Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Quasi-Bayesian Inference for Latent Variable Models with External Information: Application to generalized linear mixed models for biased data

Takahiro Hoshino, Ryosuke Igari

28 April, 2017
DP2017-014
https://ies.keio.ac.jp/en/publications/7666/

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

Quasi-Bayesian Inference for Latent Variable Models with External Information: Application to generalized linear mixed models for biased data

Takahiro Hoshino, Ryosuke Igari

Keio-IES DP2017-014
28 April, 2017
JEL classification: C11; C15; C35
Keywords: Latent Variable Modeling; Quasi-Bayes; Population-Level Information; Markov chain Monte Carlo; Data Augmentation

Abstract

There is a vast literature proposing non-Bayesian methods for making inferences incorporating auxiliary information such as population-level marginal moments. However, it is not feasible to apply these methods directly to latent variable models because the data augmentation approach, in which latent variables are treated as incidental parameters and then generated, is not developed. In this paper, we propose a Markov Chain Monte Carlo (MCMC) algorithm with data augmentation for latent variable models for cases in which we have both a sampled dataset and additional information such as population level moments. The resulting quasi-Bayesian inference with auxiliary information is very straightforwaed to implement, and consistency and asymptotic variance of the quasi-Bayesian posterior mean estimators from the MCMC outputs are shown in this paper. The proposed method is especially useful when the dataset is biased but we have an unbiased large sample for some variables or population marginal moments in which it is difficult to correctly specify the sample selection model. For illustrative purposes, we apply the proposed estimation method to generalized linear mixed models for biased data both in simulation studies and in real data analysis. The proposed method can be used to make inferences in non/semi-parametric latent variable models by incorporating the existing semi-parametric Bayesian algorithms such as the Blocked Gibbs sampler in the MCMC iteration.

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

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

Quasi-Bayesian Inference for Latent Variable Models with External Information: Application to generalized linear mixed models for biased data

Takahiro Hoshino and Ryosuke Igari

TAKAHIRO HOSHINO

Department of Economics, Keio University

Institute of Statistical Mathematics

RIKEN Center for Advanced Intelligence Project

E-mail: bayesian@jasmine.ocn.ne.jp

ABSTRACT

There is a vast literature proposing non-Bayesian methods for making inferences incorporating auxiliary information such as population-level marginal moments. However, it is not feasible to apply these methods directly to latent variable models because the data augmentation approach, in which latent variables are treated as incidental parameters and then generated, is not developed. In this paper, we propose a Markov Chain Monte Carlo (MCMC) algorithm with data augmentation for latent variable models for cases in which we have both a sampled dataset and additional information such as population level moments. The resulting quasi-Bayesian inference with auxiliary information is very straightforwaed to implement, and consistency and asymptotic variance of the quasi-Bayesian posterior mean estimators from the MCMC outputs are shown in this paper. The proposed method is especially useful when the dataset is biased but we have an unbiased large sample for some variables or population marginal moments in which it is difficult to correctly specify the sample selection model. For illustrative purposes, we apply the proposed estimation method to generalized linear mixed models for biased data both in simulation studies and in real data analysis. The proposed method can be used to make inferences in non/semi-parametric latent variable models by incorporating the existing semi-parametric Bayesian algorithms such as the Blocked Gibbs sampler in the MCMC iteration.

Keywords: Generalized Linear Mixred Models; Latent Variable Modeling; Quasi-Bayes; Population-Level Information; Dirichlet Process Mixture; Markov Chain Monte Carlo; Data Augmentation

1 Introduction

Recently, quasi-Bayesian inference methods or the Bayesian GMM (Generalized method of moments) method have been developed and applied in various studies (Kim, 2002; Chernozhukov and Hong 2003; Hoshino 2008; Yin 2009; Yang and He 2012). Most existing applications of the quasi-Bayesian estimation method emphasize the robust estimation without full model specification in contrast with traditional Bayesian methods (e.g., Li and Jiang, 2016); however, various other semi-parametric Bayesian estimation methods, such as the Dirichlet process mixtures model, have been proposed and applied to weaken the parametric assumptions in various fields (e.g., Hoshino, 2013). The distinct feature of (non-)Bayesian GMM methods is that they can easily include external information such as population-level moments in the estimation of parameters.

There is a vast literature proposing non-Bayesian methods for making an inference incorporating auxiliary information; these methods include the traditional GMM methods (Hansen, 1982; Imbens and Lancaster, 1994; Nevo, 2003) and the combined empirical likelihood method (Qin and Zhang, 2007; Chaudhuri et.al, 2008). However, it is not feasible to apply these methods directly to latent variable models because the data augmentation approach has not been considered. For inference in latent variable models, especially with mixed outcomes and/or clustered/multilevel data, MCMC algorithms with data augmentation by sampling latent variables as incidental parameters have been developed and employed in various applied studies because of the flexibility of the modeling framework and estimation procedure (e.g., Tanner and Wong 1987; Albert and Chib 1993; Dunson, 2000).

In this paper, we propose a Markov Chain Monte Carlo (MCMC) algorithm with data augmentation for latent variable models for cases in which we have both a sampled dataset and additional information, such as population-level moments. We show consistency and asymptotic variance of the quasi-Bayesian posterior mean estimators from the MCMC outputs. The proposed methods are especially useful when the dataset used can be biased, but it is difficult to correctly specify the sample selection model, and we have an unbiased large sample for some of the variables or population moments. The existing semi-parametric Bayesian algorithms such as the Blocked Gibbs sampler (Ishwaran and James, 2001) in the MCMC iteration do not easily incorporate auxiliary information, while the proposed method can be easily extended to make inferences in non/semi-parametric latent variable models.

