

© 2009

Huiying Zhang

ALL RIGHTS RESERVED

**ESSAYS ON SEMIPARAMETRIC COX PROPORTIONAL
HAZARD MODELS**

by

HUIYING ZHANG

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

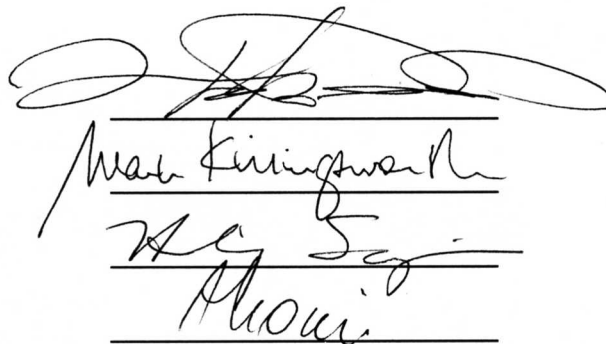
Doctor of Philosophy

Graduate Program in Economics

written under the direction of

Professor Hiroki Tsurumi

and approved by



Three handwritten signatures are displayed on horizontal lines. The top signature is the most prominent and appears to be 'H. Zhang'. The middle signature is 'Mar Kingworth'. The bottom signature is 'H. Tsurumi'.

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

ESSAYS ON SEMIPARAMETRIC COX PROPORTIONAL HAZARD MODELS

by HUIYING ZHANG

Dissertation Director: Hiroki Tsurumi

In this dissertation I study different versions of the semiparametric proportional hazard duration model and their practical applications under both frequentist and Bayesian econometrics frameworks. I use the unemployment spell data set that is created from the Panel Study of Income Dynamics (PSID).

In Chapter 1 I study the effects of unemployment compensation and other important sociodemographic factors on unemployment duration. Whether duration dependence follows a particular function form is also examined. Discrete, semiparametric, proportional hazard models are used and compared among different specifications. I allow for nonparametric estimation of the effect of time on the unemployment exit rate. Because unobserved individual heterogeneity has the potential to bias the estimation results, we also consider gamma heterogeneity as an additional source of error in the hazard model (i.e., the so called mixed proportional hazard model, MPH). I find that the nonparametric baseline hazard estimations capture very well the shape of the empirical duration, which often does not belong to a specific parametric family; and unemployment insurance and

socio-demographic aspects have significant impacts on the unemployment spell.

In the second chapter I test whether different ways to resume work, such as new job and recall, have different duration behaviors. Hence a semiparametric dependent competing risks proportional hazard model is specified. Identifiability of such model is also discussed. By assuming linearity on the baseline hazard at each time interval, I allow for unrestricted correlation between the competing risks. My model guarantees that the unobserved failure occurs later than the observed failure at any possible time point, and censored observations are accommodated explicitly in the model specification. The estimated correlation coefficient suggests that recall duration and new job duration have a positive relationship that may not be negligible. We also find that there is significant difference in the hazard structure of returning to the same employer and a different employer.

Different from the first two chapters, in the third chapter I investigate the ordered probit duration model semiparametrically using the Bayesian Markov Chain Monte Carlo (MCMC) methods. I develop and estimate the model without considering unobserved heterogeneity, and noninformative priors are assumed for both the baseline hazard and regressor parameters. Hybrid Metropolis-Hastings/Gibbs sampler is employed to speed up chain mixture. Convergence of the chains is assessed by the Gelman-Rubin scale reduction factor. Applications on the PSID unemployment duration data demonstrate that the proposed model and estimation method perform well.

Dedication

To my late father,

Zemin Zhang.

The best father I can ask for in the world.

谨以此论文献给我的父亲

Acknowledgements

I want to express my great gratitude toward Dr. Hiroki Tsurumi, my dissertation advisor, for his tremendous support, guidance and patience during the years of my study at Rutgers. I am especially appreciative for the support of the members of my committee, Dr. Mark Killingsworth, Dr. Hilary Sigman and Dr. Alex Voicu.

I would like to thank Ms. Dorothy Rinaldi, whose help made my study at Rutgers a pleasant experience.

Last, but certainly not least, I would like to thank my parents, Meixiang Chen and Zemin Zhang, my husband Hao, and my brother and sister for their love and devotion.

Table of Contents

Abstract	ii
Dedication	iv
Acknowledgements	v
List of Tables	viii
List of Figures	ix
Introduction	1
1. Semiparametric Estimation of Single Risk Proportional Hazard Models with Applications in Unemployment Duration	6
1.1 Introduction	6
1.2 Ordered logit/probit duration models without and with unobserved heterogeneity	10
1.3 Data description	18
1.4 Estimation results	22
1.5 Conclusion remarks and future directions	27
Appendix 1	45
2. Semiparametric Competing Risks Proportional Hazard Model and Unemployment Duration	49
2.1 Introduction	49
2.2 Identification	52
2.3 Competing risks proportional hazard model	55

2.4 Application and empirical results	61
2.5 Conclusion remarks and future directions	67
3. Bayesian Estimation of the Ordered Probit Unemployment Duration Model without Unobserved Heterogeneity	73
3.1 Introduction	73
3.2 Bayesian ordered probit duration model without unobserved heterogeneity	75
3.3 The standard Gibbs sampler and the hybrid Metropolis-Hastings/Gibbs sampler .	80
3.3.1 Standard sampling scheme for the ordered probit model without heterogeneity	80
3.3.2 Hybrid Metropolis-Hastings/Gibbs sampler for the ordered probit model without heterogeneity	83
3.4 Application and results	86
3.5 Conclusions and future research directions	89
References	103
Curriculum Vitae	111

List of Tables

1.1 Variable definitions and mean-standard deviations of the PSID unemployment duration sample	31
1.2 Unemployment duration in weeks and number of observations for different failure types	32
1.2 Unemployment duration in weeks and number of observations for different failure types (cont.)	33
1.3 Failures, censorings, and the Kaplan-Meier empirical hazard	34
1.4 Hazard model estimates	35
1.5 Baseline hazard estimates from model specifications in Table 1.4.	36
1.6 Hazard model estimates	37
1.7 Baseline hazard estimates from model specifications in Table 1.6.	38
2.1 Competing risks hazard model estimates	70
2.2 Competing risks baseline hazard estimates from model specification in Table 2.1	71
3.1 The ordered probit model w/o heterogeneity MLE and posterior estimates	91

List of Figures

1.1	Subsample weekly hazard for resuming work, single risk model	39
1.2	Sample monthly hazard functions for re-employment, single risk model	39
1.3	Sample monthly hazard rates for re-employment, dual-risk model	40
1.4	Sample weekly hazard functions for re-employment, dual-risk model---individuals who receive UI.	40
1.5	Sample monthly hazard functions for re-employment, dual-risk model---individuals who receive UI.	41
1.6	Sample weekly hazard functions for re-employment, dual-risk model---individuals who do not receive UI	41
1.7	Sample monthly hazard functions for re-employment, dual-risk model---individuals who do not receive UI	42
1.8	Estimated monthly cumulative baseline hazard in single risk for various models (based on Table 1.4)	42
1.9	Estimated monthly cumulative baseline hazard in single risk for various models (based on Table 1.6).	43
1.10	Estimated monthly cumulative baseline hazard in single risk for various models (based on Tables 1.4 and 1.6)	43
1.11	Estimated weekly baseline hazard under different model specifications.	44
2.1	Estimated weekly baseline hazard in competing risks model	72
2.2	Estimated monthly cumulative baseline hazard in competing risks model	72
3.1	Iteration trajectories of the covariate parameters in the standard Gibbs	

sampling scheme	92
3.2 Iteration trajectories of three representative cutpoints in the standard Gibbs sampling scheme	93
3.3 Evolution of the Gelman-Rubin factors for the covariate parameters in the standard Gibbs sampling scheme	94
3.4 Evolution of the Gelman-Rubin factors for three representative cutpoints in the standard Gibbs sampling scheme	95
3.5 Iteration trajectories of the covariate parameters in the hybrid Gibbs sampling scheme	96
3.6 Iteration trajectories of three representative cutpoints in the hybrid Gibbs sampling scheme	97
3.7 Evolution of the Gelman-Rubin factors for the covariate parameters in the hybrid Gibbs sampling scheme	98
3.8 Evolution of the Gelman-Rubin factors for three representative cutpoints in the hybrid Gibbs sampling scheme	99
3.9 Estimates of the marginal posterior pdf for the regressor coefficients	100
3.10 Estimates of the marginal posterior pdf for the first, fifth and last cutpoints	101
3.11 Estimated monthly cumulative baseline hazard for the ordered probit model without heterogeneity under the MLE and Bayesian paradigms	102

Introduction

Over the last few decades, duration models have been broadly used to analyze event history data in a wide range of disciplines. Originated from biostatistics and engineering fields, duration analysis has become a main subject of econometrics. There are hundreds of empirical studies applying duration models to examine economics issues. For example, in labor economics, duration models are used to study the unemployment duration and the duration of jobs (e.g., Meyer 1990, Baker and Rea 1998); in financial economics, duration models are used to study credit risks and the time period between stock market share transactions (e.g., Deng, Quigley and Order 2000, Engle and Russell 1998); in macroeconomics, duration models are used to study the duration of business cycles (e.g., Diebold and Rudebusch 1990); in political economics, duration models are used to study adverse selection in insurance markets (e.g., Finkelstein and Poterba 2004); and in health economics, duration models are used to examine the effect of ill health on retirement decisions (e.g., Disney, Emmerson and Wakefield 2006).

Among these applications, a reduced-form duration model, or the so-called Proportional Hazard (PH) model has received extensive attention. Basically, the PH model has two to three multiplicative components: the baseline hazard that determines the shape of duration, the particular function form that highlights the effects of observed explanatory variables on the hazard rate, and the heterogeneity term that represents unobserved

differences across sample units. When heterogeneity is added on, the PH model is often referred as the MPH (Mixed Proportional Hazard) model.

Applications of lifetime distributions to labor economics have become widely since about 1980; early references include Flinn and Heckman (1982), Lancaster (1979), Butler and McDonald (1986), Butler and Worrall (1985), Moffitt (1985). The proportional hazard model is broadly used in the empirical analysis of unemployment duration. Three key issues of these analyses are:

- 1) How unemployment insurance policies (and other factors) affect the duration of unemployment, especially at the time points close to benefits exhaustion?
- 2) When there are two or more ways of exiting unemployment, what is the duration behavior for each type of exit? Are they similar to each other, or quite different?
- 3) Whether it is important to consider unobserved heterogeneity when there is nonparametric baseline specification? Is gamma distribution accurate enough to describe the unobserved heterogeneity? What could we find if both the baseline hazard and the unobserved heterogeneity are estimated nonparametrically?

Han and Hausman (1990) specified and estimated a semiparametric (mixed) proportional hazard model to find answers to the first two questions and part of the third issue. They concluded that the baseline hazard does not follow a monotonic trend, and hazards become higher around unemployment insurance exhaustion points which are usually at 26 and 39 weeks. Besides, the duration behaviors of finding a new job and returning to a former employer are very different. Meyer (1990) also studied the first and

the third issues using time-variant covariates and nonparametric baseline hazard specification. The conclusions he drew are close to those in Han and Hausman (1990). Other works in the large literature that try to address issues related to the above three include Fallick and Ryu (2006), Baker and Rea (1998), Hausman and Woutersen (2005), Heckman and Singer (1984), and Van den Berg (2001).

The reason why the proportional hazard model becomes so popular is largely due to the common characters that exist in most empirical data and several estimation advantages inherent in the model itself. Duration data (or event history data) often provide appropriate empirical information on a particular process under study. Through application of appropriate estimation methods, one is able to describe the process of change, to discover the causal relationships in one event or among events, and to evaluate their importance. The popularity of the proportional hazard model comes from the fact that it has been recognized as a suitable approach to studying such data.

First, by the baseline hazard specification, we can separate duration dependence from individual regressors' effects. Furthermore, one can capture changes in hazard over the spell given a flexible manner of time effect. This is very useful if we want to check how some factors (such as unemployment compensation) affect a unit's spell (e.g., unemployment duration). Unlike the simple assumption that the baseline hazard belongs to a specific parametric family, non-parametric baseline specification will allow the shape of duration determined by information in the empirical datasets.

Second, duration data usually include censored data. Proportional hazard duration model, different from conventional regression models (see Flinn and Heckman 1982 for a convincing critique of the application of regression methods in duration studies), can

accommodate censored spells without too restrictive baseline assumption. Covariate coefficients estimated under the semiparametric methodology will be consistent even if the distribution of baseline hazard is unknown.

Third, the proportional hazard model allows the inclusion of time dependent regressors. In many situations, time-variant covariates are important since their changes will reflect current or future changes in hazards.

Fourth, unobserved heterogeneity is easy to be added on in the duration model. In empirical studies, we are not always able to have information on all important factors. Moreover, sometimes we do not know what is important. Therefore, introducing heterogeneity into the duration model will help us capture unobserved individual differences and hence improve our model estimations.

Given the advantages and potentials of the proportional hazard model, many important works have been done to improve its estimation power. Mainly the purpose is to put as less restrictive distribution assumptions as possible on the hazard shape and the heterogeneity so that valuable information can be released from dataset without interruption. These works focus on either nonparametric baseline hazard, or nonparametric heterogeneity, or both. See Heckman and Singer (1984), Meyer (1990), Han and Hausman (1990), McCall (1996), Horowitz (1999), Hong and Chernozhukov (2002), Hausman and Woutersen (2005), Honoré and Lleras-Muney (2006), Abbring and Van den Berg (2007), and Khan and Tamer (2007) among others for references on the refinement of estimation methods.

This dissertation analyzes unemployment duration data by using semiparametric classical and Bayesian estimation approaches. In the maximum likelihood estimation

chapters (Chapter 1 and 2), we use the duration models specified in Han and Hausman (1990), Meyer (1990) and Sueyoshi (1992) to examine the effect of unemployment compensation on duration hazard. The baseline hazard is estimated nonparametrically, and gamma heterogeneity is considered in the mixed proportional hazard model. Competing risks duration model is also studied for different exiting paths, new job and recall.

In Chapter 3, the Bayesian estimation chapter, we develop a hierarchical single risk proportional hazard model without considering unobserved heterogeneity. Priors on the baseline hazard and the covariate coefficients are both assumed to be noninformative in order to maintain objectivity in the results.

Currently we are working on the MPH model with nonparametric baseline hazard and either normal or nonparametric heterogeneity. Dirichlet process prior is applied to specify heterogeneity so that different sources of random effects will be taken into account. Although there are related works in the literature (e.g., Campolieti 2001), based on our best knowledge, our Bayesian model specification seems more flexible, especially on the baseline hazard.

Chapter 1

Semiparametric Estimation of Single Risk Proportional Hazard Models with Applications in Unemployment Duration

1.1 Introduction

In duration modeling, there are three categories of estimation methods: nonparametric estimation, parametric estimation and semiparametric estimation.

Situations sometimes arise in which there is no covariate when analyzing duration data. In that case, it is appropriate to estimate the model using nonparametric methodology, such as the Kaplan-Meier estimate (also known as product limit estimate) and Nelson-Aalen estimate (sometimes called the empirical cumulative hazard function), based on all the observations. However, situations often arise in which there are some covariates, and people wonder how to allow observed covariates to be linked to subsequent outcomes. Therefore, it is not surprising to see that over the last 20 years, hazard rate models rather than nonparametric estimates, have increasingly been used in practical research for analyses of duration data.

On the other hand, fully parametric hazard models suffer from several drawbacks. First, although selecting a specific parametric model based on its tractability and how well it fits the data seems to be a good strategy, rather large samples are often needed before one can tell the superiority of one model over another, and severe right censoring usually limits

the comparison of models (see Lawless 2003). Also, a particular model chosen based on comparisons may be hard to apply generally on other studies because different data sets might have different underlying true distributions. Second, duration models (especially the proportional hazard model) usually have two components: the baseline hazard and a particular function form that includes effects of the observed covariates. Fully parametric hazard models typically impose a smooth shape on the baseline hazard function. This commonly leads to models that do not adequately capture features of the lifetime distribution that are apparent in empirical data, hence they are inappropriate for testing hypotheses on the effects of interested covariates. Third, available duration data often represent discrete observations of a continuous process and it may be important to take the discreteness into account. Therefore, nonparametric estimation of the baseline hazard rate may be recommended.

Semiparametric models specify the dependence of time on covariates parametrically, but leave the actual distribution of duration unconstrained. This is realized by using a nonparametric baseline hazard function specification. There are several event history models under semiparametric framework (e.g., Cox and Snell 1989, McCullagh and Nelder 1989, and Honoré and Lleras-Muney 2006), including log-location-scale (accelerated failure time) models, logistic models and discrete-time hazard-based models. But the best-known semiparametric lifetime regression model may be the proportional hazard (PH) model introduced by Cox (1972). The Cox model, also called Relative Risk model, is a reasonable compromise between the nonparametric estimators and the possibly excessively restricted parametric models. It can be extended for multi-spell data, and for data with multiple origin and destination states.

Cox (1972) and Cox (1975) developed the partial likelihood method to estimate covariate parameters without having to consider the baseline hazard explicitly. Although this method has great simplicity and usefulness, it is theoretically and computationally difficult for the correct treatment of tied, completed durations. In addition, useful information about the shape of the hazard function, which is given by the baseline hazard, is important for the study of duration behavior.

After the seminal work on unemployment duration by Lancaster (1979), many researches in labor economics have been done with applications of the proportional hazard duration model. Among them, Katz (1986) specified a one-parameter Weibull distribution for the baseline hazard, while Ham and Rea (1987) assumed that the baseline hazard follows a six-order polynomial function and they also investigated the robustness of their results with a step function. Another two outstanding works are Meyer (1990) and Han and Hausman (1990). Their techniques resolve the deficiency inherent in Cox's partial likelihood method. That is, the parameters of the (log-integrated) baseline hazard are estimated flexibly and simultaneously with the covariate parameters. In their papers, the continuous duration variable is discretized into time intervals and tied observations are easy to deal with.

In this chapter I study the effects of unemployment compensation and other important sociodemographic factors on unemployment duration. Whether duration dependence follows a particular function form is also examined by estimating the baseline hazard nonparametrically.

The data set used in this study is created from the Panel Study of Income Dynamics (PSID). It includes unemployment duration in weeks for each observation, information on

whether one received unemployment insurance when unemployed, and other personal information such as age, gender and so on.

Discrete, semiparametric, proportional hazard models are used and compared among different specifications. Following Meyer (1990), Han and Hausman (1990) and Sueyoshi (1992), we allow for nonparametric estimation of the effect of time on the unemployment exit rate. This specification is more realistic than the monotonic assumption (such as Weibull) on duration dependence. Cutler (1995) also indicated that nonparametric specification will generate true baseline hazard, therefore more accurate hazard function. By diagnosing estimation results of the baseline hazard, we are able to test the null hypothesis that duration dependence distribution belongs to a specific parametric family.

Because unobserved individual heterogeneity has the potential to bias the estimation results, both Meyer (1990) and Han and Hausman (1990) considered gamma heterogeneity as an additional source of error in the hazard model (i.e., the so called mixed proportional hazard model, MPH). Though assuming gamma distribution for unobserved heterogeneity produces a likelihood function with a closed form, it is more reasonable to estimate the distribution of heterogeneity nonparametrically (see, for example, Horowitz 1999, Hausman and Woutersen 2005). This is beyond the scope of this chapter and awaits future research.

There are a number of findings from this empirical study. Our nonparametric baseline hazard estimations have the ability to represent perceived features of the hazard function. This is very important since without providing a good description of the observed data, a model may not be considered as a good one. One of the perceived features in the

analysis of empirical duration data is positive jumps near the point of UI benefits exhaustion. These jumps are reflected in the estimated baseline hazard. In addition, UI benefits are found to have a strong negative effect on the probability of leaving unemployment. Some sociodemographic characters, such as race and marital status, also have strong effects on this probability. Besides, we find that it is very important to account for unobserved heterogeneity in the regression specification if one wants to reveal informative duration distribution.

The plan of this chapter is as follows. Section 1.2 presents the (mixed) proportional hazard model specifications for single risk unemployment durations. In Section 1.3 we describe the data set used in this dissertation and report Kaplan-Meier estimates of the hazard function. Section 1.4 provides analyses of model estimation results. Concluding remarks and future research directions are given in Section 1.5.

1.2 Ordered logit/probit duration models without and with unobserved heterogeneity

Suppose that the time axis is partitioned into a finite number of disjoint intervals, say $1, 2, \dots, T$, and each interval represents its corresponding week, i.e., t means t^{th} week. And we observe a set of failure times for a group of individuals $i, i = 1, 2, \dots, N$. Denote the failure times with time intervals, $t = 1, 2, \dots, T$, we now have an event history/duration data set with the failure time serves as dependent variable. The dependent variable studied in this paper, T_i , is unemployment duration, which is measured as the length of time the unemployment lasts until one gets employed or is censored.

Very often observations of event histories are censored. Censorship occurs when the information about the duration is not completely recorded. There are different types of censoring (see, for example, Lawless 2003 and Hougaard 2000). In our study, we do not have a problem with left censoring since we know the beginning of an individual's unemployment spell; but we must be concerned with right censoring. There are individuals who are still not employed on the date interviewed. For these persons, unemployment duration is right censored because we do not know the end of their unemployment spells. This kind of censoring is also called "type I censoring" (Kalbfleisch and Prentice 1980) and is unproblematic so it can be handled with event history methods.

Let $\lambda_i(t)$ be the hazard function, the probability that person i with characteristics X_i exits unemployment status at time t , conditional upon staying unemployed until t . The equation for $\lambda_i(t)$ is (see Prentice and Gloeckler 1978):

$$\lambda_i(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T_i < t + \Delta t | T_i \geq t)}{\Delta t} = \lambda_0(t) \exp(-X_i(t)\beta). \quad (1.1)$$

The above function is called the proportional hazard model. It assumes that hazard function $\lambda_i(t)$ can be decomposed into a baseline hazard $\lambda_0(t)$ and a "shift factor" $\exp(-X_i(t)\beta)$. The shift factor represents the proportional shift in the hazard caused by explanatory covariates X_i with unknown parameters β . Values of independent variable vector X_i could change over time intervals. But here we assume X_i is constant over time.

If we let

$$\delta_i = \log \int_0^t \lambda_0(\tau) d\tau, \quad t = 1, \dots, T, \quad (1.2)$$

denote the log-integrated baseline hazard, the proportional hazard model in (1.1) becomes

$$\varepsilon_i = \delta_i - x_i\beta, \quad (1.3)$$

where $t = 1, 2, \dots, T, i = 1, 2, \dots, N$ and $\varepsilon_i = \log \int_0^t \lambda_i(\tau) d\tau$.

Suppose that the probability density function of T is $f(t)$, then the probability that the length of a duration falls between $[t-1, t)$ is

$$P(t-1 \leq T_i < t) = \int_{t-1}^t f_i(\tau) d\tau. \quad (1.4)$$

Another interpretation of (1.4) is that it represents the probability of failure in period t for person i . Now replace $f(t)$ with $\lambda(t)S(t)$ and transform the integration bounds accordingly. It is easy to see that

$$P(t-1 \leq T_i < t) = P(\delta_{t-1} - X_i\beta \leq \varepsilon_i < \delta_t - X_i\beta) = \int_{\delta_{t-1} - X_i\beta}^{\delta_t - X_i\beta} f(\varepsilon) d\varepsilon, \quad (1.5)$$

where $f(\varepsilon) = \exp[\varepsilon - \exp(\varepsilon)]$. This means if followed strictly from the definition of hazard function, ε has a standard minimum extreme value distribution. Hence (1.5) can be rewritten as

$$P(t-1 \leq T_i < t) = \exp[-\exp(\delta_{t-1} - X_i\beta)] - \exp[-\exp(\delta_t - X_i\beta)]. \quad (1.6)$$

The likelihood function for a sample of N individuals can be written as a function of terms such as (1.6).

$$L(\delta, \beta) = \prod_{i=1}^N \prod_{t=1}^T \left\{ \exp[-\exp(\delta_{t-1} - X_i\beta)] - \exp[-\exp(\delta_t - X_i\beta)] \right\}^{1-C_i} \times \{1 - \exp[-\exp(\delta_t - X_i\beta)]\}^{C_i} y_{it}, \quad (1.7)$$

where y_{it} is an indicator variable that equals one if individual i fails or is censored in period t and zero elsewhere; C_i is another indicator variable that equals one if individual i is censored and zero otherwise. Under this specification, we assume censorings occur at the end of time intervals. δ_t is constant in its corresponding time period $[t-1, t)$. That is, T is

assumed to have a piecewise-constant hazard function. X_i is a $l \times k$ vector of observed explanatory variables, and β is a $k \times l$ vector of unknown coefficients.

The corresponding log-likelihood function is

$$\log L(\delta, \beta) = \sum_{i=1}^N \sum_{t=1}^T y_{it} \{ (1 - C_i) \log \{ \exp[-\exp(\delta_{t-1} - X_i \beta)] - \exp[-\exp(\delta_t - X_i \beta)] \} + C_i \log \{ 1 - \exp[-\exp(\delta_t - X_i \beta)] \} \}. \quad (1.8)$$

Since we know the intervals during which spells end, though exact spell lengths are unknown, the log-likelihood function (1.8) is correct for both failure and censored data. Survivor functions (one minus the cumulative distribution function) are used for censored spells to describe that person i survives longer than t if the spell is censored at some point in $[t-1, t]$.

