Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Bayesian Data Combination Approach for Repeated Durations under Unobserved Missing Indicators: Application to Interpurchase-Timing in Marketing

Ryosuke Igari , Takahiro Hoshino

28 April, 2017
DP2017-015
https://ies.keio.ac.jp/en/publications/7670/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
28 April, 2017

Bayesian Data Combination Approach for Repeated Durations under Unobserved Missing Indicators: Application to Interpurchase-Timing in Marketing
Ryosuke Igari, Takahiro Hoshino
Keio-IES DP2017-015
28 April, 2017
JEL classification: C11; C41; M31
Keywords: Intermittent Missingness; Quasi-Bayes; Latent Variable Modeling; Population-Level Information; Dirichlet Process Mixture

Abstract

In this study, we focus on intermittent missingness in repeated duration analysis, which is common in applied studies but has not rigorously been considered in statistics. Under intermittent missingness, whether any missing events exist between two observed events is unknown. In other words, the missing indicators are never observed. Thus, if there exist any missing events between two observed events, we observe only the cumulated duration between two or more events.
We propose a quasi-Bayes estimation method that utilizes population-level information to identify unobserved intermittent missingness. The proposed model consists of the following: (1) latent variable model, (2) latent missing indicator model which separates true and composite duration, (3) mixtures of duration models and (4) moment restriction from population-level information to deal with nonignorable intermittent missingness. We use a new estimation procedure that combines objective functions of likelihood and GMM simultaneously with latent variables, which we call Bayesian data combination. We apply the proposed model to analyze interpurchase-duration in database marketing using purchase-history data in Japan, which capture purchase incidences and purchase stores.

Ryosuke Igari
Graduate School of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
igariryosuke@keio.jp

Takahiro Hoshino
Faculty of Economics, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
hoshino@econ.keio.ac.jp

Bayesian Data Combination Approach for Repeated Durations under Unobserved Missing Indicators: Application to Interpurchase-Timing in Marketing

Ryosuke IGARI
Graduate School of Economics, Keio University
and
Takahiro HOSHINO
Department of Economics, Keio University

Abstract

In this study, we focus on intermittent missingness in repeated duration analysis, which is common in applied studies but has not rigorously been considered in statistics. Under intermittent missingness, whether any missing events exist between two observed events is unknown. In other words, the missing indicators are never observed. Thus, if there exist any missing events between two observed events, we observe only the cumulated duration between two or more events. We propose a quasi-Bayes estimation method that utilizes population-level information to identify unobserved intermittent missingness. The proposed model consists of the following: (1) latent variable model, (2) latent missing indicator model which separates true and composite duration, (3) mixtures of duration models and (4) moment restriction from population-level information to deal with nonignorable intermittent missingness. We use a new estimation procedure that combines objective functions of likelihood and GMM simultaneously with latent variables, which we call *Bayesian data combination*. We apply the proposed model to analyze interpurchase-duration in database marketing using purchase-history data in Japan, which capture purchase incidences and purchase stores.

Keywords: Intermittent Missingness; Latent Variable Modeling; Quasi-Bayes; Population-Level Information; Dirichlet Process Mixture

1 Introduction

Duration analysis, which captures the time-to events, is widely used in various areas including biostatistics, engineering, economics, and marketing. We focus on duration analysis for repeated events (Andersen and Gill 1982; Sinha 1993; Sinha and Dey 1997; Bijwaard *et al.* 2006), such as clinical trials, unemployment or interpurchase-timing. Especially in repeated measurement data, missing data often become a problem. In duration analysis studies, this problem has been extensively considered. The major issues regarding the missingness in repeated measurements including duration analysis are censoring, missing covariates, and dropouts (Little 1995; Diggle and Kenward 1994; Ibrahim *et al.* 2001; Klein and Moeschberger 2005). However, there is a small number of studies that deal with intermittent missing data in repeated measurement (Gad and Ahmed 2007; Wand *et al.* 2010; Qin *et al.* 2016). Previous studies, which focus on the linear or binomial regression models, consider cases in which researchers can observe the missing indicators that reveal the presence of missing events between two observed events. However, as we show in detail, in many application settings, missing indicators are not observed, which can yield severely biased estimates especially in duration analysis, because some distinct true durations may be summed up to one observed duration. Despite its importance in application studies, the intermittent missingness in repeated duration analysis is not adequately considered and studied. In this study, we focus on the intermittent missingness in duration analysis with repeated measurements.

Figure 1 depicts the intermittent missingness in repeated duration analysis. In the figure, the black cross-marks show observed events and the gray cross-marks show unobserved events, called intermittent missingness. Under these conditions, researchers can observe only black cross-marks, and not gray cross-marks. If any missing events exist between two observed events, the observed duration is not the true duration. In other words, we observe only the cumulated duration for two or more events. Therefore, ignoring intermittent missingness will lead to biased estimates and incorrect interpretations about the effects of some important covariates. The intermittent missing problem can arise in various areas such as medical research (e.g., a mild spasm or taking over-the-counter drugs outside hospitals) and animal ecological studies (e.g., reproductive behavior), which cannot be known to researchers. For illustrative purposes, we now deal with the interpurchase-timing model in marketing. Consumers purchase various products from different companies or stores such as supermarkets, drugstores, and corner stores. By analyzing the purchase histories including interpurchase timing, marketers plan various marketing interventions such as coupons or direct mail. However, the purchase data consist of only customers' behaviors from the company's own stores. Therefore, data on consumers' behaviors in competing stores are unavailable. As shown in detail in Section 5, output from analyzing this incomplete data may lead to biased estimates and incorrect decision-making.

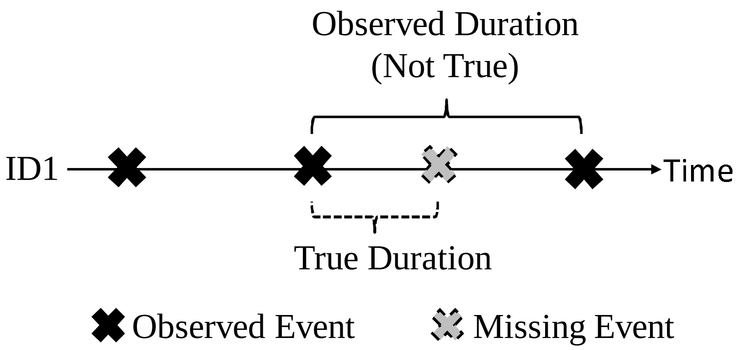

Figure 1: Intermitternt Missingness in Repeated Duration Analysis

Under intermittent missingness in repeated duration analysis, neither the covariates nor the incidences of events can be observed. These features result in difficulty in dealing with intermittent missingness. There is one study which deals with intermittent missingness in repeated duration analysis by using latent class model (Lin *et al.* 2004). Their model is a generalization of Little(1993)'s pattern mixture model, and they called their model the latent pattern mixture model (LPMM). It should be noted that Lin *et al.* (2004) merely modeled the joint distribution of the longitudinal outcomes and the number of events by using the latent class model, in which the parameters of duration analysis differ among classes. It is not guaranteed that the method will provide the class without the intermittent missingness, so Lin *et al.* (2004) simply refer to the intermittent missingness from the outputs of estimated classes retroactively. Intermittent missingness cannot be fixed by the latent class model which has common membership indicators among the same individual, because the intermittent missingness does not depend only on individuals but on the both individual and each of their events. Moreover, the latent class model itself is not able to identify intermittent missingness in repeated duration analysis because of the individual's heterogeneity in the length of durations. That is to say, it is likely that both individuals who have averagely long durations without intermittent missingness and individuals who have averagely short durations with some intermittent missingness exist. Under these conditions, even if we consider individual's heterogeneities in latent class model like Lin *et al.* (2004), we are unable to identify the differences of the length of durations caused by individual's heterogeneities or intermittent missingness. It is easily shown that the existing approach, including Lin *et al.* (2004), can yield severe biased estimates (see Section 4).