For illustrative purposes, we apply the proposed estimation method to generalized linear mixed models (GLMM) for biased data both in simulation studies and in real data analysis. Several estimation methods for biased data specify the selection mechanism model (e.g., Heckman, 1979), but it is difficult to correctly specify the model. Other methods employ non-parametric model formulation for the selection model (e.g., Lee and Berger, 2001; Hoshino, 2013). Instead, in this study, we use marginal population-level information.

In medical sciences, a low degree of generalizability or external validity (Shadish et.al. 2002) of the results obtained in randomized controlled trials due to biased sampling of subjects has recently attracted significant attention (e.g., Cole and Stuart, 2010; van Poucke et.al. 2016), and some methods to deal with the problem have been developed (e.g., Hartman et.al. 2015). The proposed method is especially useful when the population-level moment information or a large sample dataset such as the national medical database is available without the assumption of selection models.

The remainder of the paper is organized as follows. Section 2 presents the model setup. Section 3 describes the existing method which is not available in the model we considered, and the proposed method. The algorithm and the asymptotic properties are also shown in Section 3. In Section 4 we show the detaied algorithm for the GLMM model, and provide a brief summary of the simulation study. A simple real data analysis is also shown.

2 Model Setup: Latent Variable Models with External Information

We consider the following latent variable models,

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = \int p(\boldsymbol{y}|\boldsymbol{f},\boldsymbol{\theta})p(\boldsymbol{f}|\boldsymbol{\theta})d\boldsymbol{f}, \tag{1}$$

where $\boldsymbol{y}$ is the dependent variable vector, $\boldsymbol{f}$ is the latent variable vector, and $\boldsymbol{\theta}$ is a $q \times 1$ parameter vector of interest in the data generating process.

This model includes various submodels such as factor analytic models (Kunda and Dunson, 2014), random effect models, multilevel models (Goldstein, 2010), and generalized random utility models (Walker and Ben-Akiva, 2002) for discrete variables. In this study, we will focus on GLMMs, among others. Note that our model specification is applicable not only to full parametric models but also to non-parametric or semi-parametric models via the Dirichlet process mixture (DPM) models or Probit Stick-breaking Process mixture (PSBPM) models (Chung and Dunson, 2009; Kleinman and Ibrahim, 1998; Kyung et.al.2010; Hjort et.al.2010).

Consider that we have a sample, $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n$, of n independent and identically distributed (i.i.d.) random vectors. The log-likelihood for the marginalized model is

$$L_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \int p(\boldsymbol{y}_i|\boldsymbol{f}_i, \boldsymbol{\theta}) p(\boldsymbol{f}_i|\boldsymbol{\theta}) d\boldsymbol{f}_i, \tag{2}$$

while the log-likelihood for MCMC methods with data augmentation is expressed as

$$L_n(\boldsymbol{\theta}|\boldsymbol{f}) = \sum_{i=1}^{n} \log p(\boldsymbol{y}_i|\boldsymbol{f}_i, \boldsymbol{\theta}), \tag{3}$$

and the distribution of the latent variables is treated as a prior distribution of incidental parameters, $p(\boldsymbol{f}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\boldsymbol{f}_i|\boldsymbol{\theta})$.

Moreover, consider that we have another source of information such as the r-dimensional population-level information vector regarding moments of $\boldsymbol{y}$

$$E[m^*(\boldsymbol{y}|\boldsymbol{\theta})] = 0. \tag{4}$$

Note that if the model of interest is the linear regression model $E(y|\boldsymbol{\theta}, \boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\theta}$, then the population-level information is expressed directly as the function of the parameter vector, and the target of inference will just be the reparameterized model. However, this information is not easily utilized through reparametrization, parameter constraints, or prior distribution in a general model setup unless we use linear regression models (see Section 4 for detail).

3 Quasi-Bayesian Inference for Latent Variable Models with External Information

For a marginalized model, $p(\boldsymbol{y}|\boldsymbol{\theta})$, the quasi-Bayesian posterior distribution is

$$q(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{exp\{L_n(\boldsymbol{\theta})\}p(\theta)}{\int_{\boldsymbol{\Theta}} exp\{L_n(\boldsymbol{\theta})\}p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto exp\{L_n(\boldsymbol{\theta})\}p(\boldsymbol{\theta}), \tag{5}$$

where $p(\boldsymbol{\theta})$ 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 for various estimation methods such as a GMM, M-estimator, or empirical likelihood instead of a log-likelihood function. If we use the GMM-type objective function, $Q_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta})$, the function is defined as follows.

$$Q_n(\boldsymbol{\theta}) = -\frac{n}{2}\Big[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta})\Big]^T \boldsymbol{\Omega_n}\Big[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta})\Big], \tag{6}$$

where $\boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta})$ is the unbiased moment restriction vector and $\boldsymbol{\Omega_n}$ is a weight matrix converging to $E\big(\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})^T\big)^{-1}$. It is shown that under mild regularity conditions, the posterior means are consistent and asymptotically normally distributed (Kim, 2002; Chernozhukov and Hong 2003; Yin 2009; Yang and He 2012). The quasi-Bayesian methods are mainly applied in order to weaken the model assumption (Li and Jiang, 2016), but by using the GMM-type objective function, we can incorporate population-level information as a subvector of $\boldsymbol{m}(\boldsymbol{y_i}|\boldsymbol{\theta})$ as Imbens and Lancastor (1994) did in a non-Bayesian GMM estimation.

However, it is difficult to directly draw samples of parameters from Equation (5) for latent variable models, such as GLMMs, in which computation of the likelihood of marginalized models requires numerical integrations.