Because the standard normal distribution is very similar to the standard extreme value distribution except in the extreme tails, unless we are dealing with large enough data set, we can hardly tell the difference on the estimation (e.g., Greene 2000). Therefore, we suppose that in our study, standard normal ε will be a good approximation to the extreme value ε .

Accordingly, the log-likelihood function is given by

$$\log L(\delta, \beta) = \sum_{i=1}^N \sum_{t=1}^T y_{it} \left\{ (1 - C_i) \log \left[\int_{\delta_{t-1} - X_i \beta}^{\delta_t - X_i \beta} f(\varepsilon) d\varepsilon \right] + C_i \log \left[1 - \int_{-\infty}^{\delta_t - X_i \beta} f(\varepsilon) d\varepsilon \right] \right\}. \quad (1.9)$$

The only difference between log-likelihood functions in (1.8) and (1.9) is that in (1.9), ε follows a standard normal rather than standard extreme value distribution and therefore the integration has no closed form. Both y_{it} and C_i are indicator variables with same functions

as those in (1.8). Based on the definition of δ_l in (1.2), the δ s are in order. That is, $\delta_0 = -\infty < \delta_1 < \dots < \delta_T < \delta_{T+1} = \infty$. Therefore, the log-likelihood (1.8) takes an ordered logit form and log-likelihood (1.9) takes an ordered probit form.

To make sure hazard rate in the last duration interval T (in this paper $T = 40$) reasonable, one must censor any observations still ongoing after T . Here, observations lasting longer than 41 weeks are censored at 40 because the last time interval of interest is 40th week. About 10.86 percent of the spells survive until the end of our sample duration of 40 weeks. The log-likelihood functions (1.8) and (1.9) are now functions of β and 40 δ , where $\delta = [\delta_1, \delta_2, \dots, \delta_{40}]$. Here we write $\log L(\beta, \delta)$ rather than $\log L(\beta, \lambda_0)$ because it is simpler to work with δ , the unknown logarithm of baseline cumulative hazard function, in a maximum likelihood approach. Interpretation of δ is easy since these estimates are parameters of a continuous time hazard model. The probability of either failing in or surviving each time period lies in the unit interval. Meanwhile, with these model specifications, the problem of treating ties is eliminated; β are estimated consistently with nonparametric baseline hazard parameters δ . Another advantage is that we can examine whether the baseline hazard indeed falls in any parametric distribution family once we have the estimates and asymptotic covariance matrix of the δ .

Importance of considering unobservable differences across individuals is discussed in literature (for example, Heckman and Singer 1984, Horowitz 1999). Suppose heterogeneity is present and takes a multiplicative form the hazard model is then written as

$$\lambda_i(t) = \lambda_0(t) \exp(-X_i \beta) \eta_i, \quad (1.10)$$

where η_i denotes the unobserved heterogeneity and is a random variable that is assumed to be independent of X_i (see the Hausman's test for fixed or random effects 1978, and Greene 2000). Now the probability of failure in period t for person i is

$$P(t-1 \leq T_i < t) = S_i(t-1) - S_i(t), \quad (1.11)$$

where S represents unconditional survivor function. To find these unconditional survivor functions, one must integrate conditional survivor functions over the distribution of η , the unobserved heterogeneity. That is

$$S_i(t) = \int S_i(t | \eta) f_\eta(\eta) d\eta = \int \exp[-\exp(\delta_i - X_i \beta) \eta] f_\eta(\eta) d\eta. \quad (1.12)$$

Let η follow a gamma distribution with expectation $E(\eta) = 1$ (see Hougaard 2000).

As implied by this assumption, the density function of η becomes

$$f_\eta(\eta) = \frac{\theta^\theta \eta^{\theta-1}}{\Gamma(\theta)} \exp(-\theta\eta), \quad \eta > 0, \quad \theta > 0. \quad (1.13)$$

The variance of this distribution is $\text{var}(\eta) = \frac{1}{\theta}$. The calculation is now easy for (1.12) (see

Lancaster 1990) and the unconditional survivor function is

$$S_i(t) = [1 + 1/\theta \cdot \exp(\delta_i - X_i \beta)]^{-\theta}. \quad (1.14)$$

As we can see, choosing gamma distribution as the distribution of unobservable heterogeneity avoids numerical integrations that may be necessary if the assumed distribution is normal. Unfortunately, this choice is justified on the grounds of simplicity, the availability of likelihood based inference procedures and ease of use for description, prediction and decision (e.g., Heckman and Singer 1984, Lawless 2003). Hence, though gamma distribution may be a good choice, more general distribution assumptions and estimation methods are desired to control for unobserved heterogeneity.

Next we obtain the likelihood function for a sample of N units

$$L(\delta, \beta, \theta) = \prod_{i=1}^N \prod_{t=1}^T \{ [S_i(t-1) - S_i(t)]^{1-C_i} \times [S_i(t)]^{C_i} \}^{y_{it}}, \quad (1.15)$$

where y_{it} is an indicator variable that equals one if individual i fails or is censored in period t and zero elsewhere; C_i is another indicator variable that equals one if individual i is censored and zero if failure. Again here we assume censorings occur at the end of time intervals.

The corresponding log-likelihood function is

$$\begin{aligned} \log L(\delta, \beta, \theta) = & \sum_{i=1}^N \sum_{t=1}^T y_{it} \left\{ (1 - C_i) \log \left[1 + 1/\theta \cdot \exp(\delta_{t-1} - X_i \beta) \right]^{-\theta} \right. \\ & \left. - [1 + 1/\theta \cdot \exp(\delta_t - X_i \beta)]^{-\theta} \right\} + C_i \log \left\{ [1 + 1/\theta \cdot \exp(\delta_t - X_i \beta)]^{-\theta} \right\}. \end{aligned} \quad (1.16)$$

By assuming unobserved heterogeneity with a gamma distribution, we have a closed form expression for the log-likelihood without dealing with numerical integration.

The usual asymptotic property of the maximum likelihood estimator (MLE) (see Han and Hausman 1986) $\hat{B} = (\hat{\beta}, \hat{\delta}, \hat{\theta})$ is

$$\sqrt{N}(\hat{B} - B_0) \xrightarrow{A} N \left(0, \left[\lim_{N \rightarrow \infty} \frac{1}{N} \frac{\partial^2 \log L}{\partial B \partial B'} \right]^{-1} \right), \quad (1.17)$$

where $\log L$ are log-likelihood functions specified in (1.8), (1.9) and (1.16), respectively. The covariance matrix in (1.17) is just the inverse of the Hessian of the log-likelihood function. Based on (1.17), we can derive large sample inference for the unknown parameters.

Conventional approaches usually assume parametric forms for the hazard term that determines duration dependence. However, estimated covariate coefficients β are often sensitive to such restrictions and become inconsistent if the assumed baseline hazard

distribution is incorrect (e.g., Moffitt 1985). Furthermore, it may be difficult to judge which assumed baseline is correct due to the lack of theoretical support for any particular shape.

Unlike traditional methods, all the model specifications in (1.8), (1.9) and (1.16) estimate the baseline hazard nonparametrically using standard maximization techniques. Estimates of β are consistent even though the distribution of baseline hazard is unknown. One can check whether the baseline hazard follows some particular functional form after getting estimation results for δ . Moreover, estimated δ can even help detect possible neglected variables (see Meyer 1990). Specified as discrete time duration models, models (1.8), (1.9) and (1.16) are not only easy to be maximized but also retain interpretations as incompletely observed continuous time hazard models. Compared to the results found in Heckman and Singer (1984), inclusion of heterogeneity and choice of its distribution seem to have few effects on the estimation results when there is nonparametric baseline hazard (see both Han and Hausman 1990 and Meyer 1990). But to make sure both δ and β can be estimated consistently, a better choice may be leaving the distribution of η unknown too. Hausman and Woutersen (2005) argued that if the true distribution for heterogeneity is not gamma, estimates for all parameters could be biased. Interestingly, Abbring and Van den Berg (1998) propose that under mild regularity conditions, the heterogeneity distribution among individuals at long durations can be approximated well by a gamma distribution. We report and compare estimates without heterogeneity and with gamma distributed heterogeneity in this paper. Further researches on normal distributed or nonparametric heterogeneity using the Bayesian estimation method are currently in progress.

1.3 Data description

Our data set is extracted from waves 15 and 16 of the Panel Study of Income Dynamics (PSID) conducted by University of Michigan. Similar but different data sets are used in Katz (1986) and Han and Hausman (1990). We choose waves 15 and 16 because these two are the most recent data that distinguish new job transitions from recalls, which is key to the construction of a competing risks model. Note that in this chapter only single risk models are considered, studies on competing risks model could be found in next chapter.

The dependent variable is the unemployment duration in weeks for each individual, no matter the spell finally ends through recall (that is, returning to former employer), taking a new job or is censored at the interview date. Thus, we consider only complete or right-censored spells to avoid the problems caused by left censorship and other initial conditions issues (see, for example, Heckman and Singer 1984). Given this sampling scheme, the formulation of the likelihood functions should cover both types of unemployment durations.

Except information on the most recent unemployment duration, waves 15 and 16 include other detailed questions on socio-demographic characteristics, such as age, sex, race, years of education, number of dependents, marital status. These characteristics serve as explanatory variables in our model. An extra explanatory variable is whether an individual received unemployment compensation when he was unemployed most recently. Unemployment compensation is a key part of the social insurance system in the US. Like any insurance program, it can potentially influence the behavior of insured persons.

However, we can only decide if an individual received unemployment compensation (UI) or not. Calculation of weekly UI benefit rate, a variable that is commonly used in job search models and other related researches, is unavailable based on the PSID information. Though we cannot examine the theory that weekly UI leads to longer duration of unemployment, we expect the estimated UI coefficient to be positive since receiving unemployment compensation may reduce an individual's motivation to get employed again. Another potential problem in our data is that long-duration spells will be over-sampled due to the PSID sampling frame---only those unemployment spells at least partially contained in 1981 and at least partially contained in 1982 are included. Katz (1986) discussed the problems and concluded that they were too minor to affect his results. Since the same interview scheme is adopted to collect the data sets in Katz (1986) and the data applied in this paper, we would like to draw a same conclusion.

The following conditions are to be satisfied if a unit of observation (i.e., a subject's last spell of unemployment) is included in our sample:

1. The observation is on a household head;
2. The household head is the same head as in previous year;
3. The age of the household head is between 20 and 65;
4. The household head belongs to labor force participants;
5. Observations with missing values are deleted;
6. (a) The duration ended through returning to the former employer;

Or

- (b) The duration ended through finding a new job and the last unemployment spell was initiated by layoff or firing;

Or

(c) The duration was censored and the household head claimed him/her on temporary layoff or unemployed but looking for work at the time of interview.

The unemployment in either case was initiated by layoff or firing.

To summarize the data, we present variable definitions and statistical descriptions in Table 1.1. In Table 1.2, unemployment duration in weeks are listed one by one with corresponding number of observations for each type of ending (recall, new job or censored).

The data set consists of 1114 observations. Like the sample used in Han and Hausman (1990), only about 50 percent of individuals claim they are whites. This oversampling of non-whites is due to the sample frame adopted by the PSID. The most important way to exit unemployment is returning to one's previous employer (i.e., recall). Except 53 percent of the unemployment spells end in recall, 25 percent end in finding a new job and 22 percent are censored at the date interviewed. UI receiving rate is 63.7 percent for the whole sample. This is kind of low as Burtless (1983) and others pointed out. The possible reasons for such low coverage are about half size of this sample are non-whites and either they do not apply for compensation or are not qualified for this benefit. And some workers, no matter they are white or not, have poor employment history hence it is impossible for them to get unemployment insurance when they became unemployed again.

Next, we present Kaplan-Meier estimates of the hazard functions (Kaplan and Meier, 1958). Table 1.3 gives the empirical hazard for the data, and Figures 1.1 and 1.2 plot weekly and monthly hazards for single risk (meaning new job and recall are considered as

one failure), respectively. The hazard H_t in Table 1.3 is simply the ratio of the number of spells ending in week t , D_t , to the number of spells lasting at least t weeks, R_t . Failures include both new job and recall. The number of baseline hazard parameter estimated in this paper is 40, one for each week. Therefore, 121 observations that are still ongoing between 41 and 70 weeks have to be censored. The hazard is higher in weeks 26 and 30 and then weeks 35 and 39. These jumps possibly highlight the UI exhaustion effect.

The high hazard in the first several weeks is mainly caused by the high frequency of recalls when unemployment spells are in their early stage. It is easy to observe this when one separates the new job hazard from the recall hazard in Figure 1.3. Shorter length of eligible benefits for some people may also contribute to the hazard spikes that happened before the UI exhaustion points. Typical job search theories may assume that the new job hazard is monotonically increasing and the recall hazard is monotonically decreasing. However, from Figure 1.3, we see new job finding rate is not strong positive duration dependent. Recall hazard is basically decreasing but involves fluctuations during the process. Such details could not be shown if one just puts some parametric restrictions (e.g., Weibull distribution) on the baseline hazard.

Figures 1.4 and 1.5 display hazard functions of new job and recall when people receive UI benefits while unemployed; figures 1.6 and 1.7 are for people who do not receive UI benefits. The jumps appeared around benefit exhaustion points are less obvious for people without receiving UI, especially in the recall hazard. There are several spikes between 26 and 39 weeks in Figure 1.5. Employers may want to keep their skilled workers when their businesses shrink, so they dismiss their employees temporarily and get them back later by asking workers to take advantage of the UI benefits during their

unemployment. In such case, length of unemployment duration is predetermined and people get back to work before benefit is exhausted. Meanwhile, different states have different lengths of regular benefits when this data set is collected. The minimum is 26 weeks, and in some states whose unemployment rates were above a trigger level, the maximum could be 39 weeks (see Meyer 1990 for an example of the variety). But the length of benefits often changes in the course of unemployment duration. This may help explain why there are several jumps between 26 and 39 weeks. It is because there is variability in benefit lengths across states and people.

In some situations, one can get additional benefit from other programs. For example, the Federal Supplementary Compensation program started to provide up to 62 weeks of benefits in the fall of 1982. Possibly the big jump around 15 months (60 weeks) in Figure 1.5 is caused by the extra length of benefits since part of our data overlap the fall of 1982.

For UI recipients who finally find a new job, basically the hazard is climbing up while for non-UI receivers there is no such trend (see Figures 1.5 and 1.7). This might reflect the fact that even if workers receiving no benefits have stronger incentives to become employed, they are not competitive with UI receivers who might be more skilled or might have other advantages.

1.4 Estimation results

Estimation results for ordered logit and ordered probit single risk models without heterogeneity are reported in Table 1.4 and Table 1.5. Table 1.4 gives covariate coefficient estimates. The effects of unemployment insurance are measured using a dummy variable

“UI” which is equal to one if an individual receives compensation when he is unemployed. Theories predict that high benefits are expected to lower the hazard because the opportunity cost of having no job is decreased. The UI coefficients in all the three model estimations have the expected signs and are well estimated. These coefficient estimates indicate that receiving benefits during unemployment is associated with around 24 percent decrease in the hazard.

Table 1.4 also shows that except the coefficient of UI, parameters for socio-demographic variables, particularly race and marital status, are significant. If an individual is non-white, he has 28 percent less chance of exiting unemployment. Married persons have more incentives to find jobs so their spells are shorter. Both race and marital status affect the hazard at almost the same magnitude as UI does. This again proves the importance of considering such characters when studying unemployment duration. The coefficients on age and education have the expected signs and are significantly different from zero. Older workers may have more experiences than younger ones; hence for them it is easier to resume employment. In ordered probit model (1), the implied effect of a 10 percent increase in education is 0.1 percent increase in the hazard.

Unemployment durations look longer for the family heads that have more dependents. One possible explanation is that they may have received special welfare for families with several children. To test this assumption, one must have other information such as how many benefits in total a family receives during the head's unemployment. Another problematic coefficient is related to sex. In ordered probit model (2), it indicates that females become employed more quickly than males. This probably is due to the addition of the 'dependents' variable in the model. Note the variance of the sex coefficient

is much larger than the other two in ordered probit model (1) and the ordered logit model. Therefore, we would like to predict that males are more possible than females to quit unemployment status.

The nonparametric baseline hazard parameters are in Table 1.5. For each model specification, there are 40 weekly estimated baseline hazards. These estimates are in right order, meaning the values increase over time. Their variances fell well within one unit. Figure 1.8 plots the monthly cumulative baseline hazard using the estimates from three different model specifications. Compare Figure 1.8 with Figure 1.2, the Kaplan-Meier monthly cumulative hazard for single risk case, we find that our baseline hazard estimations capture very well the shape of the duration. There are hazard jumps in 5, 7 and 10 months, reflecting the effects of benefit exhaustions and possible recall or new job arrangements that are made before benefits run out. The basic trend of going up and down exists similarly in both the Kaplan-Meier estimates and the semiparametric proportional hazard duration models estimates.

Different locations of the three baseline hazards in Figure 1.8 are mainly caused by the different estimated values of “shift factor” $\exp(-X_i\beta)$. Directly, this can be observed for the ordered probit (1) and (2) models. For ordered logit model, the fact that the distribution of ε is extreme value but not standard normal, may also contribute to the difference in scale. For similar reasons, we may not be able to do model selection by simply comparing their log-likelihood values, which are shown in Table 1.4, because ordered probit and ordered logit have different likelihood functions. But if we assume ordered probit is a very good approximation for ordered logit except in the extreme tails, and the sample size is not

large enough to tell the difference, then ordered probit model (2) would be the one that wins.

Next, we allow unobserved heterogeneity in our maximum likelihood estimations. The results are given in Table 1.6 and 1.7.

Note that all the estimated proportional effects of explanatory variables on the population hazard rate have the expected signs except 'sex' when the 'dependents' variable is also included. This problem is similar to that one happens earlier in Ordered Probit (2) model and may need further examination or adjustment. Another point of interest is these coefficients tend to be larger in absolute values than those in the models specified without accounting for unobserved heterogeneity. For example, the estimate of the elasticity of unemployment duration (i.e., the logarithmic derivatives of the hazard, see Meyer 1990) with respect to the UI is about 0.39 while previous estimate is about 0.24. This is not dissimilar to the findings reported in Lancaster (1979) who observed that the regression coefficients in the heterogeneity specification are all greater in modulus than those without considering unobserved errors. He also concluded that failure to account for omitted regressors would make the covariates' proportionate effect diminish in modulus over time. Simply based on the log-likelihood values, one might say the improvement on the model fit is negligible for the heterogeneity consideration; however, the heterogeneity variance is respectively 0.9641 and 0.8197, both are significantly different from zero. We shall examine the baseline hazard estimates before discussing more on whether it is necessary to identify unobserved heterogeneity.

Table 1.7 presents estimates of the baseline hazard arising from the heterogeneity specification. Calculated monthly cumulative baseline hazard based on Table 1.7 is plotted

in Figure 1.9. Spikes in the hazard remain at 7 and 10 months, reflecting the benefit exhaustion points at 26 and 39 weeks. However, if we put both Figure 1.8 and 1.9 in one figure---Figure 1.10, we will see that the time patterns derived from the heterogeneity model are different with those we estimate earlier. Even though still exhibiting rises and falls at the same time points, the estimated baseline hazard becomes definitely upward sloping when gamma distributed heterogeneity is allowed.

These results confirm the intuitive argument in Lancaster (1979). He pointed out that though it is important to allow the data to inform us about the baseline hazard functional form, omitting regressors would cause the estimated/observed duration dependence rate to fall more rapidly than that in a model in which there is a complete regressor list. Consequently the estimated (“apparent”) hazard rate will decrease faster or increase less fast than the “true” hazard over the duration. Similar results are given in Figure 1.11, in which the baseline hazard estimates under different model specifications in Meyer (1990) Table VII have been plotted. Also see Hausman and Woutersen (2005) for discussions on this issue.

As we mention earlier, our estimated behavior of baseline hazard in Figure 1.8 captures the shape of the Kaplan-Meier hazard in Figure 1.2 very well; compare Figure 1.11 with Figure 1.3, which also shows Kaplan-Meier empirical hazard, in Meyer (1990), we again find that the nonparametric baseline hazards estimated under models 1 and 3 (both do not account for unobserved heterogeneity) show very close duration distribution to that in Meyer’s Figure 3. This interesting observation may be due to the fact that none of these models allowing for omission of regressors. In both cases, Meyer’s and mine, the

shape of the estimated falling hazard changes dramatically when the heterogeneity is considered, and we can no longer simply contribute the difference to a shift factor effect.

In the process of resuming work, individuals in the sample may adopt different ways to find their jobs, they may face different environment, hold different levels of reservation wages, and even for people with same observed regressors, they may return to work at diverse times. Hence, it is obvious that many sources of variation between individuals are blind to the researchers. Furthermore, unless one can prove that such variations are of negligible magnitude, we ought to include an error term to allow for the influence coming from possibly individually small but collectively significant causal factors.

One typical example is that some workers may resume employment faster than the rest because of unobserved regressors. For these persons, they have higher failure rates. Their returning to work will cause the average hazard of the survivors to fall more rapidly with time. Since the estimated hazard rate at any point of time will be an average of the failure rates of the surviving individuals at that time, it is unsurprised to see that, without accounting for unobserved heterogeneity, the baseline hazard distribution falls much faster through time. Such spurious duration dependence is possibly uninformative about the policy and experience of particular searching individuals. Therefore, to avoid erroneous inferences, it is necessary to make allowance for unobserved heterogeneity.

1.5 Conclusion remarks and future directions

In this chapter we have applied a semiparametric maximum likelihood approach to proportional hazard modeling for duration data and illustrated it for a publicly available

data set from the Panel Study of Income Dynamics (PSID). The methodology enables relaxation of restrictive parametric assumption on the baseline hazard, so that informative results can be attained from the data with as little deviation as possible. Gamma distributed unobserved heterogeneity is ready to be allowed in such model framework.

In both no heterogeneity and heterogeneity model specifications, estimates appear to capture the qualitative effects of explanatory variables, though those in the heterogeneity specification have larger absolute values. However, even both sets of baseline hazard estimates reflect similar patterns of jumps and falls, when accounting for unobserved heterogeneity, the estimated baseline hazard may better approximate the true shape of duration distribution. This implies that if we are to obtain inferences with confidence, we had better attempt to allow for error in our duration model specification because of the quite possibly omitted regressors.

Based on this consideration, estimates from the first heterogeneity specification (the left column in Table 1.6) are used to summarize the results. The coefficient on the UI reception has the expected sign and indicates that receiving unemployment compensation will lower the possibility of resuming work for about 39 percent. The estimated elasticity of unemployment duration with respect to race and marital status is around 0.47 and -0.42 , respectively. Older, more experienced workers and more educated workers will return to work faster than those who are younger and less educated. And quite possibly males will take shorter time periods to resume employment compared to females.

In literature it has been suggested, either that inclusion of unobserved error is negligible when the baseline hazard is nonparametrically estimated (e.g., Han and Hausman 1990), or that heterogeneity plays an important role in duration model if

assuming functional form baseline (see Heckman and Singer 1984). The estimation results in this chapter (also refer to Meyer 1990 and Lancaster 1979) point out need for allowing an error term in the specification even when the baseline hazard is estimated without functional form restrictions (for example, Sueyoshi 1992).

The omission of regressors not only affects the time variation in the probability of individuals leaving unemployment, but also produces less pronounced estimates of the effects of observed regressors. Inferences derived in such case may be misleading and could not provide valuable information on policy making and individual job search behavior (such as, the estimated hazard keeps falling while in reality, individuals resume work with increasing probability). Unfortunately, misspecified error term may still lead to biased estimation results (Hausman and Woutersen 2005). Therefore, the development of nonparametric methods on both baseline hazard and heterogeneity seems necessary if we want to have more precise estimations of the duration dependence shape and the effect of causal factors. Further work on this issue with MCMC-based Bayesian approach is currently ongoing.

We also attempt to extend single risk model to competing risks model. Since different unemployment exiting ways, such as new job and recall, have different time patterns as one can see from the Kaplan-Meier estimates, it is meaningful to consider competing risks duration model. Studies of dependent bivariate competing risks model will be reported in next chapter.

Other possible extensions of the models specified in this paper include:

- (1) Allowing for time-varying covariates. Due to limitations on data availability, regressors are treated constant over the spell in this dissertation. But extensions to

include time-dependent covariates are straightforward by using the model specifications in Meyer (1990) and Sueyoshi (1992). When time-varying covariates are considered, it is especially important to specify the time pattern of the hazard correctly, because the coefficients of such regressors are more likely to be biased than other coefficients due to their dependence on time. In this case, it may be desirable to introduce regression error into the specification. Moreover, allowing for unobserved heterogeneity will make the proportionate effect of time dependent regressors on the probability of returning to work not diminish over time (see Lancaster 1979).

- (2) Allowing for time-varying covariate coefficients and dealing with dependent multivariate competing risks duration model (for example, McCall 1996).
- (3) Allowing for the examination of multiple-duration data (as in Van den Berg 2001 and many other works).

Table 1.1 Variable definitions and mean-standard deviations of the PSID unemployment duration sample (n=1114)

Variable	Description	Mean (standard deviation)
Duration	= observed unemployment duration in weeks	16.847 (16.658)
Age	= age of individuals in years	34.232 (10.420)
Sex	= 1 if female	0.175 (0.380)
Education	= years of schooling	11.388 (2.250)
Dependents	= number of dependents	3.105 (1.654)
Race	= 1 if non-white	0.496 (0.500)
UI	= 1 if individual received compensation during spell	0.637 (0.481)
Married	= 1 if married	0.618 (0.486)

Source: Author's calculation from the PSID sample.