In this study, we propose a Bayesian estimation method to deal with intermittent missingness when missing indicators are not observed. Concretely, we incorporate population-level information into individual-level duration analysis to identify the observations that do not have any intermittent missingness. In real data, the observed individual-level data are likely to be biased for various reasons including selection bias and nonignorable missingness, causing the resulting estimates to be biased. Therefore, complete data, without any biases, are occasionally unavailable, and analysis from biased data would lead to biased estimators. On the other hand, researchers can sometimes obtain population-level information from other sources, such as government statistics or research institutions. Population-level information given by other sources is limited to summary statistics such as averages or proportions of variables, and the parameters which shows the relationships between variables cannot be given to researchers. This information cannot be incorporated into the prior distribution of Bayesian modeling, except for simple probability distribution models, like normal or binomial distributions in which averages or proportions mean the parameters of models directly. In this setup, some studies use population-level information to strengthen the accuracy of individual-level data models. Imbens and Lancastor (1994) and Hellerstein and Imbens (1999) propose the method of incorporating the population-level information into individual-level models using the generalized method of moment (GMM). Similarly, Qin (2000) and Chaudhuri *et al.* (2008) propose empirical likelihood approaches that include population-level information in individual-level modeling. We can regard these methods as a kind of data combinations especially used in economic fields (Ridder and Moffit 2007). From the viewpoint of missing data analysis, Nevo (2003), Qin (2000) and Qin and Zhang (2007) use population-level information to strengthen the analysis of data with missing responses, using GMM or empirical likelihood. However, their models deal with missing data problems in which missing indicators are observed. Thus, no studies deal with repeated duration analysis with unobserved missing indicators. We use population-level information to deal with missing data problems that have unobserved missing indicators.

However, problems exist with the related literature regarding data combinations. First, there are no studies which deal with latent variables in the model. Unobserved heterogeneities, such as multi-level models, random effect models, generalized linear mixed models or frailty models, and hierarchical Bayes models are very important in individual-level modeling. Latent variable modeling is very important in recent empirical studies, and therefore we include latent variables in data combinations. Further, there is a need to use latent variables to identify the durations without intermittent missingness. Second, there are computational problems in the complex models. For the estimation of flexible models, Bayesian estimation using Markov chain Monte Carlo (MCMC) is effective and widely used (Gelman *et al.* 2013; Koop *et al.* 2007). However, moment restrictions such as Imbens and Lancastor (1994) are not utilized in Bayesian methods. In such a situation, Chernozhukov and Hong (2003) propose the quasi-Bayes method using MCMC, which can use different types of objective functions such as GMM, M-estimators or empirical likelihoods instead of likelihood function. Similarly, Yin (2009) and Li and Jiang (2016) proposed Bayesian GMM as the same framework as Chernozhukov and Hong (2003). In this study, we use a new estimation procedure which can incorporate latent variables into the model and combine the likelihood and objective function of GMM, which enables us to estimate the repeated duration model with unobserved missing indicators by MCMC.

The remainder of the article is organized as follows. In Section 2, we develop a repeated duration model with nonignorable intermittent missingness. Section 3 provides an estimation procedure for Bayesian data combinations mixing objective functions of likelihood and GMM with latent variables using MCMC. Section 4 provides a summary of the simulation study. Section 5 presents application of the proposed model to interpurchase-timing in marketing. Here we show that analysis of the database that records the purchase histories of only their own stores underestimates the effects of price for interpurchase-timing. A concluding remark is given in Section 6.

2 Model

2.1 Whole Structure

In this study, we express the distribution of observed durations as a mixture distribution that consists of two distributions: (1) the distribution without intermittent missingness between two observed events and (2) the distribution with intermittent missingness between two observed events. This idea is similar to the LPMM by Lin *et al.* (2004), but is different in that our model identifies whether each observed duration includes intermittent missingness using data combinations, and estimate parameters with imposing the restrictions on moment conditions in GMM.

We define the latent missing indicator z_{ij} for individual i's$(i = 1, 2, ..., n)$ jth$(j = 1, 2, ..., J_i)$ event.

$$z_{ij} = \begin{cases} 1 & \textit{There exists no intermittent missingness between observed event } j-1 \textit{ and } j \\ 0 & \textit{There exists intermittent missingness between observed event } j-1 \textit{ and } j \end{cases}$$

Note that z_{ij} can be regarded as the membership indicator in the latent class model. We call the duration in $z_{ij} = 1$ *true duration*, and the duration in $z_{ij} = 0$ *composite duration* hereafter.

Let individual i's jth time-to-event y_{ij}, the latent missing indicator z_{ij} , time-varying covariates vector $\boldsymbol{x_{ij}}, \boldsymbol{w_{ij}}$, individual-level covariates vector $\boldsymbol{d_i}$, latent variable f_i, and structural parameters $\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\xi}, \boldsymbol{\eta}$,

$$\begin{aligned} p(y_{ij}|\boldsymbol{x_{ij}}) = \int p(f_i|\boldsymbol{d_i}, \boldsymbol{\eta}) \Big[& p(z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta}) \\ & + p(z_{ij} = 0|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) p(y_{ij}|z_{ij} = 0, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\xi}) \Big] df_i. \end{aligned} \tag{1}$$

The proposed model consists of four components: (1) latent variable model $p(f_i|\boldsymbol{d_i}, \boldsymbol{\eta})$, (2) latent missing indicator model $p(z_{ij}|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma})$ which separates true and composite durations, (3) mixtures of duration model $p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})$ and $p(y_{ij}|z_{ij} = 0, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\xi})$

and (4) moment restriction for some group-level information(eg., gender or age ranges). We now explain each model.

2.2 Submodels

Latent Variable Model

Latent variable f_i is included in $p(z_{ij}|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma})$, $p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})$ and $p(y_{ij}|z_{ij} = 0, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\xi})$. We assume that f_i is explained by individual-level covariates $\boldsymbol{d_i}$,

$$f_i \sim N(\boldsymbol{d_i'}\boldsymbol{\eta}, \sigma_f^2). \tag{2}$$

Here we set $\sigma_f^2 = 1$ simply.

Latent Missing Indicator Model to Separate True and Composite Duration

The probability that individual i's jth event belongs to the latent missing indicator $z_{ij} = 1$ is modeled via a logistic regression model using the time-varying covariates $\boldsymbol{w_{ij}}$ and the latent variable f_i,

$$p(z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) = \frac{exp(\gamma_0 + \boldsymbol{w_{ij}'}\boldsymbol{\gamma_1} + f_i)}{1 + exp(\gamma_0 + \boldsymbol{w_{ij}'}\boldsymbol{\gamma_1} + f_i)}. \tag{3}$$

Here we set $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma_1'})'$. However, this model cannot be estimated appropriately with only observed data because of the identification problem. Therefore, it is necessary to include the population-level information in our model and estimate this model with imposing the restrictions on the moment conditions.

Duration Model

We consider two models for the duration model, the cases of $z = 1$ and $z = 0$. We interpret the parameter $\boldsymbol{\theta}$ of the latent missing indicator $z = 1$ as a result without any bias by intermittent missingness. We are not interested in the parameter $\boldsymbol{\xi}$ in $z = 0$ and do not interpret it.

(i)Duration without Intermittent Missingness ($z = 1$)

We assume the parametric hazard model for the time-to-event y_{ij} of the individual i's jth event using the time-varying covariates vector $\boldsymbol{x_{ij}}$, latent variable f_i, and parameter $\boldsymbol{\theta}$.

The general hazard function is

$$h(y_{ij}|z_{ij}=1,\boldsymbol{x_{ij}},f_i,\boldsymbol{\theta})=h_0(y_{ij})exp(\beta_0+\boldsymbol{x'_{ij}}\boldsymbol{\beta_1}+\rho f_i). \tag{4}$$

If we employ a parametric baseline hazard function for $h_0(y_{ij})$, this model accommodates exponential, Weibull, and log-normal survival model. We can regard this model as a Bayesian *frailty model*, which is common in survival analysis (Clayton 1991; Sinha 1993; Ibrahim *et al.* 2005; Dunson and Herring 2005; Pennell and Dunson 2006). Additionally, we assume $\rho > 0$ as a theory constraint.

