draws was judged first by checking that the plots of the MCMC draws for each parameter exhibits randomness without drift. Then I applied the filtered fluctuation test (FT) and the filtered Kolmogorov-Smirnov test (KST). I accepted the null hypothesis of convergence at the 5% significance level. The summary statistics are given in Table 1.

Table 4.1 Here.

This study examined the time-series characteristics such as volatility, mean reversion, and kurtosis. The parameter α_0 for the GARCH process in equation (4.4) indicates volatility: the larger the value of α_0, the larger is the conditional variance, σ_t^2. The sum of α_i and β_i indicates the persistence of shocks: if $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i)$ is unity, then we have an I-GARCH process. The regression coefficient b in equation (4.1) shows whether y_t is mean reversion or not. The parameter c determines if the short run asset price y_t follows a particular process. For example, if $c = 0.5$, then we have the CIR process.

The parameter α of the EPD distribution shows if we have a leptokurtic, mesokurtic, or platykurtic process.

Let me examine volatility in terms of α_0. The summary statistics for α_0 are given as

Summary Statistics for α_0			
	Mean	Std	95% HPDI
CDX	.6601	.2130	(.2737, .9969)
S&P500	.8437	.2178	(.3142, .9989)
FFR	.0036	.0024	(.0002, .0083)

Notes CDX =Credit default swap index

FFR = federal funds rate

Mean = Posterior mean

Std = Posterior standard deviation

95% HPDI = 95% highest posterior density interval

The posterior mean of α_0 for S&P500 is 0.8437 that is larger than the posterior mean of α_0 for CDX. The 95% HPDI for S&P500 overlaps with that for CDX. However, the majority of the posterior pdf of α_0 for S&P500 lies to the right of that for CDX shown in Figure **??**.

Figure 4.1 Here.

According to the results, the stock market is the most volatile, the CDS market is next, and federal funds market is the least volatile during the credit crunch. The result is surprising in that I expected that the CDS market that mainly caused the market turmoil would be the most volatile during the credit crunch. The reason that the CDS market is less volatile than the stock market during the credit crunch would be found in the opposite characteristics between the stock and CDS. During the credit crunch, the credit risk became high across all financial markets. When a probability of default on a firm increases, the stock price of the firm falls, however, the spread of the CDS of the firm increases. It is because an investor who holds the stock has no value, but an investor who holds the CDS is repaid with a notional amount without a loss when the firm bankrupts or defaults. Hence, during the financial market turmoil, CDSs become relatively safe assets to stocks. We can observe it in our results that the price fluctuation of stock is more volatile than that of CDS during the credit crunch. The posterior mean of α_0 for the federal funds rates is close to zero indicating that volatility of the federal fund rate is very low.

Duffie and Duffie (1999) argues that the failure of incorporating a GARCH effect has resulted in model specification errors for credit spreads. In order to observe a GARCH effect, I examine the posterior pdf's of

$$\alpha\beta = \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) \tag{4.10}$$

$\alpha\beta$ indicates the persistence in volatility. The posterior pdf's of $\alpha\beta$ for CDX and S&P500 are given in Figure **??**.

Figure 4.2 Here.

The posterior means of $\alpha\beta$ are 0.9521, 0.8397, and 0.8494 for CDX, S&P500, and federal funds rates, respectively. Judged by Figure 4.14, we see that $\alpha\beta$ is positive and there are high probabilities that the GARCH processes are I-GARCH processes for all three time series.

The following table gives the posterior means, standard deviations, and the 95% HPDI's of the regression coefficient b.

Summary Statistics for b			
	Mean	Std	95% HPDI
CDX	.997	.0026	(.9918, 1.0)
S&P500	−1.3048	1.6622	(−3.2642, 1.0)
FFR	.9984	.0035	(.9946, 1.0)

Notes CDX =Credit default swap index

FFR = federal funds rate

Mean = Posterior mean

Std = Posterior standard deviation

95% HPDI = 95% highest posterior density interval

We see that the posterior pdf of b for S&P500 is centered around -1.3048 with a large posterior standard deviation. It indicates that S&P500 index is not a mean-reverting process. The pdf's of b for CDX and federal funds rate series are tightly distributed around 1.0 and clearly the CDX and federal funds rate series are also not mean-reverting.

The parameters of the ARMA (p, q) error process indicate that the AR process is stationary and the MA process is invertible.

The posterior means, standard deviations, and 95% HPDI's for c are given in the table below.

Summary Statistics for c

	Mean	Std	95% HPDI
CDX	.1633	.1107	(.0001, .3651)
S&P500	.4949	.1703	(.0720, .8447)
FFR	.4168	.1454	(.1283, .7024)

Notes CDX =Credit default swap index

FFR = federal funds rate

Mean = Posterior mean

Std = Posterior standard deviation

95% HPDI = 95% highest posterior density interval

The value of c determines the type of the spot asset price model. For example, if $c = 0$, then it is the Vasecek model; if $c = 0.5$, then it is the CIR model. If $c = 1$, then it is the geometric Brownian motion model of Black and Scholes (1974). The dynamics of each time-series volatility are comparable through a parameter c in the original CKLS model. The original CKLS model is following:

$$dy = (a + by)dt + \sigma y^c dZ$$

Chan et al. (1992) empirically compare alternative short-term interest rate models based on c, which indicates an elasticity of volatility in interest rate changes. A term of $\sigma^2 y^{2c}$ presents the variance of unexpected interest rate changes. It is well known that the volatility of the short-term interest rates, i.e. the conditional variance of changes in the interest rate, is sensitive to its level. The Cox-Ingersoll-Ross square root process (CIR SR) model implies that the conditional volatility of changes in y is proportional to y, i.e. c is 0.5. The Dothan, Geometric Brownian motion (GBM), and Brennan-Schwartz models indicate that the conditional volatility of changes in y is proportional to y^2, i.e. estimated c is 1. The CIR variable-rate (VR) model shows that the conditional volatility of changes in y is proportional to y^3, i.e. c is estimated to be $\frac{3}{2}$.

Since I specify the spot asset rate model that follows the GARCH process, I cannot exactly examine whether conditional volatility is proportional to a power transformation of time series. I examine only which nested model in the CKLS can explain each time series well. For S&P500, the posterior mean of c is 0.4949 and 95% HPDI is from 0.0720 to 0.8447. Although the posterior mean is not exactly 0.5, the value of 0.5 included in 95% HPDI indicates that S&P500 follows Cox-Ingersoll-Ross (1985) square root process (CIR SR). Interestingly, 95% HPDI for c for federal funds rate also contains 0.5. Hence, both S&P500 and federal funds rate share the same model, which follows Cox-Ingersoll-Ross square root process nested in the CKLS model. However, in the case of CDX, the posterior mean of c is 0.1633 and 95% HPDI is from 0.0001 to 0.3651.

Finally, we look at a parameter α in equation (4.5), an exponential power distribution (EPD). The following table presents the posterior means, standard deviations, and 95% HPDI's for α:

Summary Statistics for α			
	Mean	Std	95% HPDI
CDX	.7962	.0767	$(.6403, .9422)$
S&P500	.4141	.0811	$(.2516, .5814)$
FFR	.5037	.0815	$(.3493, .6703)$

Notes CDX =Credit default swap index

FFR = federal funds rate

Mean = Posterior mean

Std = Posterior standard deviation

95% HPDI = 95% highest posterior density interval

By allowing the error terms to follow an EPD instead of a normal distribution, we can examine the kurtosis: if $\alpha = 2$, the distribution is mesokurtic. The parameter α can be transformed into the kurtosis, γ_4 :

$$\gamma_4 = \frac{\Gamma\left(\frac{5}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{3}{\alpha}\right)^2} \tag{4.11}$$

Table below presents the posterior means, standard deviations, and 95% HPDI's of γ_4 for CDX, S&P500, and federal funds rate.

Summary Statistics for γ_4			
	Mean	Std	95% HPDI
CDX	8.955	1.8398	$(6.0247, 12.7238)$
S&P500	91.9734	204.6683	$(13.989, 295.8221)$
Fed	29.512	15.5826	$(7.0578, 57.073)$

Notes CDX =Credit default swap index

Fed = federal funds rate

Mean = Posterior mean

Std = Posterior standard deviation

95% HPDI = 95% highest posterior density interval

After the parameter α is drawn by the MCMC algorithm, we transform α into kurtosis through equation (4.11) and the posterior pdf's of kurtosis are estimated. For

S&P500, the posterior mean of α is 0.4141, and the posterior mean of its kurtosis is transformed as 91.9734, which is much greater than 3. From the results, we observe that the S&P500 is not normally distributed and it represents leptokurtic characteristic. For CDX, the posterior mean of α is 0.7962 and the posterior mean of its kurtosis is 8.9549, which is also leptokurtic. The posterior mean of α for federal fund rates is 0.5037 and that of its kurtosis is 29.512. From the results of posterior pdf's of kurtosis, it shows that the stock market shows the strongest leptokurtic behavior, then the federal funds market is followed, and the CDS market shows the modest leptokurtic behavior.