Table 1.2 Unemployment duration in weeks and number of observations for each failure type (recall, new job or censored)*

Duration in weeks	Recall	New job	Censored	Total
1	98	6	3	107
2	108	13	3	124
3	43	8	4	55
4	56	21	4	81
5	18	3	0	21
6	25	14	1	40
7	11	5	0	16
8	34	12	1	47
9	18	7	4	29
10	13	5	1	19
11	6	2	0	8
12	23	9	0	32
13	25	13	13	51
14	3	1	2	6
15	12	5	3	20
16	5	5	3	13
17	12	19	19	50
18	3	5	4	12
19	1	4	8	13
20	10	4	2	16
21	6	5	2	13
22	7	7	10	24
23	2	2	0	4
24	2	7	4	13
25	0	1	2	3
26	12	19	27	58
27	1	0	2	3
28	1	2	2	5
29	1	0	2	3
30	6	13	11	30
31	0	1	3	4
32	3	1	4	8
33	1	0	0	1
34	1	2	3	6
35	1	10	11	22

(to be continued)

Table 1.2 (continued)

Duration in weeks	Recall	New job	Censored	Total
36	1	2	1	4
37	1	0	2	3
38	0	2	1	3
39	3	7	10	20
40	0	4	2	6
41	0	0	1	1
42	0	0	1	1
43	3	2	3	8
44	0	0	1	1
45	0	1	0	1
46	0	1	1	2
47	0	1	2	3
48	2	4	4	10
49	1	2	1	4
50	1	0	0	1
51	0	0	0	0
52	0	5	40	45
53	0	1	0	1
54	0	1	1	2
55	0	1	1	2
56	1	1	3	5
57	0	1	2	3
58	0	0	2	2
59	0	0	0	0
60	0	5	2	7
61	0	0	2	2
62	0	0	4	4
63	0	0	0	0
64	0	1	0	1
65	2	2	2	6
66	1	1	0	2
67	2	0	0	2
68	0	0	1	1
69	0	1	2	3
70	0	0	1	1
Totals	586	277	251	1114

**Here 'recall' means returning to a former employer, 'new job' means finding a position in a new employer, and 'censored' refers to right-censored durations.*

Table 1.3 Failures, Censorings, and the Kaplan-Meier Empirical Hazard

Week	Risk Set	Failures	Censorings	Hazard	Standard
t	R_t	D_t	C_t	H_t	error
1	1114	104	3	0.09336	0.0087
2	1007	121	3	0.12016	0.0102
3	883	51	4	0.05776	0.0079
4	828	77	4	0.09300	0.0101
5	747	21	0	0.02811	0.0060
6	726	39	1	0.05372	0.0084
7	686	16	0	0.02332	0.0058
8	670	46	1	0.06866	0.0098
9	623	25	4	0.04013	0.0079
10	594	18	1	0.03030	0.0070
11	575	8	0	0.01391	0.0049
12	567	32	0	0.05644	0.0097
13	535	38	13	0.07103	0.0111
14	484	4	2	0.00826	0.0041
15	478	17	3	0.03556	0.0085
16	458	10	3	0.02183	0.0068
17	445	31	19	0.06966	0.0121
18	395	8	4	0.02025	0.0071
19	383	5	8	0.01305	0.0058
20	370	14	2	0.03784	0.0099
21	354	11	2	0.03107	0.0092
22	341	14	10	0.04106	0.0107
23	317	4	0	0.01262	0.0063
24	313	9	4	0.02875	0.0094
25	300	1	2	0.00333	0.0033
26	297	31	27	0.10438	0.0177
27	239	1	2	0.00418	0.0042
28	236	3	2	0.01271	0.0073
29	231	1	2	0.00433	0.0043
30	228	19	11	0.08333	0.0183
31	198	1	3	0.00505	0.0050
32	194	4	4	0.02062	0.0102
33	186	1	0	0.00538	0.0054
34	185	3	3	0.01622	0.0093
35	179	11	11	0.06145	0.0180
36	157	3	1	0.01911	0.0109
37	153	1	2	0.00654	0.0065
38	150	2	1	0.01333	0.0094
39	147	10	10	0.06803	0.0208
40	127	4	2	0.03150	0.0155

Note: The variance of hazard is $H_t(1 - H_t)/R_t$, where R_t is the number of spells that last at least t weeks. 121 observations are censored after 40 weeks.

Table 1.4 Hazard Model Estimates^a

Variable	Ordered Probit (1)	Ordered Probit (2)	Ordered Logit
age	-0.0135 (0.0027)	-0.0136 (0.0026)	-0.0132 (0.0029)
sex	0.0101 (0.0373)	-0.0204 (0.1510)	0.0139 (0.0740)
education	-0.0116 (0.0078)	-0.0099 (0.0089)	-0.0151 (0.0092)
dependents		0.0419 (0.0226)	0.0566 (0.0256)
race	0.2838 (0.0630)	0.2645 (0.0640)	0.2835 (0.0710)
UI	0.2421 (0.0677)	0.2364 (0.0650)	0.2660 (0.0752)
married	-0.2808 (0.0693)	-0.3741 (0.1123)	-0.4100 (0.0962)
sample size	1114	1114	1114
log-likelihood	-3034.1608	-3032.5708	-3035.2983

^aStandard variances are shown in parentheses. The log-likelihood functions for each model, from right to left, are equations (1.8), (1.9) and (1.9), respectively, in Section 1.2. Baseline hazard parameters are reported in Table 1.5.

Table 1.5 Baseline Hazard Estimates from Model Specifications in Table 1.4

Week	Ordered Probit (1)		Ordered Probit (2)		Ordered Logit	
	Hazard	Std. Error	Hazard	Std. Error	Hazard	Std. Error
1	-1.842	0.087	-1.771	0.081	-2.751	0.111
2	-1.340	0.080	-1.269	0.073	-1.910	0.083
3	-1.183	0.078	-1.112	0.071	-1.674	0.078
4	-0.969	0.076	-0.898	0.069	-1.376	0.072
5	-0.914	0.076	-0.844	0.068	-1.303	0.071
6	-0.815	0.075	-0.745	0.067	-1.174	0.069
7	-0.775	0.075	-0.705	0.066	-1.123	0.068
8	-0.662	0.074	-0.591	0.065	-0.983	0.066
9	-0.601	0.073	-0.530	0.064	-0.910	0.065
10	-0.557	0.073	-0.486	0.064	-0.858	0.064
11	-0.537	0.072	-0.466	0.064	-0.835	0.064
12	-0.459	0.072	-0.388	0.063	-0.746	0.062
13	-0.366	0.071	-0.295	0.061	-0.642	0.061
14	-0.356	0.071	-0.285	0.061	-0.631	0.061
15	-0.313	0.070	-0.242	0.060	-0.584	0.060
16	-0.287	0.070	-0.216	0.060	-0.557	0.060
17	-0.207	0.069	-0.136	0.059	-0.472	0.058
18	-0.186	0.068	-0.114	0.058	-0.449	0.058
19	-0.172	0.068	-0.100	0.058	-0.434	0.058
20	-0.132	0.067	-0.060	0.057	-0.393	0.057
21	-0.100	0.067	-0.028	0.057	-0.361	0.056
22	-0.059	0.066	0.013	0.055	-0.319	0.056
23	-0.046	0.066	0.026	0.055	-0.306	0.055
24	-0.018	0.065	0.054	0.056	-0.279	0.055
25	-0.015	0.065	0.057	0.056	-0.275	0.055
26	0.086	0.068	0.158	0.059	-0.177	0.053
27	0.089	0.068	0.162	0.059	-0.173	0.052
28	0.101	0.068	0.173	0.059	-0.163	0.052
29	0.104	0.069	0.177	0.059	-0.159	0.052
30	0.178	0.070	0.251	0.061	-0.089	0.050
31	0.183	0.071	0.255	0.061	-0.085	0.050
32	0.200	0.071	0.273	0.062	-0.069	0.049
33	0.204	0.071	0.277	0.062	-0.064	0.049
34	0.218	0.071	0.291	0.062	-0.052	0.049
35	0.269	0.073	0.342	0.064	-0.005	0.047
36	0.284	0.073	0.357	0.064	0.009	0.047
37	0.289	0.073	0.362	0.064	0.014	0.048
38	0.300	0.074	0.373	0.065	0.023	0.048
39	0.353	0.075	0.427	0.066	0.071	0.050
40	0.377	0.076	0.451	0.067	0.092	0.051

Table 1.6 Hazard Model Estimates^b

Variable	Ordered Logit w/ Error (1)	Ordered Logit w/ Error (2)
age	-0.0244 (0.0065)	-0.0233 (0.0064)
sex	0.0260 (0.0809)	-0.0181 (0.0583)
education	-0.0183 (0.0161)	-0.0152 (0.0144)
dependents		0.0691 (0.0364)
race	0.4747 (0.1139)	0.4222 (0.1098)
UI	0.3920 (0.1127)	0.3668 (0.1092)
married	-0.4244 (0.1255)	-0.5577 (0.1413)
theta	1.0372 (0.1718)	1.2200 (0.3016)
sample size	1114	1114
log-likelihood	-3033.6529	-3031.9965

^bStandard variances are shown in parentheses. The log-likelihood function for these models is equation (1.16) in Section 1.2. Baseline hazard parameters are reported in Table 1.7.

Table 1.7 Baseline Hazard Estimates from Model Specifications in Table 1.6

Week	Ordered Logit w/ Heterogeneity (1)		Ordered Logit w/ Heterogeneity (2)	
	Hazard	Std. Error	Hazard	Std. Error
1	-3.178	0.306	-3.024	0.263
2	-2.261	0.275	-2.119	0.228
3	-1.990	0.263	-1.854	0.216
4	-1.634	0.246	-1.507	0.196
5	-1.544	0.242	-1.420	0.191
6	-1.383	0.233	-1.263	0.181
7	-1.319	0.229	-1.201	0.176
8	-1.136	0.219	-1.024	0.164
9	-1.037	0.213	-0.929	0.157
10	-0.966	0.209	-0.861	0.152
11	-0.935	0.208	-0.831	0.150
12	-0.810	0.201	-0.711	0.142
13	-0.661	0.194	-0.569	0.133
14	-0.645	0.193	-0.553	0.132
15	-0.576	0.191	-0.488	0.128
16	-0.535	0.189	-0.449	0.126
17	-0.406	0.186	-0.327	0.121
18	-0.371	0.185	-0.294	0.120
19	-0.349	0.185	-0.273	0.119
20	-0.285	0.184	-0.212	0.118
21	-0.233	0.184	-0.164	0.118
22	-0.166	0.184	-0.101	0.118
23	-0.146	0.184	-0.082	0.118
24	-0.101	0.185	-0.039	0.119
25	-0.096	0.185	-0.035	0.119
26	0.068	0.191	0.119	0.131
27	0.074	0.191	0.125	0.131
28	0.093	0.193	0.142	0.133
29	0.099	0.193	0.148	0.134
30	0.221	0.204	0.262	0.147
31	0.228	0.205	0.268	0.148
32	0.257	0.208	0.295	0.152
33	0.264	0.209	0.302	0.153
34	0.286	0.212	0.323	0.156
35	0.372	0.222	0.402	0.168
36	0.397	0.226	0.426	0.172
37	0.406	0.227	0.434	0.173
38	0.424	0.229	0.450	0.176
39	0.515	0.243	0.535	0.191
40	0.556	0.250	0.572	0.198

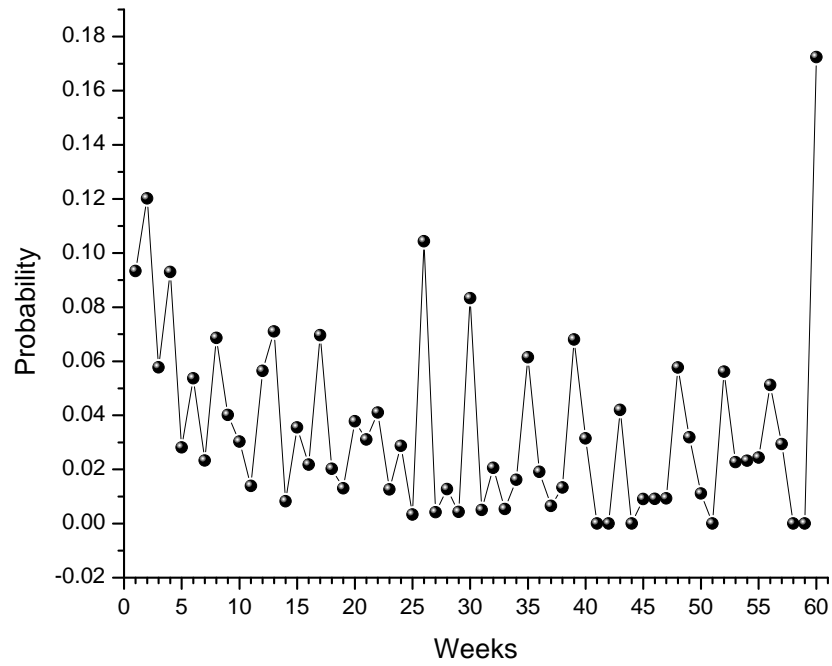


Figure 1.1 Subsample (from weeks 1 to 60) weekly hazard for resuming work. Single risk model (i.e., recall risk not distinguished from new job risk).

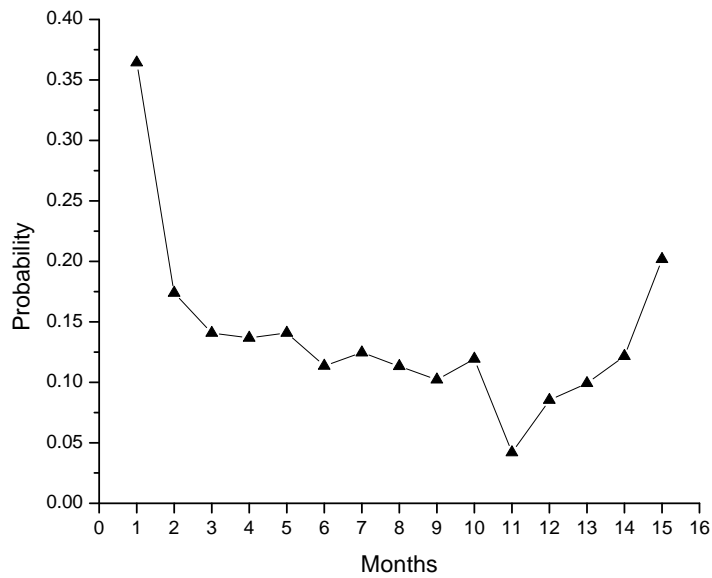


Figure 1.2 Sample monthly hazard functions for re-employment, single risk model.

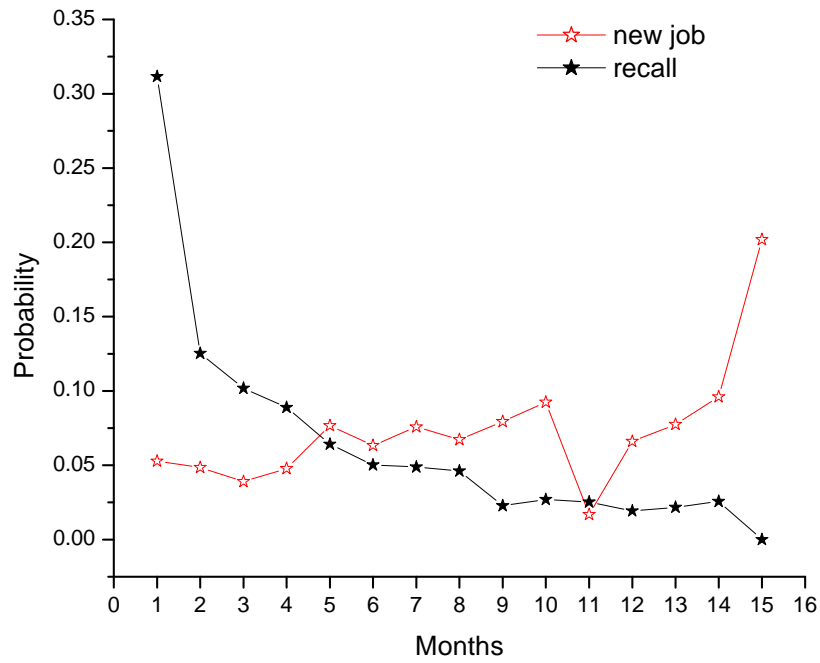


Figure 1.3 Sample monthly hazard rates for re-employment, dual-risk model.

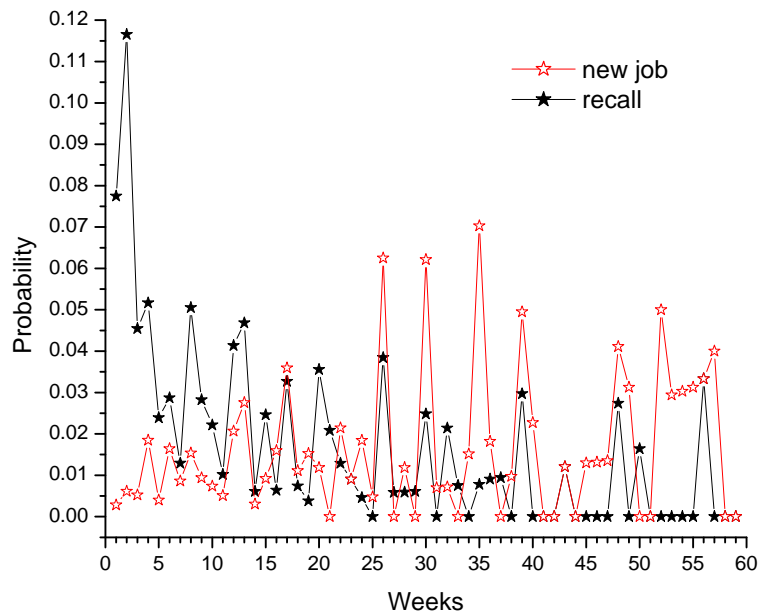


Figure 1.4 Sample weekly hazard functions for re-employment, dual-risk model---individuals who receive UI.

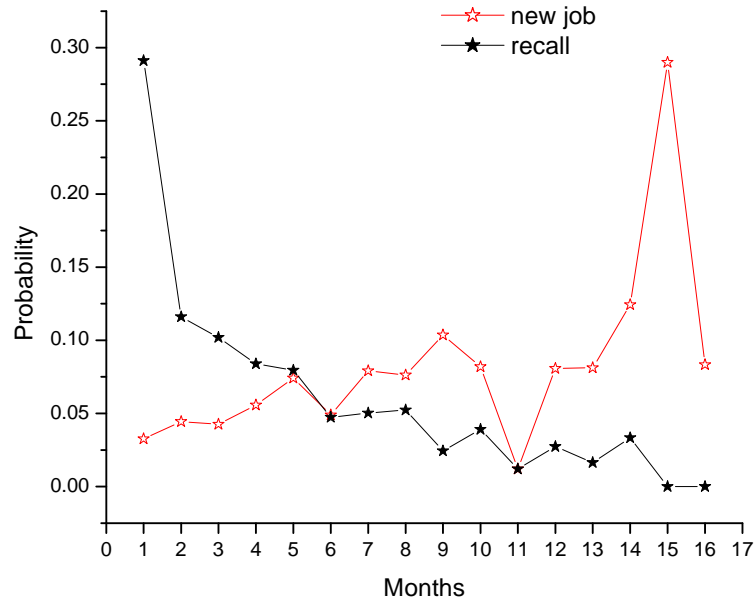


Figure 1.5 Sample monthly hazard functions for re-employment, dual-risk model---individuals who receive UI.

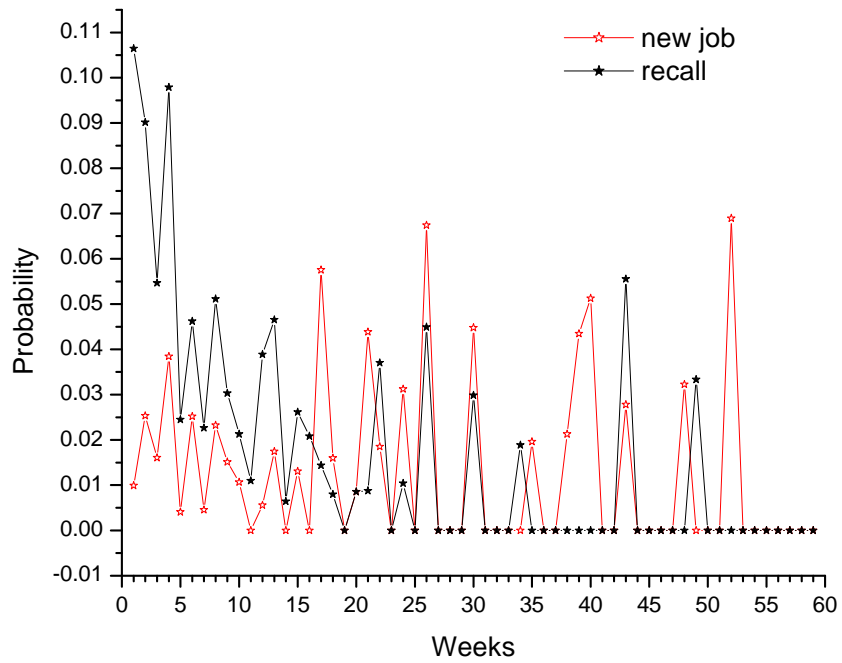


Figure 1.6 Sample weekly hazard functions for re-employment, dual-risk model---individuals who do not receive UI.

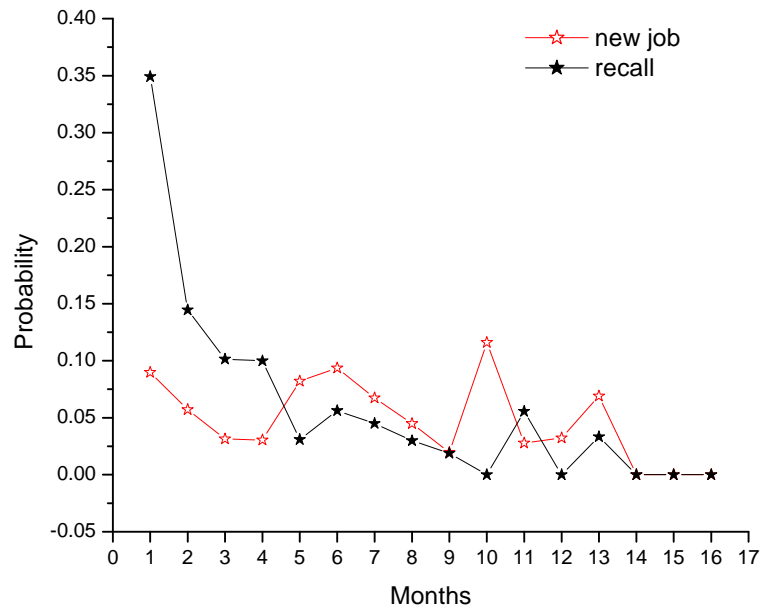


Figure 1.7 Sample monthly hazard functions for re-employment, dual-risk model---individuals who do not receive UI.

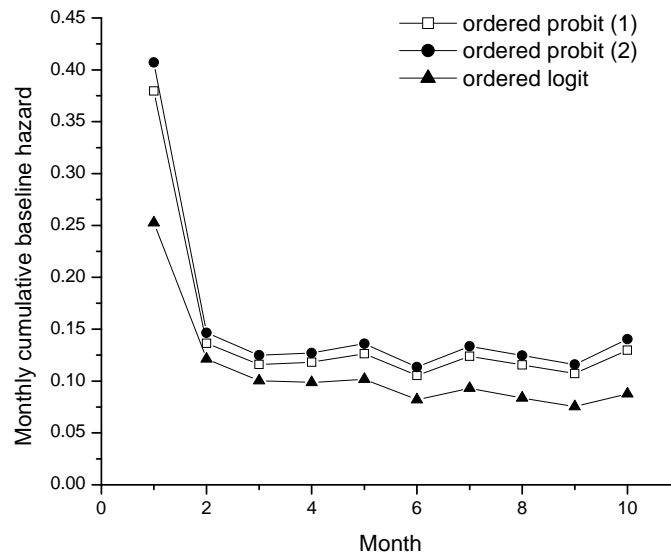


Figure 1.8 Estimated monthly cumulative baseline hazard in single risk for various models. See Table 1.4 for differences on model specifications.

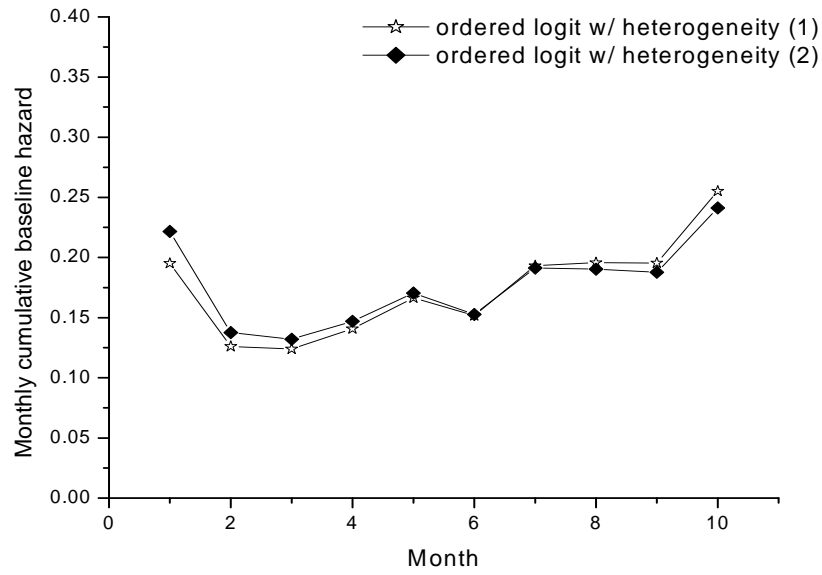


Figure 1.9 Estimated monthly cumulative baseline hazard in single risk for various models. See Table 1.6 for differences on model specifications.