The probability density function (pdf) is represented by the multiplication of hazard function and survival function.

$$p(y_{ij}|z_{ij}=1,\boldsymbol{x_{ij}},f_i,\boldsymbol{\theta})=h(y_{ij}|z_{ij}=1,\boldsymbol{x_{ij}},f_i,\boldsymbol{\theta})S(y_{ij}|z_{ij}=1,\boldsymbol{x_{ij}},f_i,\boldsymbol{\theta}) \tag{5}$$

(ii)Duration with Intermittent Missingness ($z=0$)

The observed durations y_{ij} in $z=0$ may be cumulated between two or more events, but how many missing events there are is not known. Therefore, the simple duration model $p(y_{ij}|z_{ij}=0,\boldsymbol{x_{ij}},f_i,\boldsymbol{\xi})$ may cause model misspecification. We assume nonparametric Bayes model to avoid model misspecification by using the Dirichlet process mixture (DPM) model (Ishwaran and James 2001; Dunson *et al.* 2007; Dunson and Park 2008; Chung and Dunson 2009; Hoshino 2013).

$$p(y_{ij}|z_{ij}=0,\boldsymbol{x_{ij}},f_i,\boldsymbol{\xi})=\sum_{k=1}^{\infty}\pi_k p(y_{ij}|z_{ij}=0,z_{ij}^D=k,\boldsymbol{x_{ij}},f_i,\boldsymbol{\xi_k}) \tag{6}$$

Here, k means a component in DPM and does not show the number of unobserved intermittent missingness. Additionally, π_k is a probability weight for component k and z_{ij}^D is the individual i's jth membership indicator which belongs to k-th component when $z_{ij}=0$. This hierarchical structure for the latent missing indicator model and DPM is similar to a two-stage or multilevel latent class model. We use Ishwaran and James (2001)'s types of DPM. The stick-breaking representation is by letting,

$$\pi_1=V_1,\ \pi_k=V_k\prod_{h=1}^{k-1}(1-V_h),\ V_k\sim Be(1,r). \tag{7}$$

Here, we let $r=1$ (Ishwaran and James 2001; Gelman *et al.* 2013).

2.3 Incorporating Population-Level Information

For identification of the true duration model, we use two types of population-level information, the proportion without intermittent missingness z^* and average duration y^* by using moment restrictions (Imbens and Lancastor 1994; Chaudhuri *et al.* 2008). In this study, z^* and y^* are considered to be known exactly, but the proposed method is easily generalized to deal with the case by using statistics from external surveys.

Let the moment restrictions $\boldsymbol{m}(z_{ij}|\boldsymbol{\gamma})$ using the proportion without intermittent missingness for group $s(s = 1, 2, ..., S)$ z_s^*

$$\boldsymbol{m}(z_{ij}|\boldsymbol{\gamma}) = \begin{cases} I_{ij}^1 \left[z_1^* - E[z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}] \right] \\ \cdots \\ I_{ij}^S \left[z_S^* - E[z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}] \right] \end{cases},$$

here $I_{ij}^s = 1$ when individual i's jth event belongs to group s.

Next, we use the auxiliary information y_s^* to identify the parameters of the true duration distribution. The moment restrictions $\boldsymbol{m}(y_{ij}|\boldsymbol{\theta})$ are by letting

$$\boldsymbol{m}(y_{ij}|\boldsymbol{\theta}) = \begin{cases} I_{ij}^1 \left[y_1^* - E[y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})] \right] \\ \cdots \\ I_{ij}^S \left[y_S^* - E[y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})] \right] \end{cases}.$$

3 Estimation Method

Here we introduce a new method for Bayesian data combinations. We propose the new quasi-Bayes method and MCMC algorithms, which can deal with latent variables and have hybrid posteriors with likelihood functions and moment restrictions. See the supplementary material for proof of consistency and asymptotic properties of the estimator.

3.1 Hybrid Posterior Combining Likelihood and Objective Function of GMM

The usual quasi-Bayes method is represented as follows.

$$p(\boldsymbol{\theta}|\boldsymbol{Y}) = \frac{exp\{L_n(\boldsymbol{\theta})\}p(\theta)}{\int_{\boldsymbol{\Theta}} exp\{L_n(\boldsymbol{\theta})\}p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto exp\{L_n(\boldsymbol{\theta})\}p(\boldsymbol{\theta}) \tag{8}$$

where $p(\boldsymbol{\theta}|\boldsymbol{Y})$ is a posterior distribution for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{\theta}$, $\boldsymbol{\Theta}$ is the parameter space of $\boldsymbol{\theta}$ and $L_n(\boldsymbol{\theta})$ is an objective function such as GMM, M-estimators, or empirical likelihoods instead of log-likelihood functions (Chernozhukov and Hong 2003; Hoshino 2008; Yin 2009; Yang and He 2012).

The posterior expectations of quasi-Bayes are represented as follows.

$$\hat{\boldsymbol{\theta}} = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{Y}) d\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} \left(\frac{exp\{L_n(\boldsymbol{\theta})\} p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} exp\{L_n(\boldsymbol{\theta})\} p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \right) d\boldsymbol{\theta} \tag{9}$$

If we use the GMM type of objective function, the objective function is defined as follows.

$$L_n(\boldsymbol{\theta}) = -\frac{n}{2} \Big(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta}) \Big)' \boldsymbol{\Omega_n} \Big(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta}) \Big) \tag{10}$$

Here, $\boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta})$ is a moment restriction and $\boldsymbol{\Omega_n}$ is a weight matrix, $\boldsymbol{\Omega_n} = E\big(\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})'\big)^{-1}$. We can incorporate population-level information into $\boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta})$ like Imbens and Lancastor (1994).

However, the previously proposed quasi-Bayes estimation methods do not consider latent variable modeling (LVM). How to incorporate latent variables is not known nor is the validity of resulting method. In traditional Bayesian estimation with LVM, by treating latent variables as incidental parameters and drawing samples from the full conditional distribution of the latent variable, we can avoid evaluation of marginal likelihood in which latent variables are integrated out (e.g., Tanner and Wong 1987; Albert and Chib 1993). It would also be useful to treat latent variables as incidental parameters in quasi-Bayes computation. Additionally even if we can incorporate latent variables into a GMM-type quasi-Bayes method, the method would have a heavy computational load. It is not realistic in practice because we must compose several moment restrictions in accordance with the number of latent variables, which is proportional to the sample size n. If we can express the full quasi-posterior distribution as the product of the parametric likelihood $p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{f})$ and the function regarding moment restriction $L_n(\boldsymbol{\theta})$, it is easy to draw the MCMC runs. That is, we set quasi-Bayes posterior distribution $q(\boldsymbol{\theta}, \boldsymbol{f}|\boldsymbol{Y})$ with latent variable $\boldsymbol{f}$ as follow:

$$q(\boldsymbol{\theta}, \boldsymbol{f}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{f}) \times exp\big\{L_n(\boldsymbol{\theta})\big\} \times p(\boldsymbol{\theta}) \times p(\boldsymbol{f}|\boldsymbol{\theta}). \tag{11}$$

Here, $p(\boldsymbol{f}|\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{f}$. We call this format *hybrid posterior* thereafter and this form enables us to conduct *Bayesian data combinations.* Thus, we propose hybrid methods of full-Bayes and quasi-Bayes that include latent variables in the model.

3.2 Moment Restriction with Latent Variables

The expectations of each parametric hazard model are shown in Klein and Moeschberger (2005). Here, we integrate out the latent variable in moment restriction.

