Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

A Bayesian Gamma Frailty Model Using the Sum of Independent Random Variables: Application of the Estimation of an Interpurchase Timing Model

Ryosuke Igari、Takahiro Hoshino

17 December, 2018
DP 2018-021
https://ies.keio.ac.jp/en/publications/10558/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
17 December, 2018

A Bayesian Gamma Frailty Model Using the Sum of Independent Random Variables: Application of the Estimation of an Interpurchase Timing Model

Ryosuke Igari、Takahiro Hoshino

Keio-IES DP2018-021
17 December, 2018
JEL Classification: C11; C41; M31
Keywords: Survival Analysis; Random Effects; Auxiliary Information; Quasi-Bayesian Inference; Markov Chain Monte Carlo

Abstract

In statistics, researchers have rigorously investigated the reproductive property, which maintains that the sum of independent random variables with the same distribution follows the same family of distributions. However, even if a distribution of the sum of random variables demonstrates the reproductive property, estimating parameters appropriately from only summed observations is difficult. This is because of identification problems when component random variables have different parameters. In this study, we develop a method to effectively estimate parameters from the sum of independent random variables with different parameters. In particular, we focus on the sum of Gamma random variables composed of two types of distributions. We generalize the result according to Moschopoulos(1985) to a proportional hazard model with covariates and a frailty model to capture individual heterogeneities. Additionally, to estimate each parameter from the sum of random variables, we incorporate auxiliary information using quasi-Bayesian methods, and we propose the estimation procedure by Markov chain Monte Carlo. We confirm the effectiveness of the proposed method through a simulation study and apply it to the interpurchase timing model in marketing.

Ryosuke Igari
Faculty of Business Administration, Hosei University
2-17-1 Fujimi, Chiyoda-ku, Tokyo
r-igari@hosei.ac.jp

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

Acknowledgement : This work was supported by JSPS KAKENHI 18H05689 and 18H03209 .

A Bayesian Gamma Frailty Model Using the Sum of Independent Random Variables: Application of the Estimation of an Interpurchase Timing Model

Ryosuke Igari

Faculty of Business Administration, Hosei University

RIKEN Center for Advanced Intelligence Project

E-mail: r-igari@hosei.ac.jp

Takahiro Hoshino

Graduate School of Economics, Keio University

RIKEN Center for Advanced Intelligence Project

E-mail: hoshino@econ.keio.ac.jp

Abstract

In statistics, researchers have rigorously investigated the reproductive property, which maintains that the sum of independent random variables with the same distribution follows the same family of distributions. However, even if a distribution of the sum of random variables demonstrates the reproductive property, estimating parameters appropriately from only summed observations is difficult. This is because of identification problems when component random variables have different parameters. In this study, we develop a method to effectively estimate parameters from the sum of independent random variables with different parameters. In particular, we focus on the sum of Gamma random variables composed of two types of distributions. We generalize the result according to Moschopoulos (1985) to a proportional hazard model with covariates and a frailty model to capture individual heterogeneities. Additionally, to estimate each parameter from the sum of random variables, we incorporate auxiliary information using quasi-Bayesian methods, and we propose the estimation procedure by Markov chain Monte Carlo. We confirm the effectiveness of the proposed method through a simulation study and apply it to the interpurchase timing model in marketing.

Keyword: *Survival Analysis; Covariates; Random Effects; Auxiliary Information; Quasi-Bayesian Inference; Markov Chain Monte Carlo*

1 Introduction

1.1 Background

In statistics, the "reproductive property," which holds that the sum of random variables that follow a distribution also follow the same family of distributions, is widely known. For example, in a normal distribution, when $X_i \sim N(\mu_i, \sigma_i^2)$, the sum of random variables follows the normal distribution, that is $\sum_{i=1}^n X_i \sim N\big(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\big)$. If $(\mu_1, \sigma_1^2)^T = (\mu_2, \sigma_2^2)^T = \cdots = (\mu_n, \sigma_n^2)^T \equiv (\mu, \sigma^2)^T$, the parameters $(\mu, \sigma^2)^T$ can be estimated from the sum of random variables, and the confidential intervals for sample mean $\overline{X}$ can be calculated as $\overline{X} \sim N\big(\mu, \sigma^2/n\big)$. However, when $(\mu_1, \sigma_1^2)^T \neq (\mu_2, \sigma_2^2)^T \neq \cdots \neq (\mu_n, \sigma_n^2)^T$, it is not possible to identify each parameter from only the sum of observations without additional information. In this study, we propose a method to appropriately estimate each parameter from only the sum of random variables, where each random variable follows the same distribution family but each distribution has different parameters. In particular, we focus on the Gamma distribution that is composed of two types of distributions with different parameters, and we apply it to survival analysis (e.g., Klein and Moeschberger, 2003; Ibrahim *et al.*, 2001) with covariates in which a Gamma distribution is employed as a baseline hazard function. Moreover, we expand our model to the Gamma frailty model.

In marketing, Gamma distribution is widely used in lifetime value analysis (e.g., Schmittlein, 1987; Moe and Fader, 2004; Fader and Hardie, 2009). Additionally, survival analysis, including the Gamma hazard model, has been widely used to capture interpurchase timing (e.g., Allenby *et al.*, 1999; Seetharaman *et al.*, 2003; Bijwaard *et al.*, 2006). Interpurchase timing models can predict the next time a customer will purchase products or can estimate the effects of marketing variables such as price. Generally, interpurchase timing models focus on purchase duration analysis in a store or on one product category purchase in-store. That is to say, the general interpurchase timing model deals with univariate survival times in a store or product category. However, we should consider purchase behavior not only within one store or product category but also among stores (competing stores, for example), because customers can purchase products in more than one store. When considering stores, we should model a multivariate survival analysis including competing stores. Leszczyc *et al.* (2000) proposed the interpurchase timing model to consider competing stores using multi-state hazard models. The parameters in these multi-state models are estimated differently for each event (store), that is, their model assumes that each store has a different hazard function. Thus, naturally, each realized duration value is generated from other distributions with different parameters, whereby the idea of the reproductive property of a probability distribution is necessary.

However, the purchase histories in competing stores cannot be obtained in general marketing environments, including in database marketing (e.g., Blattberg *et al.*, 2008), and only accumulated interpurchase timing is observed in the database (Igari and Hoshino, 2018). In these circumstances, general survival analysis (e.g., a proportional hazard model) cannot be applied as is. Then, Chen *et al.* (2012) proposed a method to estimate parameters from the sum of independent duration times. They assumed that each random variable follows an independent and identical exponential distribution, and they deal with the sum of durations as an Erlang distribution based on the sum of durations. However, they held the very strong assumption that each random variable follows the independent and identical distribution, which goes against the proposal of Leszczyc *et al.* (2000) in which the parameters of each store are different. Moreover, when component random variables have different parameters, it is difficult to estimate parameters from only cumulative observations because of identification problems, even if the distribution of the sum of the random variables has the reproductive property. In that case, auxiliary information is needed to accurately estimate each parameter from the sum of random variables.

1.2 Data Combination Approach

Here, we review a method to use auxiliary information in the modeling from individual-level data. It is widely known that observed individual-level data are likely to be biased for various reasons, for example, from selection bias and nonignorable missingness. That is to say, it is difficult to obtain complete individual-level data that is free from biases. Researchers can sometimes obtain auxiliary information such as population-level information or estimated means from large-scale surveys. This information is generally available from government statistical databases or research institutions. However, this auxiliary information is often limited to summary statistics such as variable averages or proportions, whereas parameters that show relationships between dependent and independent variables are not typically available.

In this situation, some studies use auxiliary information to strengthen the accuracy of their individual-level data modeling. Imbens and Lancaster (1994) and Hellerstein and Imbens (1999) suggested using the generalized method of moments (GMM) to incorporate macro-level auxiliary information obtained from governmental surveys into individual-level models. Additionally, Igari and Hoshino (2017) expanded Imbens and Lancaster (1994)'s method to Bayesian GMM and applied it to survival analysis using incomplete data. Igari and Hoshino (2018) also employed Bayesian estimation for repeated durations with intermittent missingness. They solved the problem of missing indicators not being observable in the intermittent missingness by using the Bayesian approach. Similarly, Qin (2000), Chaudhuri *et al.* (2008), and Huang *et al.* (2016) proposed empirical likelihood approaches that include auxiliary information in individual-level modeling. These kinds of data combinations are commonly used in economics (Ridder and Moffit, 2007). Data combination involves the piecing together of data that is obtained from different sources. Combining individual-level data and auxiliary information is one type of data combination; thus, we refer to it as data combination in this study. We propose a Bayesian estimation procedure from the sum of Gamma random variables integrating auxiliary information.