4.3 Lead-Lag Relationship

So far, we have examined time-series characteristics of the CDX.NA.IG index, S&P500 index, and the federal funds rates respectively. In this section, I analyze the intertemporal relationship among the markets. Recently, the market participants have addressed the intertemporal relations between the corporate bond or stock markets and the credit derivatives market. In studying lead-lag relations, Longstaff et al. (2003) quotes the December 5, 2002 Wall Street Journal: "then, because the young market has something of a reputation as an early warning signal for spotting corporate debt problems, the higher insurance prices can cause other investors to worry and thus push a company's bond and share prices even lower."

There can be two hypotheses: all financial markets react instantaneously when information on credit conditions of the firms arrives, or the credit derivatives market will respond to the credit (or default) information earlier than other financial markets. The idea that credit derivatives market will react before other bond and stock markets react can be induced because credit derivatives are determined solely by credit risk, and the credit derivatives market is comprised of a large number of sophisticated participants, dominated largely by banks and hedge funds. However, empirical studies have provided mixed results showing a leading role of the credit derivatives market.

Longstaff et al. (2003) examines the lead-lag relations between the credit derivatives, corporate bond, and equity markets in the U.S. based on weekly observations of 67 single-name CDSs for March 2001 to October 2002. Using a vector autoregression (VAR) framework employing the three variables with two lags, the corporate bond spread is predictable for 37 out of the 67 cases. However, the number of the cases to significantly predict the CDS premium is reduced to 12 and to significantly predict the stock return is also decreased to 7 out of the 67 cases. Interestingly, among the three variables, the CDS premium frequently leads the corporate bond spread, specifically 21 of the 67 firms; and the stock return leads the corporate bond spread for also 21 of the 67 firms. In addition, the CDS premium leads the stock return for 10 out of the 67 firms; and the stock return leads the CDS premium for 12 out of the 67 firms. Hence, they conclude that credit information tends to flow first into the credit derivatives and equity markets, and then into the corporate bond markets for the period from 2001 to 2002.

Norden and Weber (2004) analyze the empirical relationship among CDS, bond, and stock markets during the period 2000 to 2002. They apply the same VAR framework as Longstaff et al. (2003) to weekly and daily time series of 58 international firms. Using the VAR framework with the daily data, the number of firms whose lagged stock returns significantly explain changes in the CDS spread are 39 of the 58 firms, while changes in the CDS spread Granger-cause stock returns for the 5 firms. With regard to changes in the bond spread, the lagged CDS spreads lead changes in the bond spread for 33 of the 58 firms, and with the lagged stock returns it reduces to 21 firms. Norden and Weber find that changes in the CDS spread are frequently able to forecast bond spread, which is the same as findings in Longstaff et al. However, Norden and Weber (2004) find a definite lead of the stock market relative to the CDS market and this is in opposition to the findings in Longstaff et al. Norden and Weber explain that one reason for this difference may be the sample composition. Longstaff et al. (2003) exclusively analyze the 67 U.S. firms, whereas Norden and Weber (2004) use an international sample in which 35 out of 58 firms are European firms and remaining 23 are the U.S. firms. Interestingly, the evidence for the leading role of the CDS market with respect to the bond market is stronger in the case of the U.S. companies than in the case of European companies.

Forte and Pena (2006) construct homogeneous measures of credit risk for each bond, CDS, and stock, and the measures are based on bond spreads, CDS spreads, and implied stock market credit spreads, respectively. Then, they employ a VAR model to analyze the lead-lag relations among the changes in the measures. For the stock market, the lagged CDSs Granger-cause the current changes in the stock returns for 5 of the 65 firms. However, the lagged changes in the stock returns Granger-cause the CDSs in 24 of the 65 firms. Overall, they conclude that the stock market is leading the corporate bond and CDS markets in most cases. Their results are for international markets; therefore, it is not clear if their findings are applicable to the U.S. market.

Fung et al. (2008) examine the market-wide relations between the U.S. stock and the CDS market during the period of 2001 to 2007. Unlike the previous three studies, Fung et al. (2008) employ the indices from the stock and CDS markets and focus on the information flow by the market-wide risk. In addition, they construct new stock indices. Although the S&P 500 index is comprised of companies that are generally of high quality credit, there could be a mismatch of credit quality of the index components between the S&P 500 and CDX indices. The new stock indices that Fung et al. construct are based on the returns of the matching firms of the CDX IG and HY index components. Since the CDX indices are equally weighted by their underlying single name CDS contracts, the new stock indices with the investment grade and high yield grade are also constructed equally weighted. Through constructing the new stock indices with matching firms as in the CDX indices, they differentiate credit information content according to the credit quality. The results draw different conclusions depending on the credit quality. First, the stock market leads the investment grade CDS market. In contrast, more importantly, the high yield CDS market has the ability to lead the stock market first; then, the stock market has affected the CDS market. They find the two-way interaction between the stock and high yield CDS markets. The result is consistent with the notion that the stock and high yield CDS markets provide complementary information, which is subsequently incorporated in the other market.complementary information, which is subsequently incorporated in the other market.

4.3.1 Model

I analyze the intertemporal relationship between the CDS and stock markets by employing multivariables in the CKLS model. Although time-series characteristics have been examined in the previous section with respect to the three markets with CDX.NA.IG index, S&P500 index, and the federal funds rate, lead-lag relations in this section will be analyzed with only two markets such as the CDS and stock markets. The federal funds market is excluded because I found in the preliminary analysis that federal funds rate was too stable to be affected by information flow. Hence, I could not find any significant lead-lag relationship among the markets when employing federal funds rate.

The CKLS models incorporating both stock and CDS markets are the following:

$$CDS_t = a_c + b_c CDS_{t-1} + \sum_{i=1}^{l} b_{si} Stock_{t-1} + CDS_{t-1}^{c_c} u_t \tag{4.12}$$

$$Stock_t = a_s + \sum_{i=1}^{k} b_{ci} CDS_{t-1} + b_s Stock_{t-1} + Stock_{t-1}^{c_s} u_t \tag{4.13}$$

The CKLS model is different from the VAR model in three aspects: (1) the error term follows an ARMA-GARCH processes, (2) the error term is specified by $CDS_{t-1}^{c} u_t$, not by u_t and (3) the error term, ε_t in equation (3), follows the EPD distribution.

If the CDS market has affected the stock market, the coefficients on the lagged CDS in equation (4.13) will be significantly different from zero. However, if the stock market leads the CDS market, then the coefficients on the lagged Stock in equation (4.12) will be significant. In addition, both of the coefficients will be significantly different from zero when the markets have a two-way interaction.

4.3.2 Empirical Results

Using the same sample as before from January 2007 to September 2009, I have estimated the CKLS model employing the two variables, CDX.NA.IG index and S&P500 index.

By using daily data, I hope to see if different markets may respond differently to new information. I ran 50,000 MCMC iterations, discarded the first 10,000 draws and kept every 20th draw. The convergence of the MCMC draws was checked by same methods as in the previous section. The posterior summary statistics of the coefficients are given in Table 2.

Table 4.2 Here.

In Norden and Weber (2004), they indicate that a maximal lag of order 5 (indicating one week) seems reasonable for the daily data. As suggested by the study, I estimate each model in (4.12) and (4.13) with different lags from order 5 to order 1. Then, I choose a lag of order 1 for the case of the CDS market of (4.12), and a lag of order 3 for the stock market of (4.13).

During the sample period from 2007 to 2009, the relationship between the CDS and stock markets is different from what the previous studies have found: (1) the posterior mean of b_s on the lagged stock in (4.12) is -0.0006, which is not significantly different from zero. The posterior density for b_s is highly skewed to the left: a probability for the coefficient on the lagged stock to be less than or equal to zero is about 75 percent i.e. $prob(b_s \leq 0) = 0.7495$. (2) The posterior mean of b_{c1} on the first lagged CDS in ($\geq$) is 0.0178, which is close to zero. However, the posterior mean of b_{c2} on the second lagged CDS is 0.4001 and that of b_{c3} on the third lagged CDS is 0.3432, both of which are much greater than zero. The posterior distribution for the sum of the three coefficients $(b_{c1} + b_{c2} + b_{c3})$ is highly skewed to the right of zero with a probability of about 77 percent i.e. $prob((b_{c1} + b_{c2} + b_{c3}) \geq 0) = 0.769$. This result indicates that the CDS market led the stock market in the U.S. during the credit crunch from 2007 to 2009, and the result is different from the previous studies. For example, Longstaff et al. (2003) do not find any clear lead-lag relations between the CDS and stock markets of the U.S. Norden and Weber (2004), and Forte and Pena (2006) find that individual stock returns significantly lead CDS spread, and Fung et al. (2008) find that there are lead-lag relations such that the stock market leads the CDS market in terms of the CDX.NA.IG index. However, although this study employs the CDX.NA.IG index, the result supports a finding of the Fung et al. with respect to the CDX.NA.HY index that credit information arrives first in the high yield CDX market, and affects the stock market after.