To be more concrete, in the estimation of latent variable models using marginalized likelihood (Equation (2)) and external information (Equation (4)), the GMM-type function should be employed, because the dimension of moments to be considered is $q+r$ and the dimension of parameters is q. Then, the quasi-Bayesian posterior distribution in this setup is expressed as Equation (5), and the objective function is Equation (6), where the $q+r$ dimensional moment restriction vector is expressed as

$$m(\boldsymbol{y}|\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial \boldsymbol{\theta}^T}\log\int p(\boldsymbol{y}|\boldsymbol{f},\boldsymbol{\theta})p(\boldsymbol{f}|\boldsymbol{\theta})d\boldsymbol{f} \quad m^{*T}(\boldsymbol{y}|\boldsymbol{\theta})\Big)^T. \tag{7}$$

It is very difficult to draw MCMC samples from the resulting quasi-Bayesian posterior distribution, which is proportional to the exponential of the quadratic form of the above function m containing high dimensional integrals.

3.1 Hybrid Posterior Combining Likelihood and GMM-type Objective Function

It would be better to employ the data augmentation approach by treating latent variables as incidental parameters in a quasi-Bayesian computation. In this study, we propose the following quasi-Bayesian joint posterior distribution for parameter vector $\boldsymbol{\theta}$ and latent variable vector $\boldsymbol{f}$

$$q(\boldsymbol{\theta},\boldsymbol{f}|\boldsymbol{y})_{QB*} = \frac{\{\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})}{\int\int\{\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})d\boldsymbol{f}d\boldsymbol{\theta}} \quad (8)$$

to sample the random draws of $\boldsymbol{\theta}$ and $\boldsymbol{f}$, where

$$Q_n^*(\boldsymbol{\theta}) = -\frac{n}{2}\Big[\frac{1}{n}\sum_{i=1}^n m^*(\boldsymbol{y}_i|\boldsymbol{\theta})\Big]^T \boldsymbol{\Omega}_n^* \Big[\frac{1}{n}\sum_{i=1}^n m^*(\boldsymbol{y}_i|\boldsymbol{\theta})\Big], \quad (9)$$

and $\boldsymbol{\Omega}_n^*$ is a matrix converging to $E[m^*(\boldsymbol{y}|\boldsymbol{\theta})m^*(\boldsymbol{y}|\boldsymbol{\theta})^T]^{-1}$.

Note that the quasi-Bayesian posterior distribution (Equation (8)) is proportional to the likelihood $\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})$ times the following quasi-Bayesian posterior distribution, conditional on the external information of the moment $m^*(\boldsymbol{y}|\boldsymbol{\theta})$,

$$q(\boldsymbol{\theta}|m^*)_{QB*} = \frac{\exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})}{\int\{\exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (10)$$

3.2 Algorithm

In our augmentation approach, the algorithm to drawn samples of $\boldsymbol{\theta}$ from Equation (8) is very straightforward.

Sampling $\boldsymbol{\theta}$

The samples of $\boldsymbol{\theta}$ are drawn from

$$q(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{y})_{QB*} = \frac{\{\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})}{\int\{\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (11)$$

In this setup it is difficult to draw samples $\boldsymbol{\theta}$ directly from the above distribution, then we use the Metropolis-Hastings algorithm by drawing the candidate of $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$ from the ordinal posterior distribution

$$p(\boldsymbol{\theta}|\boldsymbol{f},\boldsymbol{y}) = \frac{\{\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\} \times p(\boldsymbol{\theta})}{\int \{\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\} \times p(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{12}$$

and accept the value with the following probability:

$$min\Big(\frac{\exp Q_n^*(\boldsymbol{\theta}^*)}{\exp Q_n^*(\boldsymbol{\theta}^o)}, 1\Big), \tag{13}$$

where $\boldsymbol{\theta}^o$ is the value obtained in the former iteration.

Sampling $\boldsymbol{f}$

$\boldsymbol{f}$ is drawn from

$$p(\boldsymbol{f}|\boldsymbol{\theta},\boldsymbol{y}) = \frac{\{\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\}}{\int \{\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{f}_i,\boldsymbol{\theta})p(\boldsymbol{f}_i|\boldsymbol{\theta})\}d\boldsymbol{f}}, \tag{14}$$

which is the ordinal full conditional posterior distribution of the latent variables. See section 4 for the detailed algorithm in the case of generalized linear mixed models.

Note that for non/semi-parametric models via DPM or PSBPM, we can express $p(\boldsymbol{y}|\boldsymbol{f},\boldsymbol{\theta})p(\boldsymbol{f}|\boldsymbol{\theta})$ as the infinite mixtures and employ the related algorithms in our formulation because the proposed quasi-Bayesian posterior distribution is proportional to the product of the likelihood and the function regarding the auxiliary information, whereas there is no algorithm in the existing quasi-Bayesian inference (Equation (5)) with external information.

When the external information is stochastic

We considered the case that has population-level information, and the proposed method is easily generalized to deal with the case by using statistics from unbiased external surveys. For such cases, we can combine the stochastic information obtained in the external surveys by adding the statistics to function m in Equation (7) and its variance matrix to the relevant part of $\boldsymbol{\Omega_n}$. To be more concrete, Equation (9) is replaced with

$$Q_n^{*s}(\boldsymbol{\theta}) = -\frac{n}{2}\Big[\frac{1}{n}\sum_{i=1}^{n} m^{*s}(\boldsymbol{y}_i|\boldsymbol{\theta})\Big]^T \boldsymbol{\Omega}_n^{*s}\Big[\frac{1}{n}\sum_{i=1}^{n} m^{*s}(\boldsymbol{y}_i|\boldsymbol{\theta})\Big], \tag{15}$$

where M is the sample size of the external information, M/N converges to some constant k, $m^s(\boldsymbol{y}|\boldsymbol{\theta})$ is the moment or estimating equation for solving $\boldsymbol{\theta}$ in the external information source, and