1.3 The Purpose of This Study

In this study, we deal with survival analysis using observations with summed durations. Igari and Hoshino (2018) proposed a method to estimate parameters from summed durations in the form of survival analysis with repeated events using auxiliary information. However, Igari and Hoshino (2018) focused on finding durations that were not summed under unobserved missing indicators. They estimated the original parameters of Weibull hazard models using only the durations that were not summed. In other words, they did not estimate parameters directly from summed durations.

In this study, we propose a method to estimate parameters directly from the sum of independent random variables with different parameters. In particular, we consider summed durations from two types of distributions. We show a diagram of this study in Figure 1. In the figure, Y_{comp} is a random variable from a distribution with parameter θ_{comp}, and Y_{own} is a random variable from a distribution with parameter θ_{own}. However, we can obtain only the sum $Y = Y_{comp} + Y_{own}$, and we must estimate both θ_{comp} and θ_{own} only from observed Y. Particularly, we assume a Gamma distribution, and we estimate each parameter from the sum of durations following the Gamma distributions with different parameters. The Gamma distribution becomes an exponential distribution when the shape parameter equals 1; it also becomes an Erlang distribution when the shape parameter is an integer greater than or equal to 2. That is, the model using the Gamma distribution incorporates Chen et (2012)'s model, in which the Erlang distribution was assumed. Additionally, the Gamma distribution has the reproductive property when the scale parameters are common among all random variables (see Section 2).

Subsequently, Moschopoulos (1985) introduced the probability density function (pdf) for the sum of Gamma distributions whose parameters differ from one another. In this study, we generalize Moschopoulos's (1985) result to a proportional hazard model with covariates. Moreover, we expand the Gamma hazard model to a frailty model, in which the latent variables are used to capture individual heterogeneities. Unobserved heterogeneities, which can be considered random effect models, generalized linear mixed models, or hierarchical Bayes models, are important in individual-level modeling. In marketing, consumer heterogeneities have also been modeled (e.g., Allenby, and Rossi, 1998; Allenby *et al.*, 1999). However, even if the pdf of the sum of durations follows the Gamma distributions with different parameters, it is difficult to estimate each parameter from only summed durations because of identification problems. Therefore, we incorporate the auxiliary information from one distribution to estimate each parameter using the quasi-Bayesian method, as in Hoshino and Igari (2017) and Igari and Hoshino (2018). Furthermore, we propose the estimation procedure by Markov chain Monte Carlo (MCMC). The originality of our method compared to Chen *et al.* (2012) or Igari and Hoshino (2018) is that we estimate parameters from only the sum of durations following the Gamma distribution with different parameters. Though our proposed approach can be expanded to the more general model that includes more than durations, we focus only on the case in which the sum of durations is composed of two kinds of distributions. This is because we aim to apply our method to database marketing, in which one distribution is in its own database and the other refers to competing stores.

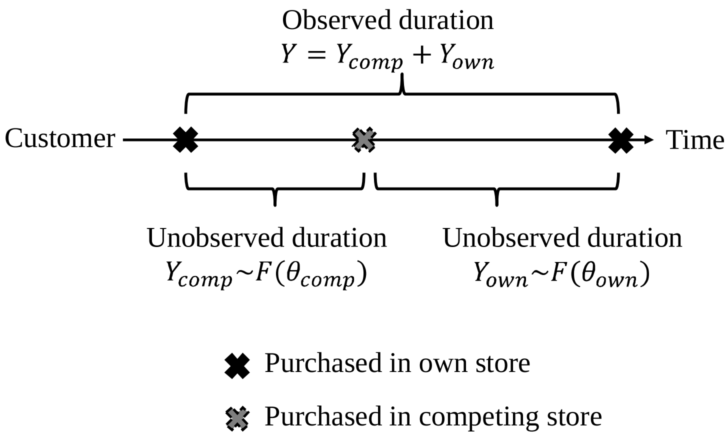

Figure 1: Sum of Durations Following Different Distributions

The remainder of the paper is organized as follows. Section 2 covers a literature review for Gamma distribution and the methods of Moschopoulos (1985). Section 3 provides a model generalizing Moschopoulos's (1985) result to a proportional hazard model with covariates and an estimation procedure by quasi-Bayesian inference using MCMC. Section 4 summarizes the simulation study and applies the proposed model to interpurchase timing in marketing.

Section 5 concludes the paper.

2 Literature Review

2.1 Gamma Distribution and Reproductivity

First of all, we introduce the Gamma hazard model. Let the y be a time-to-event following the Gamma distribution $Ga(\alpha, \lambda)$ of which the pdf is

$$p(y|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)\lambda^{\alpha}} y^{\alpha-1} exp(-y/\lambda), \tag{1}$$

where α is a shape parameter and λ is a scale parameter that are $\alpha > 0$ and $\lambda > 0$. If we incorporate the covariate vector $\boldsymbol{x}$ into the model, the scale parameter is represented as $\lambda = exp(\boldsymbol{x}^T\boldsymbol{\beta})$, which is considered a proportional hazard model. The expectation of the Gamma hazard model is $E[y|\alpha, \lambda] = \alpha\lambda$.

Here, we consider the sum of random variables following the Gamma distribution. The Gamma distribution has the reproductive property, and the sum of random variables follows the Gamma distribution only when the scale parameters λ are common among all distributions (e.g, Krishnamoorthy, 2006). When $y_k (k = 1, ..., K)$ follows $Ga(\alpha_k, \lambda)$ with $\alpha_1 \neq, ..., \neq \alpha_K$, $y = \sum_{k=1}^{K} y_k$ follows $Ga(\sum_k \alpha_k, \lambda)$. However, even when the sum of the Gamma random variables follows the Gamma distribution, the shape parameters $\alpha_1, ..., \alpha_K$ cannot be accurately estimated because of an identification problem. Besides, the scale parameter λ is assumed to be common for each distribution. It is natural that Gamma distributions that have different shape parameters have different scale parameters as $\lambda_1 \neq, ..., \neq \lambda_K$. Moreover, if we incorporate covariates in the model, the scale parameter becomes $\lambda_k = exp(\boldsymbol{x}^T\boldsymbol{\beta}_k)$, which follows a different distribution than the others. However, it is known that a Gamma distribution does not have the reproductive property when the scale parameters differ.

In response, Moschopoulos (1985) and Sim (1992) proposed the pdf for Gamma distributions with different shape and scale parameters. Moschopoulos (1985) introduced the pdf of the sum of random variables with different parameters in the Gamma distributions from the moment-generating function. We will introduce the details of Moschopoulos (1985) in the next subsection. Sim (1992) also proposed the pdf for the sum of Gamma random variables that are correlated. Johnson *et al.* (1994) provided a detailed review of the distribution of the sum of the Gamma random variables. Additionally, Nadarajah (2008) also reviewed the sum of random variables from distributions, including the Gamma distribution. In this study, we will employ the method of Moschopoulos (1985) and generalize it to the proportional hazard model and frailty model.

2.2 Moschopoulos's (1985) Method

Now we consider the time-to-events $y = \sum_{k=1}^{K} y_k$ with $y_k \sim Ga(\alpha_k, \lambda_k)$ where $\alpha_1 \neq \alpha_2 \neq ... \neq \alpha_K$ and $\lambda_1 \neq \lambda_2 \neq ... \neq \lambda_K$. Moschopoulos (1985) represented the pdf for y as follows:

$$p(y|\alpha_1, ..., \alpha_K, \lambda_1, ..., \lambda_K) = C \sum_{r=0}^{\infty} \frac{\delta_r y^{\rho+r-1} exp(-y/\omega)}{\omega^{\rho+r}\Gamma(\rho+r)}, \tag{2}$$

where $\omega = min(\lambda_1, ..., \lambda_K)$, $\rho = \sum_{k=1}^{K} \alpha_k$, $C = \prod_{k=1}^{K} (\omega/\lambda_k)^{\alpha_k}$, and

$$\begin{aligned} \delta_{r+1} &= \frac{1}{r+1} \sum_{s=1}^{r+1} s\gamma_s \delta_{r+1-s}, \quad r = 0, 1, ..., \quad \delta_0 = 1, \\ \gamma_r &= \sum_{k=1}^{K} \alpha_k (1 - \omega/\lambda_k)^r / r, \quad r = 1, 2, ... \ . \end{aligned} \tag{3}$$

From the above, we can construct the likelihood function. However, even if the pdf for the sum of durations is represented, the individual parameters cannot be estimated appropriately because of the identification problem. We will describe this in details in the simulation section. Thus, in this study, we propose an estimation method for the different parameters of the Gamma distribution using auxiliary information.