In order to examine the specific reason why this study draws different results from the studies of Norden and Weber (2004), and Forte and Pena (2006), I need to consider special characteristics of the CDS. One of the important characteristics is that CDS is basically an insurance contract on corporate default. We can infer that informed traders who would like to make profits from the likelihood of default on a company's bonds or to insure against such default may prefer to hold CDS instead of stocks. In other words, the higher is likelihood of default on a company, the more informed traders who have prior knowledge on the default do investment on CDS of the company. In addition, since CDS is a contract rather than a security, the notional amount of CDS is not limited by supply and demand, i.e. CDS can be created as long as market makers exist. It is important to note here that CDS trade refers to a notional, the quantity of the underlying asset or benchmark to which the derivatives contract applies. It is similar as the amount of insurance bought, not the premium paid. In sum, if informed traders in aggregate tend to trade CDSs than stocks without limitation of the notional amount, then CDS spread should lead the stock prices. In this study, I observe the circumstance that the CDS market leads the stock market, but Norden and Weber (2004), and Forte and Pena (2006) observe the opposite circumstance that the stock market leads the CDS market.

I could find one possible reason in Fung et al. (2008) that explain about the different results. The result of the study of Fung et al. (2008) shows that lead-lag relationship between two markets are contingent upon the credit quality of the underlying entities. Fung et al. (2008) examine lead-lag relations between the stock index and two CDX indices differentiated by credit qualities: CDX.NA.IG and CDX.NA.HY. First in the case of the investment grade CDX, their result is same as the previous studies that a price movement in the stock market has an ability to affect the CDS market. However, in the case of high yield grade CDX, the result shows the CDS market leads the stock market. These different results contingent upon credit qualities of the underlying entities might explain the different results drawn between this study and the studies of Norden and Weber (2004), and Forte and Pena (2006). During a stable period as in Norden and Weber (2004), 2000-02, and Forte and Pena (2006), 2001-03, investors have less incentive to insure against corporate default and prefer stocks to CDSs. It might cause the stock market to reflect information earlier than the CDS market.

However, my sample period includes the credit crunch starting from 2007. The financial market turmoil led by the credit crunch caused most of the companies, regardless of their credit grades, to be at risk of degrading credit qualities. Hence, investors would have liked to prevent the degrade risk or bet on the risk by holding CDS. In other words, during the period, speculative traders might prefer the CDS market to the stock market in order to bet on the likelihood of default on corporate bonds. And traders who held corporate bonds might prefer to insure those against default risk that became very high during the credit crunch. It might cause the CDS market to be faster to reflect information than the stock market. This is why we can see the results that investment grade CDX index leads the stock index in our study. In sum, different sample periods might allow the result of Fung et al. that can be observed with low quality CDSs to appear in our study in which high quality CDSs are used. Hence, different sample periods might be one reason why our study and the studies of Norden and Weber, and Forte and Fena have the opposite results.

Another possible reason rather than the sample periods could be due to model specification or estimation used in the studies. In order to see differences induced by the different model specification, I compare the results between the modified CKLS and VAR models. I run the VAR model with the same samples employing 1 and 3 lags, respectively. The following is the VAR model when $k = 1$ or 3:

$$CDS_t = c_1 + \sum_{j=1}^{k} b_{1j} CDS_{t-j} + \sum_{j=1}^{k} c_{1j} Stock_{t-j} + e_1 \quad (4.14)$$

$$Stock_t = c_2 + \sum_{j=1}^{k} b_{2j} CDS_{t-j} + \sum_{j=1}^{k} c_{2j} Stock_{t-j} + e_2 \quad (4.15)$$

where the error terms in the VAR model, e_1 and e_2, are normally distributed such that $E(e_i) = 0$ and $var(e_i) = \sigma_i^2$, $i = 1$ or 2. The results are provided in Tables 4.3 and 4.4.

Table 4.3 and 4.4 Here.

According to the results from equation (4.14) in Table 3, the t-value of the coefficient on the lagged Stock is 0.33, and the t-value of the coefficient on the lagged CDS in equation (4.15) is 1.07. When one lag is employed, both t-values confirm that the lagged Stock and CDS variables are not significant to explain the paired variables, which are CDS and Stock respectively. However, when 3 lags are employed, the t-values in Table 4 show that each of the CDS_{t-2} and CDS_{t-3} is significant to predict the current value of Stock. But any of the lagged Stock do not have an impact on explaining CDS.

For testing the Granger causality, I also conduct the Wald test that the coefficients on other lagged variables are jointly different from zero, that is, we can exam whether the individual markets have predictive power by applying the Wald test to the sets of parameter restrictions, for example, $c_{11} = c_{12} = c_{13} = 0$ and/or $b_{21} = b_{22} = b_{23} = 0$ if $k = 3$. With one lag, both hypotheses that the stock market does not Granger-cause the CDX market and vice versa cannot be rejected. However, when employing 3 lags, the hypothesis that the CDX market does not Granger-cause the stock market is strongly rejected at the 5% significance level with p-value of 0.0084. The reverse case that the stock market does not Granger-cause the CDX market cannot be rejected with p-value of 0.84. In sum, the results from the VAR model indicate that the CDX market leads the stock market, which is same as the results from the modified CKLS model. Hence, I can conclude that the model specification between the VAR and modified CKLS models cannot be the reason why our study and the studies of Norden and Weber, and Forte and Fena have the opposite results.

Last, but not least, we need to observe signs on the coefficients of b_{c2} and b_{c3}, and b_s in (4.13). In the stock market, the coefficient on the lagged stock index itself displays a negative value. However, both of the coefficients on the lagged CDX index display positive values. During the credit crunch, the lagged stock index negatively correlated with the current stock index, but the lagged CDX indices are positively correlated with the current stock index. We can observe that the lagged stock and CDX indices have opposite impacts on the stock market.

4.4 Conclusions

This study investigated (1) the time-series characteristics of the CDS, stock, and federal funds markets and (2) the market-wide relations between the CDS and stock markets, using daily index data from January 2007 to September 2009.

First, this study examined the distinct dynamic movements of each time series through estimating the CKLS model by employing the MCMC method. Since the original CKLS model is considered to be restricted in terms of the constant conditional variance, we generalize the model by allowing the error term to have level and time-varying conditional heteroscedasticity effects by employing ARMA-GARCH processes. Especially by employing GARCH process, my modified CKLS model allows volatility to be more general and flexible rather than fixed as a constant. In addition, by allowing the error term to be followed by the exponential power distribution (EPD), I specify the excess kurtosis of the data instead of fixing with mesokurtic from a normal distribution. Importantly, the modified CKLS model is estimated by MCMC algorithms based on the Bayesian method. Employing MCMC methods might provide more significant estimators than GMM, originally used in Chan et al. (1992), since Qian and Tsurumi (2005) prove that the Bayesian and MLE estimators dominate GMM estimator with respect to the mean absolute deviation (MAD) and the sums of relative mean absolute deviation (SMAD).

The time-series characteristics of the three distinct markets show the similar movements during the sample period except volatility according to a constant parameter α_0 in the GARCH process 4.4. Given the three time-series data, the stock market is most volatile and the CDS market is followed. However, all three time series show a similar pattern on the persistence of volatility: according to equation 4.10, there are high probabilities that all three series have I-GARCH processes. In terms of c in equation

4.1, both stock and federal funds markets share the same process, Cox-Ingersoll-Ross square root process nested in the CKLS model. Finally, the MCMC results show that all of the three market show the strong leptokurtic behavior. Among them, the stock market shows the strongest leptokurtic behavior, them the federal funds market is next, and the CDS market is the least.

Second, this study analyzed the intertemporal relationship between the CDS and stock markets in the U.S. through estimating the CKLS-ARMA-GARCH-EPD model employing multivariables as a vector autoregression model (VAR). There have been several studies to examine the lead-lag relations between the markets employing the VAR framework. However, I examine the market-wide relations by employing more sophisticated model specification of the modified CKLS model instead of the VAR framework.

Results from the CKLS model show that the CDS market appears to lead the investment grade CDX market. Focusing on the sample period including the credit crunch of 2008, the results imply that when the market is highly volatile, informed traders have affected the CDS market first, and then information on credit risk arrives at the stock market later. The significant interaction between investment grade CDS index and stock index suggests that investors should examine more carefully the dynamic information between the CDS and stock markets when the credit market is in a credit and liquidity crunch as 2008.

Usually, economic time series exhibit dramatic breaks in the behavior when it is associated with events such as financial crisis or abrupt changes in government policy. The CDX index, stock index, and federal funds rate also show breaks caused by the credit crunch of 2008. The time series have become volatile in an oscillatory manner and there may be some regime changes. While the modified CKLS model captures the behaviors of the time series reasonably well, we need to compare the model with models incorporating regime changes especially since the sample periods contain the highly volatile years of 2007-08.