$$m^{*s}(\boldsymbol{y}_i|\boldsymbol{\theta}) = \left(m^{*T}(\boldsymbol{y}|\boldsymbol{\theta}) \quad m^{sT}(\boldsymbol{y}|\boldsymbol{\theta})\right)^T,$$

$$\lim_{n\to\infty} \boldsymbol{\Omega}_n^{*s} = \begin{pmatrix} E[m^*(\boldsymbol{y}|\boldsymbol{\theta})m^*(\boldsymbol{y}|\boldsymbol{\theta})^T]^{-1} & 0 \\ 0 & kE[m^s(\boldsymbol{y}|\boldsymbol{\theta})m^s(\boldsymbol{y}|\boldsymbol{\theta})^T]^{-1}, \end{pmatrix} \tag{16}$$

3.3 Asymptotic Properties

We define the quasi-Bayesian posterior mean estimator as

$$\hat{\boldsymbol{\theta}}_{QB*} = \int\int \boldsymbol{\theta} q(\boldsymbol{\theta}, \boldsymbol{f}|\boldsymbol{y})_{QB*} d\boldsymbol{\theta} d\boldsymbol{f}, \tag{17}$$

where $q(\boldsymbol{\theta}, \boldsymbol{f}|\boldsymbol{y})_{QB*}$ is defined as Equation (8). We can easily obtain the following theorem.

Theorem 1. *$\hat{\boldsymbol{\theta}}_{QB*}$ is consistent for estimating $\boldsymbol{\theta}$ and is asymptotically normally distributed as*

$$A_n^{-1/2}(\boldsymbol{\theta}_0)B(\boldsymbol{\theta}_0)[\sqrt{N}(\hat{\boldsymbol{\theta}}_{QB*} - \boldsymbol{\theta}_0)] \xrightarrow{d} N(0, \boldsymbol{I}), \tag{18}$$

where $\boldsymbol{\theta}_0$ is true value of the parameter vector,

$$A_n(\boldsymbol{\theta}_0) = \frac{1}{n}\Big[\frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta}_0)\Big]\Big[\frac{\partial}{\partial \boldsymbol{\theta}} R_n(\boldsymbol{\theta}_0)\Big]^T, \tag{19}$$

and

$$B_n(\boldsymbol{\theta}_0) = -\frac{1}{n}\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} R_n(\boldsymbol{\theta}_0), \tag{20}$$

where

$$R_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(\boldsymbol{y}|\boldsymbol{\theta}) + Q_n^*(\boldsymbol{\theta}). \tag{21}$$

Proof. See Appendix. □

4 Application to the Generalized Linear Mixed Model

We apply the proposed method to the GLMM for biased data.

4.1 Generalized Linear Mixed Model

The GLMM, which includes latent variables in a generalized linear model (GLM), is a major model that can express various types of responses, such as linear, binomial, count, and multinomial.

Let individual $i(1,...,n)$'s $t(1,...,T_i)$-th event response be y_{it}; then, the probability density function (pdf) is as follows.

$$p(y_{it}|\boldsymbol{\theta}, f_i) = exp\left\{\frac{y_{it}\xi_{it} - b(\xi_{it})}{a(\phi)} + c(y_{it}, \phi)\right\}. \tag{22}$$

$$g(\xi_{it}) = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{w}_{it}^T\boldsymbol{\gamma} + f_i \tag{23}$$

Here, $\boldsymbol{x_i}$ is individual i's covariate vector; $\boldsymbol{w}_{it}$ is individual i's t-th time-varying covariate vector; α is an intercept; $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are coefficients, which are common among individuals, and we set $\boldsymbol{\theta} = [\alpha\ \boldsymbol{\beta}^T\ \boldsymbol{\gamma}^T]^T$; f_i is a latent variable that differs for each individual. Besides, $a()$, $b()$, and $c()$ are known functions; ϕ is a dispersion parameter that may or may not be known; and $g()$ is a known link function.

Let the latent variable f_i follow normal distribution with mean 0 and variance σ^2.

$$f_i \sim N(0, \sigma^2) \tag{24}$$

The GLMM covers various models such as linear regression, logistic regression, Poisson regression, or parametric hazard models by setting an error function and link structure.

4.2 Simulation Study

Here, we will show the performance of the proposed model by simulation study for biased data using a logistic regression model with random effects.

$$p(y_{it}|\boldsymbol{\theta}, f_i) = p(y_{it} = 1|\boldsymbol{\theta}, f_i)^{y_{it}} p(y_{it} = 0|\boldsymbol{\theta}, f_i)^{1-y_{it}} \tag{25}$$

$$logit\left[p(y_{it} = 1|\boldsymbol{\theta}, f_i)\right] = \alpha + x_i\beta + w_{it}\gamma + f_i \tag{26}$$