3 Proposed Method

3.1 The Gamma Frailty Model for Sum of Durations

We consider a Gamma frailty model for repeated events. Let individual $i(i = 1, 2, ..., n)$'s $j(j = 1, 2, ..., J_i)$-th time-to-event be y_{ij} with $y_{ij} = \sum_{k=1}^{K} y_{(k)ij}$ where $y_{(k)ij} \sim Ga(\alpha_k, \lambda_{(k)ij})$. Only y_{ij} can be observed and $y_{(1)ij}, ..., y_{(K)ij}$ cannot be observed. Here, α_k is a shape parameter that is $\alpha_1 \neq \cdots \neq \alpha_K$, and $\lambda_{(k)ij}$ is a scale parameter that is $\lambda_{(1)ij} \neq \cdots \neq \lambda_{(K)ij}$. We incorporate the covariate and latent variables in the scale parameter as $\lambda_{(k)ij} = exp(b_k f_i + \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_k)$. Then, f_i is a latent variable with $f_i \sim N(\mu, \sigma^2)$, $\boldsymbol{x}$ is a covariate vector excluding the intercept term, $\boldsymbol{\beta}_k$ is a coefficient vector that is $\boldsymbol{\beta}_1 \neq \cdots \neq \boldsymbol{\beta}_K$, and b_k is the coefficient for the latent variable f_i that is $b_1 \neq \cdots \neq b_K$. Here, one of b_k is fixed to 1 (e.g., $b_K = 1$). This gives one of the frailty models. We set $\boldsymbol{y}_i = (y_{i1}, ..., y_{iJ_i})^T$, $\boldsymbol{y} = (\boldsymbol{y}_1^T, ..., \boldsymbol{y}_n^T)^T$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_K)^T$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_K^T)^T$ and $\boldsymbol{b} = (b_1, ..., b_K)^T$.

The likelihood function of the proposed model is

$$
\begin{aligned}
p(\boldsymbol{y}|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{b}) &= \prod_{i=1}^{n}\int p(\boldsymbol{y}_i|\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{b},f_i)p(f_i|\mu,\sigma^2)df_i \\
&= \prod_{i=1}^{n}\int \prod_{j=1}^{J_i}\left[C_{ij}\sum_{r=0}^{\infty}\frac{\delta_{ij,r}y_{ij}^{\rho+r-1}exp(-y_{ij}/\omega_{ij})}{\omega_{ij}^{\rho+r}\Gamma(\rho+r)}\right]p(f_i|\mu,\sigma^2)df_i
\end{aligned} \tag{4}
$$

where $\omega_{ij} = min(\lambda_{(1)ij},...,\lambda_{(K)ij})$, $\rho = \sum_{k=1}^{K}\alpha_k$, $C_{ij} = \prod_{k=1}^{K}(\omega_{ij}/\lambda_{(k)ij})^{\alpha_k}$, and

$$
\begin{aligned}
\delta_{ij,r+1} &= \frac{1}{r+1}\sum_{s=1}^{r+1}s\gamma_{ij,s}\delta_{ij,r+1-s}, \quad r=0,1,...., \quad \delta_{ij,0}=1. \\
\gamma_{ij,r} &= \sum_{k=1}^{K}\alpha_k(1-\omega_{ij}/\lambda_{(k)ij})^r/r, \quad r=1,2,..... .
\end{aligned} \tag{5}
$$

However, it is difficult to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{b}$ and f_i appropriately. Thus, we estimate them using the auxiliary information.

3.2 Incorporating Auxiliary Information

The parameters can be identified to incorporate auxiliary information y^* in Moschopoulos (1985). Thus, we incorporate the auxiliary information, similar to Imbens and Lancaster(1994), Chaudhuri *et al.* (2008), and Igari and Hoshino (2017, 2018). Generally, when there are K-th types of distributions, the auxiliary information for $K-1$ distribution is needed. In this study, we consider $K=2$ hereafter. We use the auxiliary information y^* for one kind of distribution.

We set a moment condition from the auxiliary information y^* for $Ga(\alpha_1,\lambda_{(1)ij})$ by the GMM, Then, the moment condition is

$$
\boldsymbol{m}(y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1) = y^* - E[y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1] \tag{6}
$$

In practice, we use the Monte Carlo method:

$$
\begin{aligned}
E[y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1] &= \int E[y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1,f_i]p(f_i)df_i \\
&\simeq \frac{1}{L}\sum_{l=1}^{L}E[y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1,f_i^l]
\end{aligned} \tag{7}
$$

where L is the number of Monte Carlo simulation and f_i^l is generated from $f_i^l \sim N(\mu,\sigma^2)$. The expectation of the proposed model is $E[y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,f_i] = \alpha_1 exp(b_1 f_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta}_1)$.

The objective function of the moment condition is

$$L_n^* = -\frac{N}{2}\left(\frac{1}{N}\sum_{i,j}\boldsymbol{m}(y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1)\right)^T\boldsymbol{\Omega}\left(\frac{1}{N}\sum_{i,j}\boldsymbol{m}(y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1)\right), \tag{8}$$

where N is a total number that is $N=\sum_{i=1}^n J_i$ and $\boldsymbol{\Omega}$ is an optimal weight matrix that is $\boldsymbol{\Omega} = E\Big[\boldsymbol{m}(y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1)\boldsymbol{m}(y_{(1)ij}|\alpha_1,\boldsymbol{\beta}_1,b_1)^T\Big]^{-1}$. This form is a kind of quasi-Bayesian inference (e.g., Chernozhukov and Hong, 2003; Hoshino, 2008) and Bayesian GMM (Yin, 2009; Igari and Hoshino, 2017).

3.3 Quasi-Bayesian Estimation

We use the method by Hoshino and Igari (2017) and Igari and Hoshino (2018), which expanded the quasi-Bayesian inference (Chernozhukov and Hong, 2003; Hoshino, 2008; Yin, 2009). The quasi-Bayesian method permits the objective functions such as GMM or M-estimator instead of the likelihood function.

Hoshino and Igari (2017) developed the quasi-Bayesian posterior distribution by dividing the objective function of quasi-Bayes into the likelihood and the additional moment conditions from auxiliary information. Let $\boldsymbol{\theta}$ be a parameter vector for a model. Then the quasi-Bayesian posterior is

$$q(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}) \times exp\{L_n^*\} \times p(\boldsymbol{\theta}), \tag{9}$$

where $p(\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{\theta}$, Please refer to Hoshino and Igari (2017) for an introduction to this method. From this form, we can use the likelihood function to draw some parameters efficiently.

Additionally, Hoshino and Igari (2017) showed that the latent variables can be incorporated into the moment conditions. When the parameter is $\boldsymbol{\theta}$ and the latent variable is $\boldsymbol{f}$, the quasi-Bayesian posterior distribution $q(\boldsymbol{\theta},\boldsymbol{f}|\boldsymbol{y})$ with latent variable $\boldsymbol{f}$ is generally represented as

$$q(\boldsymbol{\theta},\boldsymbol{f}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{f}) \times exp\big\{L_n(\boldsymbol{\theta})\big\} \times p(\boldsymbol{\theta}) \times p(\boldsymbol{f}|\boldsymbol{\theta}), \tag{10}$$

where $p(\boldsymbol{f}|\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{f}$.

3.4 Sampling using MCMC

We draw the samples for quasi-Bayesian posterior distribution by using the MCMC.