Alexander and Kaeck (2008) indicate that CDS spreads display a pronounced regime-specific behavior. Based on this observation, they employ a Markov switching model using the iTraxx Europe indices in order to examine the determinants of changes of the indices at each regime.

Another way to specify the regime switching is through the threshold ARMA model (TARMA). Using data on realized volatility, Goldman et al. (2009) show TARMA capture the high volatility regime well and that the persistence of volatility is short lived compared to the persistence in the low volatility regime. The difference between the Markov switching model and the TARMA model is that in the Markov switching model the movement of the observation from the previous time to the current time is determined probabilistically, whereas in the TARMA model this movement is determined by the threshold. It will be interesting to compare our model with the Markov switching model and the TARMA model to see which model explains the time-series data well or predicts better.

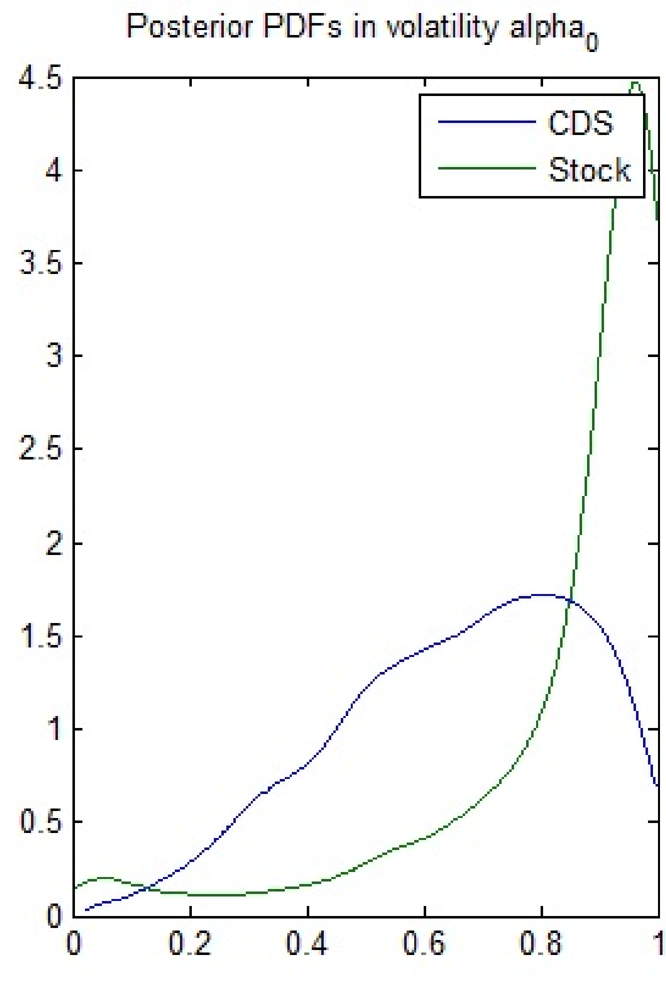

Figure 4.1: Posterior pdf's of α_0 for S&P500 and CDX

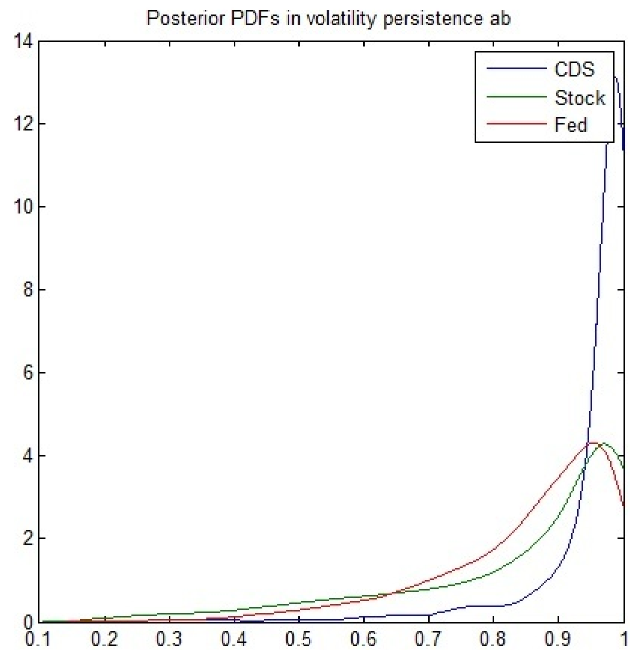

Figure 4.2: Posterior pdf's of $\alpha\beta$

Table 4.1: CKLS-ARMA-GARCH-EPD Models of CDX, SP500, and Federal Funds rates

	CDX	SP500	FFR
a	.2273 (.2071)	1.1805 (7.3083)	.0022 (.0067)
b	.997 (.0026)	-1.3048 (1.6622)	.9984 (.0035)
c	.1633 (.1107)	.4949 (.1703)	.4168 (.1454)
ϕ_1	$-.4085$ (.1146)	.6482 (.6210)	$-.4317$ (.1280)
ϕ_2		.2734 (.5830)	
θ_1	.4951 (.2769)	.5024 (.2757)	.4902 (.2649)
α_0	.6601 (.2130)	.8437 (.2178)	.0036 (.0024)
α_1	.4399 (.1154)	.5795 (.1988)	.6323 (.1696)
β_1	.5122 (.1402)	.2601 (.1544)	.2171 (.1172)
α	.7962 (.0767)	.4141 (.0811)	.5037 (.0815)
$Kurtosis$	8.9549 (1.8398)	91.9734 (204.6683)	29.5120 (15.5826)

Notes
1. Figures without parentheses are ”posterior means”
2. Figures in parentheses are ”posterior standard deviation”
3. FFR=federal funds rate

Table 4.2: Adjusted CKLS model with two variables

	CDX	SP500
a	1.2602 (1.5326)	5.5137 (10.452)
b_{c1}	.9955 (.0039)	.0178 (1.011)
b_{c2}		.4001 (1.3341)
b_{c3}		.3432 (1.5314)
b_s	$-.0006$ (.0009)	-1.0517 (.9467)
c	.1693 (.1348)	.4822 (.1601)
ϕ_1	$-.401$ (.1028)	.6709 (.5901)
ϕ_2		.3208 (.5727)
θ	.4979 (.2758)	.5056 (.2746)
α_0	.5789 (.2683)	.88 (.17)
α_1	.4067 (.1208)	.5999 (.1848)
β_1	.5471 (.1485)	.2524 (.1424)
α	.806 (.0792)	.4207 (.0762)

Note 1. Figures without parentheses are "posterior means"
2. Figures in parentheses are "posterior standard deviation"

Table 4.3: VAR Models with one lag

	CDX_t	$SP500_t$
c	$-.147150$ (2.74299) [$-$.05365]	12.53797 (9.42582) [1.33017]
CDX_{t-1}	.996249 (.00678) [146.877]	$-.024833$ (.02331) [$-$1.06544]
$SP500_{t-1}$	.000529 (.00159) [.33352]	.992063 (.00545) [182.136]

Note 1. Figures in parentheses ( ) are "standard errors"
2. Figures in brackets [ ] are "t-statistics"

Table 4.4: VAR Model with 3 lags

	CDX_t	$SP500_t$
c	.058891 (2.76646) [.02129]	16.50390 (9.32287) [1.77026]
CDX_{t-1}	1.060621 (.05170) [20.5162]	−.181197 (.17422) [−1.04007]
CDX_{t-2}	−.096338 (.07377) [−1.30601]	.658071 (.24859) [2.64725]
CDX_{t-3}	.031218 (.04996) [.62481]	−.514339 (.16838) [−3.05469]
$SP500_{t-1}$	−.013144 (.01515) [−.86741]	.789364 (.05107) [15.4579]
$SP500_{t-2}$	.011101 (.01897) [.58529]	.181122 (.06391) [2.83382]
$SP500_{t-3}$	.002471 (.01478) [.16725]	.019506 (.04979) [.39173]

Note 1. Figures in parentheses ( ) are ”standard errors”
2. Figures in brackets [ ] are ”t-statistics”

Appendix

In this appendix, MCMC algorithms employed by this study are explained first. Also the basic concepts of credit default swap and credit default swap index are explained, and history of regulation and deregulation of the U.S. financial markets are presented.