The expectations of the proportion without intermittent missingness is

$$\begin{aligned} E[z_{ij}=1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}]] &= \int \frac{exp(\gamma_0+\boldsymbol{w'_{ij}\gamma_1}+f_i)}{1+exp(\gamma_0+\boldsymbol{w'_{ij}\gamma_1}+f_i)} p(f_i|\boldsymbol{\eta},\sigma^2)df_i \\ &\simeq \frac{1}{L}\sum_{l=1}^{L}\frac{exp(\gamma_0+\boldsymbol{w'_{ij}\gamma_1}+f_i^l)}{1+exp(\gamma_0+\boldsymbol{w'_{ij}\gamma_1}+f_i^l)}. \end{aligned} \tag{12}$$

We draw the latent variable by using Monte Carlo simulations, that is $f_i^l \sim N(\boldsymbol{d'_i\eta}, \sigma^2)$.

Similarly, when we use Weibull hazard model, the expectation is

$$\begin{aligned} E[y_{ij}|z_{ij}=1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})] &= \int \Gamma\big(1+\alpha^{-1}\big)exp\Big\{-\frac{\beta_0+\boldsymbol{x'_{ij}\beta_1}+\rho f_i}{\alpha}\Big\}p(f_i|\boldsymbol{\eta},\sigma^2)df_i \\ &\simeq \frac{1}{L}\sum_{l=1}^{L}\Gamma\big(1+\alpha^{-1}\big)exp\Big\{-\frac{\beta_0+\boldsymbol{x'_{ij}\beta_1}+\rho f_i^l}{\alpha}\Big\}. \end{aligned} \tag{13}$$

Here, $\Gamma()$ means the gamma function and set $\boldsymbol{\theta} = (\alpha, \beta_0, \boldsymbol{\beta'_1}, \rho)'$.

The objective functions of GMM for moment restriction parts are given by

$$\begin{aligned} L_N^z(\boldsymbol{\gamma}) &= -\frac{N}{2}\left(\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\boldsymbol{m}(z_{ij}|\boldsymbol{\gamma})\right)'\boldsymbol{\Omega_N^z}\left(\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\boldsymbol{m}(z_{ij}|\boldsymbol{\gamma})\right) \\ L_{N_{z=1}}^y(\boldsymbol{\theta}) &= -\frac{N_{z=1}}{2}\left(\frac{1}{N_{z=1}}\sum_{i=1}^{n}\sum_{j=1}^{J_i}I(z_{ij}=1)\boldsymbol{m}(y_{ij}|\boldsymbol{\theta})\right)'\boldsymbol{\Omega_{N_{z=1}}^y}\left(\frac{1}{N_{z=1}}\sum_{i=1}^{n}\sum_{j=1}^{J_i}I(z_{ij}=1)\boldsymbol{m}(y_{ij}|\boldsymbol{\theta})\right). \end{aligned} \tag{14}$$

Here, $\boldsymbol{\Omega_N^z}$ and $\boldsymbol{\Omega_{N_{z=1}}^y}$ are the optimal weight matrix.

$$\boldsymbol{\Omega_N^z} = \left[\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\boldsymbol{m}(z_{ij}|\boldsymbol{\gamma})\boldsymbol{m}(z_{ij}|\boldsymbol{\gamma})'\right]^{-1}, \ \boldsymbol{\Omega_{N_{z=1}}^y} = \left[\frac{1}{N_{z=1}}\sum_{i=1}^{n}\sum_{j=1}^{J_i}I(z_{ij}=1)\boldsymbol{m}(y_{ij}|\boldsymbol{\theta})\boldsymbol{m}(y_{ij}|\boldsymbol{\theta})'\right]^{-1} \tag{15}$$

Here, $N = \sum_i^n \sum_j^{J_i} I$, $N_{z=1} = \sum_i^n \sum_j^{J_i} I(z_{ij}=1)$.

3.3 Total Objective Function

The objective function combining likelihood and GMM for the proposed model is

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{x}) \propto \int \prod_{i=1}^{n} p(f_i|\boldsymbol{d_i}, \boldsymbol{\eta}) \prod_{j=1}^{J_i} \Big[& p(z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta}) \\ & + p(z_{ij} = 0|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) \sum_{k=1}^{\infty} \pi_k p(y_{ij}|z_{ij} = 0, z_{ij}^D = k, f_i, \boldsymbol{\xi_k}) \Big] df_i \\ & \times exp\big\{L_N^z(\boldsymbol{\gamma}))\big\} \times exp\big\{L_{N_{z=1}}^y(\boldsymbol{\theta})\big\}. \end{aligned} \tag{16}$$

3.4 MCMC Implementation

We draw samples using MCMC.

Draw $\boldsymbol{\theta}(z = 1)$

The posterior distribution is combined with likelihood and GMM ($z = 1$) and draw $\boldsymbol{\theta}$ from hybrid posterior using the Metropolis-Hastings algorithms.

$$\begin{aligned} q(\boldsymbol{\theta}|y_{ij}, z_{ij} = 1, \boldsymbol{x_{ij}}, f_i) = & \prod_{i=1}^{n} \prod_{j=1}^{J_i} \Big\{ p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})^{I(z_{ij}=1)} \Big\} \\ & \times p(\boldsymbol{\theta}) \times exp\big\{L_{N_{z=1}}^y(\boldsymbol{\theta})\big\} \end{aligned} \tag{17}$$

Here $p(\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{\theta}$.

Draw $\boldsymbol{\gamma}$

The posterior distribution is combined with likelihood and GMM and draw $\boldsymbol{\gamma}$ from hybrid posterior using the Metropolis-Hastings algorithms.

$$q(\boldsymbol{\gamma}|z_{ij}, \boldsymbol{w_{ij}}, f_i) = \prod_{i=1}^{n} \prod_{j=1}^{J_i} \Big\{ p(z_{ij}|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) \Big\} \times p(\boldsymbol{\gamma}) \times exp\big\{L_N^z(\boldsymbol{\gamma}))\big\} \tag{18}$$

Here $p(\boldsymbol{\gamma})$ is a prior distribution for $\boldsymbol{\gamma}$.

Draw f_i

The posterior distribution is simple likelihood and the objective function of GMM is integrated out, thus we draw f_i from simple posterior using the Metropolis-Hastings algorithms.

$$\begin{aligned} p(f_i|\cdot) = \prod_{j=1}^{J_i} \Big[& p(z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta}) \\ & + p(z_{ij} = 0|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) \sum_{k=1}^{\infty} \pi_k p(y_{ij}|z_{ij} = 0, z_{ij}^D = k, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\xi_k}) \Big] p(f_i|\boldsymbol{d_i}, \boldsymbol{\eta}) \end{aligned} \tag{19}$$

Draw z_{ij} and z_{ij}^D

The posterior distribution is simple likelihood and the objective function of GMM is integrated out.

$$\begin{aligned} p(z_{ij} = 1|\cdot) &= \frac{p(z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta})}{p^*(y_{ij}, z_{ij})} \\ p(z_{ij}^D = k|\cdot) &= \frac{p(z_{ij} = 0|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) \pi_k p(y_{ij}|z_{ij} = 0, z_{ij}^D = k, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\xi_k})}{p^*(y_{ij}, z_{ij})} \end{aligned} \tag{20}$$

Here,

$$\begin{aligned} p^*(y_{ij}, z_{ij}) = & p(z_{ij} = 1|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) p(y_{ij}|z_{ij} = 1, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\theta}) \\ & + p(z_{ij} = 0|\boldsymbol{w_{ij}}, f_i, \boldsymbol{\gamma}) \sum_{k} \pi_k p(y_{ij}|z_{ij} = 0, z_{ij}^D = k, \boldsymbol{x_{ij}}, f_i, \boldsymbol{\xi_k}) \end{aligned} \tag{21}$$

Draw Other Parameters

Drawing $\boldsymbol{\xi_k}$ and V_k is the same as those of the general DPM (Ishwaran and James 2001; Gelman *et al.* 2013) and drawing $\boldsymbol{\eta}$ is the same as the usual Gibbs sampling.

In the Bayesian latent class model, label switching problems are often pointed out, and researchers usually set the restriction of one parameter, such as $\theta_1^1 < ... < \theta_k^1 < ... < \theta_K^1$ (Frühwirth-Schnatter 2001), where $k = (1, ..., K)$ means each class. In the proposed model, the moment restriction $L_N^z(\boldsymbol{\gamma})$ and $L_{N_{z=1}}^y(\boldsymbol{\theta})$ play a role on label switching restrictions. Thus, any label switching restriction is not required in the proposed model.