For sampling the parameters for the Gamma duration model, $\boldsymbol{\theta}^* = \big(log(\boldsymbol{\alpha})^T, \boldsymbol{\beta}^T, \boldsymbol{b}^T\big)^T$, we use the random-walk Metropolis–Hastings (MH) algorithm to draw samples as follows:

$$\boldsymbol{\theta}^{*(cand)} \sim N(\boldsymbol{\theta}^{*(old)}, \boldsymbol{\Psi}), \tag{11}$$

where $\boldsymbol{\theta}^{*(old)}$ is the previous value in the MCMC iteration, and $\boldsymbol{\Psi}$ is a variance-covariance matrix adjusted to an appropriate acceptance rate.

Then, the probability of accepting the new candidate sample $\boldsymbol{\theta}^{*(cand)}$ is

$$min\left\{1, \frac{q(\boldsymbol{\theta}^{*(cand)}|\boldsymbol{y})}{q(\boldsymbol{\theta}^{*(old)}|\boldsymbol{y})}\right\}. \tag{12}$$

In the calculation of $q(\boldsymbol{\theta}^{*(cand)}|\boldsymbol{y})$, the latent variable in the moment condition is integrated out, and we calculate $E[y_{(1)ij}|\alpha_1, \boldsymbol{\beta}_1, b_1]$ using the Monte Carlo method in each MCMC iteration.

For sampling the latent variable f_i, we can use the general data augmentation method (Tanner and Wei, 1987; Albert and Chib, 1993),

$$p(f_i|\boldsymbol{\theta}^*, \boldsymbol{y}_i) \propto \prod_{j=1}^{J_i} p(y_{ij}|f_i, \boldsymbol{\theta}^*)p(f_i|\mu, \sigma^2), \tag{13}$$

where the term $exp\{L_n^*\}$ is canceled out when sampling f_i because f_i is integrated out in the moment condition $E[y_{(1)ij}|\alpha_1, \boldsymbol{\beta}_1, b_1]$. That is to say, we can use the usual MH algorithms here. Please refer to Hoshino and Igari (2017) for details on the quasi-Bayesian inference with latent variables.

For sampling μ and σ^2, the term $exp\{L_n^*\}$ is canceled out, and μ and σ^2 can be drawn from a full-conditional posterior distribution using a usual Gibbs sampling approach (e.g., Gelman *et al.*, 2013).

4 Application

4.1 Simulation Study

In this section, we introduce two types of simulation studies: (1) the basic Gamma hazard model without latent variables, and (2) the Gamma frailty model with latent variables. In particular, we show the performance of the proposed model in comparison with existing models. That is, we construct three models in both simulations as follows:

(1) Moschopoulos's (1985) model without auxiliary information (Non-Macro)

(2) The proposed model with one piece of auxiliary information (Proposed NMR =1)

(3) The proposed model with three pieces of auxiliary information (Proposed NMR=3) where models (2) and (3) are the proposed models, but they differ in the amount of auxiliary information. NMR means the number of moment restriction.

Simulation 1: Basic Gamma Hazard Model

In simulation 1, we evaluate a basic Gamma hazard model without latent variables, and one observation is given for each individual, that is $J_1 = \cdots = J_n \equiv 1$. Now, we explain a procedure for generating simulation data. In simulation 1, we assume the sum of durations comprises one duration from distribution A and one from distribution B: $y_{(A)i} \sim Ga(\alpha^A, \lambda_{(A)i}), \lambda_{(A)i} = exp(\beta_0^A + x_i\beta_1^A)$ and $y_{(B)i} \sim Ga(\alpha^B, \lambda_{(B)i}), \lambda_{(B)i} = exp(\beta_0^B + x_i\beta_1^B)$ where $\alpha^A \neq \alpha^B$, $\beta_0^A \neq \beta_0^B$, and $\beta_1^A \neq \beta_1^B$. Here, we generate $y_{(A)i} \sim Ga(\alpha^A, \lambda_{(A)i})$ and $y_{(B)i} \sim Ga(\alpha^B, \lambda_{(B)i})$ once for each individual i, and set $y_i = y_{(A)i} + y_{(B)i}$. The covariate x is generated as $x_i \sim N(0, 1)$, and we assume that the covariate x is common among distributions A and B. That is, the x should be a stable variable such as demographics. However, we can observe only y_i and x_i. The true values of parameters are set to $\alpha^A = 1.2$, $\alpha^B = 1$, $\beta_0^A = 1$, $\beta_1^A = 0.4$, $\beta_0^B = 1.2$, $\beta_1^B = 1$.

Additionally, in Proposed NMR=1 and Proposed NMR=3, we use auxiliary information for distribution A. Concretely, in Proposed NMR=1, we use the mean value for distribution A. On the other hand, in Proposed NMR=3, we use three pieces of auxiliary information: (i) the total average duration for distribution A, (ii) the average duration for distribution A, that is $x > 0$, (iii) the average duration for distribution A, that is $x < 0$. If we use more than one piece of auxiliary information, we generalize the moment condition in equation (6) to the S-th dimensional vector as

$$\boldsymbol{m}(y_{(A)i}|\alpha^A, \beta_0^A, \beta_1^A) = \begin{cases} I_i^1\left[y_1^* - E[y_{(A)i}|\beta_0^A, \beta_1^A)]\right] \\ \cdots \\ I_i^S\left[y_S^* - E[y_{(A)i}|\beta_0^A, \beta_1^A]\right] \end{cases}, \tag{14}$$

where $I_i^s = 1$ when individual i belongs to group s (e.g., gender or age range), and $E[y_{(A)i}|\beta_0^A, \beta_1^A]] = \alpha^A exp(\beta_0^A + x_i\beta_1^A)$. In simulation 1, the Monte Carlo integration in equation (7) is not required, because there are no latent variables in the model.

We set the simulation number to 200 and draw a total of 80,000 MCMC samples after 20,000 burn-in phases, and we set the thinning interval per 10. The resulting number of MCMCs after thinning and burn-in was 8,000. We set the sample size as $n = 100$ and $n = 500$. We confirm the convergence of the MCMCs by the Geweke (1992) method. We show the summary of simulation results in Table 1. The table shows the results from an average of 200 simulation data sets. The "Average" shows mean values for each posterior mean and the "MSE $\times 10^2$" shows the mean squared error (MSE) which shows performance of the models. The "MSE Ratio" is fixed when the MSE of Proposed NMR=3 equals one; when "MSE Ratio" is greater than 1, the accuracy of the model is worse than that of Proposed NMR=3. The table shows that the results of Proposed NMR=3 are the best of all models in terms of MSEs. Additionally, the results of Proposed NMR=1 are worse than those of Proposed NMR=3 but better than those of Non-Macro. Next, we show the boxplot of results for each posterior mean in Figure 2. The figure shows that Non-Macro cannot effectively estimate parameters but that the Proposed NMR=3 can. From the results of simulation study 1, we understand that even if the pdf for the sum of durations is represented by the Moschopoulos (1985) method, the individual parameters cannot be estimated because of lack of identification. However, using auxiliary information, the parameters can be estimated appropriately.

Table 1: Simulation Results (Simulation 1)

	True Value	Average			MSE×10^2		
		Non-Macro	Proposed (NMR=1)	Proposed (NMR=3)	Non-Macro	Proposed (NMR=1)	Proposed (NMR=3)
n=100							
α^A	1.2	1.014	0.808	1.349	103.31	36.56	11.59
α^B	1	1.101	1.472	1.008	104.63	39.14	7.06
β_0^A	1	5.044	3.181	0.959	4199.49	2425.53	6.26
β_1^A	0.4	0.530	0.755	0.410	107.85	117.08	0.03
β_0^B	1.2	5.123	1.170	1.267	4352.95	3.12	7.10
β_1^B	1	0.814	0.797	0.992	121.37	5.18	1.79
MSE Ratio					761.44	644.31	1.00
n=500							
α^A	1.2	1.050	1.000	1.191	45.38	10.26	2.62
α^B	1	1.126	1.223	1.033	43.05	11.20	1.91
β_0^A	1	1.501	1.107	1.027	107.77	4.46	1.95
β_1^A	0.4	0.740	0.633	0.402	15.94	9.32	0.00
β_0^B	1.2	1.394	1.179	1.188	60.11	1.70	1.59
β_1^B	1	0.731	0.824	0.995	11.43	4.84	0.31
MSE Ratio					2863.63	1661.82	1.00