A.1 MCMC Algorithms

The original CKLS (1992) model for the short-term interest rate is estimated by the generalized methods of moments (GMM). In the Qian and Tsurumi (2005), the Bayesian estimation methods for the CKLS model is presented, and the results from the Bayesian method are compared with other estimation methods as Maximum Likelihood Estimation (MLE) and GMM, which are widely used. Qian and Tsurumi (2005) obtain the Bayesian inference of the parameters and conduct Monte Carlo experiments to compare the proposed Bayesian estimator with MLE and GMM estimators. The simulation results show that Bayesian and MLE estimators dominate GMM estimator in terms of the mean absolute deviation (MAD) and the sums of relative mean absolute deviation (SMAD). Based on the study of Qian et al., Li (2009) developed the estimation method for the modified CKLS model employed in my study. Since the original CKLS model is approximated by a discrete-time process following a series of standard Gaussian variables, there is a limitation to capture the time-varying volatility changes. In order to overcome the limitation, the modified CKLS model allows the error term to follow ARMA-GARCH processes and the Bayesian estimation method is modified accordingly.

In Bayesian inference, we need to derive the marginal pdf of a parameter. However, it is difficult to evaluate the multiple integral analytically in our model. We employ a numerical integration method since there is no closed form solution to the multiple integration. There are various ways to carry out numerical integration, for example, a quadrature formula, importance sampling, and Markov chain Monte Carlo (MCMC) algorithm. Among them, we apply MCMC algorithms. MCMC algorithm is a stochastic numerical integration method. It attempts to simulate direct draws from some complex distributions of interest. MCMC approaches are named because it is based on the property of Markov chains, that is, one uses the previous sample values to randomly generate the next sample value, generating a Markov chain as the transition probabilities between sample values are only a function of the most recent sample value. Then, Markov chains converge to steady state probabilities under certain conditions; the steady state probabilities are the marginal distributions of our interest.

In order to achieve the proposal density for parameters (α_i, β_i), we follow the approximation of Nakatsuma (1998). They develop a new MCMC method for Bayesian estimation and inference of the $ARCH(p,q) - GARCH(r,s)$ model. To generate a Monte Carlo sample from the joint posterior distribution, a Markov chain sampling with the Metropolis-Hasting algorithm is employed. The proposal distributions for the parameters are based on an approximated GARCH model:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{l} (\alpha_i + \beta_i)\, \varepsilon_{t-i}^2 - \sum_{i=1}^{s} \beta_i w_{t-i} + w_t, \quad w_t\tilde{\ }N\left(0, 2\sigma_t^2\right) \tag{4.16}$$

where $w_t = \varepsilon_t^2 - \sigma_t^2,\ l = \max(r,s),\ \alpha_i = 0\ for\ i > r,\ \beta_i = 0\ for\ i > s,\ \varepsilon_t^2 = 0$ and $w_t = 0\ for\ t \leq 0$. The proposal density (4.16) is derived by that GARCH (r,s) model (4.4) is expressed as an ARMA (l,s) process of $\left\{\varepsilon_t^2\right\}_{t=1}^n$:

$$\begin{aligned} e_t^2 &= \alpha_0 + \sum_{j=1}^{l} (\alpha_j + \beta_j)\, e_{t-j}^2 + \tilde{w}_t - \sum_{j=1}^{s} \beta_j \tilde{w}_{t-j} \\ \tilde{w} &= e_t^2 - \sigma_t^2 \end{aligned} \tag{4.17}$$

Since $\tilde{w}_t = \left(\frac{e_t^2}{\sigma_t^2} - 1\right)\sigma_t^2 = \left(\chi^2(1) - 1\right)\sigma_t^2$, the conditional mean of $\tilde{w}_t$ is $E\left[\tilde{w}|F_{t-1}\right] = 0$, and the conditional variance is $var\left[\tilde{w}_t|F_{t-1}\right] = 2\sigma_t^4$. By replacing $\tilde{w}_t$ with $w_t\tilde{\ }N\left(0, 2\sigma_t^4\right)$, (4.16) is derived.

Nakatsuma (1998) uses non-linear least square estimation to draw parameters in MA and GARCH processes. Qian and Tsurumi (2005) use random walk draws for all parameters. Li (2009) generally follows Qian and Tsurumi (2005)'s modification. Different from Nakatsuma (1998), who uses the linear regression model, we need to draw both non-linear parameters c in the CKLS model and the parameter α in EPD. Different from Qian and Tsurumi (2005) whose model is the CKLS with normally distributed error terms, we use efficient jump algorithms to draw $\{\theta_i\}_{i=1}^{q}$, parameters of MA process, $\{\beta\}_{i=1}^{s}$, parameters for GARCH error term, and also α in EPD.

I follow the methods of Li (2009):

1. To draw c in the CKLS model, the modified efficient jump method proposed in Tsurumi and Shimizu (2008) is employed.

2. While Qian and Tsurumi (2005) use random walk draws for all parameters in the CKLS model with errors following a normal distribution, this study employs efficient jump algorithms for $\{\theta_i\}_{i=1}^{q}$, parameters of MA process, $\{\beta\}_{i=1}^{s}$, parameters for GARCH error term, and also α in EPD.

3. All parameters, except c, $\{\theta_i\}_{i=1}^{q}$, $\{\beta\}_{i=1}^{s}$, and α, are drawn by random walk draws followed by Qian and Tsurumi (2005).

A.2 Credit Default Swap

Experiences of credit events such as the 1997 Asian Financial Crisis, the Russian bond default, the collapse of Long-Term Capital Management (LTCM), and the Enron and WorldCom defaults highlighted importance of assessing and managing credit risk. And moreover, the credit risk has been paid attention significantly through the recent large scale incidents occurred in 2008 including the collapse of Bear Sterns, the bankruptcy of Lehman Brothers, and a federal bailout plan for American International Group (AIG). Although it is hard to avoid the blame that credit default swaps (CDSs) exacerbated the 2008 credit crunch by hastening the demise of the financial companies, benefits from credit derivatives also cannot be overlooked. With the benefits from credit derivatives we can purchase, sell or restructure credit risk or credit default risk. And importantly Credit Default Swap (CDS) is one of the tools to manage credit risk.

An original form of CDS had been existed in the past. However, a beginning of the modern form of CDS was invented in 1997 by a team working for J.P.Morgan Chase. Its invention was revolutionary in terms of a purpose of shifting risk out of a company's balance sheet by separating the default risk from the loans themselves. The CDS market has grown dramatically over a short period of time. As shown in Figure 4.3, the International Swap and Derivatives Association (ISDA) indicated that the CDS notional outstanding was approximately \$632 billion in 2001 and expanded to \$63 trillion in the second half of 2007. However, it is reduced to about \$31 trillion in mid of 2009 since the credit crunch of 2008.

As dealers serve the role of liquidity providers, the CDS market has been dominated by the dealer community since an inception. Goldman Sacks, Deutsche Bank, J.P.Morgan and Morgan Stanley are regarded as the dominant dealer participants in the CDS market. As the CDS market evolved, insurance companies became the second most active group of participants. During the past several years, the hedge fund community has begun to play more significant role in the CDS market, supplanting the role of the insurers.

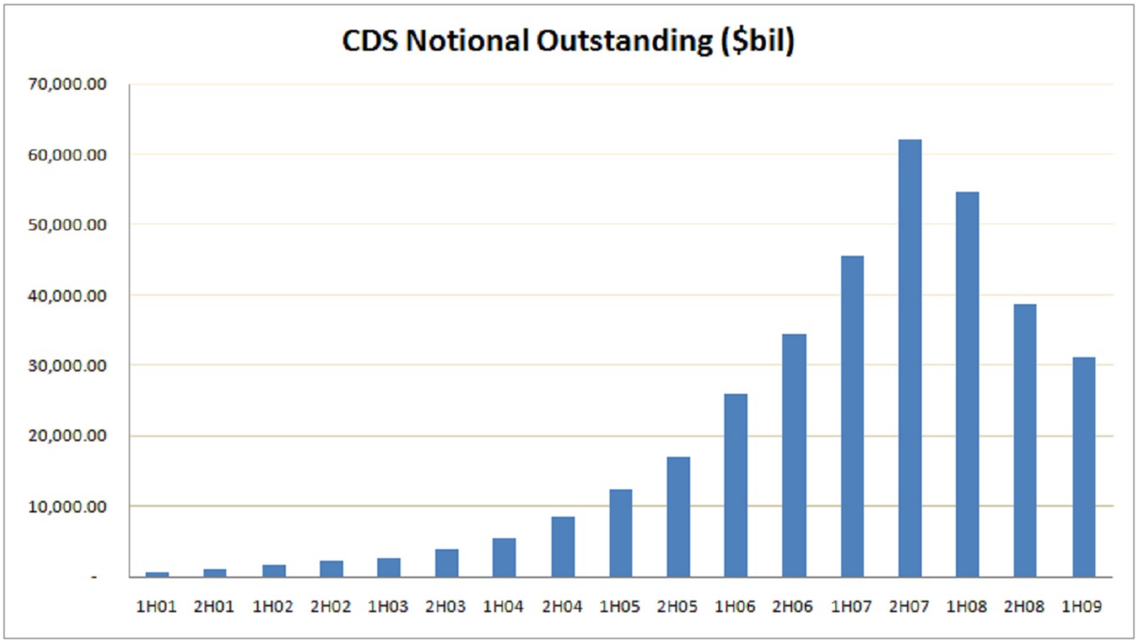

Figure 4.3: Semi-annual data for CDS notional amounts outstanding (Source: ISDA 2009)

CDS is basically an insurance contract. With the CDS contract, the protection buyer pays a fee, the swap premium, to the protection seller in exchange for the right to receive a payment conditional upon the occurrence of a credit event with respect to a reference entity for which credit protection is being sold. The swap premium is quoted in basis points per annum of the contract's notional value and is usually paid quarterly.