4 Simulation Study

In this section, we show the performance of the proposed model using simulation data. Here, we assume a Weibull distribution for baseline hazard function(Ibrahim *et al.* 2005; Klein and Moeschberger 2005),

$$h(y_{ij}|z_{ij} = 1, x_{ij}, f_i, \boldsymbol{\theta}) = \alpha y_{ij}^{\alpha-1} exp(\beta_0 + x_{ij}\beta_1 + \rho f_i) \tag{22}$$

where α is a shape parameter of the Weibull distribution and β_0 and β_1 are the coefficients for covariates. We set $\boldsymbol{\theta} = (\alpha, \beta_0, \beta_1, \rho)'$. The pdf for the Weibull hazard model is represented,

$$p(y_{ij}|z_{ij} = 1, x_{ij}, f_i, \boldsymbol{\theta}) = \alpha y_{ij}^{\alpha-1} exp(\beta_0 + x_{ij}\beta_1 + \rho f_i) exp\Big\{ - exp(\beta_0 + x_{ij}\beta_1 + \rho f_i) y_{ij}^{\alpha} \Big\}. \tag{23}$$

Here we estimate four models to show the performance of the proposed model: (1) *Frailty*, (2) *LPMM*, (3) *DPM* and (4) *Proposed* models. *Frailty* is the general Weibull hazard model with unobserved heterogeneity, ignoring the intermittent missingness. *LPMM* is the latent pattern mixture model (Lin *et al.* 2004) with two classes of which membership indicator is common in each individual across different events, that is $z_{i1} = z_{i2} = ... = z_{iJ_i} \equiv z_i$. In *LPMM*, we incorporate latent variables into the latent class model like Lin *et al.*(2004). We pick up estimated parameters from the latent class which estimates the shortest duration in all classes. *DPM* is the Dirichlet process mixture model without the population-level information. *Proposed* is the full model with population-level information, which has hybrid posterior and DPM.

Next, we set the conditions of the simulation: (1) sample size $n = 500$, (2) the average number of events $\bar{J}_i = 10$, (3) missing rate (MR)$= 20\%, 40\%, 60\%$, (4) the 200 simulation sets in each model.

We show the MSEs of simulation study in Table 1. In Table 1, we show only the results of the estimators of the parameters in the latent missing indicator $z = 1$ of *DPM* and *Proposed*. Table 1 also shows the average MSEs and ratio of MSEs that are scaled based on *Proposed* ($MSE_{Proposed} = \alpha, \beta_0, \beta_1$), which are common to each model. The proposed model outperforms the existing models and is the only one that yields unbiased estimates.

Next, we show the box plots of common parameters α, β_0, β_1 for each model in Figure 2. From this, we can show reproducibility of parameters. In the *Frailty* model, parameters cannot be estimated appropriately in each missing rate. The results show that ignoring intermittent missingness will lead to biased estimates. In *LPMM*, the results are not close to true parameters even in low missing rates. As the intermittent missingness does not depend on only individuals but on both the individual and each of thier events, *LPMM* does not work appropriately. Though we show the results of two classes for LPMM, we also simulate this model from two to ten classes and confirm LPMM does not work appropriately. The results of the *DPM* model are close to those of *Proposed*, but results show that even when low missing rate, the estimated parameters have biases from the true values. The reason why the *DPM* model does not work appropriately is lack of identification for the latent missing indicator models and distributions without intermittent missingness. Finally, the *Proposed* model performs the best of the four models. *Proposed* can estimate parameters appropriately even with a high missing rate. From this, we can determine that previous models are not able to deal with intermittent missingness appropriately.

Table 1: Simulation Results (MSE$\times 10^2$)

		Frailty	LPMM	DPM	Proposed
$MR = 20\%$	α	17.45	24.31	0.58	0.05
	β_0	56.54	137.54	106.45	2.50
	β_1	16.46	251.45	2.20	0.06
	average MSE	30.15	57.42	36.41	0.87
	MSE Ratio	214.56	1620.58	30.64	1.00
$MR = 40\%$	α	18.88	2.18	0.35	0.19
	β_0	45.35	29.82	136.41	1.25
	β_1	20.95	315.84	1.90	0.16
	average MSE	28.39	10.90	46.22	0.53
	MSE Ratio	88.55	655.45	40.74	1.00
$MR = 60\%$	α	15.92	0.38	0.38	0.36
	β_0	11.32	83.46	173.45	1.87
	β_1	21.97	307.42	1.97	0.24
	average MSE	16.41	28.06	58.60	0.82
	MSE Ratio	46.90	435.88	34.00	1.00