Simulation 2: Gamma Frailty Model

Next, we consider the Gamma frailty model for repeated durations. Let individual i's j-th time-to-event be y_{ij} with $y_{ij} = y_{(A)ij} + y_{(B)ij}$. We consider two types of Gamma frailty mod-

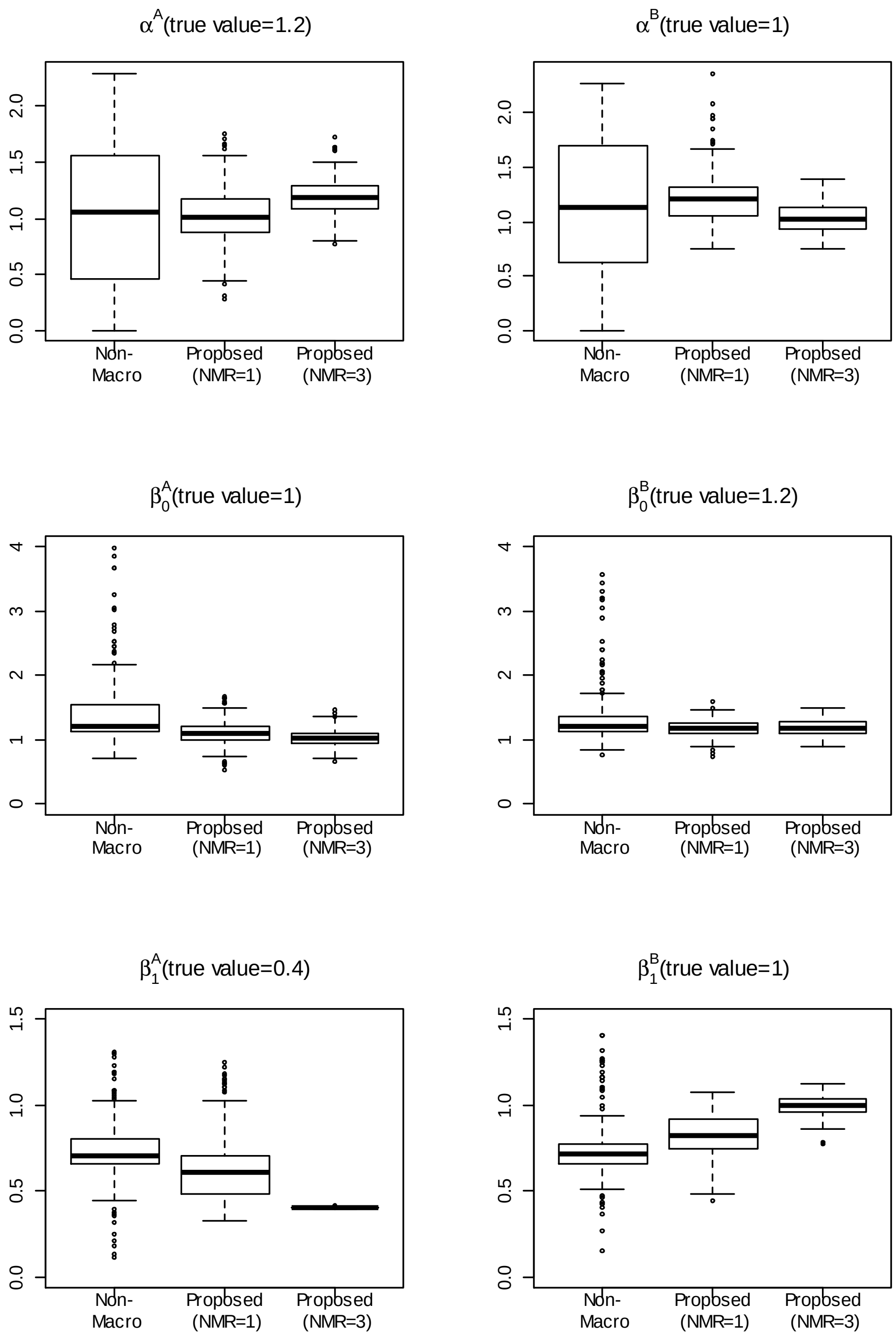

Figure 2: Boxplot in Simulation 1 ($n = 500$)

els: $y_{(A)ij} \sim Ga(\alpha^A, \lambda_{(A)ij}), \lambda_{(A)ij} = exp(x_{ij}\beta^A + f_i)$ and $y_{(B)ij} \sim Ga(\alpha^B, \lambda_{(B)ij}), \lambda_{(B)ij} = exp(x_{ij}\beta^B + b_2 f_i)$. Then $f_i \sim N(\mu, \sigma^2)$. However, only y_{ij} and x_{ij} can be observed in the analysis. The true values of parameters are set to $\alpha^A = 1.2$, $\alpha^B = 1$, $\beta^A = 0.4$, $\beta^B = 1$, $b_2 = 1.2$, $\mu = 1$, $\sigma^2 = 0.5$, and b_1 is fixed at 1. The procedures for generating data sets and auxiliary information in simulation 2 are the same as in simulation 1. In simulation 2, we set the number of repeated events at 3, that is $J_1 = \cdots = J_n \equiv 3$. The MCMC algorithms for estimating the model in simulation 2 are shown in Section 3.

We set the simulation number to 100, and the sample size at $n = 200$. The number of MCMCs and the convergence are the same as in simulation 1. We show the summary of simulation results in Table 2 and the boxplot of results for each posterior mean in Figure 3. The interpretations of the table are the same as in simulation 1. The table shows that the results of Proposed NMR=3 are the best of all models in terms of MSEs, while the results of Proposed NMR=1 are worse than those of Proposed NMR=3 but better than those of Non-Macro. Next, the boxplot shows that Non-Macro cannot estimate parameters appropriately but that the proposed model Proposed NMR=3 can. From the results of simulation studies 1 and 2 we can understand that, even if there are some latent variables (as in the frailty model), the proposed model using auxiliary information can estimate each parameter effectively.

Table 2: Simulation Results (Simulation 2)

	True Value	Average Non-Macro	Average Proposed (*NMR*=1)	Average Proposed (*NMR*=3)	$MSE\times10^2$ Non-Macro	$MSE\times10^2$ Proposed (*NMR*=1)	$MSE\times10^2$ Proposed (*NMR*=3)
α^A	1.2	1.152	0.738	1.104	50.28	25.30	2.73
α^B	1	1.018	1.427	1.124	44.91	23.80	3.04
β_1^A	0.4	0.637	0.660	0.401	122.45	98.26	0.13
β_1^B	1	0.937	0.745	0.941	354.74	15.93	0.76
b_2	1.2	1.286	1.013	1.162	76.59	15.44	1.89
μ	1	-1.465	1.043	1.076	4458.28	249.42	2.59
σ^2	0.5	1.100	0.750	0.593	155.11	10.93	1.77
MSE Ratio					468.19	127.73	1.00

4.2 Application to Interpurchase Timing Model in Marketing

Model for Empirical Analysis

As we previously mentioned, interpurchase timing models using the proportional hazard model have been widely studied in marketing (e.g., Jain and Vilcassim, 1991; Helsen, and Schmittlein, 1993; Allenby *et al.*, 1999; Seetharaman *et al.*, 2003; Bijwaard *et al.*, 2006; Igari