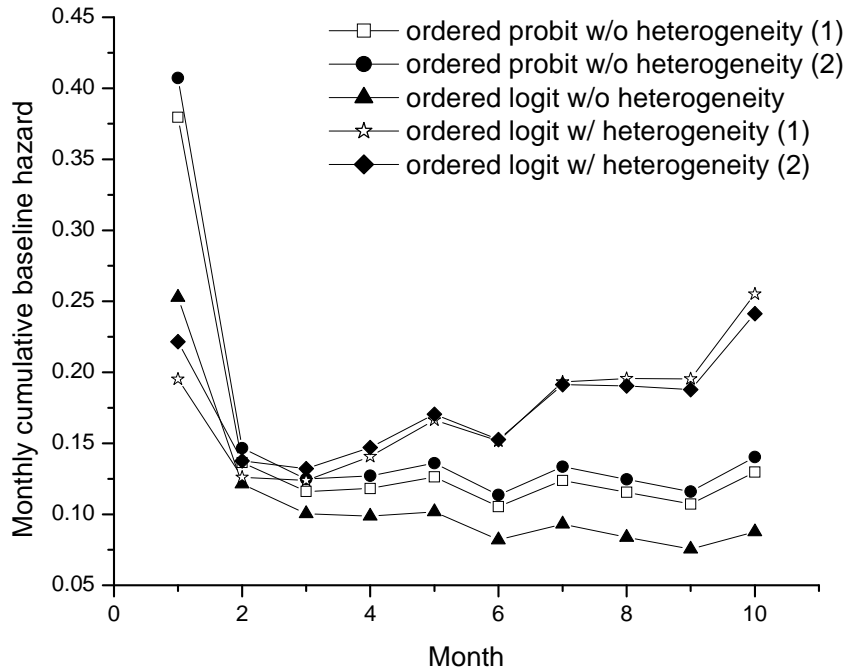


Figure 1.10 Estimated monthly cumulative baseline hazard in single risk for various models. See Tables 1.4 and 1.6 for differences on model specifications.

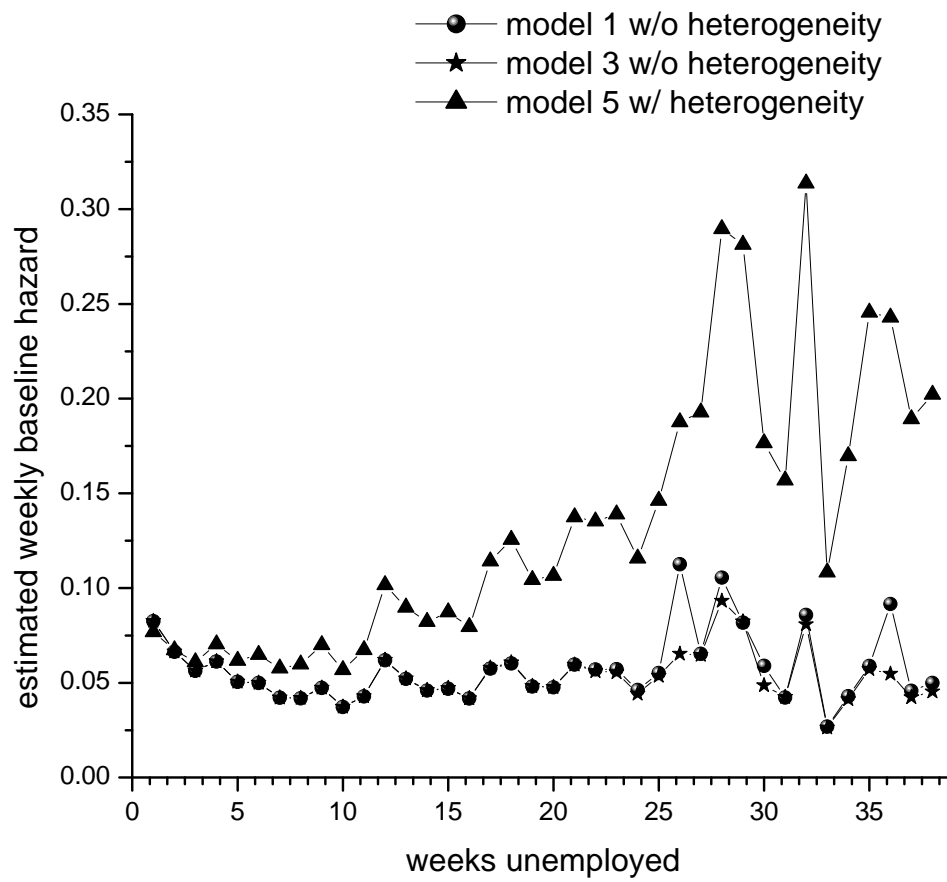


Figure 1.11 Estimated weekly baseline hazard under different model specifications. Data for this figure comes from Table VII in Meyer (1990).

Appendix 1. Summary of Gauss Maxlik program for the ordered probit duration model without heterogeneity

Step I. Generate log-likelihood function for failure data

1. First, we select observations with observed failure from the whole sample 1114 observations. Observations that have durations equal or longer than 41 weeks are all treated as censored no matter their true endings are censored or not. Therefore, in failure data the maximum spell length is 40 weeks. If one observation, here means unemployment duration, ends with going back to former employer or finding a new job then it is considered as failure data.

2. Next, we generate indicator matrix for individual failure order, i.e., matrix indicating at which time period t person i failed. Here for simplicity, we explain the process by using a small sample with size 9. Suppose there are four failure observations (implying that the other five should be treated as censored data), person 1 fails at week 1, person 2 fails at week 2, person 3 fails at week 3, and person 4 fails at week 4. The generated indicator matrix is $A_{4 \times 4}$:

$$A_{4 \times 4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The row of the matrix represents the number of individuals in the failure dataset and the column number corresponds to the longest duration in weeks. The process is in week 1,

only person 1 fails therefore $A_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{4 \times 1}$; in week 2, person 2 fails, so $A_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}_{4 \times 1}$ and so

on. Each time (meaning for each week) there is a column vector generated until the final week, here say week 4. Merge the four columns horizontally and we have the complete indicator matrix $A_{4 \times 4}$ for the 4 observations in the failure data set.

3. Now we calculate failure probabilities for each person at each week. Suppose the model is ordered probit then in week 1 we have the column vector

$B_1 = \begin{bmatrix} \ln[\text{cdfn}(\delta_1 - X_1\beta)] \\ \vdots \\ \vdots \\ \ln[\text{cdfn}(\delta_1 - X_4\beta)] \end{bmatrix}_{4 \times 1}$; in week 2 we have a column vector

$B_2 = \begin{bmatrix} \ln[\text{cdfn}(\delta_2 - X_1\beta) - \text{cdfn}(\delta_1 - X_1\beta)] \\ \vdots \\ \vdots \\ \ln[\text{cdfn}(\delta_2 - X_4\beta) - \text{cdfn}(\delta_1 - X_4\beta)] \end{bmatrix}_{4 \times 1}$. Repeating the same procedure until

week 4 and merging B_1 to B_4 horizontally, we have the complete failure probability matrix

$B_{4 \times 4}$:

$B_{4 \times 4} = \begin{bmatrix} \ln[\text{cdfn}(\delta_1 - X_1\beta)] & \cdots & \ln[\text{cdfn}(\delta_4 - X_1\beta) - \text{cdfn}(\delta_3 - X_1\beta)] \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \ln[\text{cdfn}(\delta_1 - X_4\beta)] & \cdots & \ln[\text{cdfn}(\delta_4 - X_4\beta) - \text{cdfn}(\delta_3 - X_4\beta)] \end{bmatrix}_{4 \times 4}$.

Penalty function is added if $\delta_t > \delta_{t+1}$, $t = 1, 2, 3$, to make sure $\delta_1 \leq \delta_2 \leq \delta_3 \leq \delta_4$. Note that the values of δ come from Maxlik output in each iteration. Starting values of δ are provided as one of the inputs required by the Maxlik procedure.

4. Next we calculate the log-likelihood vector for failure data. To do so, first multiply matrix $A_{4 \times 4}$ with $B_{4 \times 4}$ element-by-element. Now for each observation we have its probability of failure at the time the failure occurs and zero in other time periods. Let $C_{ij} = A_{ij} \times B_{ij}$, $i=j=1, 2, 3, 4$, and do summation over the rows in matrix C , we have a column vector $D_{4 \times 1}$:

$$D_{4 \times 1} = \begin{bmatrix} \ln[\text{cdfn}(\delta_1 - X_1\beta)] \\ \ln[\text{cdfn}(\delta_2 - X_2\beta)] - \ln(\text{cdfn}(\delta_1 - X_2\beta)) \\ \ln[\text{cdfn}(\delta_3 - X_3\beta)] - \ln(\text{cdfn}(\delta_2 - X_3\beta)) \\ \ln[\text{cdfn}(\delta_4 - X_4\beta)] - \ln(\text{cdfn}(\delta_3 - X_4\beta)) \end{bmatrix}.$$

This vector is ready to serve as the input argument “vector of log-likelihood” in the Maxlik procedure once it is merged with the corresponding vector for censored data vertically.

Step II. Generate log-likelihood function for censored data

1. First, we select observations that are censored. In our explanatory sample here, if we want to estimate 4 weekly δ , then all observations lasting longer than 4 weeks are treated as censored at week 4.

2. Suppose persons 5-8 are censored accordingly in weeks 1-4, and observation 9 fails in week 6, then the indicator matrix for censored data is

$$E_{5 \times 4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$E_{5 \times 4}$ is generated with ways similar to those in failure data.

3. Probabilities of surviving longer than the censored time for person i , $i=5, 6, 7, 8, 9$, are generated in matrix $F_{5 \times 4}$:

$$F_{5 \times 4} = \begin{bmatrix} \ln[1 - \text{cdfn}(\delta_1 - X_5\beta)] & \cdots & \ln[1 - \text{cdfn}(\delta_4 - X_5\beta)] \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \ln[1 - \text{cdfn}(\delta_1 - X_9\beta)] & \cdots & \ln[1 - \text{cdfn}(\delta_4 - X_9\beta)] \end{bmatrix}.$$

Again, penalty function is added if $\delta_t > \delta_{t+1}$, $t = 1, 2, 3$, to make sure $\delta_1 \leq \delta_2 \leq \delta_3 \leq \delta_4$.

4. Now multiply matrix $E_{5 \times 4}$ with $F_{5 \times 4}$ element by element and follow the other steps in step I-4, we have a column vector

$$G_{5 \times 1} = \begin{bmatrix} \ln[1 - \text{cdfn}(\delta_1 - X_5\beta)] \\ \ln[1 - \text{cdfn}(\delta_2 - X_6\beta)] \\ \ln[1 - \text{cdfn}(\delta_3 - X_7\beta)] \\ \ln[1 - \text{cdfn}(\delta_4 - X_8\beta)] \\ \ln[1 - \text{cdfn}(\delta_4 - X_9\beta)] \end{bmatrix}.$$

Add $G_{5 \times 1}$ vertically on $D_{4 \times 1}$ and this is the return of the “log-likelihood procedure” that is one of the mandatory inputs in the Maxlik.

Step III. Call the Maxlik procedure and do the estimation

Next, call the Maxlik procedure, load the dataset, choose variables that one wants to use in the Maxlik procedure from the dataset, and set parameters’ starting values. The Maxlik will do the rest of the work and return coefficient estimates, the mean log-likelihood value, first-order gradients and second-order gradients evaluated at the estimated parameters, and a return code indicating whether the calculation converges.

Chapter 2

Semiparametric Competing Risks Proportional Hazard Model and Unemployment Duration

2.1 Introduction

In the previous chapter we consider different model specifications for single risk case. We have not distinguished between different types or modes of failure though we do allow for the possibility of random censoring. Here censoringship means withdrawal from study for some reason independent of the failure process, such as preplanned termination of unemployment duration at the interview date. We consider in this chapter the different, although related, matter of analysis when there is a single lifetime for each individual, but failure may be of various modes.

The failure modes often refer to causes of failure, for example, an unemployment spell may end with finding a new job (briefly, new job) or returning to one's former employer (briefly, recall); a marriage may end due to divorce, death of one or even both partners. In such cases the term "competing risks" is usually used. A competing risks model is a duration model where the observed duration is the minimum of k durations. Studies of multivariate failure time problems, in which situation there are two or more observed failure times on each unit, are beyond the scope of this chapter.

The ideas in competing risks have a long history and can go back to the study of the potential consequences of smallpox inoculation in Daniel Bernoulli (1760). Early

fundamental works and reviews in this literature include Gail (1975), Seal (1977), David and Moeschberger (1978), Cox (1959, 1962), Tsiatis (1975), Peterson (1976) and so on.

Significant applications of the competing risks model are to studies of human mortality, actuarial science and demography. For instance, Honoré and Lleras-Muney (2006) derived bounds for aspects of the underlying dependent competing risks distributions and then estimated changes in cardiovascular and cancer mortality since 1970; Berrington and Diamond (2000) studied age at first-time marriage or cohabitation; Cornfield (1957) considered applications of actuarial techniques in medical contexts. In economics, such applications include: Han and Hausman (1990), Katz and Meyer (1990) studied the probability of leaving unemployment through new job or recall; Flinn and Heckman (1982) examined the duration of unemployment where one can resume work either by leaving the labor force or by finding a new job; and Deng, Quigley and Van Order (2000) investigated mortgage termination via prepayment and default.

One approach to dealing with competing risks models is to assume that the risks are independent conditional on a set of observed variables. This is equivalently to study the distribution of failure time for, say, type 1 failure, and other types of failure have been eliminated. Under such assumption, estimation of competing risks duration models then amounts to estimation of duration models with random censoring.

Unfortunately, the assumption of independence may be reasonable when the different T_k refer to separate subsystems; in other cases, as for example in connection with human survival, the independence is suspect and even implausible. Estimates of the cause-specific hazard rates and of the covariate effects on those hazards may be inconsistent, when the underlying risks are indeed dependent but independence is imposed

(e.g., Honoré and Lleras-Muney 2006). Therefore, very cautious interpretation has to be given since the estimation results in this instance could be misleading.

Bearing in mind that different ways of exiting unemployment might imply different duration behaviors, which cannot be identified if only one convoluted risk is observed, we apply the competing risks model to study unemployment duration in the current chapter. Since we are not sure whether different T_k belong to separate subsystems, no dependence restriction is imposed on the joint distribution of T_k .

For simplicity, let k take values 1 and 2 corresponding to just two distinct failure types, say, type 1 and type 2. In our studies, one could define “new job” as type 1 failure, “recall” as type 2 failure, or vice versa. We also suppose that observations may be censored in the sense outlined above.

The full joint distribution of T_1 and T_2 given a set of covariates X becomes of interest in the present chapter, since we do not single out one variable as a response hence correlation will be considered. Knowledge of such joint distribution is of considerable importance and it allows us to answer policy questions that could not be answered based only on the distribution of (T, K) given X (see, Honoré and Lleras-Muney 2006).

We find that the time patterns for recall risk and new job risk are significantly different, which is not observable if they are treated as one single risk. Further, none of the duration dependence is monotonic increasing or decreasing, implying that parametric functions may not be suitable to describe such time behavior. Unemployment insurance does induce longer unemployment spell; and demographic variables also play important and intriguing roles in both risks. The correlation between recall duration and new job duration is shown to be positive.

This chapter is organized as follows. A discussion of the identification conditions of competing risks models is presented in section 2.2. Section 2.3 contains semiparametric competing risks duration model specification. The estimation results are reported in section 2.4. Section 2.5 is concluding remarks.

2.2 Identification

One main concern when analyzing competing risk data is the model identifiability. The identification problems arise because only the first occurred failure on any subject is observed, this effectively censoring the remaining time to another failure. In unemployment duration analysis this corresponds to the ways people get employed again through, so that the entire spell ends as soon as any possible way happens. More specifically, consider there are $K = 2$ possible ways of resuming employment, and let T_k denote the lifetime or failure time of risk K ($K = 1, 2$). The duration ends when the first risk happens, so its lifetime is $T = \min(T_1, T_2)$, where the failure mode K indicates which risk happens to each duration so that $T = T_k$ if $K = k$.

This framework seems promising since it appears that we can consider bivariate models $F(t_1, t_2)$ for the joint distribution of T_1 and T_2 . Unfortunately, this is entirely notional because all that is ever observed or realized is the pair (T, K) . Cox (1959, 1962) and Tsiatis (1975) claimed the nontestability of dependence in the competing risks model. Their key result is that, for any dependent distribution of (T_1, T_2) , one can find an observationally equivalent independent distribution. That is, for every distribution having nonindependent T_k , there exists a unique pair of univariate distributions with independent T_k that gives the same distribution of (T, K) (see Lawless 2003, Honoré and Lleras-Muney 2006).

Due to the fact that data of the form (T, K) do not allow one to discriminate between an independent competing risks model and an infinitude of dependent model, many authors have either assumed independence for the competing risks model, or relaxed this assumption but considered parametric models for competing risks data. Some references are listed here: Katz (1986) assumed that the time patterns for different risks are independent from each other so he specified two standard duration models; besides, he utilized Weibull baseline hazard model. Diamond and Hausman (1984) allowed for dependence by using bivariate log-normal distribution for the competing risks model. Katz and Meyer (1990) specified a semiparametric independent competing risks model.

However, as we mention earlier, if some underlying factors that cause the failure of an event are common to both risk types, assuming independence on the competing risks model would be inappropriate and the estimation results could be biased. On the other hand, the discussions in Cox (1962) and Tsiatis (1975) are based on the case when there is no covariate present in the model. Nevertheless, their “nonidentifiable dependence” conclusion can be extended to the situation with a set of covariates X .

In order to identify some features of the conditional distribution of (T_1, T_2) given the explanatory variables X , one has to impose restrictions on these conditional distributions. One extreme approach is to assume fully parametric form on the model. However, estimation results may be entirely driven by the functional form assumptions. Therefore, it will be of value to study identifiability of dependent, semiparametric competing risks models. In other words, additional assumptions are necessary to be made on identification so that one can answer questions that require knowledge of the joint distribution of T_1 and T_2 .

Following the approach of, for example, Han and Hausman (1990) and Sueyoshi (1992), we can estimate the semiparametric competing risks model relatively easily without having to assume independence. In the case that the explanatory variables are different for each risk, i.e., X_1 and X_2 do not have identical variables, the model is identified because an independent distribution that is observationally equivalent cannot be found. Unfortunately, often X_1 and X_2 are identical in empirical analyses. Han and Hausman (1988) demonstrated that even under such instance, as long as at least two covariates are continuous and some other regularity conditions are satisfied, the identification of the bivariate competing risks model is possible.

The important strategy is to assume linearity on the dynamics of the baseline hazards over the discrete time interval. This identifying assumption is similar in spirit to the bounds technique discussed in Cox and Oakes (1984). Basically, they pointed out that if one wants to determine the marginal distribution of observed failure times, the best one can hope for is bounds. Two extreme cases of the bounds are treating the unobserved failure time as being either slightly greater than the observed failure time, or being effectively infinite. Nevertheless, they noted that the bounds would often be too wide to draw conclusions. Peterson (1976) studied and formulated the bounds on the underlying distribution function. However, Honoré and Lleras-Muney (2006) found that generally the nonparametric bounds are so wide that they may have none practical value.

In our present case, the assumption of linearity on intra-interval hazard shapes is just identifying and the probabilities can be calculated exactly. But there are a couple of points to keep in mind. First, imposing the identification assumption is equivalent to placing parametric restrictions on the shapes of hazards within intervals. Hence the

baseline hazard is no longer fully nonparametric with respect to the complete time pattern. Second, the identifying assumption will be violated more easily in longer sampling intervals, such as monthly basis but not weekly basis. So the semiparametric estimator will be sensitive to the length of time intervals (e.g., Sueyoshi 1992).

Recent articles studying identifiability of semiparametric competing risks model are: Heckman and Honoré (1989), Abbring and Berg (2003), Crowder (2001) and Honoré and Lleras-Muney (2006) among others. Note that these works do not claim they have “solved” the nonidentification of competing risks model. Indeed, identification is studied under additional assumptions on the model; so do we in this chapter.

2.3 Competing risks proportional hazard model

Most of the competing risk literature define latent or conceptual failure times corresponding to each failure type. Let T_1, \dots, T_k denote such conceptual times. The actual failure time is defined to be $T = \min\{T_1, \dots, T_k\}$, and the corresponding failure type is $J = \{j | T_j \leq T_k, k = 1, \dots, k\}$. Hence, a (T, J) pair is defined for each individual. It is assumed that T is a continuous random variable and that J takes on values in the set $\{1, \dots, k\}$. Frequently, the latent failure time framework is also used to study the association between particular failure types and predetermined covariates (e.g. Kalbfleisch and Prentice 1980, David and Moeschberger 1978).

For simplicity, consider the case when there are two competing risks ($k = 1, 2$). These two risks are exiting unemployment through either new job or recall, with T representing the length of unemployment duration before any failure occurs. Suppose we observe that person i fails in time t with risk type 1, and suppose that the joint probability

density function of t_1 and t_2 is $f(t_1, t_2)$, then the probability that the length of a duration falls between $[t-1, t)$ in type 1 is

$$P(t-1 \leq T_{i1} < t, T_{i1} < T_{i2}) = \int_{t-1}^t \int_{t_1}^{\infty} f_i(t_1, t_2) dt_2 dt_1. \quad (2.1)$$

Note that if the actual failure type is 1, we must insure that the latent failure time T_2 is greater than the observed failure time T_1 , for every point in the time interval $[t-1, t)$. Covering all the possible time points type 2 failure may occur at and integrating them out, we have Eq. (2.1) as the failure probability in type 1 given there is another possible failure type 2. Also note that since we want to examine whether these two risks are dependent from each other, we consider the joint distribution of (T_1, T_2) rather than formulate two duration models for T_1 and T_2 , respectively.

Applying the same method used to derive Eq. (1.5) in Chapter I, we know that Eq. (2.1) is equivalent to

$$P(t-1 \leq T_{i1} < t, T_{i1} < T_{i2}) = \int_{\delta_{t-1}^{1-X_i\beta_1}}^{\delta_t^{1-X_i\beta_1}} \int_{g(\varepsilon_1)}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1. \quad (2.2)$$

When there are two or more competing risks, the baseline hazard δ , the covariates X and the coefficients β all may differ across failure types. Therefore, when the observed failure is type 1, the baseline hazard should be specified accordingly as δ^1 , and so as X_1 and β_1 . Nevertheless, we do not have to distinguish X_1 from X_2 in this paper based on the identification conditions proposed in Han and Hausman (1988). In other words, X_1 and X_2 can be identical.

Now the joint distribution of T_1 and T_2 is represented as the joint distribution of ε_1 and ε_2 , with $\varepsilon = \log \int_0^t \lambda(\tau) d\tau$, ε_1 and ε_2 are both distributed as standard normal, and $\text{corr}(\varepsilon_1, \varepsilon_2) = \rho$. The distributions of ε_1 and ε_2 do not come directly from the duration model

specification considered in Chapter I. We use standard normal distribution, but not extreme value distribution, for both ε_1 and ε_2 because the former one seems to be a very good approximation to the latter one based on our earlier estimation results. Because what we study presently are binomial outcomes, the similarity still exists. If there are numerous outcomes, the estimation results using probit or logit models could be very different (see Han and Hausman 1990). Hence we call (2.2) a bivariate ordered probit competing risk model and the orders/categories are defined by the failure times, of which the δ s are functions.

The $g(\varepsilon_1)$ in (2.2) is a function to insure that the required relationship between the potential and the observed failure times hold everywhere in the discrete time intervals. Now comes the problem --- what kind of function form $g(\varepsilon_1)$ should take? Finding answers to this question is just finding the identification conditions for the competing risks model. Without loss of generality, we assume that the baseline hazards, δ^1 and δ^2 , are linear in each time interval, although the change rate may be different over risks and intervals.

Suppose that type 1 failure occurs at some time point $t_1^* \in [t-1, t)$. We then have

$$\frac{\delta_t^1 - \delta_{t-1}^1}{t - (t-1)} = \frac{\delta_{t_1^*}^1 - \delta_{t-1}^1}{t_1^* - (t-1)}, \quad (2.3)$$

based on the linearity assumption on δ^1 .

Next, suppose that type 2 failure may happen at another time point $t_2^* \in [t-1, t)$.