In MCMC procedure, we set the parameters $\boldsymbol{\theta} = [\alpha\ \beta\ \gamma]^T$. The quasi-bayesian joint posterior distribution is

$$q(\boldsymbol{\theta}, \sigma^2, \boldsymbol{f}|\boldsymbol{y})_{QB*}$$
$$= \frac{\left\{\prod_{i=1}^n p(f_i|\sigma^2)\{\prod_{t=1}^{T_i} p(y_{it}|f_i, \boldsymbol{\theta})\}\right\} \times \exp\left[Q_{n^*}^*(\boldsymbol{\theta})\right] \times p(\boldsymbol{\theta}) \times p(\sigma^2)}{\int\int\int\left\{\prod_{i=1}^n p(f_i|\sigma^2)\{\prod_{t=1}^{T_i} p(y_{it}|f_i, \boldsymbol{\theta})\}\right\} \times \exp\left[Q_{n^*}^*(\boldsymbol{\theta})\right] \times p(\boldsymbol{\theta}) \times p(\sigma^2) dfd\boldsymbol{\theta}d\sigma^2}, \tag{27}$$

where

$$Q^*_{n^*}(\boldsymbol{\theta}) = -\frac{n^*}{2}\Big[\frac{1}{n^*}\sum_{i=1}^{n}\sum_{t=1}^{T_i} m^*(y_{it}|\boldsymbol{\theta})\Big]^T \boldsymbol{\Omega}^*_{n^*}\Big[\frac{1}{n^*}\sum_{i=1}^{n}\sum_{t=1}^{T_i} m^*(y_{it}|\boldsymbol{\theta})\Big], \tag{28}$$

and n^* is the total number of data size, $n^* = \sum_{i=1}^{n}\sum_{t=1}^{T_i} I$. The moment restriction $m^*(y_{it}|\boldsymbol{\theta})$ is

$$m^*(y_{it}|\boldsymbol{\theta}) = \boldsymbol{y}^* - E(y_{it}=1|x_{it},\boldsymbol{\theta}) = \boldsymbol{y}^* - \int \frac{exp(\alpha + x_i\beta + w_{it}\gamma + f_i)}{1+exp(\alpha + x_i\beta + w_{it}\gamma + f_i)} p(f_i)df_i. \tag{29}$$

Here, $\boldsymbol{y}^*$ is the r-dimensional population-level proportion of $y=1$.

Sampling $\boldsymbol{\theta}$

The quasi-bayesian conditional distribution of $\boldsymbol{\theta}$ is

$$q(\boldsymbol{\theta}|\boldsymbol{y},\sigma^2,\boldsymbol{f})_{QB*} \propto \{\prod_{i=1}^{n}\prod_{t=1}^{T_i} p(y_{it}|f_i,\boldsymbol{\theta})\} \times \exp\Big[Q^*_{n^*}(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta}). \tag{30}$$

It is difficult to draw samples $\boldsymbol{\theta}$ directly from the above distribution, then we use the Metropolis-Hastings algorithm. We redefine the conditional posterior of $\boldsymbol{\theta}$ except the part of moment restriction,

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y},\sigma^2,\boldsymbol{f}) &\propto \{\prod_{i=1}^{n}\prod_{t=1}^{T_i} p(y_{it}|f_i,\boldsymbol{\theta})\} \times p(\boldsymbol{\theta}) \\ &= \Big\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\Big[\frac{exp(\alpha + x_i\beta + w_{it}\gamma + f_i)}{1+exp(\alpha + x_i\beta + w_{it}\gamma + f_i)}\Big]^{y_{it}}\Big[\frac{1}{1+exp(\alpha + x_i\beta + w_{it}\gamma + f_i)}\Big]^{1-y_{it}}\Big\} \\ &\times exp\{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu_\theta})^T \boldsymbol{V}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu_\theta})\}, \end{aligned} \tag{31}$$

where $\boldsymbol{\mu_\theta}$ and $V_{\boldsymbol{\theta}}$ are the mean vector and variance matrix of prior distribution $p(\boldsymbol{\theta})$. Then, we use the Metropolis-Hastings algorithm by drawing the candidate of $\boldsymbol{\theta}$, $\boldsymbol{\theta}^*$ from the ordinal candidate distribution,

$$\boldsymbol{\theta}^*|\boldsymbol{y},\sigma^2,\boldsymbol{f} \sim N(\boldsymbol{\theta}^o + H(\boldsymbol{\theta}^o)T(\boldsymbol{\theta}^o), H(\boldsymbol{\theta}^o)), \tag{32}$$

where,

$$T(\boldsymbol{\theta}) = \frac{\partial logp(\boldsymbol{\theta}|\boldsymbol{y},\sigma^2,\boldsymbol{f})}{\partial\boldsymbol{\theta}},\ H(\boldsymbol{\theta}) = -\Big[\frac{\partial logp(\boldsymbol{\theta}|\boldsymbol{y},\sigma^2,\boldsymbol{f})}{\partial\boldsymbol{\theta}}\Big]^{-1}, \tag{33}$$

and accept the value with the following probability:

$$min\Big(\frac{q(\boldsymbol{\theta}^*|\boldsymbol{y},\sigma^2,\boldsymbol{f})_{QB*}\pi(\boldsymbol{\theta}^o|\boldsymbol{\theta}^*,\boldsymbol{y},\sigma^2,\boldsymbol{f})}{q(\boldsymbol{\theta}^o|\boldsymbol{y},\sigma^2,\boldsymbol{f})_{QB*}\pi(\boldsymbol{\theta}^*|\boldsymbol{\theta}^o,\boldsymbol{y},\sigma^2,\boldsymbol{f})}, 1\Big), \tag{34}$$

where $\pi()$ is pdf of candidate distribution.