The definition of a credit event, the relevant obligations, and the settlement mechanism used to determine the contingent payment are flexible and determined by negotiation between counterparties at an inception of transactions as a characteristic of the OTC market products. A credit event is most commonly defined as 1) bankruptcy or insolvency of the reference entity, 2) failure to pay an amount above a specified threshold over a specified period, and 3) financial or debt restructuring. In addition, the CDSs are typically 5 year contracts although the range of the CDS maturities may extend from six months to 3, 7, and 10 years.

A single-name CDS has a single reference entity. If no credit event occurs over the life of the swap, the protection buyer will make a swap premium payment until maturity. If a certain pre-specified credit event occurs, first, the protection buyer pays out the accrued premium from the last payment date to the time of the credit event, on a days fraction basis. After that payment, there are no further payments of the swap premium by the protection buyer to the protection seller. Second, a termination value is determined for the swap, the procedure depending on the settlement terms specified in the trade's documentation. The contingent payments are either physical settlement or cash settlement. With physical settlement the protection buyer delivers a specified amount of the face value of bonds for the reference entity to the protection seller. Then, the protection seller pays the protection buyer the face value of the bonds. With cash settlement, the protection seller pays the protection buyer an amount equal to the difference between the face value of bonds and their market value after the default.

A basket default swap has more than one reference entity. Based on when the protection seller is obligated to make a payment to the protection buyer, different types of basket default swap exist. These are classified as Nth-to-default swaps, subordinate basket default swaps, and senior basket default swaps. For example, in a first-to-default swap, the protection buyer is compensated if one asset in the basket default but receives no compensation for any subsequent defaults. In addition, in an Nth-to-default swap, the protection seller makes a payment to the protection buyer only after there has been a default for the Nth reference entity and no payment for the defaults of the first (N-1) reference entities. Once there is a payout for the Nth reference entity, the swap terminates. That is, the protection buyer absorbs losses resulting from the first (N-1) defaults and receives compensation only upon the occurrence of the Nth default.

A.3 Credit Default Swap Index (CDX)

As the CDS market increased in importance and enlarged its portion in the derivatives market, it became inevitable to provide the timely information for pricing and measuring credit risk to investors. It became possible by the Credit Default Swap Indices which measure the average CDS spread of all the index dealers. There are two primary tradable index families: CDX index and iTraxx index. Since the CDS indices are tradable now, they allow players to trade a broader range of credits at a lower cost. Both of the CDS indices are published by Markit. The CDX indices are the actively traded indices based upon North American reference entities, and the iTraxx indices are based upon European and Asian reference entities. For a transaction of the CDX or iTraxx indices, one of the counterparties must hold a Markit index license. The CDX indices have subindices as following: CDX North American Investment Grade (CDX.NA.IG), CDX North American High Volatility, CDX North American High Yield (CDX.NA.HY), CDX Crossover, CDX Emerging Market, etc.

As a synthetic collateralized debt obligation (CDO), a CDS index is also sliced into standardized synthetic tranches. The reason for slicing risk is to provide institutional investors with alternative vehicles for obtaining exposure to risk that are more acceptable to them, given their investment objectives and contraints. For CDX North American Investment Grade (CDX.NA.IG), there are five tranches: an equity tranche $0-3\%$, a junior mezzanine $3-7\%$, a senior mezzanine $7-10\%$, a senior $10-15\%$, and a super senior tranche $15-30\%$. For example, suppose an investor holds a $7-10\%$ senior mezzanine tranche of CDX.NA.IG. If there are a sufficient number of defaults for the losses to exceed the subordination of 7% over the life of the tranche, the investor will only realize a principal loss and will lose all the principal when the losses reach the upper limit of the tranche of 10%[2].

[2]The indices trade at a fixed coupon, which is paid quarterly by the buyer of protection on the index, i.e. a short index position, and upfront payment are made at initiaion and close of the trade to reflect the change in price. Correspondingly, the protection seller, or buyer of the index, receives the coupon. (Markit Credit Indices: A Primer)

The composition of each index is reconstituted and determined by member banks on March 20th and September 20th of each year semiannually. March 20th and September 20th are referred to as the index roll dates. The member banks which help compose and price the index include thirteen major international banks: Bank of America, BNP Paribas, Barclays Capital, Citibank, Credit Suisse, Duetsche Bank, Goldman Sacks, HSBC, JPMorgan, Morgan Stanley, RBS, USB, and Well Fargo. The new rollout of the CDX indices during September 2009 is designated as CDX Series 13. Although the previous series continues to trade, most of the market liquidity tends to be concentrated on the most recent series which is referred to as on-the-run. In addition, if a credit is initially included in a series, the name is not removed from a given series as long as the CDS protection credit event has not been triggered.

The most actively traded index among the CDX indices is the CDX North American Investment Grade (CDX.NA.IG). The Investment Grade means that the companies included in the index are highly expected to meet the payment obligations on their outstanding debt. Basically, the index measures the average CDS spreads of all the index dealers. CDX.NA.IG index consists of 125 CDSs of North American companies with an investment-grade rating. The 125 corporate names in the index are equally weighted within the index (i.e. 0.8%). If a name is excluded from a given series due to a credit event, the weight may change.

The mechanism of the CDX indices is slightly different from that of a single name CDS. In a single name CDS, a protection buyer pays a swap premium. If a credit event occurs, the swap premium payment ceases in the case of a single name CDS. However, in the CDX indices, the protection buyer also pays the protection seller the initial price of the index on a given notional amount of the index. If the index value changes over the next 90 days (a quarter), the protection buyer will make a payment to the protection seller equal amount of the present value of change in the value of the index over the remaining life of the contract. When a credit event happens, the protection buyer continues to pay the swap premium but it is based on a reduced notional amount since less reference entities are being protected, and there is typically physical settlement.

The protection buyer gives the protection seller a face amount of the defaulted debt which is equal to 0.8% of the original notional value of the index purchased, and then the protection seller will deliver an equal amount of cash.

If a new protection buyer enters into the existed CDX indices after an inception of a new series of CDX, it requires the exchange of an up-front payment representing the probability weighted present value difference between the current market value of the CDX and the initial deal value of the CDX. In addition, upon entering the CDX the protection seller pays the accrued premium from the last payment due to the settlement date in order to receive a full 90 days of premium on the next payment date.

A.4 History of Regulation and Deregulation of the U.S. Financial Markets

In 1863, the Federal Government which was short of cash as a result of Civil War formed the National Bank Act, which established a system of national charters for banks that would have authority to issue their own currency so long as it was backed by holdings in U.S. Treasury bonds. The law was completely rewritten as the National Banking Act of 1864 and the Act formed the Office of the Comptroller of the Currency (OCC) with authority to charter and examine national banks.

In 1907, a failure of the scheme to corner United Copper Company via short-selling led bank-run in the New York City banks which lent money for the scheme, and the bank-run extended across the nation. As a response to the panic, the 1913 Federal Reserve Act set the Federal Reserve System (Fed) as a central bank and lender of last resort.

During the Great Depression, beginning in 1929 and bottoming in 1933, about 5000 banks failed and the U.S. stock market crashed. As a response to the failure of the financial market, the regulatory structures were erected by President Roosevelt–including the creation of the Securities and Exchange Commission (SEC) as the primary regulator of the U.S. security market, the establishment of banking oversight, the guaranteeing of bank deposits by creating the Federal Deposit Insurance Corporation (FDIC) and the passage of the Glass-Steagall Act.

The Glass-Steagall Act of 1933 was denoted as a comprehensive piece of regulation reform which separated commercial banking from investment banking since the bankers invested vigorously in the market, loaned money to speculators to invest, and even loaned their depositors' money to companies in which they were invested (Mullin). Through several legislations mainly based on these regulatory structures, commercial banks, investment banks, and insurance companies had been separate, and they had oversight from separate regulators – the Fed and OCC for commercial banks, the SEC for investment banks, and state regulators for insurance companies. The financial regulations did not clarify the definition of derivatives as securities 1) because during the 1930's, futures markets were not significant, and existed primarily for transacting agricultural products, and 2) because the emergence of the active over-the-counter (OTC) derivatives markets were not anticipated in that era.