Similarly, we have

$$\frac{\delta_t^2 - \delta_{t-1}^2}{t - (t-1)} = \frac{\delta_{t_2^*}^2 - \delta_{t-1}^2}{t_2^* - (t-1)}. \quad (2.4)$$

Now if failure type 1 rather than type 2 is observed, it implies that $t_2^* \geq t_1^*$. Hence from (2.3) and (2.4), we know

$$\frac{\delta_t^2 - \delta_{t-1}^2}{\delta_t^1 - \delta_{t-1}^1} \geq \frac{\delta_t^1 - \delta_{t-1}^1}{\delta_t^1 - \delta_{t-1}^1}. \quad (2.5)$$

After some manipulation on (2.5) and replace δ with $X_i\beta + \varepsilon$, the function $g(\varepsilon_1)$ is defined as

$$\varepsilon_2^* \geq \frac{\delta_t^2 - \delta_{t-1}^2}{\delta_t^1 - \delta_{t-1}^1} (\varepsilon_1^* + X_i\beta_1 - \delta_{t-1}^1) + \delta_{t-1}^2 - X_i\beta_2 = g(\varepsilon_1^*). \quad (2.6)$$

Inequality (2.6) gives the support of ε_2 when the observed failure is type 1. Note that ε_2^* is a function of ε_1^* . That is, we make sure that for any possible value of ε_1 in $(\delta_{t-1}^1 - X_i\beta_1, \delta_t^1 - X_i\beta_1)$, its corresponding value of t is less than that of ε_2 .

Hence, the probability of type 1 failure in period t for person i is

$$\int_{\delta_{t-1}^1 - X_i\beta_1}^{\delta_t^1 - X_i\beta_1} \int_{\delta_{t-1}^2 - X_i\beta_2 + h_1}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1, \quad (2.7)$$

where $h_1 = \frac{\delta_t^2 - \delta_{t-1}^2}{\delta_t^1 - \delta_{t-1}^1} (\varepsilon_1 - \delta_{t-1}^1 + X_i\beta_1)$.

Let $\lambda_1 = \frac{\delta_t^2 - \delta_{t-1}^2}{\delta_t^1 - \delta_{t-1}^1}$. Note that when $t = 1$, $\lambda_1 = 1$; and for $t = 2, \dots, T$, $\lambda_1 = \frac{\delta_t^2 - \delta_{t-1}^2}{\delta_t^1 - \delta_{t-1}^1}$.

Accordingly, the probability of type 2 failure in period t for person i is

$$\int_{\delta_{t-1}^2 - X_i\beta_2}^{\delta_t^2 - X_i\beta_2} \int_{\delta_{t-1}^1 - X_i\beta_1 + h_2}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2, \quad (2.8)$$

where $h_2 = \frac{\delta_t^1 - \delta_{t-1}^1}{\delta_t^2 - \delta_{t-1}^2} (\varepsilon_2 - \delta_{t-1}^2 + X_i\beta_2)$.

Let $\lambda_2 = \frac{1}{\lambda_1}$. Again, if $t = 1$, $\lambda_2 = 1$; $\lambda_2 = \frac{\delta_t^1 - \delta_{t-1}^1}{\delta_t^2 - \delta_{t-1}^2}$ for $t = 2, \dots, T$.

The linearity assumption provides advantages in computational simplicity. The spirit here is to use such assumption/function to tighten the bounds of the parameters of interest. Without assuming linearity, one would not be able to identify the competing risks model. Though strictly speaking, the estimated baseline hazard is no longer fully nonparametric. Nevertheless, the unrestricted pattern of dependence between two different risks is allowed through the joint density specification $f(\varepsilon_1, \varepsilon_2)$.

If an observation is censored in period t , then it means that neither type 1 nor type 2 failure occurs before the duration is censored. Alternatively, we could say that failure times of both risk modes are greater than t . Therefore, the probability of an observation censored at time t is

$$\int_t^\infty \int_t^\infty f(t_1, t_2) dt_1 dt_2, \quad (2.9)$$

or equivalently

$$\int_t^\infty \int_t^\infty f(t_1, t_2) dt_2 dt_1.$$

Remember that when a subject is censored in $[t-1, t)$, we consider the censoring time is at the end of an interval.

We rewrite (2.9) in a way similar to that for (2.1) as

$$\int_{\delta_t^2 - X_t \beta_2}^\infty \int_{\delta_t^1 - X_t \beta_1}^\infty f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2, \quad (2.10)$$

or equivalently

$$\int_{\delta_t^1 - X_t \beta_1}^\infty \int_{\delta_t^2 - X_t \beta_2}^\infty f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1.$$

We do not consider heterogeneity in the competing risks model. This does not mean that the individuals in our sample form a homogeneous population. Heterogeneity is suppressed tentatively due to computation complications. Another possible way to deal

with heterogeneity is to replace ε with $\varepsilon + V$, where V summarizes the unobserved regressors (Sueyoshi 1992). However, Sueyoshi found that the biases associated with assuming bivariate normality on $(\varepsilon_1 + V_1, \varepsilon_2 + V_2)$ are quite large, and so future work on correct specification of the bivariate distribution is necessary.

Equations (2.7), (2.8) and (2.10) may be used to formulate a likelihood function which corresponds to the competing risks model specification.

$$\begin{aligned}
L(\delta^1, \delta^2, \beta_1, \beta_2, \rho) = & \prod_{i=1}^N \prod_{t=1}^T \left\{ \left[\int_{\delta_{t-1}^1 - X_i \beta_1}^{\delta_t^1 - X_i \beta_1} \int_{\delta_{t-1}^2 - X_i \beta_2 + h_1}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1 \right]^{1-a_i} \right. \\
& \times \left. \left[\int_{\delta_{t-1}^2 - X_i \beta_2}^{\delta_t^2 - X_i \beta_2} \int_{\delta_{t-1}^1 - X_i \beta_1 + h_2}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \right]^{a_i} \right\}^{1-c_i} \\
& \times \left[\int_{\delta_t^2 - X_i \beta_2}^{\infty} \int_{\delta_t^1 - X_i \beta_1}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \right]^{c_i} \}^{y_{it}},
\end{aligned} \tag{2.11}$$

where

$$\begin{aligned}
\delta_0^1 &= -\infty < \delta_1^1 \leq \delta_2^1 \leq \dots \leq \delta_T^1 < \delta_{T+1}^1 = \infty, \\
\delta_0^2 &= -\infty < \delta_1^2 \leq \delta_2^2 \leq \dots \leq \delta_T^2 < \delta_{T+1}^2 = \infty,
\end{aligned}$$

and h_1 and h_2 are defined as above. Here y_{it} is an indicator variable that equals one if individual i fails (in either type 1 or type 2) or is censored in period t and zero otherwise; c_i is another indicator variable that is equal to one if subject i is censored and zero if failure with type 1 or type 2; and a_i is a third indicator variable that equals one if the failure type is 2 and zero if it is type 1 failure. So this is the likelihood function associated with observations that are obtained on n independent individuals in the form $(T(1), K(1); X(1)), \dots, (T(n), K(n); X(n))$, where $T(i)$ and $K(i)$ represent individual i 's duration length and failure type respectively, and $X(i)$ is a vector of explanatory variables for the i th individual.

The log-likelihood function is

$$\begin{aligned}
& \log L(\delta^1, \delta^2, \beta_1, \beta_2, \rho) \\
&= \sum_{i=1}^N \sum_{t=1}^T y_{it} \times \left\{ (1 - c_i) \times \left\{ (1 - a_i) \times \log \left[\int_{\delta_{t-1}^1 - X_i \beta_1}^{\delta_i^1 - X_i \beta_1} \int_{\delta_{t-1}^2 - X_i \beta_2 + h_i}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1 \right] \right. \right. \\
& \quad \left. \left. + a_i \times \log \left[\int_{\delta_{t-1}^2 - X_i \beta_2}^{\delta_i^2 - X_i \beta_2} \int_{\delta_{t-1}^1 - X_i \beta_1 + h_2}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \right] \right\} + c_i \times \log \left[\int_{\delta_i^2 - X_i \beta_2}^{\infty} \int_{\delta_i^1 - X_i \beta_1}^{\infty} f(\varepsilon_1, \varepsilon_2) d\varepsilon_1 d\varepsilon_2 \right] \right\}. \tag{2.12}
\end{aligned}$$

The form (2.12) arises because each individual in the sample contributes to the likelihood a factor, no matter whether a specific failure type is observed or not. That means censored observations, which is usually an important feature in duration study, are also considered explicitly.

According to Han and Hausman (1986), large sample inferences for the maximum likelihood estimators $\hat{B} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\delta}^1, \hat{\delta}^2, \hat{\rho})$ follow from

$$\sqrt{N}(\hat{B} - B_0) \xrightarrow{A} N\left(0, \left[\lim_{N \rightarrow \infty} \frac{1}{N} \frac{\partial^2 \log L}{\partial B \partial B'} \right]^{-1} \right), \tag{2.13}$$

where ρ is a finite parameter that characterizes the correlation of the error terms ε_1 and ε_2 .

2.4 Application and empirical results

We employ the same data set, which is described in Section 1.3 of Chapter I, to examine the effect of measurable differences on the unemployment duration under different hazards. The log-likelihood function used to do the maximization is equation (2.12) in previous section. The joint distribution of $(\varepsilon_1, \varepsilon_2)$ is assumed to be bivariate ordered probit not only because, based on our estimates in Chapter I, the ordered probit provides a good approximation to the ordered logit; but it permits unrestricted correlation between ε_1 and ε_2 (e.g., Han and Hausman 1990). However, one should keep in mind that if the observed outcomes are multinomial rather than binomial, such assumption will be

problematic since in these situations probit and logit specifications can yield very different estimates.

The results of the computation are displayed in Table 2.1 and 2.2. These two tables summarize the parameter estimates for the same employer hazard (recall) and the different employer hazard (new job). For each risk, we allow for 40 parameters to describe the duration pattern nonparametrically. The correlation coefficient, ρ , is also estimated simultaneously without restriction imposed. Figures 2.1 and 2.2 plot estimated weekly baseline hazard and monthly cumulative baseline hazard, respectively.

The estimates in Table 2.1 show some consistencies. The coefficients associated with the unemployment compensation, being white or non-white, and with the marital status, all take on the same sign across risks. Obviously the unemployment insurance has an important effect for both recalls and new jobs. And as we have expected, the estimated elasticity of unemployment duration with respect to receiving compensation is higher in the different employer hazard than that in the same employer hazard. Although unemployment insurance will induce longer time periods without having jobs, the estimates in Table 2.1 confirm the hypothesis that recalls could be prearranged by the employers hence they are less dependent on the compensation received. Meanwhile, if one is not white, this will put him at a disadvantage no matter he is seeking a new job or trying to return to his former employer. On the other hand, being married will strengthen the workers' motivation to resume employment; therefore as expected the coefficients are negative in both risks.

The age coefficient and sex coefficient are negative for the same employer and positive for the different employer, while the education coefficient is positive for the same

employer and negative for the different employer. These opposite signs may be striking at first glance. However, they might have revealed some informative inferences that are not possible to show if both risks just treated as one way to leave unemployment. Elder workers are supposed to be more experienced and master the techniques better. For such group, the employers are more likely to temporarily dismiss them because of the lack of new orders and other reasons that cause business shrinkage. On the contrary, getting older may not be a good signal for the new employer. Higher education would put the workers in an awkward position if they want to be recalled. Usually higher education implies higher salary, higher benefits and higher positions. But when businesses turn down, hiring these persons means more costs and sometimes is even unnecessary to the employers. Hence, though stronger education background will increase one's opportunity on finding a new job, it may potentially do more harm than good on the chance of getting back to his/her former employer. Additionally, females are more possible to be recalled than males; but males have much more advantages in finding new jobs.

The one variable that stands out as having the greatest impact on the recall hazard rate is marital status. The negative coefficient implies that if one is married, his possibility to be recalled will be 34 percent higher than those who are single. This is consistent with what we have noted earlier: the recall is negotiable and can be prearranged. Receiving unemployment compensation is a good strategy to keep valuable employees; but married workers will resume work quickly possibly due to the fact that there are dependents in their families. The marital status plays the most important role in making recall decision, based on our study.

The corresponding variable in the new job case is race. Whether or not a person is married now has a much smaller impact on the different employer hazard. Although unemployment insurance significantly extends the unemployment duration, its effect is not as strong as race: the positive coefficient implies that if one is non-white, he is 36 percent less likely to find a new job. This could express an informational aspect with respect to hiring. New job opportunities become available to market independent of the workers' marital status, but emphasize more on one's gender, education background, age, and especially race. For non-whites, it is difficult for them to resume their old works, and the situation will be tougher if they plan to seek new employers.

The effects of both the unemployment compensation and the demographic variables in the competing risks model have the same signs as those in Han and Hausman (1990), except 'sex', which they did not include in their competing risks model. The magnitudes of these effects are also very close, in particular those related to insurance, race, education and age. The marital status in our study has a much larger effect in recall hazard than it does in new job hazard. One explanation is that married workers are given an impetus to return to work sooner, but this can be realized much more easily in the same employer case because the employers will take this factor into serious consideration when they lay people off; while it is almost impossible for new employers to hire one simply because he is married.

The estimated correlation coefficient, ρ , is about 0.30. This means ε_1 and ε_2 are positively correlated. Based on ε 's definition in Chapter I, implicitly it indicates positive correlation between T_1 and T_2 . One reasonable explanation is that the longer one has to wait before he is recalled, the more time he is supposed to spend on finding a new job, or vice

versa. Our estimation result confirms at some level the importance to consider risks dependence. First, there are several observed risk factors that are common to both recall and new job risks. Second, there is evidence of unobserved differences across individuals with respect to their possibility of resuming work (see Chapter I). For example, if circumstances/nature of one's job changed, he might not be able to return to his former position but to get a new job instead. Hence, the decision for a new job may be correlated with the chance of being recalled, especially when the recall waiting takes longer and longer time period.

Using the estimated baseline hazard in Table 2.2, we can plot weekly baseline hazard and monthly cumulative baseline hazard in Figure 2.1 and 2.2, respectively. In Figure 2.1, the same employer baseline hazard starts at a higher value and ends at a lower value than that for the different employer, a decrease about 87.5 percent, implying a higher probability of staying with the employer at the beginning of unemployment spell but a much lower chance of eventual returning. This result could be indicative of taking advantage of unemployment insurance. Indeed, after week 30, the typical length that one can be covered by compensation, there is little chance to be recalled.

In contrast, the baseline hazard for different employer starts at a lower value and ends at a higher value. There are jumps perceived at weeks 26, 30, 35 and 39. This implies that individuals may arrange the start of a new job to be the benefits exhaustion points, since the new job hazard increases significantly after week 26 compared to those in weeks 1-25. Another possibility is that while time goes on, things change and those who initially expect to return to former employers have to find new jobs, though they do receive benefits during unemployment.

Figure 2.2 also displays that the structure of the monthly baseline hazard rates for recall and new jobs is significantly different. Basically, the recall hazard decreases over time, while the new job hazard decreases first and then increases. Such estimated basic trends are close to the patterns shown in Figure 1.3 of Chapter I, which are calculated using the Kaplan-Meier estimator. At somewhere between 4 months and 5 months, the new job hazard catches the recall hazard and becomes much higher than the latter one during the remaining unemployment spell. This is similar to what we have noticed from the correlation coefficient estimate: the longer the recall process takes, the less likely one can return to same employer, so the more possible that he would have to find a new job.

On the other hand, an individual whose purpose is to resume work from different employer may face disadvantages in the job market, such as non-white, asking for higher pay and so on. These “bad” characters may induce longer job searching period. That is why we see in Figure 2.1 and 2.2 that new job hazard becomes higher while time passes, but neither recall nor new job risk shows monotonic increasing or decreasing pattern, especially the weekly baseline hazards.

Therefore, parametric functional assumptions are not recommended because they can give inconsistent estimates of the hazard rates and the covariate effects on these rates. If the covariance matrix for the baseline hazard is available (though our calculation does show convergence), then we will be able to do some formal test to determine whether any parametric specification (e.g., Weibull) is consistent with our nonparametric baseline estimates (for example, Han and Hausman 1990)¹.

¹ Han and Hausman did not report the baseline hazard estimates and their corresponding standard errors. Instead, they presented the estimated cumulative distribution functions calculated on the baseline estimates.

Thus, we conclude that both the unemployment insurance and the demographic variables have important effects on the hazard of resuming work. Meanwhile, the estimated effects and the duration patterns are quite different between the same employer and the different employer hazards. By estimating a competing risks model without restrictions on the dependence, we are able to retrieve meaningful information from the data, which cannot be told if only single risk considered.

2.5 Conclusion remarks and future directions

This chapter studies unemployment duration using a sample drawn from the PSID and described in Section 1.3 of Chapter I. The method of analysis used is a semiparametric dependent competing risks proportional hazard model. The model is semiparametric in the sense that the log-cumulative baseline hazard is left unparameterized while the effect of the explanatory covariates takes a particular functional form. The two risks specified are: 1) returning to work for the same employer, and 2) returning to work for a different employer. Furthermore, we allow for unrestricted correlation between the competing risks. This is realized by assuming linearity on the baseline hazard at each time interval. Monte Carlo experiments performed in Han and Hausman confirmed that the linearity design is accurate enough for estimation when we use discrete interval data to approximate an underlying continuous baseline hazard process. Meanwhile, as long as at least one explanatory variable in our study is continuous, we are able to identify the competing risks model even if the covariates in both risks are identical (Han and Hausman 1990). Our model guarantees that the unobserved failure occurs later than the observed failure at any possible time point, and censored observations are accommodated explicitly in the model specification.

It is shown that unemployment benefits do induce longer durations unemployed in both risks. Further, the variables that appear to have the greatest impacts on recall hazard and new job hazard rates are marital status and race, respectively. Other variables considered (age, sex, education) indicate that they may have opposite effects on different hazard rates. For those variables that have same signs in both risks, their magnitudes of effects are not close to each other. The correlation coefficient 0.30 suggests that recall duration and new job duration have a positive relationship that may not be negligible.

We also find that there are significant differences in the hazard structure of returning to the same employer and a different employer. The probability of being recalled basically decreases over time, while chances of finding a new job decline first and eventually increase. Therefore, we conclude that it is important to consider dependent competing risks, at least when there is more than one way to resume work and we are not sure whether these ways are independent of each other (also see Honoré and Lleras-Muney 2006 for interesting findings when dependent risks are allowed). Our estimated baseline hazards suggest that nonparametric approach is preferred if one wants to capture the perceived features of duration dependence, and so to avoid biased estimates of the hazard rates and of the covariate effects.

In future work we could analyze some time interactions (such as interactions in weeks 26 and 39, two representative benefits exhaustion points) and undertake tests on a given parametric specification for the baseline hazard via a minimum chi-square framework as proposed in Han and Hausman (1990). We could also allow for unobserved heterogeneity so that the model would become more realistic. When heterogeneity is considered, the bivariate standard normal distribution may not be reasonable (e.g.,

Sueyoshi 1992) hence further work is needed to find correct specification of the combined error distribution.

Table 2.1 Competing Risks Hazard Model Estimates^a

Variable	Recall Risk	New Job Risk
Age	-0.0182	0.0140
Sex	-0.0720	0.1787
Education	0.0161	-0.1171
Race	0.2354	0.3669
UI	0.2082	0.2722
Married	-0.3439	-0.0679
ρ	0.2998	
sample size	1114	
log-likelihood	-3395.3539	

^aThe log-likelihood function for the model is equation (2.12) in section 2.3. Baseline hazard parameters are reported in Table 2.2.

Table 2.2 Competing Risks Baseline Hazard Estimates from Model Specification in Table 2.1

Week	<u>Recall Risk</u>	<u>New Job Risk</u>
	Hazard	Hazard
1	-1.826	-3.204
2	-1.346	-2.667
3	-1.200	-2.476
4	-1.024	-2.135
5	-0.970	-2.095
6	-0.895	-1.931
7	-0.863	-1.879
8	-0.761	-1.760
9	-0.708	-1.694
10	-0.669	-1.649
11	-0.651	-1.631
12	-0.581	-1.551
13	-0.505	-1.443
14	-0.495	-1.435
15	-0.457	-1.393
16	-0.441	-1.352
17	-0.401	-1.206
18	-0.390	-1.168
19	-0.386	-1.138
20	-0.348	-1.108
21	-0.324	-1.069
22	-0.296	-1.015
23	-0.287	-0.999
24	-0.279	-0.944
25	-0.279	-0.937
26	-0.223	-0.791
27	-0.218	-0.791
28	-0.212	-0.774
29	-0.207	-0.774
30	-0.172	-0.661
31	-0.172	-0.652
32	-0.153	-0.642
33	-0.146	-0.642
34	-0.140	-0.623
35	-0.132	-0.524
36	-0.125	-0.503
37	-0.117	-0.503
38	-0.117	-0.482
39	-0.091	-0.406
40	-0.091	-0.360

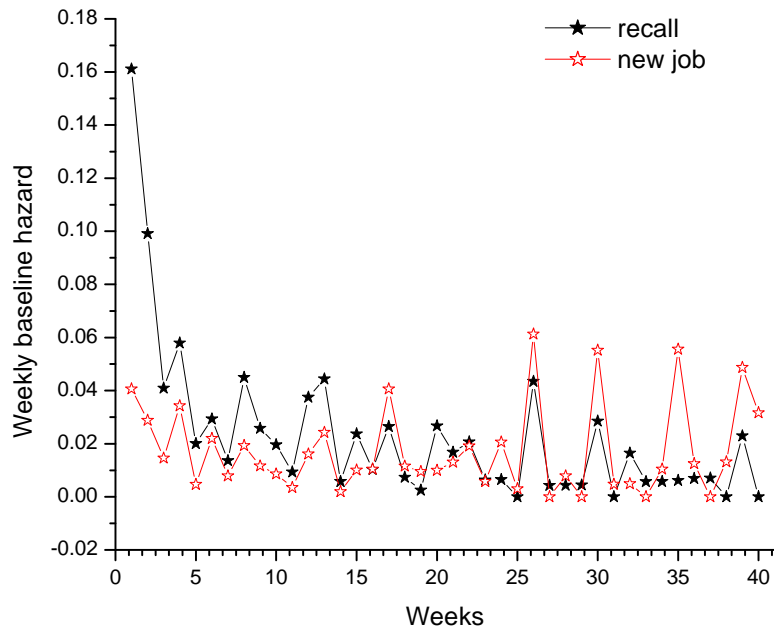


Figure 2.1 Estimated weekly baseline hazard in competing risks model; the specified risks are recall and new job.

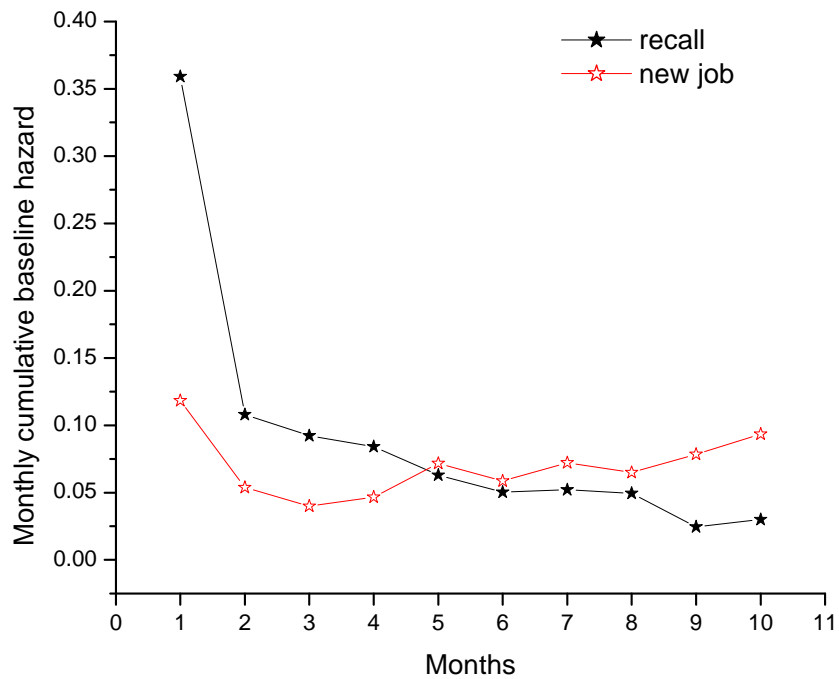


Figure 2.2 Estimated monthly cumulative baseline hazard in competing risks model; the specified risks are recall and new job.

Chapter 3

Bayesian Estimation of the Ordered Probit Unemployment Duration Model without Unobserved Heterogeneity

3.1 Introduction

As discussed in chapter I, one main issue in the study of duration models is the duration dependence, or the so-called baseline hazard. Over the past two decades there has been considerable interest to distinguish the effect of time on unemployment hazard rate from the effects of observed and unobserved heterogeneity. Two representative works are Meyer (1990) and Han and Hausman (1990), who estimated semi-parametric duration models by specifying the baseline hazard nonparametrically. Han and Hausman also demonstrated that the parametric assumption (more specifically, Weibull distribution) applied in Katz (1986) was overly restrictive and misleading.

In this chapter we try to estimate the ordered probit duration model without heterogeneity in the Bayesian paradigm. The primary purpose is to study the ordered probit model with either normal or nonparametric random effect using the Bayesian estimation procedure. Previous researches found that it is important to consider unobserved heterogeneity in the duration model (e.g., Heckman and Singer 1984). For computational convenience, some assumed that the random effect follows a gamma distribution. However, if the underlying distribution is not gamma, biased parameter inferences may be

obtained (see Hausman and Woutersen 2005). If normal random effect assumed, it will raise a number of challenges for the frequentist estimation due to the high-dimensional integration required in large sample datasets, which may lead to intractability of the likelihood function.

But many of the computational problems encountered in frequentist econometrics have been largely overcome in the Bayesian framework. Particularly, using the Bayesian simulation method, we can specify normal or nonparametric random effect. Such works are meaningful as Sueyoshi (1992) pointed out that nonparametric methods for both the baseline hazard and the unobserved heterogeneity should be developed to achieve satisfactory estimation of duration models. Studies on the duration model with random effect are currently in progress.