Sampling $\boldsymbol{f}$

The conditional distribution of f_i is

$$\begin{aligned} p(f_i|\boldsymbol{y},\boldsymbol{\theta},\sigma^2) &\propto p(f_i|\sigma^2)\{\prod_{t=1}^{T_i} p(y_{it}|f_i,\boldsymbol{\theta})\} \\ &= exp\{-\frac{1}{2\sigma^2}f_i^2\}\Bigg\{\prod_{t=1}^{T_i}\Big[\frac{exp(\alpha+x_i\beta+w_{it}\gamma+f_i)}{1+exp(\alpha+x_i\beta+w_{it}\gamma+f_i)}\Big]^{y_{it}} \\ &\times\Big[\frac{1}{1+exp(\alpha+x_i\beta+w_{it}\gamma+f_i)}\Big]^{1-y_{it}}\Bigg\}. \end{aligned} \tag{35}$$

which is the ordinal full conditional posterior distribution of the latent variables. We draw new candidate sample from

$$f_i^*|\boldsymbol{y},\sigma^2 \sim N(f_i^o,\tau^2), \tag{36}$$

where τ^2 is a variance parameter of random-walk Metropolis-Hastings algorithm, and accept the value with the following probability:

$$min\Big(\frac{p(f_i^*|\boldsymbol{y},\boldsymbol{\theta},\sigma^2)}{p(f_i^o|\boldsymbol{y},\boldsymbol{\theta},\sigma^2)},1\Big). \tag{37}$$

Sampling σ^2

σ^2 is drawn from

$$p(\sigma^2|\boldsymbol{\theta},\boldsymbol{f},\boldsymbol{y}) = \frac{\{\prod_{i=1}^n p(f_i|\sigma^2)\}\times p(\sigma^2)}{\int\{\prod_{i=1}^n p(f_i|\sigma^2)\}\times p(\sigma^2)d\sigma^2}, \tag{38}$$

which is widely known as the ordinal full conditional posterior distribution of the variance parameters like linear regression model.

Now, we explain the method of generating biased datasets. First, we generate artificial data sets for the usual GLMM using true parameters. Second, we create a biased dataset. To be more concrete, we set the probability of missing to be a decreasing function of the expectation value, $E(y_{it}|\boldsymbol{\theta}_{true})$. The resulting missing mechanism is not ignorable, and we do not use the missing mechanism (or selection) model.

We set the missing rate to be around 40%. Third, we estimate and compare the accuracy of the estimated parameters of the unconstrained (existing) and constrained (proposed) models for biased data using GLMM, in which we do not consider the missing indicator or selection biases. We show the results of the logistic and Poisson regressions.

In this simulation study, we set the sample size, $n = 500$, and average number of events, $\bar{T}_i = (5, 7, 10)$, and estimate the parameters of the unconstrained and constrained models. We draw 2,000 MCMC samples after a 2,000 burn-in phase, and confirm the convergence of MCMC using Geweke (1992)'s method. We generate 1,000 datasets for each situation. We consider one covariate, x, and one time-varying covariate, w, and obtain a total of 11 moment restrictions on x and w.

We show the mean squared errors (MSEs) and coverage from the 95% credible interval in Table 1. Table 1 also shows the average MSEs and the ratio of MSEs that are scaled based on the proposed model. The proposed model outperforms the unconstrained models without population-level information and is the only model that yields unbiased estimates. Next, we show the box plots of each parameter in Figure 1. From this, we can show the reproducibility of parameters. In the unconstrained model, parameters cannot be estimated appropriately, and the results show that analysis using biased data will lead to biased estimates. On the other hand, the constrained model can estimate parameters appropriately, and it performs better than the unconstrained model. From this, we can understand that the quasi-Bayesian method with latent variables works appropriately.

4.3 Real Data Analysis

In the empirical analysis, we analyze the economic panel data on quantity of purchase goods using a Poisson regression model. In the economics and marketing fields, this model is used for modeling purchase quantities. The Poisson regression model also is widely used for analyzing count data in biostatistics and medical statistics (e.g., Agresti & Kateri 2013; Dobson & Barnett 2008). The Poisson regression model is a GLM in which the error function is the Poisson distribution and the link function is the logarithm.

$$p(y_{it}|\boldsymbol{\theta}, f_i) = \frac{exp(-\lambda_{it})\lambda_{it}^{y_{it}}}{y_{it}!} \tag{39}$$

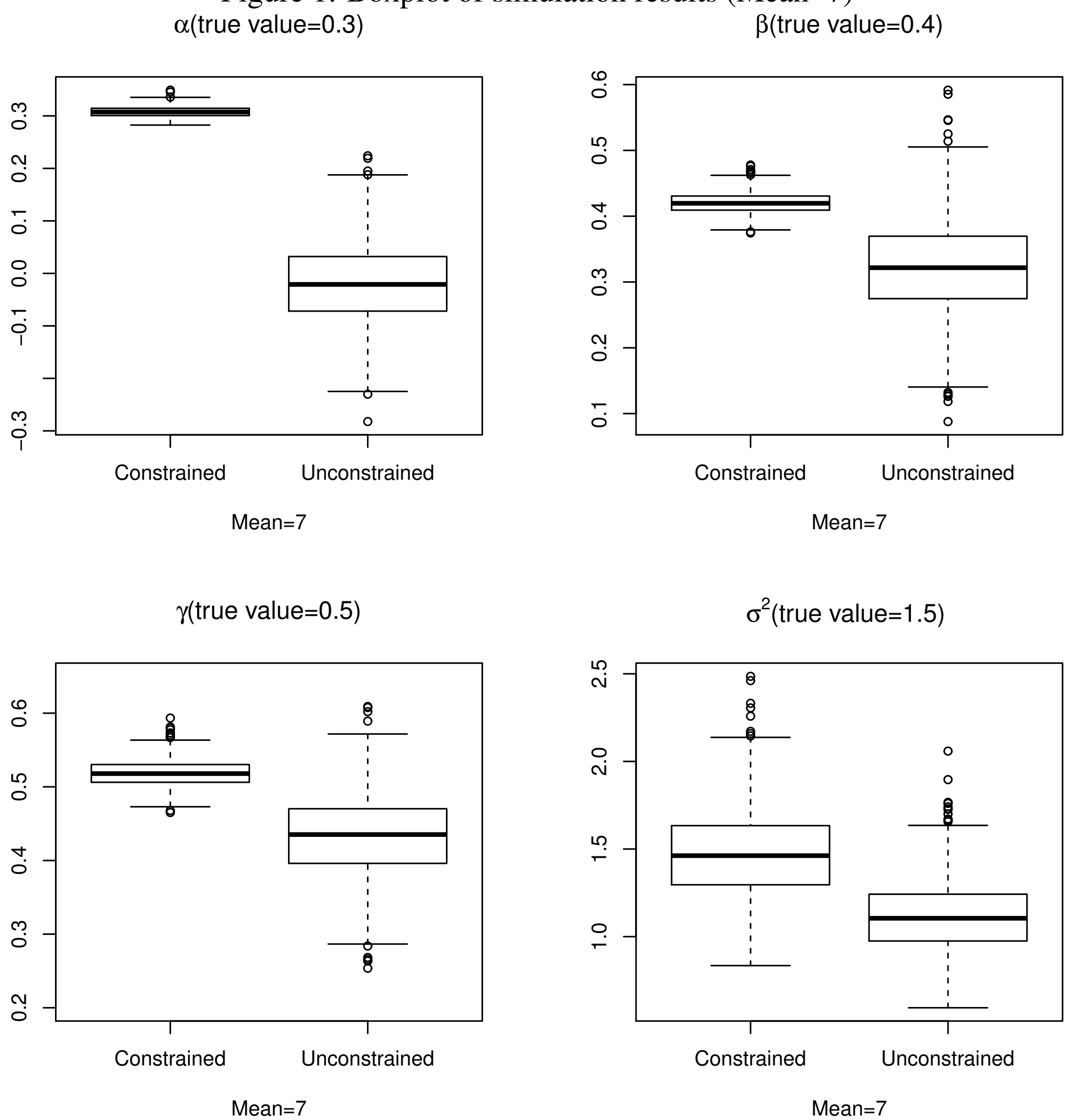
$$log(\lambda_{it}) = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{w}_{it}^T\boldsymbol{\gamma} + f_i \tag{40}$$