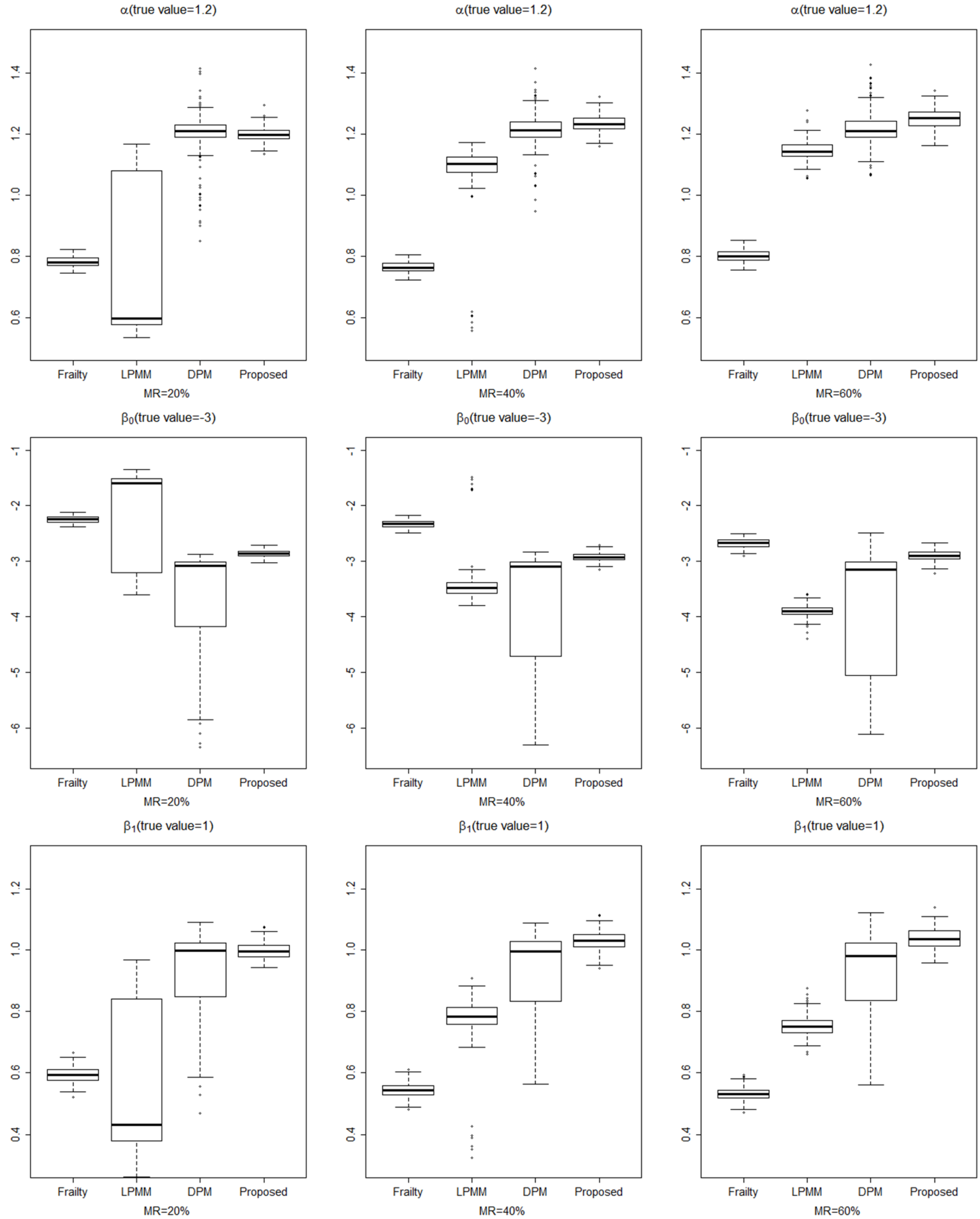

Figure 2: Box plots of Simulation Results

5 Application to Interpurchase-Timing Modeling

5.1 Interpurchase-Timing in Marketing and Incomplete Data

We apply our model to interpurchase-timing analysis in marketing. Many studies deal with interpurchase-time using duration analysis (Jain and Vilcassim 1991; Helsen and Schmittlein 1993; Allenby *et al.* 1999; Seetharaman and Chintagunta 2003; Bijwaard *et al.* 2006). The role of duration analysis in marketing is estimating the effect of marketing promotions such as price coupons and predicting the time at which consumers are highly likely to purchase products. Marketing managers use their own databases which record purchase histories in their own stores, and this is called *Database Marketing* (eg., Payne and Frow 2005). In *Database Marketing*, marketers conduct some forms of marketing activities such as offering price coupons or sending direct mails to customers, depending on each customers' previous purchase records. The goal is to stop customers from switching to competing stores or to encourage consumers to purchase more products in their own stores. However, databases usually record the purchase histories of only their own store chains and lack those of competing stores. Therefore, they are likely to be incomplete and companies need to predict consumers' purchase behavior from only available data.

Our model enables us to estimate parameters exactly using incomplete data, and helps marketing manager decide marketing activities. For clarity of explanation, we consider the two customers in Figure 3. Black cross-marks mean observed purchase for their own store in some product categories, and gray cross-marks mean unobserved purchase from competing stores. In this situation, the marketer may offer price coupons to customers who have a low frequency of purchase. Both ID1 and ID2 customers purchase products three times from observed data, but ID2 customer have purchased more, though some of purchases occur in competing stores. These purchases are unobserved by the marketer. If the company gives a price coupon to the ID1 customer, it is difficult for the ID1 customer to purchase more products, because they have no potential capacity in this product category. However, if company gives a price coupon to the ID2 customer, it is possible for the ID2 customer to purchase more products, because they have potential capacity in this product category, and coupons may cause store switching from the competing stores.

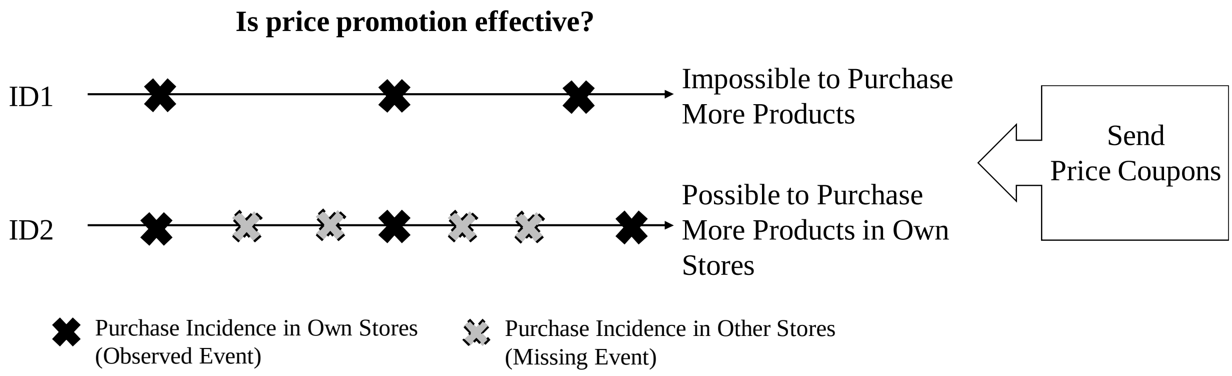

Figure 3: Intermittent Missingness in Direct Marketing

5.2 Data Description

Purchase History Data

We use the *Survey of Consumer Index* (SCI) data provided by Intage Inc in Japan. The SCI is scanner panel data that is the de facto standard for purchase panel data in the Japanese marketing field. The SCI records the purchase incidence, purchase product, number of products consumer purchases, amounts, prices, and purchase stores with date and time. The data record the store names when purchase events occur. Though the scanner panel data are recorded for purchase histories in competing store chains, we regard it to be a database from a particular store, which is incomplete and lacks information on competing stores. We assume that, though the purchase incidences in each store are observed, the purchase incidence in other competing stores can not be observed. To make inference from this incomplete data, we utilize auxiliary information by aggregating the complete data, in this case to identify unobserved intermittent missingness.

In the analysis, we use the purchase data of the haircare category that consists of shampoo, hair rinse, and hair treatment. We analyze data from the January 2015 to June 2016 period. From the purchase data, we pick up consumers who purchase products in this category more than 3 times within a period. We show histograms of observed duration in Figure 4. We choose a drugstore chain which is very familiar in Japan in analysis. We select a sample size (n = 967) and total events number (= 5157) for the estimation of parameters.

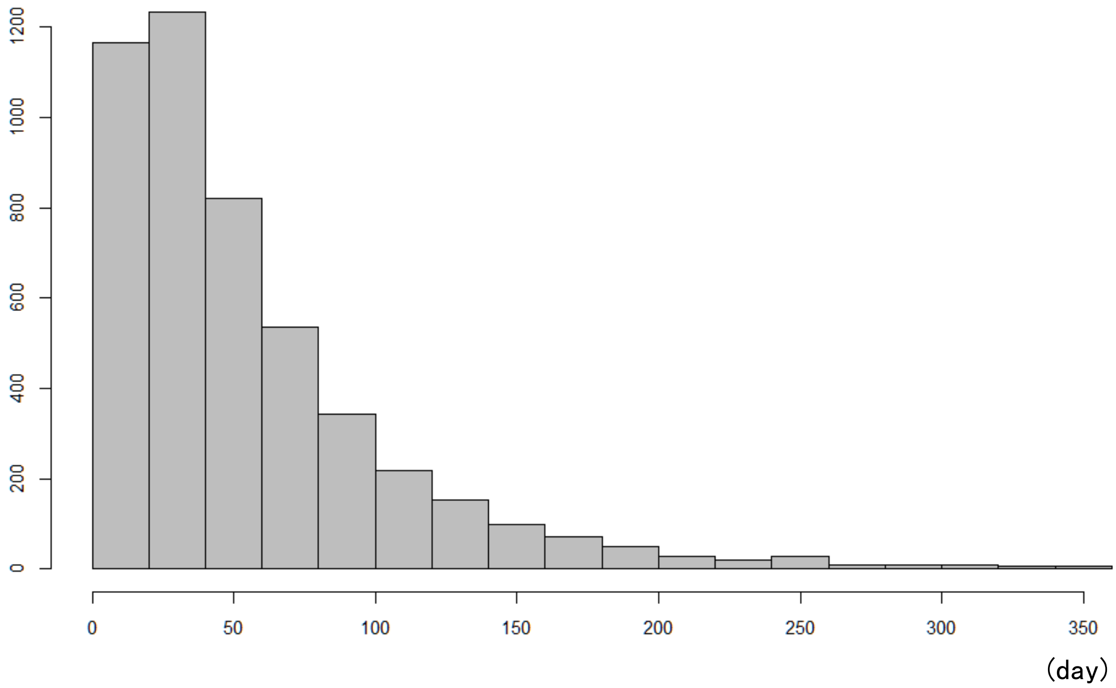

Figure 4: Histograms of Observed Duration

Covariates and Population-Level Information

Next, we define the covariates and auxiliary information used in analysis. We show the summary statistics in Table 2.