Both the standard Gibbs sampler and the hybrid Metropolis-Hastings/Gibbs sampler are used to estimate the ordered probit model without heterogeneity (Albert and Chib 1993, Cowles 1996, Johnson and Albert 1999). The standard Gibbs sampler is easy to apply. However, because at a single iteration there may be limited space for the cutpoints in our model to move, and by the very nature of the Gibbs sampler's iterative updating scheme, Gibbs sampler draws of a parameter are highly positively autocorrelated leading to slow convergence. Therefore the hybrid Metropolis-Hastings/Gibbs sampler is introduced so that the cutpoints will have more room to move and consequently speed up convergence.

To allow the data to suggest the appropriate mixing distributions, we assume uniform priors on both the regressor parameters and the cutpoints that formulate the

baseline hazard. While MLE estimates can be set as sequences' starting values, in order to explore the Bayesian method we use randomly chosen data as the initial parameter values.

Favorable features in our model specification and estimation include: uncertainty on the distributions of the baseline hazard and the covariate coefficients can be handled in the Bayesian paradigm under a tractable way; we have exact finite sample inference; extension to mixed effects ordered model is feasible and straightforward; other studies, such as time-varying covariates and unbalanced data, are easy to accommodate.

The plan of the paper is as follows. Section 3.2 presents the specification of the ordered probit duration model without random effects in the Bayesian framework. Section 3.3 is concerned with our MCMC based fitting methods. An application to the unemployment duration data retrieved from the PSID and studied in the first two chapters is reported in Section 3.4. Section 3.5 concludes.

3.2 Bayesian ordered probit duration model without unobserved heterogeneity

In this section we consider the ordered probit duration model without random effects. It is a generalization of the proportional hazard model proposed by Han and Hausman (1990), and Meyer (1990). However, we would like to study this model in the Bayesian paradigm.

Suppose that T represents the length of a spell of unemployment, and that Y_i is observed and takes one of T ordered categories or time intervals $\{1, \dots, T\}$ for $i=1, \dots, N$. Letting $p_{it} = P(Y_i=t)$ represent the probability of event occurrence in time interval t for person i , we denote the cumulative probabilities over the t time intervals of the outcome Y_i

as $P_{i,t} = P(Y_i \leq t) = \sum_{s=1}^t p_{is}$. The ordered probit model, introduced by Aitchison and Silvey (1957), defines $P_{i,t} = \Phi(\delta_t - X_i\beta)$, $i=1, \dots, N$, $t=1, \dots, T$. Here X_i is a $1 \times k$ vector of observations on a set of explanatory variables, β is a $k \times 1$ vector of unknown parameters, and δ_t is a series of unknown strictly increasing model cutpoints or thresholds (i.e., $\delta_0 = -\infty < \delta_1 < \dots < \delta_{T-1} < \delta_T < \delta_{T+1} = \infty$) that allows the cumulative response probabilities to differ from each other.

Now, the probability of ‘getting employed’ in time period t for person i is

$$p_{it} = P(Y_i = t) = P_{i,t} - P_{i,t-1} = \Phi(\delta_t - X_i\beta) - \Phi(\delta_{t-1} - X_i\beta) = \int_{\delta_{t-1} - X_i\beta}^{\delta_t - X_i\beta} f(\varepsilon) d\varepsilon, \quad (3.1)$$

where ε follows the standard normal distribution. Equation (3.1) is exactly the first duration model specified in Han and Hausman (1990), which steps from the proportional hazard model proposed in Prentice (1976) and Kalbfleisch and Prentice (1980). The δ vector is a set of unknown parameters that forms the nonparametric log-integrated baseline hazard specification. Each δ corresponds to the t^{th} of T mutually exclusive and exhaustive time intervals, and is assumed to be constant in that time period.

By imagining there is a latent variable V that underlies the generation of the ordered outcomes (e.g., Tanner and Wong 1987), we consider the latent variable representation of the ordered probit model:

$$\begin{aligned} V_i &= Age\beta_1 + Sex\beta_2 + Edu\beta_3 + Race\beta_4 + UI\beta_5 + MS\beta_6 + \varepsilon_i \\ &= X_i\beta + \varepsilon_i, \quad \varepsilon_i \sim N(0, 1), \end{aligned} \quad (3.2)$$

for $i=1, \dots, N$. Here Edu stands for years of education received, while UI and MS refer to unemployment insurance and marital status, respectively. The data sources and variable descriptions are given in chapter I of this dissertation.

Suppose V_i is linked to observed Y_i as follows:

$$Y_i^f = \begin{cases} 1 & \text{if } V_i \leq \delta_1 \\ 2 & \text{if } \delta_1 < V_i \leq \delta_2 \\ 3 & \text{if } \delta_2 < V_i \leq \delta_3 \\ \vdots & \vdots \\ T & \text{if } \delta_{T-1} < V_i \leq \delta_T \end{cases}, \quad (3.3)$$

with δ_t being a vector of cutoffs that classify an observation into an ordered category t , $t = 1, \dots, T$. The probability of a particular observed outcome, namely, the probability of failure in time period t for person i is given by the same form as Equation (3.1).

The above formulation is for failure observations only. Note that duration data often contains censored samples. For right-censored spells, the link is

$$Y_i^c = \begin{cases} 1 & \text{if } V_i \geq \delta_1 \\ 2 & \text{if } V_i \geq \delta_2 \\ 3 & \text{if } V_i \geq \delta_3 \\ \vdots & \vdots \\ T & \text{if } V_i \geq \delta_T \end{cases}. \quad (3.4)$$

Hence $Y_i^c = t$ means an unemployment duration is censored at sometime in $(t-1, t]$.

Again, we assume censorings occur at the end of time intervals. Other types of censorship are not considered in this chapter since our data set only covers complete and right-censored durations.

Now the probability for an individual agent's spell lasting longer than period t is

$$\begin{aligned} P(Y_i^c = t) &= P(V_i \geq \delta_t) = P(\varepsilon_i \geq \delta_t - X_i\beta) \\ &= 1 - \Phi(\delta_t - X_i\beta). \end{aligned} \quad (3.5)$$

By further assuming that the observed outcomes for a sample of n individuals are independent of one another given probabilities (3.1) and (3.5), we can specify a

multinomial distribution for the sample data Y_i (see, Johnson and Albert 1999). Next, we adopt the notation in Albert and Chib (1993) and formulate the likelihood function for a sample of N individuals as

$$L(\beta, \delta | y) = \prod_{i=1}^N \left\{ \sum_{t=1}^T I(Y_i = t) \times [\Phi(\delta_t - X_i \beta) - \Phi(\delta_{t-1} - X_i \beta)] \right\}^{1-c_i} \times \left\{ \sum_{t=1}^T I(Y_i = t) \times [1 - \Phi(\delta_t - X_i \beta)] \right\}^{c_i}, \quad (3.6)$$

where

$$c_i = \begin{cases} 1 & \text{if individual is censored} \\ 0 & \text{otherwise} \end{cases}$$

and

$$I(Y_i = t) = \begin{cases} 1 & \text{when failure or censoring occurs for person } i \text{ at time } t \\ 0 & \text{elsewhere} \end{cases}$$

for the ordered probit duration model without heterogeneity. Notice that because there is no constant intercept included in the regression function $X_i \beta$, we do not face the usual identification problem on the parameters δ and β (e.g., Albert and Chib 1993).

If we introduce the latent variable V_i , the likelihood function (3.6) will become

$$L(\beta, \delta, V | y) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}} \exp\left[-(V_i - X_i \beta)^2 / 2\right] \times \left[\sum_{t=1}^T I(Y_i = t) I(\delta_{t-1} < V_i \leq \delta_t) \right]^{1-c_i} \times \left[\sum_{t=1}^T I(Y_i = t) I(V_i \geq \delta_t) \right]^{c_i}. \quad (3.7)$$

The definitions for c_i and $I(Y_i = t)$ are same as those defined above. When $I(Y_i = t) = 1$, it indicates that V_i falls between the cutpoints δ_{t-1} and δ_t for failures, and for censored cases $V_i \in [\delta_t, \infty)$. Note that equations (3.6) and (3.7) actually refer to the same model from different perspectives.

The regression parameters and the cutoff points are assumed to be a priori independent, and uniform priors are taken for both δ and β . Using the latent variable representation for the likelihood function (3.7), the joint posterior density of the model parameters and the latent variables is then given by

$$\begin{aligned} \pi(\beta, \delta, V | y) \propto & \prod_{i=1}^N \phi(V_i - X_i \beta) \times \left[\sum_{t=1}^T I(Y_i = t) I(\delta_{t-1} < V_i \leq \delta_t) \right]^{1-c_i} \\ & \times \left[\sum_{t=1}^T I(Y_i = t) I(V_i \geq \delta_t) \right]^{c_i}, \end{aligned} \quad (3.8)$$

where ϕ denotes the standard normal density, and δ subject to the constraint that $\delta_0 = -\infty < \delta_1 < \dots < \delta_{T-1} < \delta_T < \delta_{T+1} = \infty$. Note that the joint posterior is simply proportional to the likelihood function (3.7).

Equation (3.6) can be used to search for the maximum likelihood estimates of the regression parameters β , and the nonparametric baseline hazard δ (Han and Hausman 1990). Another version of the same model, equation (3.7) lends itself to full Bayesian analysis by Markov Chain Monte Carlo (MCMC) methods. Samples from the joint posterior of (3.8) can be obtained via the standard Gibbs sampler or the hybrid Metropolis-Hastings/Gibbs sampler. One of the first such algorithms was proposed by Albert and Chib (1993). Here, combined with the conjugacy of the prior specification, our Gaussian-linear structure model makes all of its full conditional posteriors easily available in closed form (as normal and uniform. See next section for details). Under mild regularity conditions, the draws from the full conditional posterior distributions will converge to those from the desired joint posterior distribution. Therefore, the MCMC algorithms involving use of latent data will allow us to estimate and examine the posterior quantities

of interest straightforward without putting parametric assumptions on the baseline hazard and the covariate coefficients.

3.3 The standard Gibbs sampler and the hybrid Metropolis-Hastings/Gibbs sampler

3.3.1 Standard sampling scheme for the ordered probit model without heterogeneity

Because the full conditional posteriors for (V, β, δ) all have analytically tractable forms, as we will see next, it is possible to implement a standard Gibbs sampling approach to the simulation of the joint posterior (V, β, δ) .

(1) Generate the latent variable V

The latent variable V_i is sampled from its complete posterior conditional:

$$V_i | \beta, \delta, y_i = t \stackrel{ind}{\sim} \begin{cases} TN_{(\delta_{i-1}, \delta_i]}(X_i \beta, 1) & \text{if } fail \\ TN_{[\delta_i, \infty)}(X_i \beta, 1) & \text{if } survive \end{cases}$$

for $i=1, \dots, N$. Here the notation $TN_{(L,U]}(\mu, \omega)$ denotes a normal distribution with mean μ and variance ω truncated to the interval $(L, U]$.

(2) Generate the covariate coefficient β

The prior information for β is assumed to be diffuse. In particular, a uniform prior is taken. This choice of prior results in a multivariate normal conditional posterior for β with mean and covariance matrix that are identical to its least squares estimates, namely

$$\beta | V, \delta, y \sim N_k(D_\beta d_\beta, D_\beta),$$

where $D_\beta = (X'X)^{-1}$, $d_\beta = X'V$, X is an $N \times k$ matrix and V is an $N \times 1$ latent vector.

(3) Generate the baseline parameters δ

We also assume flat prior on the log-integrated baseline hazard parameter. For the complete posterior conditional for the cutpoints, note that when the data includes only failures,

$$\delta_t | \delta_{-t}, \beta, V, y \propto \prod_{i: y_i^f = t} I(\delta_{t-1} < V_i \leq \delta_t) \prod_{i: y_i^f = t+1} I(\delta_t < V_i \leq \delta_{t+1}).$$

It follows that

$$\delta_t | \delta_{-t}, \beta, V, y \stackrel{ind}{\sim} U[S_1, S_2],$$

where

$$S_1 = \max\left\{\delta_{t-1}, \max_{i: y_i^f = t} V_i\right\},$$

$$S_2 = \min\left\{\delta_{t+1}, \min_{i: y_i^f = t+1} V_i\right\},$$

and U denotes uniform distribution; δ_{-t} refers to all the other δ_s except δ_t .

When an observation is censored in t^{th} time interval, it means that

$$V_i \geq \delta_t \quad \text{And} \quad \delta_{t-1} < \delta_t < \delta_{t+1}.$$

Similarly, we know that δ_t should fall in

$$\left[\delta_{t-1}, \min\left(\delta_{t+1}, \min_{i: y_i^c = t} V_i\right)\right].$$

Since our data set contains both failure and right-censored observations, the full posterior conditional distribution for δ_t is

$$\delta_t | \delta_{-t}, \beta, V, y \stackrel{ind}{\sim} U\left[S_1, \bar{S}_2\right],$$

where

$$\bar{S}_2 = \min \left\{ \delta_{t+1}, \min_{i: y_i^f = t+1} V_i, \min_{i: y_i^c = t} V_i \right\}.$$

(4) Repeat steps (1) to (3) using the most recent values of the conditioning variables.

Blocking steps may be used to improve the mixing of the chain. For example, we can update all the latent variables V_i in one block. In this case we have to draw directly from a multivariate truncated normal (Geweke 1991). Such drawing is nontrivial in general, however, a Gibbs sampler itself has to be implemented in order to generate a sample from the multivariate truncated normal density. Subsequently the computational cost is very high although an improvement in convergence may be achieved. In order to reach comparable convergence results, the time needed for the grouping method increases in such a level that the overall improvement may be negligible (see Geweke 1991, 1996, Robert 1995, Muller and Czado 2005). Besides, this blocking strategy does not address a typical problem encountered in the ordered probit model when simulation-based estimation method is employed. See next for discussions on the problem and its solution. Therefore, we stick to our algorithm that samples the latent variable individually and use an efficient procedure (the so-called inverse transform method) for sampling from the univariate truncated normal.

A problem associated with the ordered probit model is very slow mixing in the standard Gibbs sampling scheme developed above. Mixing refers to the dependence between one simulated value and that of j iterations away. It is measured by the autocorrelation among the Gibbs draws. Rapid mixing means that the dependence decays quickly as j increases. Conversely, slow mixing occurs when the autocorrelation is still significant even as j becomes large. Slow chain mixing induces slow convergence. A leading cause of slow convergence is multimodality of the underlying, possibly unknown

joint posterior distribution. In this case, the Gibbs sequence may get stuck in a small subset of the sample space and many iterations are required to recover, which may represent as slow chain mixing.

The slow mixing in our case is mostly due to the following reason. We simulate the cutpoints δ from a uniform distribution with boundaries S_1 and \bar{S}_2 given before. When there is large number of observations in categories, the interval $\bar{S}_2 - S_1$ becomes very narrow, so δ_i has very little room to move in one iteration. Because of the distorted cutpoint values, convergence of the regressor coefficients β is also retarded. Several hybrid Metropolis-Hastings/Gibbs algorithms have been proposed for sampling from the posterior distribution on ordered probit regression parameters. Two notable algorithms are Cowles (1996), and Nandram and Chen (1996). We next use Cowles' method to speed up the convergence of the standard Gibbs sampler. As noted in Johnson and Albert (1999), Cowles' algorithm is relatively easy to apply, displays good mixing and is suitable for models with arbitrary constraints on the interval cutoffs. The reparameterization method suggested in Nandram and Chen (1996) can also be implemented when there are more than three bins. But this method may be more complicated so we leave it for future exercise.

3.3.2 Hybrid Metropolis-Hastings/Gibbs sampler for the ordered probit model without heterogeneity

The algorithm developed below is based on the work presented in Cowles (1996), and Johnson and Albert (1999). Revision is made to accommodate our data characteristics. This method can be applied on models with non-Gaussian link functions straightforward, though some significant changes are required.

(1) Generate a candidate δ^* for updating $\delta^{(s-1)}$

- a. In the s^{th} iteration, for $t = 1, \dots, T$, sample $\delta_t^* \sim N(\delta_t^{(s-1)}, \sigma_\delta^2)$ truncated to the interval $(\delta_{t-1}^*, \delta_{t+1}^{(s-1)})$. Note that we take $\delta_0^* = -\infty$ and $\delta_{T+1}^* = \infty$. The truncated normal is our candidate generating density and random draws are taken from it. We adopt the rule of thumb wherein $\sigma_\delta^2 = 0.05/T$. Adjustments to σ_δ^2 may be necessary so that appropriate acceptance rates for δ , approximately between 0.25 and 0.5, are obtained.

- b. Compute the acceptance probability AR according to

$$AR = \prod_{i: y_i^f} \frac{\Phi(\delta_{y_i^f}^* - X_i \beta^{(s-1)}) - \Phi(\delta_{y_i^f-1}^* - X_i \beta^{(s-1)})}{\Phi(\delta_{y_i^f}^{(s-1)} - X_i \beta^{(s-1)}) - \Phi(\delta_{y_i^f-1}^{(s-1)} - X_i \beta^{(s-1)})} \\ \times \prod_{i: y_i^c} \frac{1 - \Phi(\delta_{y_i^c}^* - X_i \beta^{(s-1)})}{1 - \Phi(\delta_{y_i^c}^{(s-1)} - X_i \beta^{(s-1)})} \\ \times \prod_{t=1}^T \frac{\Phi[(\delta_{t+1}^{(s-1)} - \delta_t^{(s-1)}) / \sigma_\delta] - \Phi[(\delta_{t-1}^* - \delta_t^{(s-1)}) / \sigma_\delta]}{\Phi[(\delta_{t+1}^* - \delta_t^*) / \sigma_\delta] - \Phi[(\delta_{t-1}^{(s-1)} - \delta_t^*) / \sigma_\delta]}.$$

The first two terms on the right hand side are the contributions from the likelihood function for both failure and right-censored observations; while the 3rd term comes from the transformation of the proposal density for δ . We assume the prior for δ still to be uniform. Of course, the ratio of the prior evaluated at candidate values and last cycle values should be multiplied if nonuniform prior on δ is employed.

- c. Set $\delta^{(s)} = \delta^*$ with probability AR .

(2) Generate the latent variable V

The full conditional density for V_i is

$$V_i | \beta, \delta, y_i = t \stackrel{\text{ind}}{\sim} \begin{cases} TN_{(\delta_{t-1}, \delta_t)}(X_i \beta, 1) & \text{if } \textit{fail} \\ TN_{[\delta_t, \infty)}(X_i \beta, 1) & \text{if } \textit{survive} \end{cases}$$

for $i = 1, 2, \dots, N$. In this revised sampling method, we use current cycle cutoff values rather than those from last cycle to define truncation boundaries for the latent variable, conditional on the acceptance of the candidates δ^* .

(3) Generate the regression coefficient β

The conditional posterior of β is unchanged and its sampling follows the same way specified in the standard Gibbs sampling.

(4) Repeat steps (1) to (3) using the most recent values of the conditioning variables.

Convergence of the chains may be assessed using the Gelman-Rubin scale reduction factor (SRF). To apply this check, one runs a number of chains simultaneously, say M chains with n realizations in each chain, and compares variations in the generated parameter values between and within chains.

Define

$$B = \frac{n}{M-1} \sum_{m=1}^M (\bar{\omega}_m - \bar{\omega})^2$$

as the variability of the parameter of interest ω between the M parallel chains. Then the within chain variance is

$$W = \frac{1}{M(n-1)} \sum_{m=1}^M \sum_{i=1}^n (\omega_{m,i} - \bar{\omega}_m)^2,$$

where

$$\bar{\omega}_m = \frac{1}{n} \sum_{i=1}^n \omega_{m,i} \quad ,$$

$$\bar{\omega} = \frac{1}{M} \sum_{m=1}^M \bar{\omega}_m \quad .$$

The posterior marginal variance, $\text{var}(\omega|y)$, is a weighted average of B and W . The estimated variance is

$$\hat{V} = \frac{n-1}{n}W + \frac{B}{n}\left(1 + \frac{1}{M}\right).$$

Then the potential scale reduction factor (PSRF) is given by

$$R = \sqrt{\frac{\hat{V}}{W}} = \sqrt{\left(1 - \frac{1}{n}\right) + \frac{B}{W \cdot n}\left(1 + \frac{1}{M}\right)}.$$

If R is close to 1, it may indicate that convergence has been reached. Note that this is indicative of convergence but does not prove it. Further details can be found in Gelman and Rubin (1992), and Brooks and Gelman (1998).

3.4 Application and results

We apply the Bayesian ordered probit duration model to study the effect of unemployment insurance, along with other socio-demographic factors, on the unemployment duration. Descriptions on the data set can be found in Section 1.3 of Chapter I. The number of time intervals is reduced to be 11 rather than 41 as used in chapters 1 and 2. Extension to arbitrary number of time intervals is feasible but also time-consuming, so we leave it for future exercise. In fact, if we use the MLE estimates as the initial values of the Markov chains, it is fairly easy to get satisfactory MCMC results for even 46 parameters. However, to better explore how to estimate the ordered probit model in the Bayesian paradigm, we use randomly generated numbers as the starting values of the sampling sequences.

Estimated results for the ordered probit duration model without considering heterogeneity are presented in Table 3.1 and Figures 3.1 to 3.11. Figures 3.1 to 3.4 come from the standard Gibbs sampling scheme for such a model. As one can directly tell from Figures 3.1 and 3.2, the convergence has not yet been reached even after 300,000

iterations. For each interested parameter, five parallel chains are run with relative diffuse beginning values. Generally speaking, all the β parameters converge better than the cutpoints, especially for the dummy variables' coefficients. Trajectory patterns for all the cutpoints δ are similar, quite possibly because there is order restriction present among them. Hence, three representative δ trajectories are reported, one from the beginning, one from the middle, and the other from the right end. None of these three shows sign of convergence.

Consequently, the Gelman-Rubin factors (following the literature, we call them R factors afterwards) in Figures 3.3 and 3.4 are all larger than 2, except the one for sex parameter, which indicates failure of convergence. It is not a good idea to compute the R factor from the very beginning of a chain. The values will be relatively large at the start and make it difficult to see what happens after the chains tend to converge because of the graph scale. Therefore, we begin to calculate the R factors after 150,000 iterations.

This experiment confirms the difficulty mentioned in Cowles (1996). She noted that standard Gibbs sampler would become inefficient in the estimation of the ordered probit model due to the slow move of the cutoffs δ_j and the more the cutpoints, the more severe the problem is.

In other words, because of the Markovian updating nature of the Gibbs sampler, the poor estimates can not be recovered even after many replications. Moreover, the sample data studied in Cowles (1996) is generated from a linear model; while our real data may not follow a linear relationship. Therefore, estimation based on the standard Gibbs sampler and data augmentation will be affected adversely given the complexity of our data and model.

Naturally we apply the hybrid Metropolis-Hastings/Gibbs sampler proposed in Cowles (1996) to improve the speed of convergence. As Cowles claimed, the autocorrelation functions for all of the series will die out quickly. So the chain mixing is faster. This can be seen from Figures 3.5 to 3.8. Now all six β converge fast, particularly the coefficients for the dummy variables. Age parameter shows mixture at around 40,000th iteration, while the one for education at about 80,000th iterate. The mixing of δ parameters is not as fast as the regression coefficients, but unlike draws simulated under the standard Gibbs sampling, they do show signs of mixture after 120,000 iterates. The R factors are also calculated to double check the convergence of the chains. At 150,000 iterates, all factors are valued less than 1.8; as iterations increase, the values continue to drop and all of them fall below 1.25, which indicate sequence convergence (Gelman and Rubin 1992). The acceptance rate in the Metropolis-Hastings step is around 0.33 for each parallel chain.

As the question of how many initial draws to discard so that sufficient mixing of the iterates is achieved, we make our choice based on both the sampling trajectory and the R factor value. That is, for each chain, the first 150,000 samples are dropped to allow for the proper amount of ‘burn-in’. Since five chains are run for each parameter, it means that our inference is then based on a total of 750,000 iterates.

The posterior means and standard deviations are presented in Table 3.1, along with the MLE estimates. All the regression parameters have the expected signs and level of significance. Aside from slight differences, the posterior means of the cutpoints have the same sign of those MLE estimates and fulfill their order constraint. Recall that such results are realized under noninformative prior assumptions for both β and δ . That is, we allow the data itself to tell the story without putting subjective directions on the estimation.

The posterior marginal densities derived from the hybrid MH/Gibbs sampling are presented in Figures 3.9 and 3.10. We report the densities for all β_s and three typical δ . The figures show that all the densities are concentrated around their respective posterior means. Moreover, the β densities display normal curve while it is difficult to assign the distribution of δ to a parametric family. Our experiment with uniform priors on both β and δ proposes that it is reasonable to assume normal prior for regressor parameters but better leave the log-integrated baseline hazard's distribution unspecified.

Once again the MCMC estimates address the important negative effect of unemployment benefit on the length of duration without a job. Race and marital status also have significant effects on the unemployment spell. Interestingly, the magnitude of the effects of the covariates does stay almost the same in spite of different time intervals chosen, which has been noted in Han and Hausman (1990). While the cutpoint estimates do change, their plots in Figure 3.11 show excellent catch of the duration shape perceived from empirical data (see Chapter I for details). This is true for both the MLE and MCMC results, though there exists a small 'shift' factor between them.

3.5 Conclusions and future research directions

In this chapter we have developed the Bayesian paradigm to estimate the ordered probit duration model without heterogeneity. The Bayesian estimation method that we propose can be used to examine models with arbitrary cutpoints and extended to study the ordered model with normal or nonparametric random effect.