Table 1: Simulation results of Logistic regression for biased data

	MSE $\times 10^2$			Coverage(95%)	
	Proposed	Existing	Ratio	Proposed	Existing
Mean=5					
α	0.024	12.843	539.333	1.000	0.010
β	0.092	1.277	13.914	0.990	0.790
γ	0.096	0.971	10.086	0.985	0.756
σ^2	8.822	21.019	2.383	0.920	0.596
Mean=7					
α	0.016	10.767	666.655	0.999	0.016
β	0.068	1.141	16.783	0.994	0.793
γ	0.066	0.749	11.313	0.990	0.758
σ^2	6.119	18.855	3.082	0.926	0.512
Mean=10					
α	0.012	8.239	694.965	1.000	0.027
β	0.055	1.031	18.895	0.998	0.776
γ	0.047	0.540	11.540	0.996	0.752
σ^2	4.744	15.540	3.276	0.886	0.457

In each condition, we generated 10,000 sets.

Figure 1: Boxplot of simulation results (Mean=7)



Here, the moment restriction for y is as follows.

$$m^*(y_{it}|\boldsymbol{\theta}) = \boldsymbol{y}^* - E(y_{it}|\boldsymbol{\theta}) = \boldsymbol{y}^* - \int exp\left\{\alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{w}_{it}^T\boldsymbol{\gamma} + f_i\right\} p(f_i)df_i \qquad (41)$$

Sampling each parameters of Poisson regression model is the same as those of logistic regression model.

In the empirical analysis, we use the *Survey of Consumer Index* (SCI) data provided by Intage Inc. in Japan. The SCI data is the de facto standard for purchase panel data in the Japanese marketing field. The SCI records the purchase incidence, purchased products, number of products purchased by consumers, amounts and prices of products, and stores in which the purchase occurred with dates and times. Although the scanner panel data record purchase histories for all the stores, we used purchase histories from particular stores and regard it to be a complete dataset, which can yield severely biased results. Here, we utilize purchase histories of corner stores. That is, we assume that, although the purchase incidences of each store type are observed, the purchase incidence of other competing stores cannot be observed. This situation is very popular in real data analysis by marketing managers of retail companies. We show the summary statistics in Table 2. We select a sample size $(n = 3,316)$ and total number of events $(= 36,978)$ for the estimation of parameters, which is limited to the purchase histories of corner stores. Corner stores have obvious purchase behavior tendencies, especially regarding quantities and independent variables such as price, compared with other stores. The average purchase quantity of a corner store is lower than that of total stores. On the other hand, the price is higher than that of the total stores, and analysis using this limited information should lead to a biased estimator. To make inferences from this incomplete data, we utilize auxiliary information by aggregating the complete data.

In the analysis, we use purchase data of the cola category from January 2015 to June 2016. We use "gender (male=1)," "age," and "family size" for individual-level covariates, $\boldsymbol{x_i}$; the logarithm of "unit price" for time-varying covariates, w_{it}; and four about "unit price": (1) all, (2) under 100 yen, (3) $100 \sim 120$ yen, and (4) over 120 yen.

The coefficients of price should be negative, because consumers are likely to purchase more products when price discounts are available. In economics and marketing fields, since the effects of price discounts are very important, we compare the coefficients of price with the unconstrained (existing) model.

Table 2: Summary statistics for purchase quantity

	complete data	selected data (corner store)
purchase quantity	1.379	1.105
unit price	101.531	122.093
age	39.373	38.403
gender(male 1)	0.519	0.696
child	0.423	0.378
family size	3.090	3.063

In each model, we draw 10,000 MCMC iterations after a 2,000 burn-in phase and confirm the convergence of MCMC using Geweke (1992)s method. We show the 95% credible intervals of "unit price" in Figure 2 and trace the plots of MCMC in Figure 3. From this, it is obvious that the effects of "unit price" are underestimated in the unconstrained model with respect to the complete data model. The results may cause marketing managers to disregard price discounts for consumers because the effects of price discounts are underestimated. On the other hand, the trace of MCMC in the constrained model is stable compared with the complete data and unconstrained model, which justifies the proposed method.