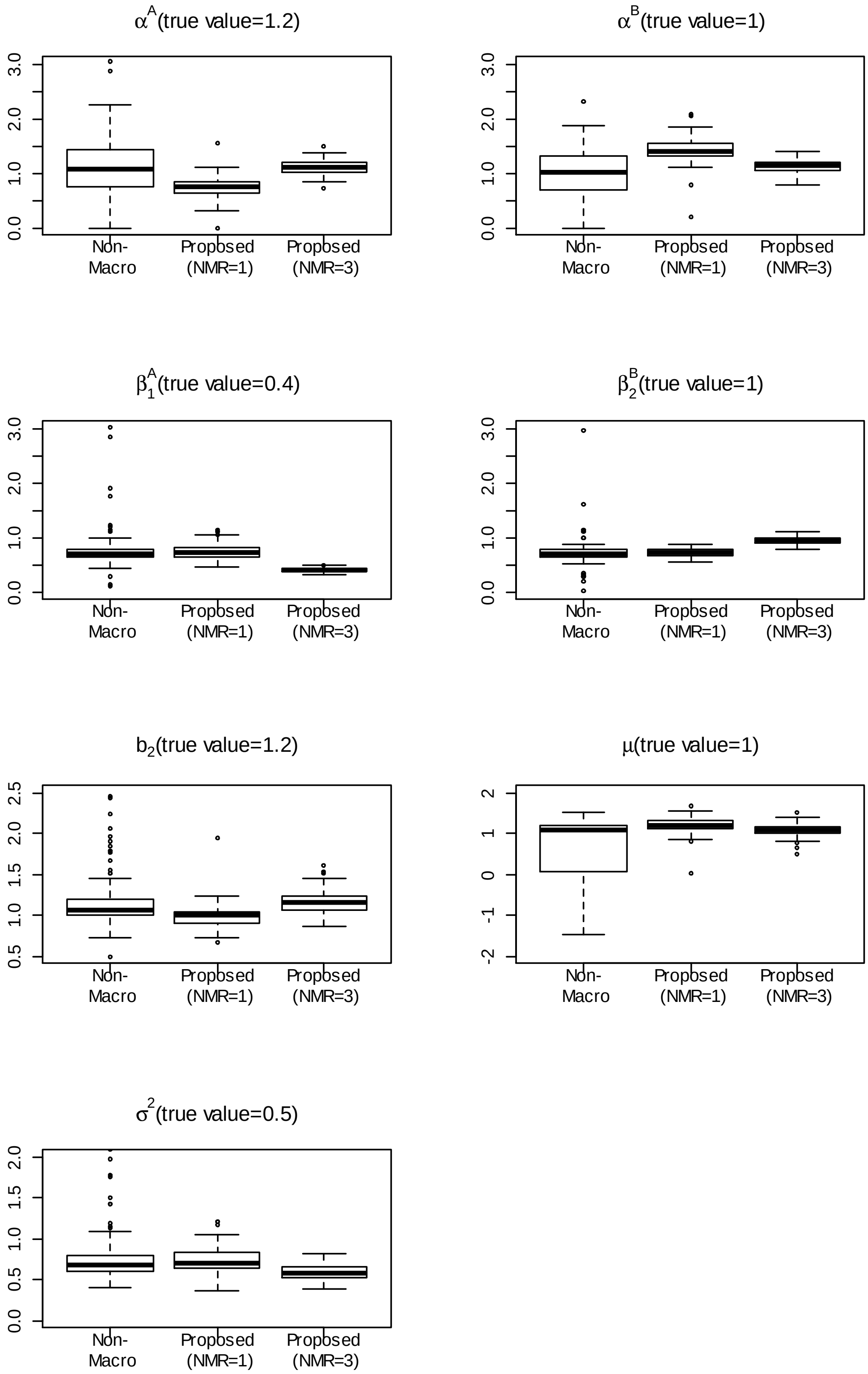

Figure 3: Boxplot in Simulation 2 ($n = 200$)

and Hoshino, 2018). In empirical analysis, only purchase histories in a company's own store can be observed, but we assume that the purchase number in competing stores can be known to researchers by using auxiliary information such as marketing surveys. We show a diagram of the analysis in Figure 4. First, we consider the two distributions: own store and competing stores. We let the consumer i's j-th interpurchase timing observed in the individual database be y_{ij}. Here, y_{ij} comprises the sum of K_{ij}-th durations, the one-duration in own store, y_{ij}^{own}, and $(K_{ij}-1)$-th durations in competing stores. That is $y_{ij} = (\sum_{k=1}^{K_{ij}-1} y_{ij}^{comp(k)}) + y_{ij}^{own}$. In the hazard function, we use covariate $\boldsymbol{x}_{ij}$ in the individual duration, but we do not use covariates in competing durations because marketing variables cannot be obtained in competing stores.

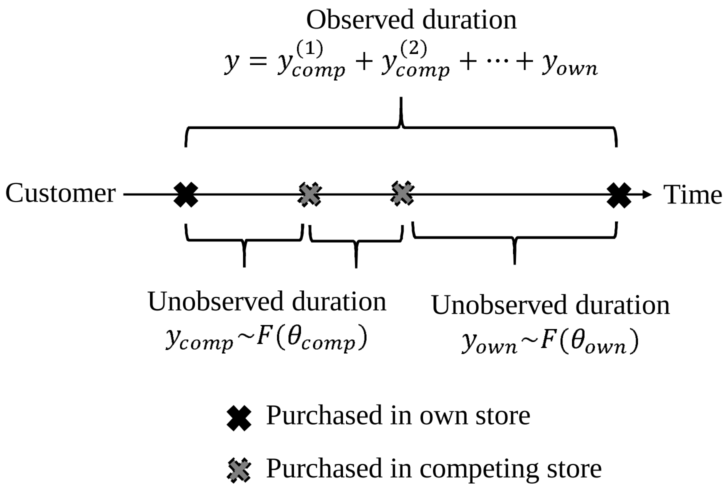

Figure 4: Interpurchase Timing Model in Empirical Analysis

The scale parameters are represented as $\lambda_{ij}^{own} = exp(b_1 f_i + \boldsymbol{x}_{ij}^T \boldsymbol{\beta})$ and $\lambda_{ij}^{comp} = exp(f_i)$, where b_2 is fixed to 1. The latent variable f_i is explained by the demographic variable $\boldsymbol{d}_i$, that is $f_i \sim N(\boldsymbol{d}_i^T \boldsymbol{\eta}, \sigma^2)$. We assume $b_1 > 0$ as a theory constraint, and we set $\sigma^2 = 1$ to simply perform the estimation.

The pdf, given latent variable f_i, is

$$p(y_{ij}|\boldsymbol{\alpha}, \boldsymbol{\beta}, b_1, f_i) = \begin{cases} \frac{y_{ij}^{\alpha_{own}-1} exp(-y_{ij}/\lambda_{ij}^{own})}{\Gamma(\alpha_{own})\lambda_{ij}^{\alpha_{own}}} \; if \; K_{ij} = 1 \\ \\ C_{ij} \sum_{r=0}^{\infty} \frac{\delta_{ij,r} y_{ij}^{\rho_{ij}+r-1} exp(-y_{ij}/\omega_{ij})}{\omega_{ij}^{\rho_{ij}+r} \Gamma(\rho_{ij}+r)} \; if \; K_{ij} > 1 \end{cases}, \tag{15}$$

where $\omega_{ij} = min(\lambda_{ij}^{own}, \lambda_{ij}^{comp})$, $\rho_{ij} = \alpha_{own} + (K_{ij} - 1)\alpha_{comp}$, and

$$C_{ij} = (\omega_{ij}/\lambda_{ij}^{own})^{\alpha_{own}} \times (\omega_{ij}/\lambda_{ij}^{comp})^{(K_{ij}-1)\alpha_{comp}}$$

$$\delta_{ij,r+1} = \frac{1}{r+1}\sum_{s=1}^{r+1} s\gamma_{ij,s}\delta_{ij,r+1-s}, \quad r = 0, 1, ..., \quad \delta_{ij0} = 1. \tag{16}$$

$$\gamma_{ij,r} = \alpha_{own}(1 - \omega_{ij}/\lambda_{ij}^{own})^r/r + (K_{ij} - 1)\alpha_{comp}(1 - \omega_{ij}/\lambda_{ij}^{comp})^r/r, \quad r = 1, 2, ...$$

As there are latent variables f_i in the model, the Monte Carlo calculation in equation (7) is required, and f_i are drawn from equation (13).

Purchase History Data

We use the *Syndicated Consumer Index* (SCI) data provided by Intage, Inc. in Japan. The SCI scanner panel data is the de facto standard for purchase panel data in the Japanese marketing field. The SCI records purchases along with the kinds, quantities, and prices of products purchased, the stores where items were purchased, and the date and time of purchase. Although the scanner panel data records purchase histories within competing store chains, we treat it as the database for a particular store, which is incomplete and lacks information about competing stores. We choose a familiar general merchandise store chain in Japan for the analysis. We also use purchase data for haircare items such as shampoo, conditioners, and treatment products. We analyze the data for the period from January 2015 to June 2016. From the purchase data, we consider consumers who purchase products in this category more than three times during the period. We select the sample size ($n = 200$) and total events number ($= 1214$) for the parameter estimation. The histogram of the observed duration is shown in Figure 5.