The estimation procedure for the ordered probit model includes two MCMC methods. The first method we employ is the standard Gibbs sampler. However, under such

method the sampling sequences will not converge after 300,000 iterates. Hence we conduct our analysis via the hybrid Metropolis-Hastings/Gibbs sampler. Both simulation trajectory and the Gelman-Rubin factor show that convergence does speed up once the refined sampler is used. As the prior on the parameters of the ordered probit model, we propose uniform priors to avoid the problem of improper information and meanwhile make sure the model is identifiable.

The ordered probit model without heterogeneity is applied to analyze the effect of unemployment insurance, along with other economic variables, on the unemployment duration. We find that race, unemployment compensation and one's marital status all have significant effects on the length of unemployment spell, though the first two characteristics will prolong the duration while the last one will make the length shorten. The kernel density estimates of the parameters show that assuming normal prior on β_s may be reasonable. However, to estimate δ_s correctly, we should not assign them any parametric prior. That says, if we let the data itself tell us the story, we can capture very well the perceived feature of the unemployment duration.

Currently we are studying the ordered probit hazard model with heterogeneity. The prior for the unobserved heterogeneity is assumed to be either normal or Dirichlet Process. The latter specification allows us to estimate both the baseline hazard and the unobserved heterogeneity nonparametrically, which is an important issue in the study of the mixed proportional hazard (MPH) model. We believe that the proposed MPH models will be appropriate to detect the clustering existing among subjects.

Table 3.1 The Ordered Probit Model w/o Heterogeneity MLE and Posterior Estimates^a

Variable	MLE results	MCMC posterior results	
	mean	mean	std. error
age	-0.0155	-0.0161	0.0032
sex	-0.0017	-0.0038	0.1068
education	-0.0153	-0.0193	0.0142
race	0.2825	0.2766	0.0685
UI	0.2551	0.2530	0.0697
married	-0.2679	-0.2740	0.0841
Delta1	-1.0660	-1.1491	0.2251
Delta2	-0.7572	-0.8397	0.2244
Delta3	-0.5547	-0.6359	0.2243
Delta4	-0.3853	-0.4652	0.2240
Delta5	-0.2334	-0.3121	0.2236
Delta6	-0.1211	-0.1979	0.2233
Delta7	-0.0039	-0.0786	0.2234
Delta8	0.0941	0.0220	0.2235
Delta9	0.1768	0.1082	0.2235
Delta10	0.2674	0.2028	0.2239

^aThe estimates are for the ordered probit model without considering heterogeneity. The sampling scheme used is the hybrid Metropolis-Hastings/Gibbs sampler presented in section 3.3.2. MLE estimates are presented on the left column for the purpose of comparison.

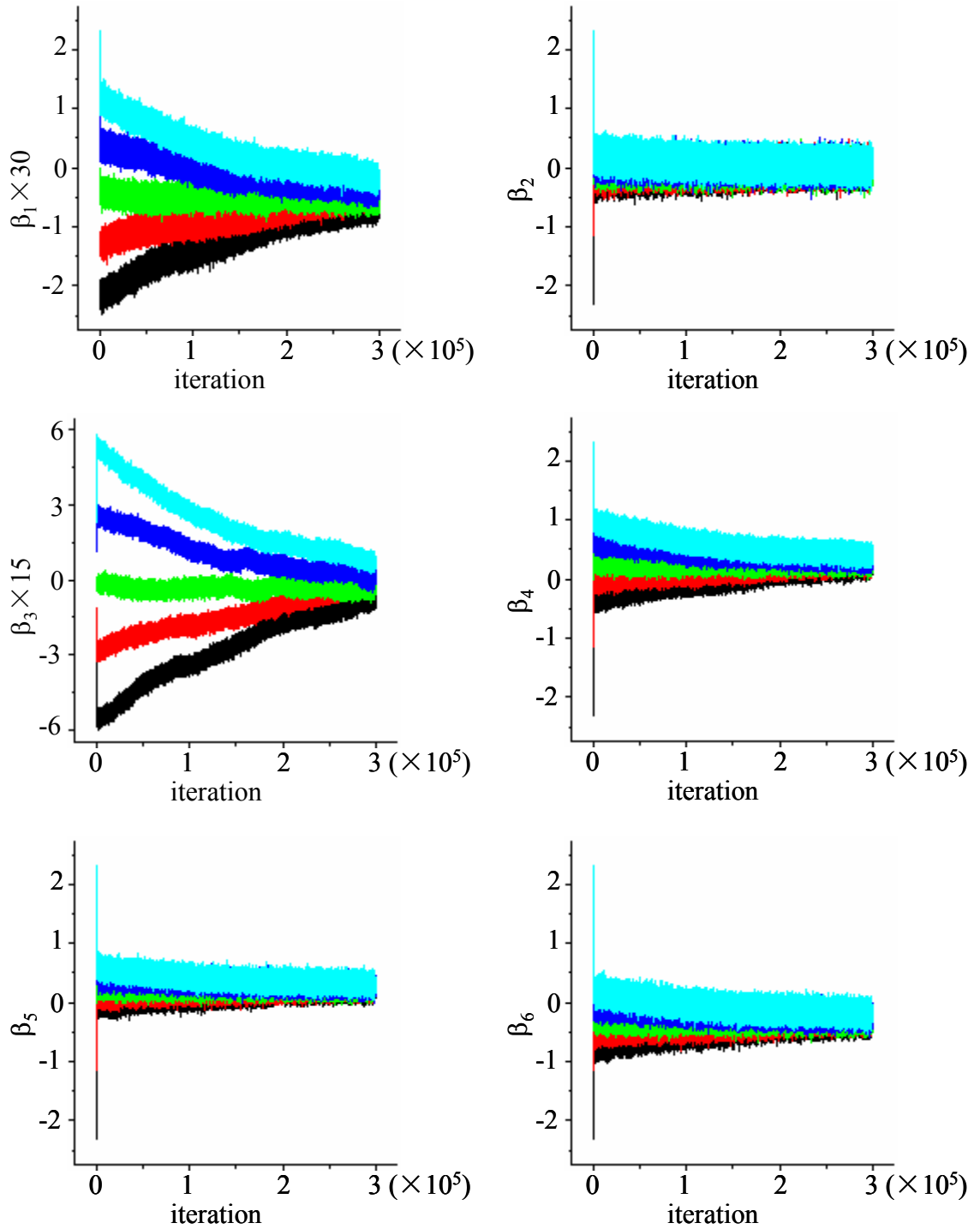


Figure 3.1 Iteration trajectories of the covariate parameters in the standard Gibbs sampling scheme. The covariates respective to β_1 to β_6 are age, sex, education, race, unemployment insurance and marital status.

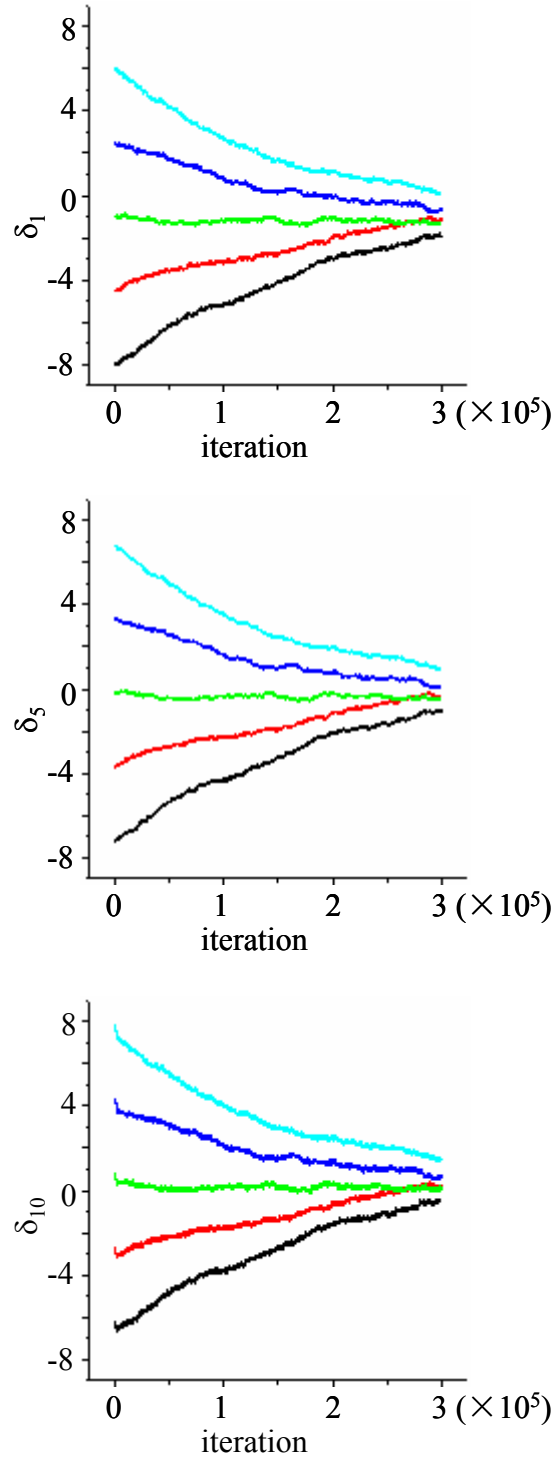


Figure 3.2 Iteration trajectories of three representative cutpoints in the standard Gibbs sampling scheme.

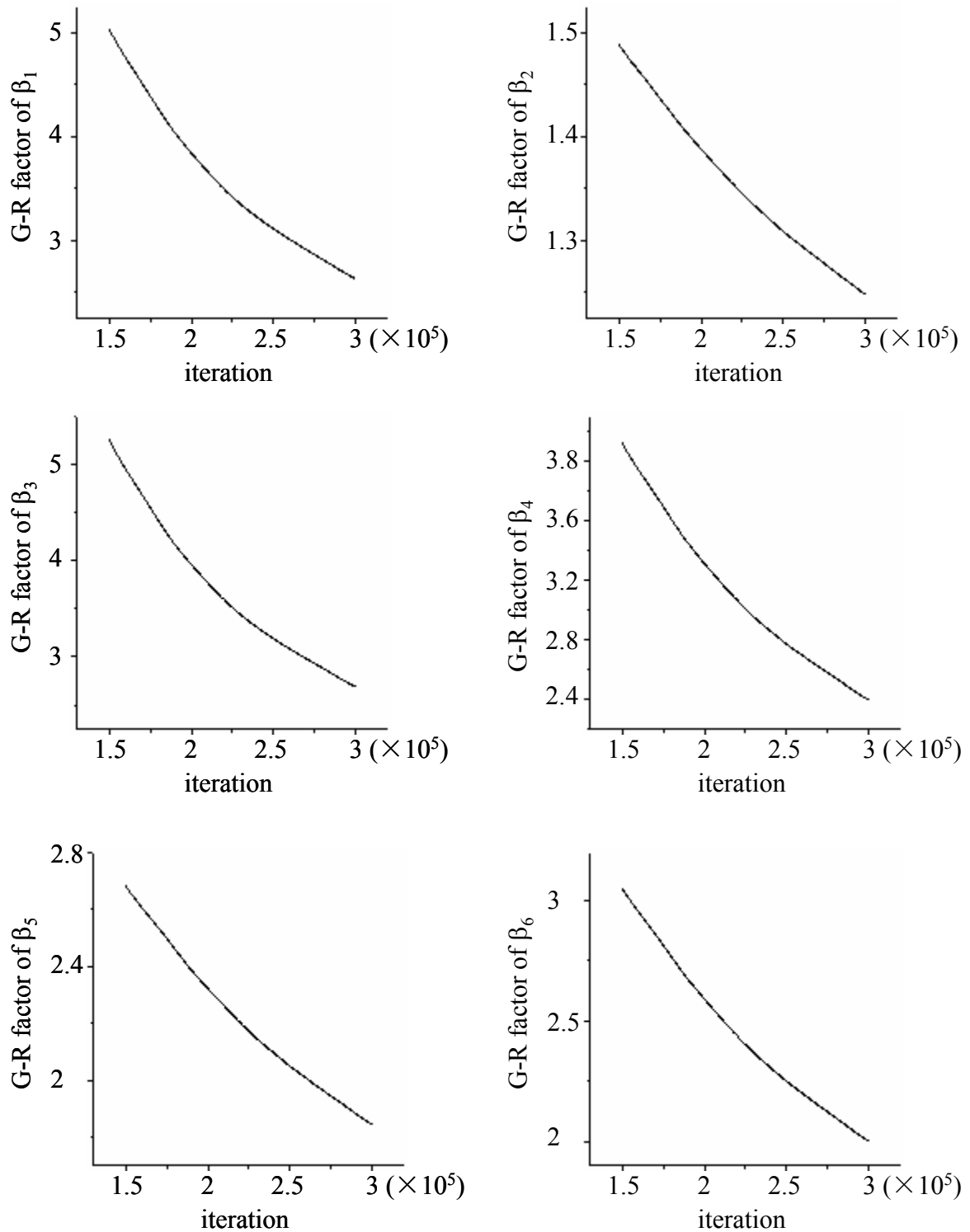


Figure 3.3 Evolution of the Gelman-Rubin factors for the covariate parameters in the standard Gibbs sampling scheme. The covariates respective to β_1 to β_6 are age, sex, education, race, unemployment insurance and marital status.

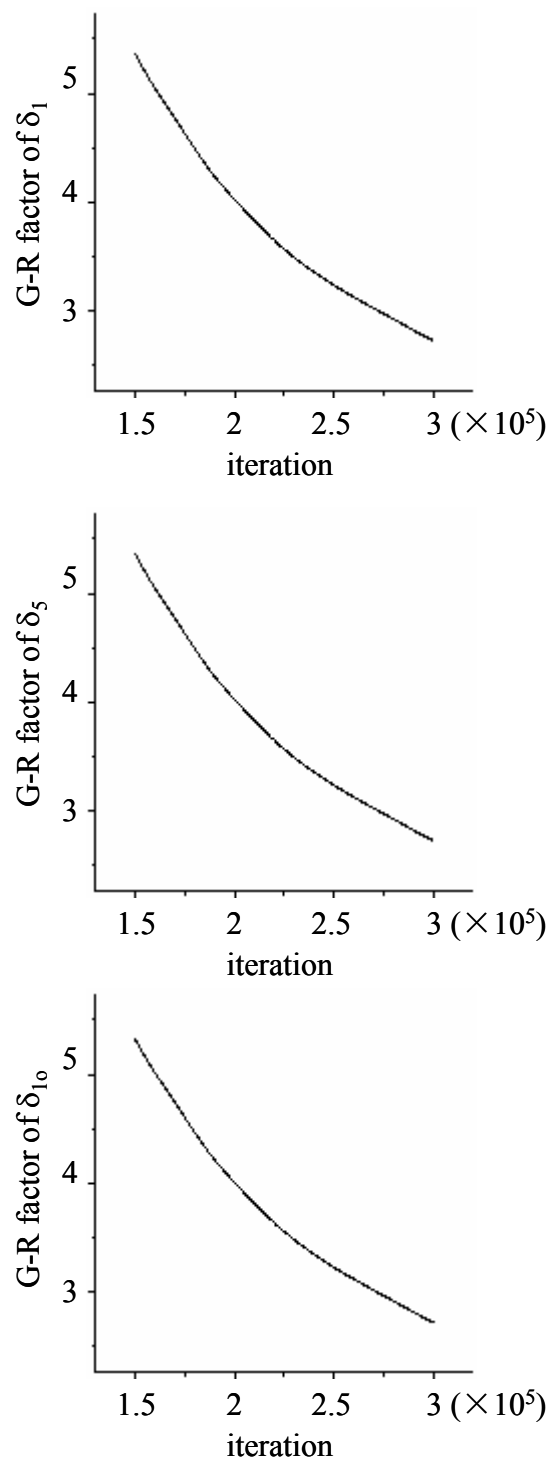


Figure 3.4 Evolution of the Gelman-Rubin factors for three representative cutpoints in the standard Gibbs sampling scheme.

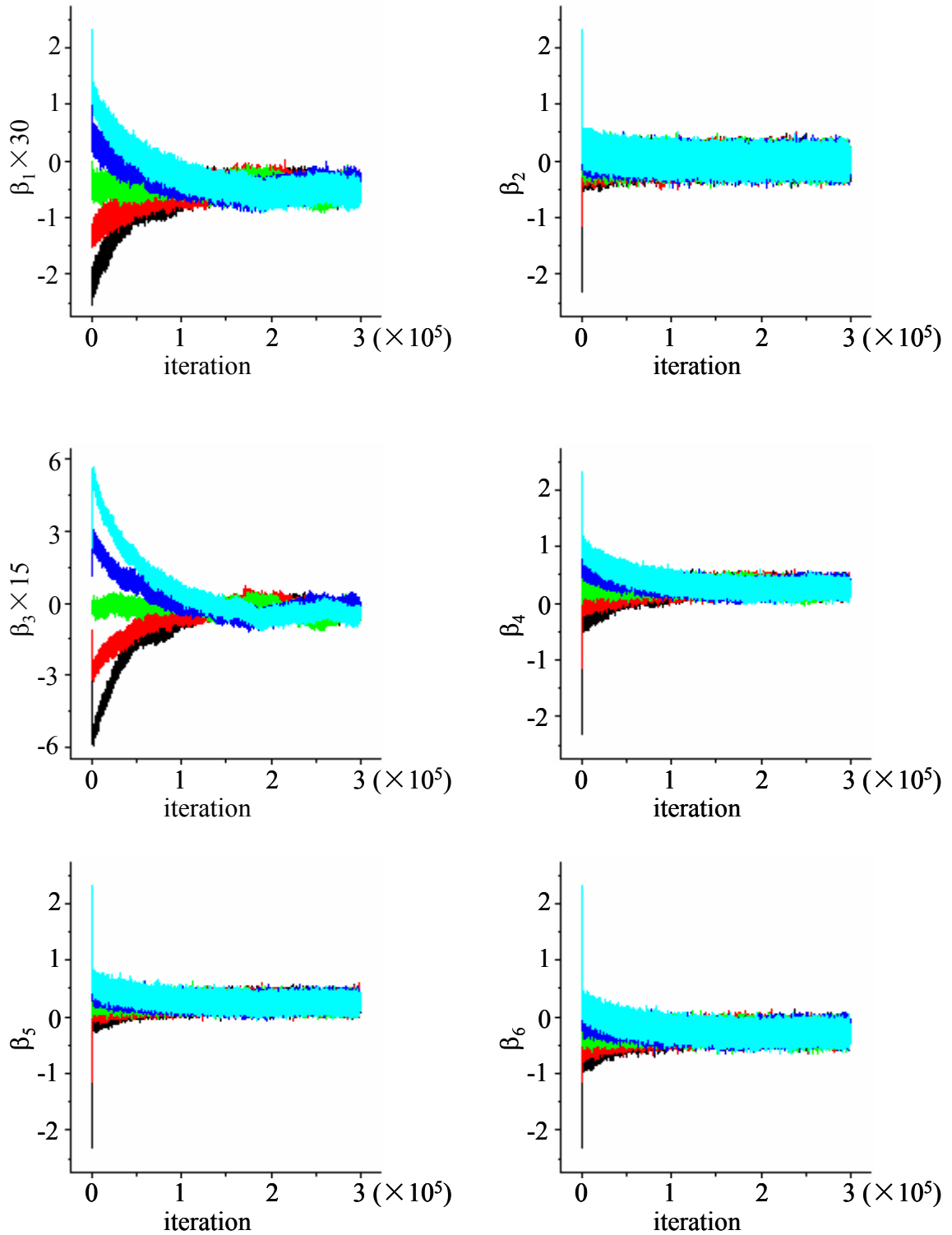


Figure 3.5 Iteration trajectories of the covariate parameters in the hybrid Gibbs sampling scheme. The covariates respective to β_1 to β_6 are age, sex, education, race, unemployment insurance and marital status.

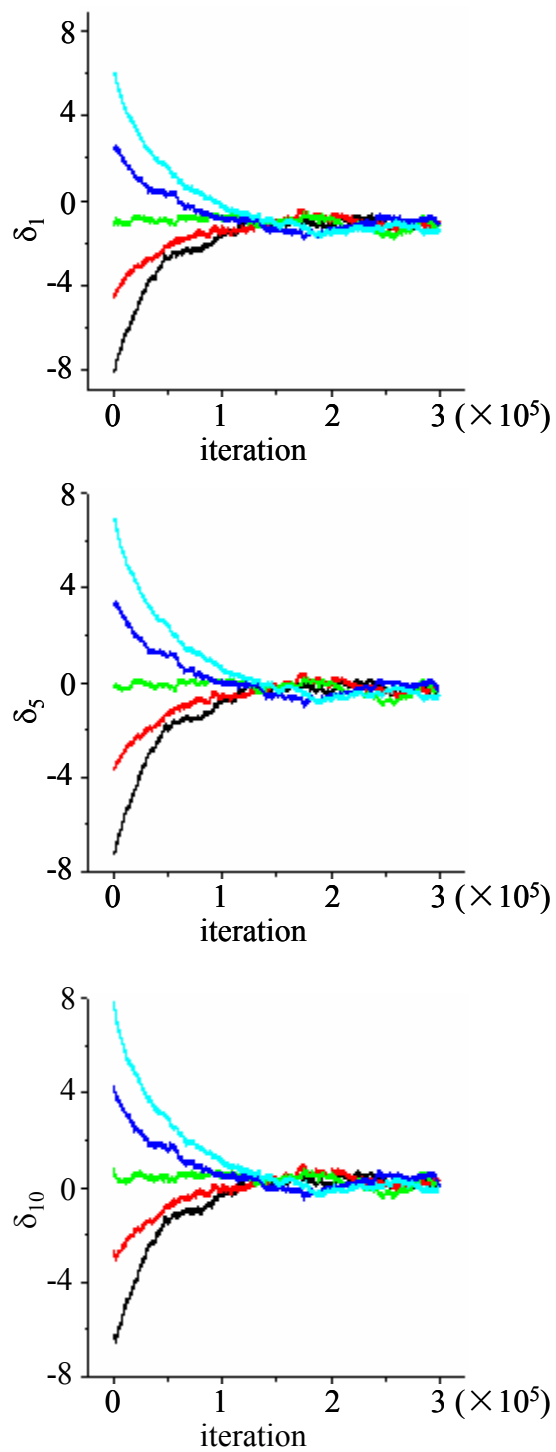


Figure 3.6 Iteration trajectories of three representative outpoints in the hybrid Gibbs sampling scheme.

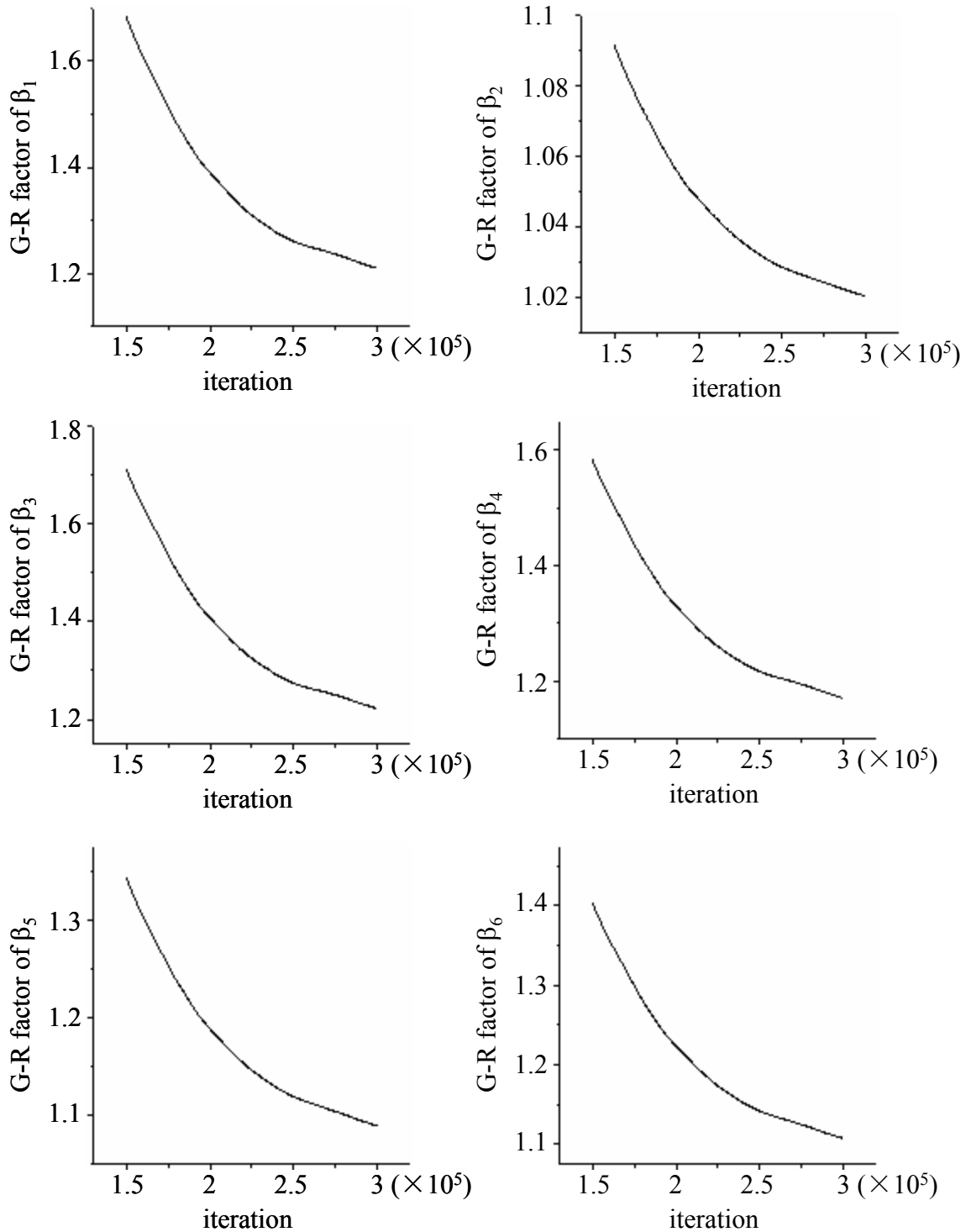


Figure 3.7 Evolution of the Gelman-Rubin factors for the covariate parameters in the hybrid Gibbs sampling scheme. The covariates respective to β_1 to β_6 are age, sex, education, race, unemployment insurance and marital status.