First, we show the covariates $\boldsymbol{x}$ in the repeated duration model.

$$\boldsymbol{x} = \begin{bmatrix} Price \\ log(Previous\ Amount) \\ log(Previous\ Duration) \end{bmatrix}.$$

Here "Price " is scaled and equals 1 when the price in the purchase incidence equals to regular price. The coefficients of "Price" should be negative, because consumers are likely to visit stores when price discounts are available. In marketing fields, since the effects of price discounts are very important, we compare "Price" coefficients with competing models.

Second, we show the covariates $\boldsymbol{w}$ in a logistic regression model. We use the previous visit frequency in this store chain (not only for this category but for all products), logarithms for the average of the previous duration for this category, and store loyalty.

$$\boldsymbol{w} = \begin{bmatrix} Previous\ Visit\ Frequency \\ log(Average\ of\ Previous\ Duration) \\ Store\ Loyalty \end{bmatrix}.$$

Third, we show the individual-level covariates $\boldsymbol{d}$. We use gender (male 1), age, family size (1,2, etc.), child and job(fulltime 1).

$$\boldsymbol{d} = \begin{bmatrix} Gender(Male = 1) \\ Age \\ Family\ Size \\ Child \\ Job(Fulltime = 1) \end{bmatrix}.$$

Finally, we show the population-level information for proportions without intermittent missingness z_s^* and duration y_s^*. We use 6 total demographic groups for ranges of age.

$$s = \begin{bmatrix} All \\ Age\ 20s \\ Age\ 30s \\ Age\ 40s \\ Age\ 50s \\ Age\ 60s\ or\ more \end{bmatrix}.$$

5.3 Results

For real data analysis, we use an exponential hazard model, which is a special type of the Weibull hazard model, in which the shape parameter α is fixed to one. We estimated five models: (1) *Complete Data*, (2) *Frailty*, (3) *LPMM*, (4) *DPM* and (5) *Proposed* models. In addition to four models in the simulation study, we estimated *Complete Data* of which the results from analyzing the data without any intermittent missing. Since we have purchase records from competing stores, we can check the accuracy of the estimated parameters by comparing them with the *Complete Data*. In each model, we draw 5,000 MCMC iterations

Table 2: Summary Statistics

		Mean
Basic Information		
(Observed)	Average Frequency	5.3
	Observed Duration	56.6
Auxiliary Information		
	True Duration	43.3
	Observed Rate	0.72
Covariates x		
	Price Scale	1.23
	Previous Amounts	686.1
	Previous Duration	56.9
Covariates w		
	Frequency of Previous Visit	28.6
	Average of Previous Duration	57.5
	Chain Loyalty	0.72
Demographic d		
	gender(Male=1)	0.10
	Age	41.1
	Family Size	3.3
	Child	0.52
	Job(Fulltime=1)	0.31

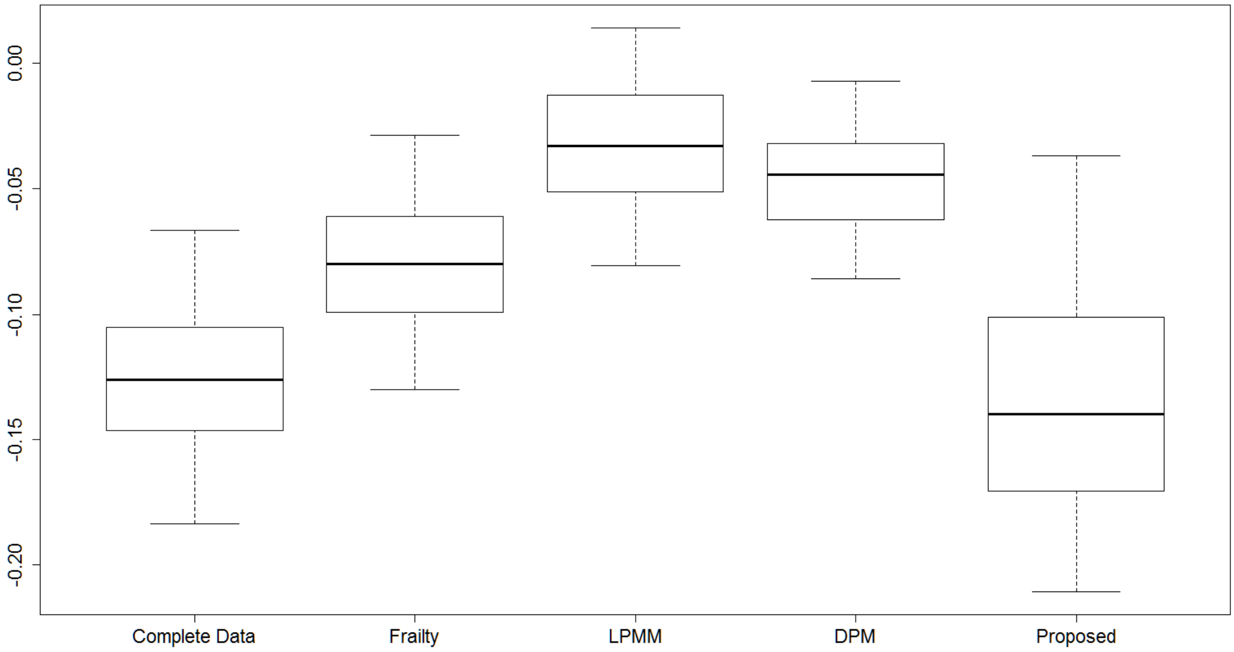

Figure 5: Estimated Price Coefficients

after 15,000 burn-in phase. In the *DPM* and *Proposed* models, we set $M = 10$ as the number of components for the Dirichlet process mixture model. We confirm the convergence of each model by the Geweke (1992) method.

We show the box plots for "Price" coefficients in each model in Figure 5. From this, it is obvious that the effects of "Price" are underestimated, except for in the *Proposed* model with respect to *Complete Data*. The results may cause marketing managers to disregard price discounts for consumers because effects of price promotions are underestimated.

Next, we confirm that moment restrictions function appropriately. We show the auxiliary information z^* and estimated proportion without intermittent missingness of *Proposed* in Table 3. We use the posterior means for the estimated values. Though the estimated proportions without intermittent missingness are not completely consistent with auxiliary information z^*, the estimated values are very close to population-level information.

For prediction of the intermittent missingness, we use the results of logistic regression models. We show the coefficients of $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ in Table 4. Table 4 shows the posterior mean, posterior standard deviation, and significant test from a 95% credible interval, which shows "*" when the 95% credible interval does not include zero. In the logistic regression model,

Table 3: Auxiliary Information z^* and Estimated Proportions without Intermittent Missingness in *Proposed*

	Auxiliary Information	Estimated Proportions
All	0.706	0.754
Age 20s	0.767	0.779
Age 30s	0.714	0.754
Age 40s	0.667	0.758
Age 50s	0.700	0.713
Age 60s	0.763	0.693

coefficients of all covariates are significant. The effect of "Previous Visit Frequency" is estimated to be negative. We interpret this as frequent consumers in this store chain make high-involvement purchases in this category, may cause them to visit other competing stores frequently and cause intermittent missingness. On the other hand, the effects of "Store Loyalty" are estimated to be positive. We can understand that higher the store loyalty of consumers, the lower the proportions with intermittent missingness. Additionally in the effects of demographic variable, male consumers, young consumers and fulltime-working consumers are likely to visit other store chains.