As time went on, dramatic growth in the over-the-counter (OTC) derivatives market induced a need for clarifying a regulatory authority to oversee the undefined OTC derivatives market under the Glass-Steagall Act. A 1982 amendment to the Securities Exchange Act of 1934, known as the Shad-Johnson Accord, specified that options on securities or baskets of securities were to be regulated by the SEC; however, this left the parts not regulated by the SEC such as forwards, swaps, derivatives on interest rates or foreign exchange. The Commodity Futures Trading Commission (CFTC), which was established under the 1974 Commodity Exchange Act (CEA), took a role to oversee the left derivatives; forwards, swaps, derivatives on interest rates or foreign exchange. The CFTC assumed exclusive jurisdiction to regulate commodity futures and options. Hence, according to 1982 amendment, Congress clarified the jurisdictions of the SEC and the CFTC over security-based options and futures. The accord granted to the SEC sole authority to regulate options on securities, certificates of deposit and stock groups, and to the CFTC an authority to regulate futures and options on futures on exempted securities and broad-based indices. In spite of the expanded definition of a commodity through the Accord, the SEC and the CFTC have still battled over holding jurisdictional reins. Futures on individual stocks is one of examples which created a conflict for jurisdictions for the SEC and the CFTC since futures on individual stocks were not clearly defined in which institutions should oversee. In order to allow the SEC and the CFTC the time to resolve regulatory and philosophical differences, the Accord prohibited the sale of futures on single stocks and on narrow-based indices.

Since it has not been clear whether OTC financial derivatives fall under the jurisdiction of the CFTC, the CFTC hesitated to regulate the OTC derivatives market, and the CFTC's intervening to the market was not welcomed by market participants. In the early 1990s, the CFTC exempted swaps from regulation; however, the questions from unclarified terms of the OTC derivatives was still not resolved.

As the financial market developed, not only did the OTC derivatives market explode, but the Glass-Steagall separation of investment and commercial banking was also gradually eroded. By the Financial Services Modernization Act of 1999, known as the Gramm-Leach-Bliley Act, the Act created Financial Holding Companies (FHCs) which may hold commercial banks, investment banks and insurance companies. Removing the barrier between investment and commercial banking, maintained since the Glass-Steagall Act of 1933, was dramatic transformation; however, the regulatory transformation to oversee the new structure was not created. With the previous acts, each industry was under the regulation of separate regulators; however, under the act, instead of a overall regulator which oversee the new FHCs that combined all three industries, a functional approach was adopted. The Fed and OCC oversee the commercial banking functions of FHCs. The SEC oversees the investment banking function, and state insurance regulators oversee the insurance functions.

Banning on the sale of single-stock futures and futures on narrow-based stock indices since the Shad-Johnson Accord of 1982 led concerns that it would drive the market overseas such as Sydney, Hong Kong, OM Stockholm and Montreal whose exchange offered single stock futures products. And it needed to resolve the legal ambiguity of jurisdictions between the SEC and the CFTC. In 2000, Congress passed the Commodity Futures Modernization Act (CFMA) which provided resolutions for concerns about the competitiveness of the U.S. financial market and the ambiguous jurisdictions. Most importantly, the CFMA was from a call for setting out the liberated market conditions under which derivative financial products could be legally traded in the OTC market with lax regulations. The legislations have evolved through transforming markets into regulated markets or free markets with reflecting different ideas between paternalism and libertarianism repeatedly. When the Securities Act of 1933, the Securities Exchange Act of 1934, and the Glass-Steagall Act of 1933 were enacted during the Great Depression, the market was more under the government's protection; when the Gramm-Leach-Bliley Act, which repealed the separation between commercial and investment banking since the Glass-Steagall Act, was signed into law during the boom of 1999, the market was liberated from regulations. The Gramm-Leach-Bliley Act and the CFMA were a result of reflecting the idea intrinsically that we can believe the market adjusts itself.

The idea that the market can regulate itself goes back to the Reagan administration which had a strong belief about the free market, deregulation, and balanced budgets. The two representative acts of the deregulation in the financial market, the Gramm-Leach-Billey Act of 1999 and the Commodity Futures Modernization Act (CFMA) of 2000, are the products reflecting the trend in economics and business philosophy that the free market will bring efficiency and self-regulation through competition, and that the regulation is an impediment to the working of the free market. Although the Garn-St. Germain Depository Institutions Act, signed by Reagan, deregulating savings and loan institutions and providing the institutions more flexibility in operations in order to vitalize the depressed housing industry, turned out to be one of main factors that led to the savings and loans crisis of the 1980s, the trend of the free market and deregulation gained a great weight backed by many powerful people in the financial industry e.g. Alan Greenspan, a former Chairman of the Federal Reserve, Arthur Levitt Jr., a former Chairman of the SEC, Phil Gramm, a former U.S. senator, and so on.

The CFMA clarified the regulatory and supervisory roles of the SEC and the CFTC by excluding all OTC derivatives from the CFTC's jurisdiction. Hence, the OTC derivatives market was to remain largely unregulated. The sale of single stock futures and futures on narrow-based indices, prohibited by the Shad-Johnson Accord of 1982, was allowed. In order to clarify the jurisdiction about security futures products, the CFMA granted the SEC and the CFTC the joint rules over futures on single stocks and narrow-based stock indices: futures contracts on broad-based indices remained under the exclusive jurisdiction of the CFTC. It is obvious that the CFMA gave the derivatives market flexibility that required to foster innovation: the total OTC derivatives market had exploded to \$ 600 trillion, increasing 826% in 10 years according to BIS statistics.

The CFMA created a shadow banking system, which consists of non-banking financial institutions outside the government oversight that play an increasingly critical role in lending businesses the money necessary to operate. Shadow banking institutions typically act as intermediaries between investors and borrowers. For example, an institutional investor like a pension fund may be willing to lend money, while a corporation may be searching for funds to borrow. The shadow banking institution will channel funds from the investors to the corporation, profiting either from fees or from the difference in interest rate. Since shadow institutions do not accept deposits like a depository bank and therefore are not subject to the same regulation. Familiar examples of shadow institutions are Bear Stearns and Lehman Brothers. Other complex legal entities comprising the system include hedge funds, Securitized Investment Vehicles, conduits, and investment banks. According to Mr. Geithner, a secretary of the U.S. Treasury, by 2007 more than half of America's banking was being handled by a "shadow banking" of largely unregulated institutions. And the unregulated shadow banking system, together with the CDSs which are the primary financial instruments behind the near collapse of AIG and Bear Stearns, has been assumed to contribute to the credit crunch.

Bibliography

Acharya, V. V. and T. C. Johnson (2007): "Insider trading in credit derivatives," *Journal of Financial Economics*, 84, 110–141.

Aitchison, J. and S. D. Silvey (1957): "The generalization of probit analysis to the case of multiple responses," *Biometrika*, 44, 131–140.

Albert, J. H. and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

——— (2001): "Sequential ordinal modeling with applications to survival data," *Biometrics*, 57, 829–836.

Alexander, C. and A. Kaeck (2008): "Regimes dependent determinants of credit default swap spreads," *Journal of Banking and Finance*, 32, 1008–1021.

Arrieta, A. and L. B. Russell (2008): "Effects of leisure and non-leisure physical activity on mortality in U.S. adults over two decades," *Annals of Epidemiology*, 18(12), 889–895.

Banks, J., A. Muriel, and J. P. Smith (2010): "Attrition and health in ageing studies: Evidence from ELSA and HRS," *IZA Discussion paper*, No.5161.

Berry, B. (2007): "Does money buy better health? Unpacking the income to health association after midlife," *Health: An interdisplinary journal for the social study of health, illness and medicine*, 11(2), 199–226.

Black, F. and M. Scholes (1974): "The effects of dividend yield and dividend policy on common stock prices and returns," *Journal of Financial Economics*, 1 (1), 1–22.

BOSMA, M., M. MARMOT, H. HEMINGWAY, A. C. NICHOLSON, E. BRUNNER, AND S. A. STANSFELD (1997): "Low job control and risk of coronary heart disease in Whitehall II (prospective cohort) study," *British Medical Journal*, 314, 558–565.

CHAN, K., G. KAROLYI, F. LONGSTAFF, AND A. SANDERS (1992): "An empirical comparision of alternative models of the short-term interest rate," *Journal of Finance*, 47, 1209–1227.

CHANG-QUAN, H., Z. XUE-MEI, D. BI-RONG, L. ZHEN-CHAN, Y. JI-RONG, AND L. QING-XIU (2010): "Health status and risk for depression among the elderly: a meta-analysis of published literature," *Age and Ageing*, 39, 23–30.

CHEN, M.-H. AND D. K. DEY (2000): *Bayesian analysis for correlated ordinal data models, In Generalizede Linear models: A Bayesian Perspective*, New York: Mrcel Dekker.

COTT, C., M. GIGNAC, AND E. BADLEY (1999): "Determinants of self-rated health for Canadians with chronic disease and disability," *Journal of Epidemiology and Community Health*, 53, 731–736.