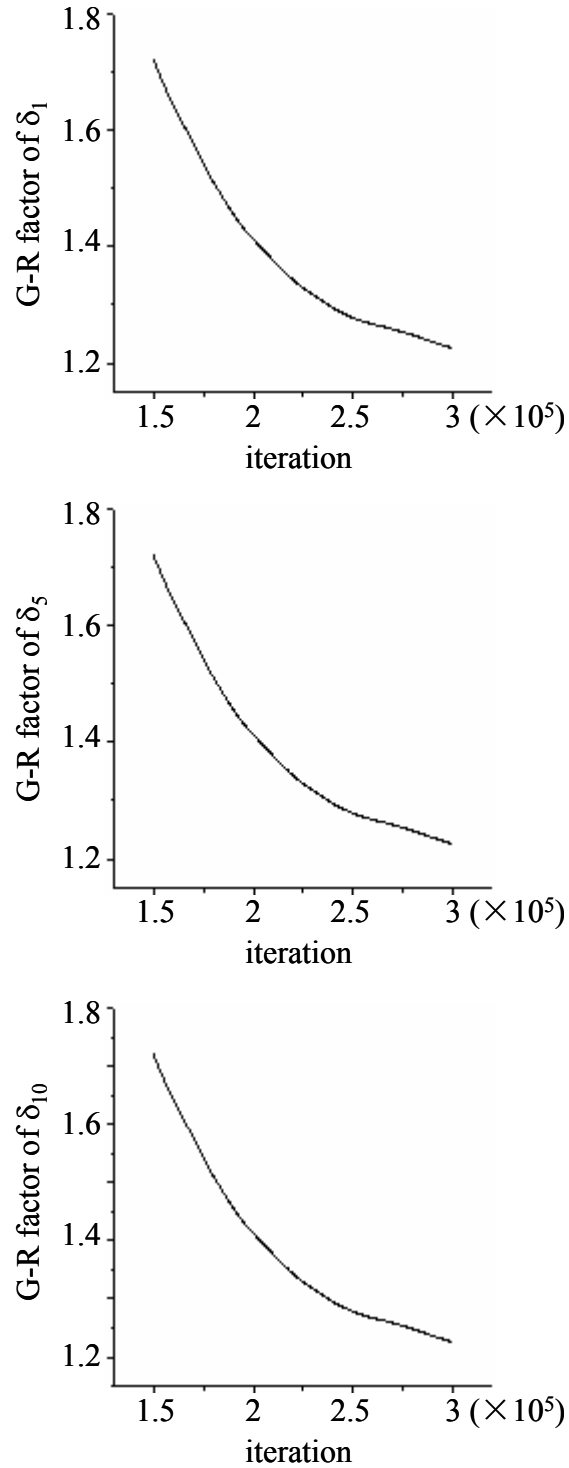


Figure 3.8 Evolution of the Gelman-Rubin factors for three representative cutpoints in the hybrid Gibbs sampling scheme.

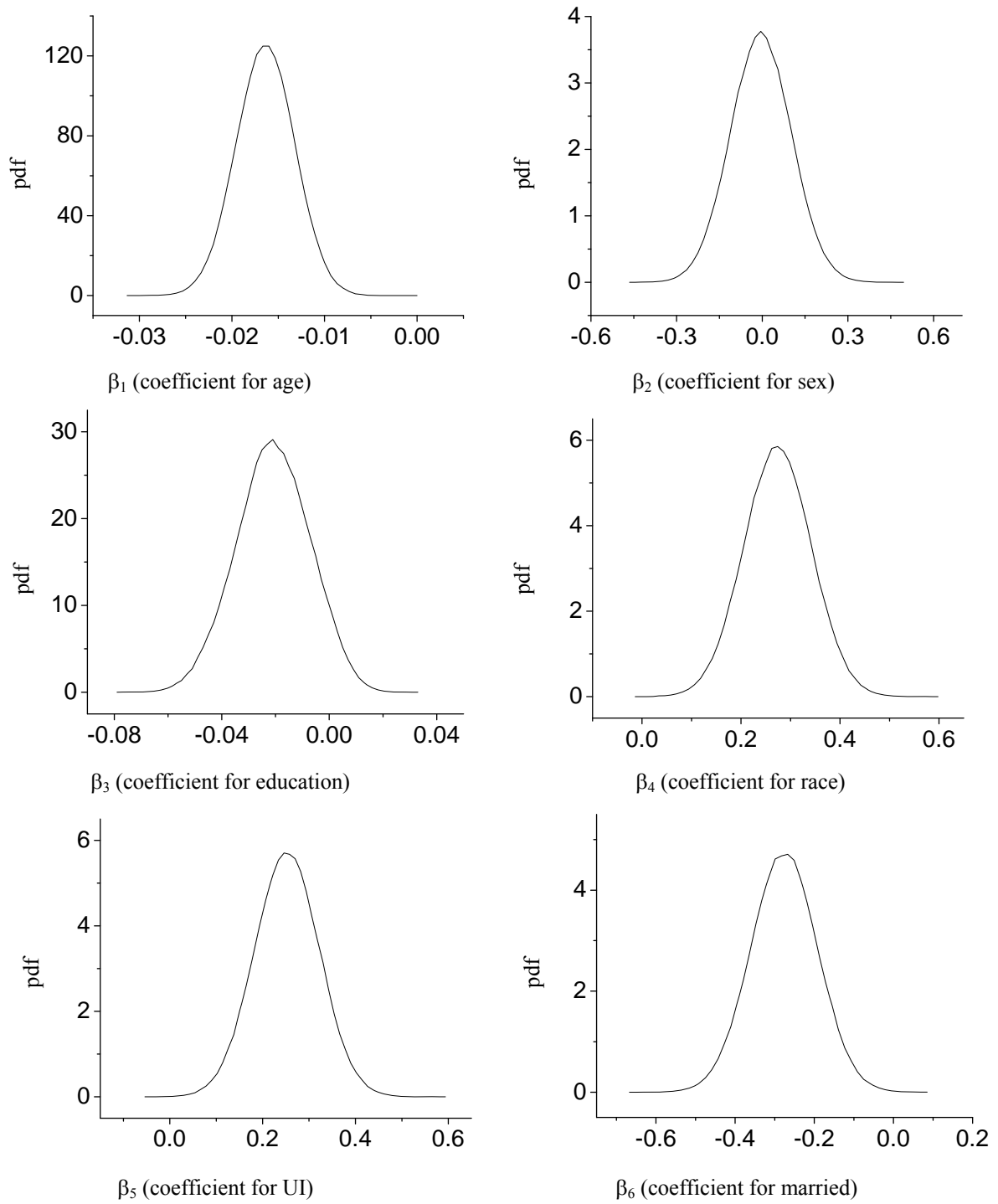


Figure 3.9 Estimates of the marginal posterior pdf for the regressor coefficients.

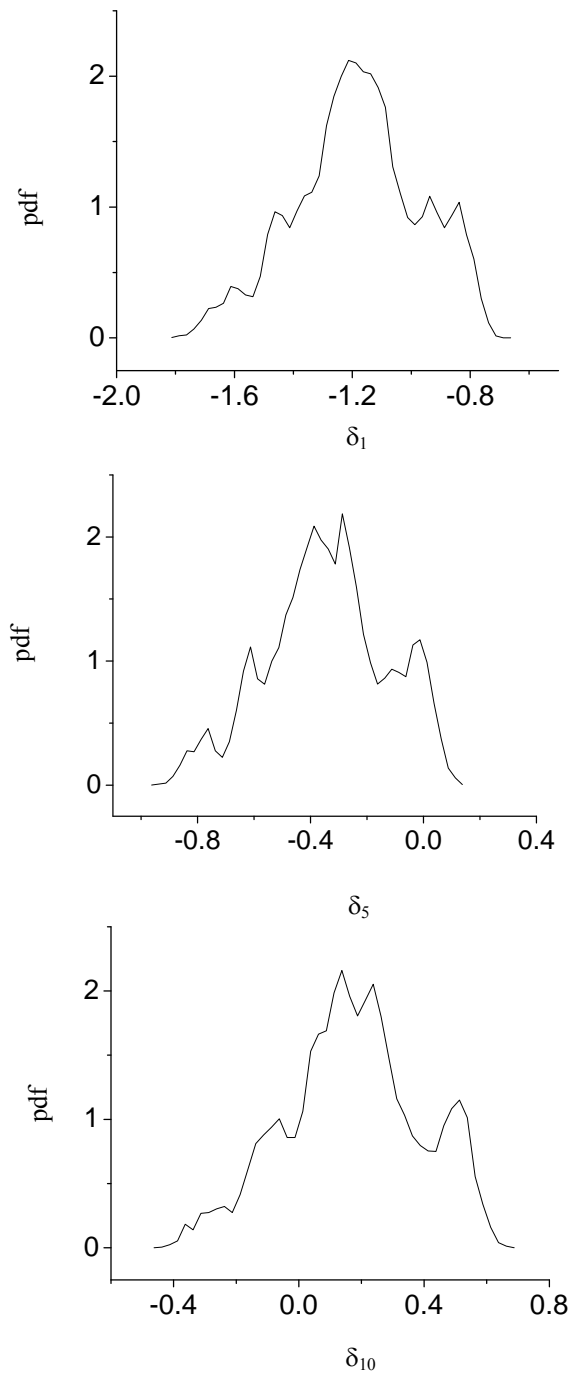


Figure 3.10 Estimates of the marginal posterior pdf for the first, fifth and last cutpoints.

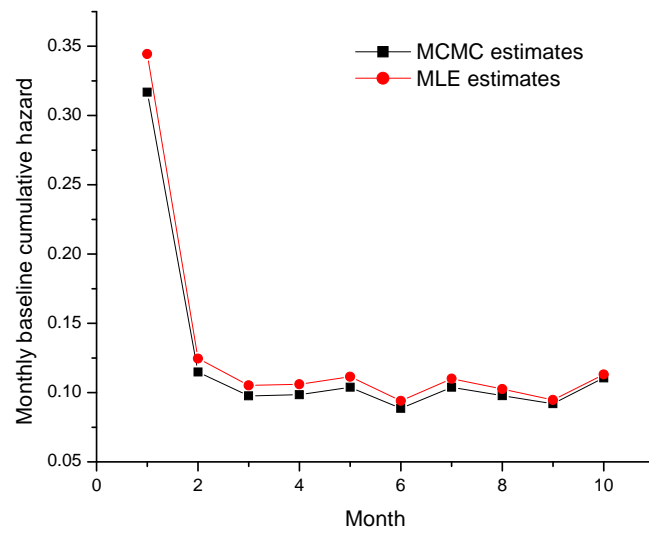


Figure 3.11 Estimated monthly cumulative baseline hazard for the ordered probit model without heterogeneity under the MLE and Bayesian paradigms.

References

- [1] Abbring, J.H. and G.J. van den Berg (1998). Limit results for unobserved heterogeneity distributions in mixed proportional hazard models. Working paper, Free University, Amsterdam.
- [2] Abbring, J.H. and G.J. van den Berg (2003). The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society, Ser. B*, 65, 701–710.
- [3] Abbring, J.H. and G.J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika*, March 2007, 94, 87 - 99.
- [4] Abrevaya, Jason and J. Hausman (1999). Semiparametric estimation with mismeasured dependent variables: an application to duration models for unemployment spells. *Annales D’Economie et de Statistique*, 55–56, pp. 243–75.
- [5] Aitchison, J. and S.D. Silvey (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika*, 44, 131-140.
- [6] Albert, J.H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669-679.
- [7] Albert, J.H. and S. Chib (1995). Bayesian residual analysis for binary response regression models. *Biometrika*, 82, 747-759.
- [8] Albert, J.H. and S. Chib (2001). Sequential ordinal modeling with applications to survival data. *Biometrics*, 57, 829–836.
- [9] Baba, N. and H. Goko (2006). Survival analysis of hedge funds. Bank of Japan Working Paper Series.
- [10] Baker, M. and S. Rea (1998). Employment spells and unemployment insurance eligibility requirements. *Review of Economics and Statistics*, 80, pp. 80–94.
- [11] Baker, M. and A. Melino (2000). Duration dependence and nonparametric heterogeneity: A Monte Carlo study. *Journal of Econometrics*, 96, 357-393.
- [12] Basu, S., and S. Chib (2003). Marginal likelihood and Bayes factors for Dirichlet process mixture models. *Journal of the American Statistical Association*, 98, 224–235.

- [13] Bearnse, P., J. Canals and P. Rilstone (1998). Consistent standard errors for semiparametric duration models with unobserved heterogeneity. *Economics Letters*, 59, 153–156.
- [14] Berrington, A. and I. Diamond (2000). Marriage or cohabitation: a competing risks analysis of the first-partnership formation among the 1958 Birth Cohort. *Journal of the Royal Statistical Society, Ser. A*, 163, pp. 127–152.
- [15] Brinch, C.N. (2007). Nonparametric identification of the mixed hazards model with time-varying covariates. *Econometric Theory*, 23, 349–354.
- [16] Brooks, S.P. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, Volume 7, Number 4, 434–455.
- [17] Burda, M., M. Harding and J. Hausman (2008). A Bayesian mixed logit-probit model for multinomial choice. *Journal of Econometrics*, 147, 232-246.
- [18] Burtless, G. (1983). Why is insured unemployment so low? *Brookings Papers on Economic Activity*, 1, 225-253.
- [19] Butler, R.J. and J.D. Worrall (1985). Work injury compensation and the duration of non-work spells. *Economic Journal*, 85, pp. 714-724.
- [20] Butler, R.J. and J.B. McDonald (1986). Trends in unemployment duration data. *The Review of Economics and Statistics*, 68, 545-557.
- [21] Campolieti, C. (2001). Bayesian semiparametric estimation of discrete duration models: an application of the Dirichlet process prior. *Journal of applied econometrics*, 16, 1–22.
- [22] Campolieti, C. (2003). On the estimation of hazard models with flexible baseline hazards and nonparametric unobserved heterogeneity. *Economics Bulletin*, Vol. 3, No. 24, 1–10.
- [23] Carlin, B. P. and T. A. Louis (2000). *Bayes and empirical Bayes methods for data analysis*, 2nd ed. Boca Raton: Chapman and Hall-CRC Press.
- [24] Chernozhukov, V. and H. Hong (2002). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, Vol. 97, No. 459, Theory and Methods, 872-882.
- [25] Chernozhukov, V. and H. Hong (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115, 293-346.

- [26] Chib, S. and E. Greenberg (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician*, 49, 327–335.
- [27] Chib, S. and E. Greenberg (1998). Analysis of multivariate probit models. *Biometrika*, 85, 347-361.
- [28] Chib, S. and B.P. Carlin (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, 9, 17–26.
- [29] Chib, S. and B. Hamilton (2000). Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics*, 97, 25–50.
- [30] Chib, S. and E. Greenberg (2007). Semiparametric modeling and estimation of instrumental variable models. *Journal of Computational and Graphical Statistics*, Vol. 16, No. 1, 86–114.
- [31] Chib, S., E. Greenberg and I. Jeliazkov (2008). Estimation of semiparametric models in the presence of endogeneity and sample selection. Washington University and University of California, Irvine. Working paper.
- [32] Cornfield, J (1957). The estimation of the probability of developing a disease in the presence of competing risk. *American J Public Health*, 47, pp. 601-607.
- [33] Cowles, M.K. (1996). Accelerating Monte Carlo Markov Chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, 6, 101-111.
- [34] Cox, D.R. (1959). The analysis of exponentially distributed lifetimes with two types of failure. *Journal of the Royal Statistical Society, Series B*, 21, pp. 411–421.
- [35] Cox, D.R. (1962). *Renewal theory*. London: Methuen.
- [36] Cox, D.R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, 34, pp. 187-202.
- [37] Cox, D.R. (1975). Partial likelihood. *Biometrika*, 62, pp. 269-276.
- [38] Cox, D.R. and D. Oakes (1984). *Analysis of survival data*. Monographs on Statistics and Applied Probability. London: Chapman and Hall.
- [39] Cox, D.R. and E.J. Snell (1989). *Analysis of binary data*. 2nd ed. London: Chapman & Hall.
- [40] Crowder, M. (2001). *Classical competing risks*. New York: Chapman and Hall/CRC.

- [41] Cutler, D.M. (1995). The incidence of adverse medical outcomes under prospective payment. *Econometrica*, 63:1, pp. 29-50.
- [42] David, H.A. and M.L. Moeschberger (1978). *The theory of competing risks*. New York: Macmillan.
- [43] Deng, Y., J.M. Quigley and R. Van Order (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68, 275–307.
- [44] Diamond, P. and J. Hausman (1984). Retirement and unemployment behavior of older men. In H. Aaron and G. Burtless (eds.), *Retirement and Economic Behavior*, Washington: Brookings Institution, 97-135.
- [45] Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89, 268–277.
- [46] Escobar, M.D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577–588.
- [47] Escobar, M.D. and M. West (1998). Computing nonparametric hierarchical models. *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer-Verlag, 1–22.
- [48] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1, 209–230.
- [49] Flinn, C.J. and J.J. Heckman (1982). New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics*, 18, 115–168.
- [50] Gail, M.H. (1975). A review and critique of some models used in competing risk analysis. *Biometrics*, 31, pp. 209-222.
- [51] Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.
- [52] Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In *Computing Science and Statistics, Volume 23, Proceedings of the 23rd Symposium on the Interface* (E. M. Keramidas and S. M. Kaufman, eds.), 571-578.
- [53] Geweke, J. (1996). Bayesian inference for linear models subject to linear inequality constraints. In J. C. Lee, W. O. Johnson, and A. Zellner (Eds.), *Modeling and prediction honoring seymour geisser*, 248–263. New York: Springer.
- [54] Goffe, W., G.D. Ferrier and J. Rogers (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60, 65-99.

- [55] Greene, W.H. (2000). *Econometric analysis*. Fourth edition. NJ: Prentice-Hall.
- [56] Ham, J. and S. Rea (1987). Unemployment insurance and male unemployment duration in Canada. *Journal of Labor Economics*, vol. 5, pp. 325–53.
- [57] Han, A. and J.A. Hausman (1986). Semiparametric estimator of duration and competing risk models. MIT working paper.
- [58] Han, A. and J.A. Hausman (1988). Identification of continuous and discrete competing risk models. mimeo.
- [59] Han, A. and J.A. Hausman (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5, 1-28.
- [60] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97–109.
- [61] Hausman, J.A. and T.M. Woutersen (2005). Estimating a semi-parametric duration model without specifying heterogeneity. MIT and Johns Hopkins University, working paper.
- [62] Heckman, J. and B. Singer (1984). A method for minimizing the distributional assumptions in econometric models for duration data. *Econometrica*, 52, 271-320.
- [63] Heckman, J.J. and B.E. Honoré (1989). The identifiability of the competing risks model. *Biometrika*, 76, 325–330.
- [64] Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica*, Vol. 70, No. 2, 781-799.
- [65] Honoré, B.E. and A. Lleras-Muney (2006). Bounds in competing risks models and the war on cancer. *Econometrica*, 74, 1675–1698.
- [66] Horowitz, J.L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica*, 67, 1001–1028.
- [67] Hougaard, P. (1984). Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, Vol. 71, No. 1, 75-83.
- [68] Hougaard, P. (2000). *Analysis of multivariate survival data*. New York: Springer-Verlag.
- [69] Ibrahim, J.G., M.H. Chen and D. Sinha (2001). *Bayesian survival analysis*. Springer, New York.

- [70] Imai, K., Y. Lu and A. Strauss (2006). Bayesian and likelihood inference for 2×2 ecological tables: an incomplete data approach. Princeton University and University of Colorado at Boulder, working paper.
- [71] Johnson, V.E. and J.H. Albert (1999). Ordinal data modeling. New York: Springer-Verlag.
- [72] Kalbfleisch, J.D. and R.L. Prentice (1980). The statistical analysis of failure time data. New York: Wiley.
- [73] Kalbfleisch, J.D. and R.L. Prentice (2002). The statistical analysis of failure time data. Second edition. New York: John Wiley and Sons.
- [74] Kaplan, E.L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *J Am Stat Assoc*, 53, 457-81.
- [75] Katz, L.F. (1986). Layoffs, recall and the duration of unemployment. NBER Working Paper.
- [76] Katz, L.F. and B.D. Meyer (1990). Unemployment insurance, recall expectations, and unemployment outcomes. *Quarterly Journal of Economics*, 105, 973–1002.
- [77] Kiefer, N.M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, Vol. 26, No. 2, 646-679.
- [78] Koop, G. and J. Tobias (2006). Semiparametric Bayesian inference in smooth coefficient models. *Journal of Econometrics*, 134, 283–315.
- [79] Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47, 939-956.
- [80] Lancaster, T. (1990). The econometric analysis of transition data. Cambridge: Cambridge University Press.
- [81] Lawless, J. (2003). Statistical models and methods for lifetime data. Second edition. Hoboken, NJ: Wiley.
- [82] Lindley, D.V. and A.F.M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 1, 1-41.
- [83] Liu, J.S. and C. Sabatti (2000). Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*, Vol. 87, No. 2, 353-369.
- [84] McAuliffe, J.D., D.M. Blei and M.I. Jordan (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Stat Comput*, 16, 5–14.

- [85] McCall, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica*, 64, 647-682.
- [86] McCullagh, P. and J.A. Nelder (1989). *Generalized linear models*. Second edition. London: Chapman and Hall.
- [87] Meyer, B.D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58, 757-782.
- [88] Moffitt, R. (1985). Unemployment insurance and the distribution of unemployment spells. *Journal of Econometrics*, 28, 85-101.
- [89] Montgomery, A.L., V. Zarnowitz, R.S. Tsay and G.C. Tiao (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association*, 93, 478-493.
- [90] Mosler, K. and T. Philipson (1999). Specification analysis in mixed hazard models and a test of crossing survival functions. *Statistical Papers*, 40, 37-54.
- [91] Muller, G. and C. Czado (2005). An autoregressive ordered probit model with application to high-frequency financial data. *Journal of Computational and Graphical Statistics*, Volume 14, Number 2, 320-338.
- [92] Munkin, M.K. and P.K. Trivedi (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics*, 143, 334-348.
- [93] 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.
- [94] Narendranathan, W. and M.B. Stewart (1993). How does the benefit effect vary as unemployment spells lengthen? *Journal of Applied Econometrics*, Vol. 8, No. 4, 361-381.
- [95] Ondrich J. and K. Prasad (1997). A duration model with unobserved heterogeneity as a mixture of Dirichlet processes. *Economics Letters*, Vol. 55, Issue 1, 19-25.
- [96] Peterson, A.V. (1976). Bounds for a joint distribution with fixed sub-distribution functions: application to competing risks. *Proceedings of the National Academy of Science*, 73, 11-13.
- [97] Prentice, R. and L. Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57-67.
- [98] Robert, C.P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, 5, 121-125.

- [99] Seal, H.L. (1977). Studies in the history of probability and statistics. XXXV. Multiple decrements or competing risks. *Biometrika*, 64(3), 429-439.
- [100] Sueyoshi, G. (1992). Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of Econometrics*, 51, 25-58.
- [101] Tanner, M.A. and W.H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.
- [102] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72, 20-22.
- [103] Tsurumi, H. (2001). Lecture notes on econometrics II. Rutgers University.
- [104] Van den Berg, G.J. (2001). Duration models: specification, identification, and multiple duration. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume V, Chapter 55*, pp. 3381-3460. Amsterdam: North-Holland.
- [105] Vaupel, J., K. Manton and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, Vol. 16, No. 3, 439-454.
- [106] Vaupel, J. and A.I. Yashin (1983). The deviant dynamics of death in heterogeneous populations. Research Report, RR-83-1, 1-35.
- [107] Verbeke, G. and E. Lesaffre (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91, 217-221.
- [108] Zeng, D. and D.Y. Lin (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Statist. Soc. B*, 69, Part 4, 507-564.

Curriculum Vitae

Huiying Zhang

Education

- 1996 B.A. in Tourism Economics, summa cum laude, Huaqiao University, China.
- 1997 B.A. in Business Administration, second degree, Huaqiao University, China.
- 2002 M.A. in Economics, Rutgers University, New Brunswick, New Jersey.
- 2009 Ph.D. in Economics, Rutgers University, New Brunswick, New Jersey.

Working experience

- 1997-2000 Assistant Accountant, Department of Finance, Xiamen Airlines Ltd., China.

Publication

- Comparison of Tourism Economics in Mainland China, Hong Kong and Taiwan (with Jin Chi, in Chinese), Journal of Huaqiao University (Philosophy & Social Sciences), 1997.