Next, we define the covariates and the auxiliary information used in our analysis. First, for the covariates $\boldsymbol{x}$ in the Gamma hazard model, we use two variables: "Price" and "log (Previous Amount)." Here, the "Price" is scaled and equals 1 when the price at the purchase incidence equals the regular price. In the marketing field, the effects of price discounts are very important. Second, we use five variables for $\boldsymbol{d}$: "gender (male 1)," "age," "child," "job (fulltime 1)," and "family size (1, 2, etc.)." Finally, we give the auxiliary information for the expected duration for own store y_s^*. We use a total of seven pieces of information: "all," "age under 30," "age in 30s," "age in 40s," "age in 50s or more," "price under 0.9," and "price greater than or equal to 0.9." The summary statistics are shown in Table 3.

Results

Table 3: Summary Statistics

	Average
Basic Information	
Sample Size	200
Total Events Number	1214
Duration	
Observed Cumulated Duration	63.763
The Number of Cumulated Duration	1.557
Time-Dependente Variable	
Price	0.997
Previous Amount	717.878
Demographic Variables	
Gender (Male=1)	0.085
Age	42.255
Child	0.545
Job (Fulltime=1)	0.235
Familiy Size	3.365
Auxiliary Information	
Total	44.928
Age under 30	53.307
Age 30s	37.796
Age 40s	43.035
Age 50s or more	54.244
Price under 0.9	42.104
Price greater than or equal to 0.9	45.522

In the real data analysis, we construct two models: (1) the proposed model with auxiliary information, and (2) the competing model without auxiliary information. In each analysis, we draw 5,000 MCMC iterations after 25,000 burn-in phases. We confirm the convergence of each model using the Geweke (1992) method. First of all, we check the accuracy of estimated results compared to the auxiliary information (=true value) as in Chaudhuri *et al.*(2008). We show the estimated expected duration for the own store $E[y^{own}]$ in Table 4. From this, we can understand that the competing model cannot estimate the durations of the own store but the proposed model can.

Next, we show the estimated results for parameters in the own store in Table 5. The table shows the results from the proposed model with auxiliary information and the competing model without auxiliary information. The table shows the posterior mean, posterior standard deviation, and Bayesian 95% credible intervals. Additionally, the * in the table shows significance from Bayesian 95% credible intervals. In the models, it is interpreted that if the coefficient of a covariate is positive, the duration becomes long. On the other hand,

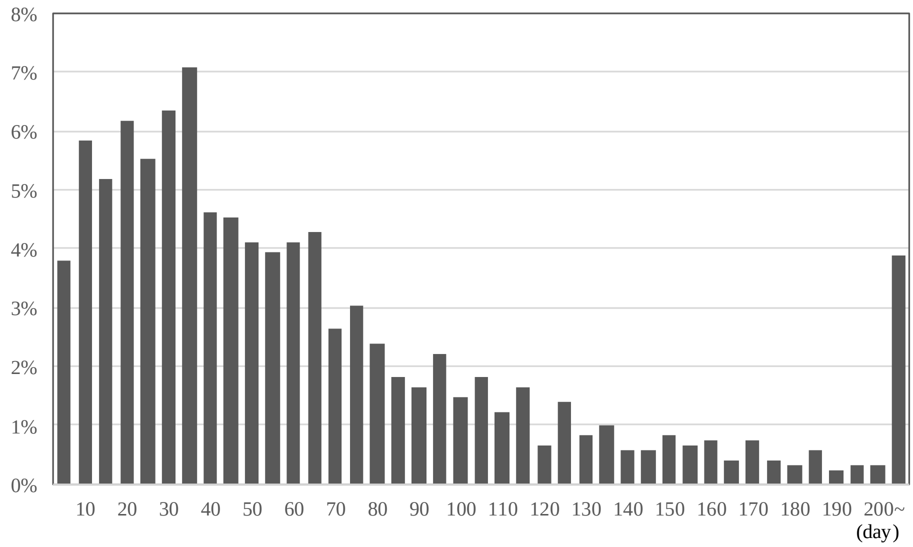

Figure 5: Histogram of Observed Duration

Table 4: Checking Accuracy of Estimated $E[y^{own}]$

	Auxiliary Information	Estimated Proposed Model	Competing Model
Total	44.93	43.58	17.22
Age under 30	53.31	45.95	17.84
Age 30s	37.80	37.40	16.31
Age 40s	43.03	43.88	17.11
Age 50s or more	54.24	50.85	18.39
Price under 0.9	42.10	40.33	11.33
Price greater than or equal to 0.9	45.52	44.26	18.45

we understand that if the coefficient of a covariate is negative, the duration becomes short. In other words, the interpretation of coefficients in the proposed model is contrary to that of the general proportional hazard model.

The results of the duration model show that the coefficient of "Price" is significantly positive. That is, when "Price" becomes higher, the duration becomes longer. In the marketing field, it is widely known that price discounts make interpurchase timing shorter (e.g., Jain and Vilcassim, 1991; Helsen and Schmittlein, 1993; Seetharaman *et al*, 2003; Igari and Hoshino, 2017, 2018). Additionally, the purchase amount in the previous (the last) event should be significantly positive, because a customer who purchased more products in the last purchase event will purchase products later more certainly than a customer who purchased fewer products.. The table shows that the estimated results in the proposed model do not contradict general marketing knowledge. A comparison of the coefficients of the two models shows that the shape parameters α are estimated to be approximately the same. However, the coefficients of "Price" differ between the models. The coefficient of "Price" in the competing model is overestimated in comparison to that of the proposed model. In contrast, the coefficient of "b" in the competing model is underestimated in comparison to that of the proposed model.

The estimated coefficients for demographic variables in the two models have the same tendency. It can be interpreted that the coefficient of gender (male) is positive, which shows the durations for male customers are longer than those for female customers. On the other hand, the coefficients of age, child, job, and family size are negative. Families with many children consume more products than those with fewer children. However, because there are no demographic variables whose coefficients are significant, the interpretation of positive or negative should be just a suggestion. Thus, the estimated results are consistent with the previous literature, and evidently, the proposed model can estimate parameters appropriately.

Table 5: Estimated Results for Own Store

		Posterior Mean	Posterior Sd	CI95 Low	CI95 High	
Proposed Model(with Auxiliary Infromation)						
α		0.7417	0.1069	0.5518	0.9952	*
β	Price	0.2627	0.1568	0.0186	0.5847	*
	log(Previous Amount)	0.3907	0.0633	0.2411	0.5016	*
b		0.4745	0.0956	0.2980	0.6414	*
η	Intercept	2.9137	0.4773	2.0048	3.8586	*
	Gender (Male=1)	0.2463	0.3080	-0.3499	0.8369	
	Age	-0.0011	0.0082	-0.0172	0.0148	
	Child	-0.3019	0.1853	-0.6684	0.0530	
	Job (Fulltime=1)	-0.0393	0.2067	-0.4471	0.3735	
	Family Size	-0.0592	0.0810	-0.2206	0.0999	
Competing Model(without Auxiliary Infromation)						
α		0.7593	0.1137	0.5843	0.9941	*
β	Price	1.6323	0.2694	1.0283	2.1282	*
	log(Previous Amount)	0.2078	0.0487	0.1118	0.3178	*
b		0.0434	0.0363	0.0020	0.1344	*
η	Intercept	3.7593	0.4477	2.8806	4.6215	*
	Gender (Male=1)	0.2929	0.3088	-0.3218	0.8919	
	Age	-0.0004	0.0081	-0.0167	0.0154	
	Child	-0.2776	0.1893	-0.6467	0.0922	
	Job (Fulltime=1)	-0.0553	0.2018	-0.4525	0.3461	
	Family Size	-0.0665	0.0841	-0.2325	0.1017	

Next, we show the distribution of expected estimated durations in competing stores, that is $E[y^{comp}]$, in the proposed model. The posterior mean of the shape parameters of competing stores $\hat{\alpha}_{comp}$ is 1.655. We show the distribution of expected durations $E[y_i^{comp}]$ in competing stores in Figure 6. Thus, the proposed model can estimate the unobserved distribution of durations in competing stores. Marketing managers may find this useful to understand consumer behavior in competing stores.