COWLES, M. K. (1996): "Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear model," *Statistics and Computing*, 6, 101–111.

DANIELS, M. J. AND J. W. HOGAN (2008): *Missing data in longitudinal studies*, Chapman & Hall/CRC.

DIGGLE, P. J. AND M. G. KENWARD (1994): "Informative dropout in longitudinal data analysis (with discussion)," *Journal of Royal Statistical Society B*, 43, 49–93.

DUFFIE, D. AND K. J. DUFFIE (1999): "Modeling the term structure of defaultable bonds," *Review of Financial Studies*, 12, 687–720.

FERRIE, J. E., P. MARTIKAINEN, M. J. SHIPLEY, AND M. G. MARMOT (2002): "Self-reported economic difficulties and coronary events in men: Evidence from the Whitehall II study," *International Journal of Epidemiology*, 34, 640–648.

Forte, S. and J. Pena (2006): "Credit spreads: theory and evidence about the information content of stocks, bonds and CDSs," *Working paper*.

Frijters, P. and A. Ulker (2008): "Robustness in health research: do differences in health measures, techniques, and time frame matter?" *Journal of health economics*, 27, 1626–1644.

Fung, H. G., G. E. Sierra, J. Yau, and G. Zhang (2008): "Are the U.S. stock market and credit default swap market related? : Evidence from the CDX indices," *Working paper*.

Gad, A. M. (2011): "A selection model for longitudinal data with non-ignorable non-monotone missing values," *Journal of Data Science*, 9, 171–180.

Goldman, E., Nam, H. Tsurumi, and Wang (2009): "Regimes and long memory in realized volatility," *working paper*, ., .

Goldman, N., C. M. Turra, L. Rosero-Bixby, D. Weir, and E. Crimmins (2011): "Do biological measures mediate the relationship between education and health: a comparative study," *Social Science & Medicine*, 72, 307–315.

Green, W. H. and D. A. Hensher (2010): *Modeling ordered choices: a primer*, Cambridge University Press.

Greene, W. H. (2002): *Econometric analysis (5th edition)*, Prentice Hall.

Gueorguieva, R., J. L. Sindelar, T. A. Falba, J. M. Fletcher, P. Keenan, R. Wu, and G. W. T. (2009): "The impact of occupation on self-rated health: cross-sectional and longitudinal evidence from the Health and Retirement Survey," *Journal of Gerontology: Social Science*, 64B(1), 118–124.

Hasegawa, H. (2009): "Bayesian dynamic panle-ordered probit model and its application to subjective well-being," *Communications in statistics-simulation and computation*, 38, 1321–1347.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47, 153–162.

IDLER, E., L. B. RUSSELL, AND D. DAVIS (2000): "Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study," *American Journal of Epidemiology*, 152, 874–883.

IDLER, E. L. AND Y. BENYAMINI (1997): "Self-rated health and mortality: a review of twenty-seven community studies," *Journal of Health and Social Behavior*, 38, 21–37.

JONES, A. M. (2000): *Health econometrics*, Elsevier.

KAPTEYN, A., P.-C. MICHAUD, J. SMITH, AND A. VAN SOEST (2006): "Effects of attrition and non-response in the Health and Retirement Study," *IZA Discussion paper*, No.2246.

KARP, F., ed. (2007): *Growing older in America: The Health and Retirement Study*, U.S. Department of health and human services.

KENWARD, M. G. (1998): "Selection models for repeated measurements with non-random dropout: an illustration of sensitivity," *Statistical Methods*, 17, 2723–2732.

KIM, J. (2011): "Socioeconomic inequalities in self-rated health among middle-aged and older adults," *Social work in health care*, 50, 124–142.

LEON, F. G., A. K. ASHTON, D. A. DMELLO, AND B. DANTZ (2003): "Depression and comorbid medical illness: therapeutic and diagnostic chellenges," *The journal of family practice*, Suppl., 19–23.

LI, L. (2009): "CKLS models for U.S. short-term interest rate," *Ph.D Dissertation, Rutgers*, 5–39.

LI, M. AND J. L. TOBIAS (2006): "Bayesian analysis of structural effects in an ordered equation system," *Studies in nonlinear dynamics and econometrics*, 10(4), Article 7.

——— (2008): "Bayesian analysis of treatment effects in an ordered potential outcomes model," *Advances in Econometrics*, 21, 57–91.

LITTLE, R. J. (1995): "Modeling the dropout mechanism in repeated-measures studies," *Journal of American Statistical Association*, 90, 1113–1121.

Little, R. J. and D. B. Rubin (2002): *Statistical analysis with missing data (2nd ed.)*, Wiley.

Liu, J. S. and C. Sabatti (2000): "Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation," *Biometrika*, 87, 353–369.

Longstaff, F., S. Mithal, and E. Neis (2003): "The credit default swap market: Is credit protection priced correctly?" *Working paper: Anderson School at UCLS.*

Marmot, M., H. Bosma, H. Hemingway, E. Brunner, and S. Stansfeld (1997a): "Contribution of job control and other risk factors to social variations in coronary heart disease incidence," *Lancet*, 350, 235–239.

Marmot, M. G., S. G. Davey, S. A. Stansfeld, C. Patel, R. North, and J. Head (1991): "Health inequalities among British civil servants: The Whitehall study," *Lancet*, 337, 1387–1393.

Marmot, M. G., C. D. Ryff, L. L. Bumpass, M. J. Shipley, and N. F. Marks (1997b): "Sociel inequalities in health: Next questions and converging evidence," *Social Science & Medicine*, 44, 901–910.

McCulloch, R. E. and P. E. Rossi (1994): "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, 64(1-2), 207–240.

McKelvey, R. D. and W. Zavoina (1975): "A statistical model for the analysis of ordinal level dependent variables," *Journal of mathematical sociology*, 4, 103–120.

Mossey, J. and E. Shapiro (1982): "Self-rated health: a predictor of mortality among the elderly," *American Journal of Public Health*, 72, 800–808.

Nakatsuma, T. (1998): "A Makov chain sampling algorithms for GARCH models," *Studies in Nonlinear Dynamics and Econometrics*, 3, .

Nandram, B. and M. H. Chen (1996): "Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence," *Journal of Statistical Computation and Simulation*, 54, 129–144.

NEYMAN, J. AND E. L. SCOTT (1948): "Consistent estimation from partially consistent observations," *Econometrica*, 16 (1), 1–32.

NORDEN, L. AND M. WEBER (2004): "The comovement of credit default swap, bond and stock markets: An empirical analysis," *Working paper*.

PINQUART, M. (2001): "Correlates of subjective health in older adults: a meta-analysis," *Psychology and Aging*, 16, 414–426.

QIAN, XUFENGAND ASHIZAWA, R. AND H. TSURUMI (2005): "Bayesian, MLE, and GMM estimation of spot rate model," *Communication in Statistics*, 34(12), 2221–2233.

RUBIN, D. B. (1976): "Inference and missing data," *Biometrika*, 63 (3), 581–592.

SMITH, J. P. AND R. KINGTON (1998): "Race, socioeconomic status, and health in late life," *RAND Corporation*.

SNELL, E. (1964): "A scaling procedure for ordered categorical data," *Biometrics*, 20, 592–607.

SPIEGELHALTER, D. J., K. R. ABRAMS, AND J. P. MYLES (2004): *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, John Wiley & Sons.

TANNER, M. A. AND W. H. WONG (1987): "The calculation of posterior distributions by data augmentation (with discussion)," *Journal of the American Statistical Association*, 82, 528–550.

TSURUMI, H. AND H. SHIMIZU (2008): "Bayesian inference of the regression model with error terms following the exponential power distribution and unbiasedness of the LAD estimator," *Communication in Statistics-Theory and Methods*, 37(2), 221–232.

Vita

Jeehyun Park

Education:

2012	Ph.D, Economics, Rutgers University	*New Jersey, U.S*
2009	M.A, Economics, Rutgers University	*New Jersey, U.S*
2005	B.A, Ewha Womans University	*Seoul, Korea*

Experience:

- Instructor in Introduction to Macroeconomics (summer 2009) and in Intermediate Macroeconomics (summer 2010, fall 2010, summer 2011, fall 2011, spring 2012) : Rutgers University
- Teaching assistant (2008 ~2011) : Rutgers University
- Research Assistant for Professor Louise Russell in Institute for Health, Health Care Policy and Aging Research (2010 ~present)
  1. Bayesian meta-analysis for evidence synthesis (Project: Cost-effectiveness analysis of maternal immunization using a potential group B streptococcal vaccine in South Africa, PIs: Anushua Sinha, MD, and Louise Russell, Ph.D)
  2. Propensity score matching analysis of the effect of physical activity on mortality in NHEFS (National Health and Nutrition Examination Survey Epidemiologic Followup Study)
  3. Analysis of American Time Use Survey data (ATUS), a Bureau of Labor Statistics Survey