6 Conclusion

In this study, we proposed a duration model with repeated events, which has unobserved intermittent missingness using hybrid posterior incorporating population-level information. The proposed model consists of four components: (1) the latent variable model, (2) the latent missing indicator model which separates true and composite duration, (3) mixtures of duration models and (4) moment restriction from population-level information to deal with intermittent missingness directly. Additionally, we propose the hybrid quasi-Bayes method to estimate parameters of the proposed model that incorporates population-level informa-

Table 4: Coefficients for $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ in *Proposed* Model

	mean	(sd)	
$\boldsymbol{\gamma}$			
Intercept	0.609	(0.046)	*
Previous Visit Frequency	-0.023	(0.009)	*
log(Average of Previous Duration)	0.056	(0.018)	*
Store Loyalty	2.040	(0.021)	*
$\boldsymbol{\eta}$			
Gender(Male=1)	-0.335	(0.203)	*
Age	-0.013	(0.006)	*
Family Size	-0.002	(0.064)	
Child	0.180	(0.155)	
Job(Fulltime=1)	-0.316	(0.134)	*

tion into duration models with latent variables. From the simulation study and real data analysis, we show that ignoring the intermittent missingness in repeated measurement data may result in severely biased estimates. To show the usefulness of our model, we applied the proposed model to interpurchase-timing in marketing, in which we can trace the whole purchase histories observed both in own stores and competing stores. We select and use only the purchase histories observed in own stores to mimic the situation in database marketing. We confirm that the proposed model can estimate the coefficients of the duration model appropriately compared to the results from the complete data.

Our model can be applied to other issues in marketing. For example, we apply our model to internet marketing using web access data. In internet circumstances, the complete data on consumer's website browsing behaviors cannot be collected because consumers visit websites of competing companies and purchase products. On the other hand, the proposed model can be applied to other research fields such as social and natural sciences. For example, in medical statistics, researchers often use longitudinal data about clinical trial for patients, but such data often record histories within the limited medical institution and patients may go to another clinic or take over-the-counter drugs. In this situation, researchers may underestimate the effects of therapy programs, since there exists unobserved events between observed events. Additionally, in economics, researchers use panel data on factors such as job employment, marriage, and wages. Here, incomplete data problems can occur in the same way. We can strengthen incomplete observed data using population-level information from government statistics or other research institutes.

In future research, we shall extend our model to include the dynamic latent missing indicator models. In this study, we assume that the latent missing-indicators are independent between each event for the same individual. We intend to relax this restriction and permit the dependence of dynamic relationships by using Markov models or hidden Markov models, which will allow us to capture the dynamics of intermittent missingness. Additionally, we shall extend our model to Bayesian Cox's proportional hazard model of which the baseline hazard function is a nonparametric model such as a gamma process model (Kalbfleisch 1978; Ibrahim *et al.* 2005).

References

Albert, J. H., and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass*, **88**, 669-679.

Allenby, G. M., Leone, R. P., and Jen, L. (1999). A Dynamic Model of Purchase Timing with Application to Direct Marketing. *J. Am. Statist. Ass*, **94** , 365-374.

Andersen, P. K., and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist*, **10** , 1100-1120.

Bijwaard, G. E., Franses, P. H., and Paap, R. (2006). Modeling Purchases as Repeated Events. *J. Bus. Econ. Stat*, **24**. 487-502.

Chaudhuri, S., Handcock, M. S., and Rendall, M. S. (2008). Generalized linear models incorporating population level information: an empirical-likelihood-based approach. *J. R. Statist. Soc. B*, **70**, 311-328.

Chernozhukov, V., and Hong, H. (2003). An MCMC approach to classical estimation. *J. Econom*, **115**, 293-346.

Chung, Y., and Dunson, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *J. Am. Statist. Ass*, **104**, 1646-1660.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, **47**, 467-485.

Diggle, P., and Kenward, M. G. (1994). Informative Drop-Out in Longitudinal Data Analysis. *Appl. Stat*, **43**, 49-93.

Dunson, D. B., and Herring, A. H. (2005). Bayesian latent variable models for mixed discrete outcomes. *Biostatistics*, **6**, 11-25.

Dunson, D. B., and Park, J. H. (2008). Kernel stick-breaking processes. *Biometrika*, **95**, 307-323.

Dunson, D. B., Pillai, N., and Park, J. H. (2007). Bayesian density regression. *J. R. Statist. Soc. B*, **69**, 163-183.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *J. Am. Statist. Ass*, **96**, 194-209.

Gad, A. M., and Ahmed, A. S. (2007). Sensitivity analysis of longitudinal data with intermittent missing values. *Stat. Methodol*, **4**, 217-226.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis 3rd.* London: Chapman & Hall/CRC.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, Bernardo, J. M., J. O. Berger, A. P. Dawid, A. F. M. Smith (ed.), *Bayesian Statistics 4* , Oxford University Press.

Hellerstein, J. K., and Imbens, G. W. (1999). Imposing moment restrictions from auxiliary data by weighting. *Rev. Econ. Stat*, **81**, 1-14.

Helsen, K, and Schmittlein, D. C. (1993). Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models. *Market. Sci*, **11**, 395-414.

Hoshino, T. (2008). A Bayesian propensity score adjustment for latent variable modeling and MCMC algorithm. *Comput. Stat. Data. Anal*, **52**, 1413-1429.

Hoshino, T. (2013). Semiparametric Bayesian estimation for marginal parametric potential outcome modeling: Application to causal inference. *J. Am. Statist. Ass*, **108**, 1189-1204.

Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2001). Missing Responses in Generalised Linear Mixed Models When the Missing Data Mechanism is Nonignorable. *Biometrika*, **88**, 551-564.

Ibrahim, J. G., Chen, M. H., Sinha, D. (2005). *Bayesian Survival Analysis*. John Wiley and Sons, Ltd.

Imbens, G. W., and Lancaster, T. (1994). Combining Micro and Macro Data in Microeconometric Models. *Rev. Econ. Stud*. **61**, 655-680.

Ishwaran, H., and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Ass*, **96**, 161-173.

Jain, D, C., and Vilcassim, N. J. (1991). Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach. *Market. Sci*, **10**, 1-23.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *J. R. Statist. Soc. B*, **40**, 214-221.

Klein, J. P., and Moeschberger, M. L. (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian Econometric Methods*. Cambridge University Press.

Li, C., and Jiang, W. (2016). On oracle property and asymptotic validity of Bayesian generalized method of moments. *J. Multivar. Anal*, **145**, 132-147.

Lin, H., McCulloch, C. E., and Rosenheck, R. A. (2004). Latent Pattern Mixture Models for Informative Intermittent Missing Data in Longitudinal Studies. *Biometrics*, **60**, 295-305.

Little, R. J. (1993). Pattern-Mixture Models for Multivariate Incomplete Data. *J. Am. Statist. Ass*, **88**, 125-134.

Little, R. J. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *J. Am. Statist. Ass*, **90**, 1112-1121.

Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *J. Bus. Econ. Stat*, **21**, 43-52.

Payne, A., and Frow, P. (2005). A strategic framework for customer relationship management. *J. Mark*, **69**, 167-176.

Pennell, M. L., and Dunson, D. B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics*, **62**, 1044-1052.

Qin, J. (2000). Combining Parametric and Empirical Likelihoods. *Biometrika*, **87**, 484-490.

Qin, L., Weissfeld, L., Levine, M., Marcus, M., and Dai, F. (2016). Latent Variable Model for Weight Gain Prevention Data with Informative Intermittent Missingness. *J. Mod. Appl. Stat. Methods*, **15**, 627-642.

Qin, J., and Zhang, B. (2007). Empirical- likelihood- based inference in missing response problems and its application in observational studies. *J. R. Statist. Soc. B*, **69**, 101-122.

Ridder, G., and Moffitt, R. (2007). The econometrics of data combination. in *Handbook of Econometrics*, **6**, 5469-5547.

Seetharaman, P. B., and Chintagunta, P. K. (2003). The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications. *J. Bus. Econ. Stat*, **21**, 368-382.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *J. Am. Statist. Ass*, **88**, 979-983.

Sinha, D., and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *J. Am. Statist. Ass*, **92**, 1195-1212.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Ass*, **82**, 528-540.

Wang, C., Daniels, M. J., Scharfstein, D. O., and Land, S. (2010). A Bayesian shrinkage model for incomplete longitudinal binary data with application to the breast cancer prevention trial. *J. Am. Statist. Ass*, **105**, 1333-1346.

Yang, Y., and He, X. (2012). Bayesian empirical likelihood for quantile regression. *Ann. Statist*, **40**, 1102-1131.

Yin, G. (2009). Bayesian generalized method of moments. *Bayesian. Anal*, **4**, 191-207.