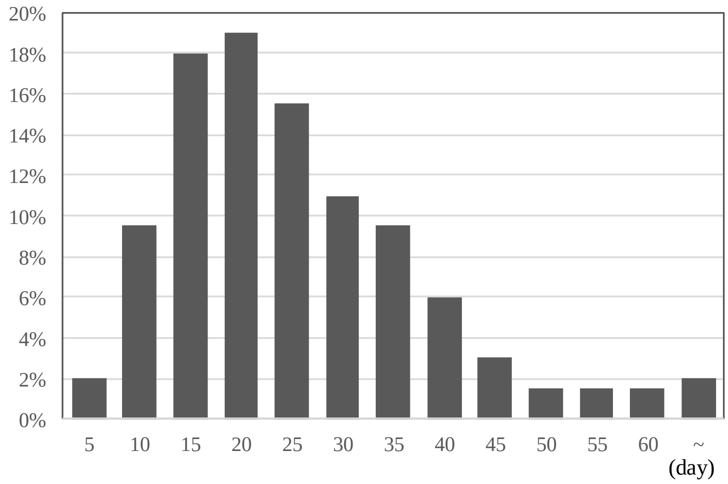

Figure 6: Expected Durations $E[y^{comp}]$ in Proposed Model

5 Conclusion

In this study, we propose a method to estimate parameters from the sum of independent Gamma durations with different parameters. Concretely, we generalized the Moschopoulos (1985) result to the proportional hazard model with covariates and to frailty models. We also incorporate auxiliary information by the quasi-Bayesian methods and propose the estimation procedure by MCMC. In two simulation studies, we show that the results of the proposed model with auxiliary information are better than those without it. The proposed model reproduces the true parameters appropriately. Additionally, we apply our model to the interpurchase timing model in marketing. From the estimated expected durations, we show the accuracy of the proposed model with auxiliary information in comparison to the competing model without auxiliary information. The results show that the proposed model can estimate the effects of price and the previous amounts, which does not contradict the findings of previous studies. Moreover, we show that the proposed model can calculate the expected duration of competing stores, which is usually unobserved.

Our approach can be applied to other proportional hazard models whose baseline hazard function is represented by exponential, log-normal distributions instead of Gamma distributions. However, specifying the pdf of the sum of durations is necessary. For example, the pdf of the sum of Weibull durations is not yet known (Nadarajah, 2008). It is also expected that the proposed method will be expanded to the Cox proportional hazard model (Cox, 1972), whose baseline hazard function is expressed in a nonparametric form. We set this as a goal for future research. Additionally, we consider only two types of distributions, but our model can be applied to three or more types of distributions. However, to estimate parameters appropriately, the auxiliary information for $K-1$-th distributions is generally required for the analysis of K-th type of distributions.

In this study, we demonstrate the effectiveness of our model through a simulation study. In real data analysis, we verify the accuracy by the estimated expected durations. Model selection using information criteria such as the Bayes factor in quasi-Bayesian inference is proposed by Li and Jiang (2016). Using this method, we can discuss the overidentification problem in GMM (e.g., Hansen, 1982) and discuss which auxiliary information should be used. We posit this as another goal for future research.

Moreover, our model can be applied to recency, monetary, frequency, and clumpiness (RMFC) analysis (Zhang *et al.*, 2013; 2015) in marketing. Clumpiness has recently emerged as an important concept to capture consumers' purchase behaviors. Researchers can capture the irregularity of interpurchase timing called clumpiness, which we think is generated from purchase behavior in competing stores. Our model can capture the causes of clumpiness as a sum of independent random variables. In the future, we will expand our model to RFMC analysis in marketing science.

Acknowledgments

We would like to acknowledge Intage Inc., who provided us the dataset we used for the empirical analysis. This work was supported by JPSP KAKENHI (18H05689 and 18H03209).

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669-679.

Allenby, G. M. and Rossi, P. E. (1998). Marketing models of consumer heterogeneity. *Journal of Econometrics*, **89**, 57-78.

Allenby, G. M., Leone, R. P. and Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, **94**, 365-374.

Bijwaard, G. E., Franses, P. H. and Paap, R. (2006). Modeling purchases as repeated events. *Journal of Business & Economic Statistics*, **24**, 487-502.

Blattberg, R, C., Kim, B, D. and Neslin, S, S. (2008). *Database marketing: Analyzing and managing customers*. Springer Science & Business Media.

Chaudhuri, S., Handcock, M. S. and Rendall, M. S. (2008). Generalized linear models incorporating population level information: An empirical - likelihood - based approach. *Journal of the Royal Statistical Society: Series B*, **70**, 311-328.

Chen, Y. and Steckel, J. H. (2012). Modeling credit card share of wallet: Solving the incomplete information problem. *Journal of Marketing Research*, **49**, 655-669.

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

Cox, D. R. (1972). Regression models and life-tables (with discussion) . *Journal of the Royal Statistical Society: Series B*, **34**, 187-220.

Fader, P. S. and Hardie, B. G. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing*, **23**, 61-69.

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

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith eds., 169-193. Oxford University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029-1054.

Hellerstein, J. K. and Imbens, G. W. (1999). Imposing moment restrictions from auxiliary data by weighting. *The Review of Economics and Statistics*, **81**, 1-14.

Helsen, K. and Schmittlein, D.C. (1993). Analyzing duration times in marketing: evidence for the effectiveness of hazard rate models. *Marketing Science*, **11**, 395-414.

Hoshino, T. (2008). A Bayesian propensity score adjustment for latent variable modeling and MCMC algorithm. *Computational Statistics & Data Analysis*, **52**, 1413-1429.

Hoshino, T. and Igari, R. (2017). Quasi-Bayesian inference for latent variable models with external information: Application to generalized linear mixed models for biased data. *Keio-IES- discussion paper series.*

Huang, C. Y., Qin, J. and Tsai, H. T. (2016). Efficient estimation of the Cox model with auxiliary subgroup survival information. *Journal of the American Statistical Association*, **111**, 787-799.

Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001). *Bayesian survival analysis.* Springer Science & Business Media.

Igari, R. and Hoshino, T. (2017). A survival analysis incorporating auxiliary information by a Bayesian generalized method of moments: Application to purchase duration modeling. *Journal of the Japanese Society of Computational Statistics.* **30**, 27-44.

Igari, R. and Hoshino, T. (2018). A Bayesian data combination approach for repeated durations under unobserved missing indicators: Application to interpurchase-timing in marketing. *Computational Statistics & Data Analysis*, **126**, 150-166.

Imbens, G. W. and Lancaster, T. (1994). Combining micro and macro data in microeconometric models. *The Review of Economic Studies*, **61**, 655-680.

Jain, D.C. and Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science*, **10**, 1-23.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate probability distributions, vol. 1*, Wiley-Interscience.

Krishnamoorthy, K. (2006). *Handbook of statistical distributions with applications.* Chapman and Hall/CRC.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data 2nd.* Springer Science & Business Media.

Leszczyc, P. T. P., Sinha, A. and Timmermans, H. J. (2000). Consumer store choice dynamics: An analysis of the competitive market structure for grocery stores. *Journal of Retailing*, **76**, 323-345.

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

Moe, W. W. and Fader, P. S. (2004). Dynamic conversion behavior at e-commerce sites. *Management Science*, **50**, 326-335.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, **37**, 541-544.

Nadarajah, S. (2008). A review of results on sums of random variables. *Acta Applicandae Mathematicae*, **103**, 131-140.

Qin, J. (2000). Combining parametric and empirical likelihoods. *Biometrika*, **87**, 484-490.

Ridder, G. and Moffitt, R. (2007). The econometrics of data combination. In *Handbook of Econometrics*, Heckman, J,J. and Leamer, E. eds., **6**, 5469-5547.

Seetharaman, P. B. and Chintagunta, P. K. (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business & Economic Statistics*, **21**, 368-382.

Schmittlein, D. C., Morrison, D. G. and Colombo, R. (1987). Counting your customers: Who-are they and what will they do next?. *Management science*, **33**, 1-24.

Sim, C. H. (1992). Point processes with correlated gamma interarrival times. *Statistics & Probability letters*, **15**, 135-141.

Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528-540.

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

Zhang, Y., Bradlow, E. T. and Small, D. S. (2013). New measures of clumpiness for incidence data. *Journal of Applied Statistics*, **40**, 2533-2548.

Zhang, Y., Bradlow, E. T. and Small, D. S. (2015). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science*, **34**, 195-208.