Figure 2: Coefficient of γ.

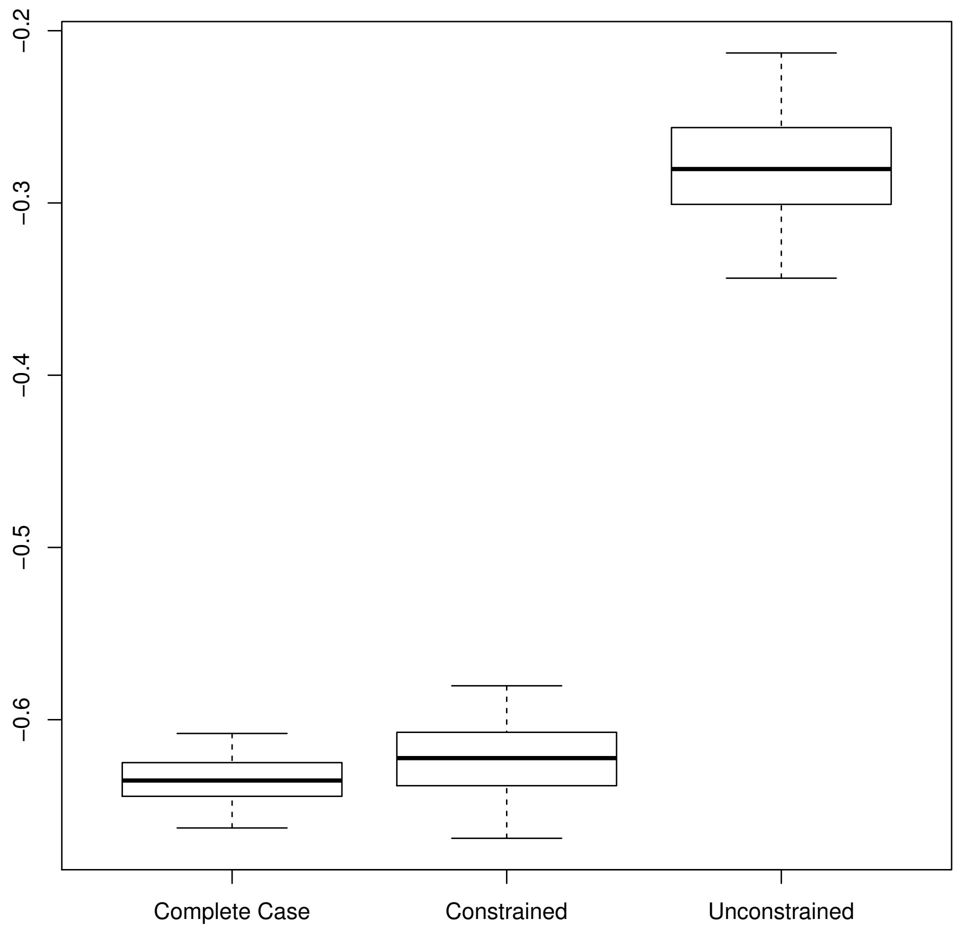
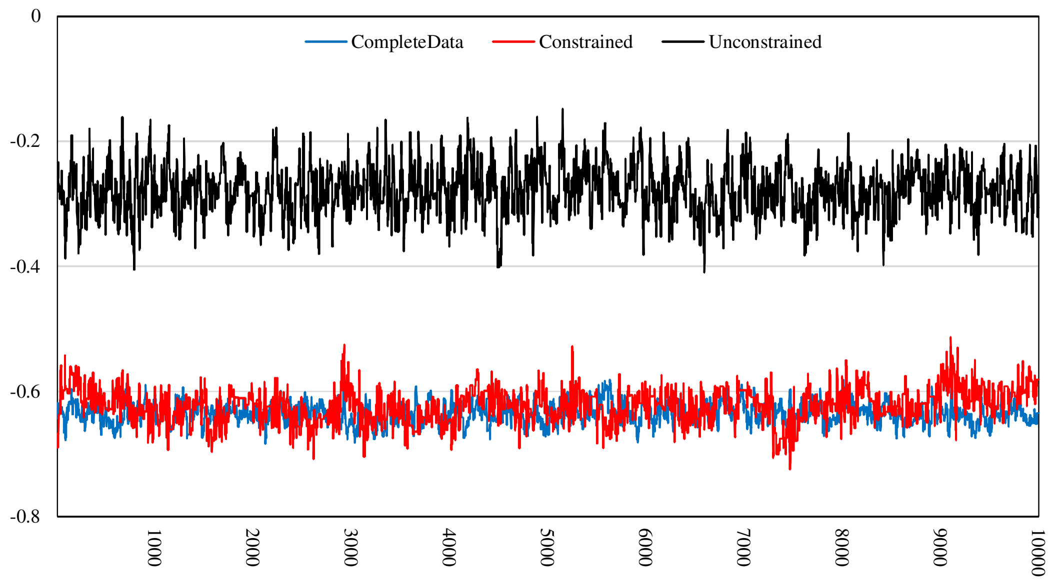


Figure 3: Trace of γ.



5 Conclusion

In this study, we proposed an MCMC algorithm with data augmentation for latent variable models with auxiliary information. We showed consistency and asymptotic variance of the quasi-Bayesian posterior mean estimators from the MCMC outputs. As we illustrated in simulaton studies and real data analysis, the proposed method is especially useful when our dataset is biased. It is usually difficult to correctly specify the sample selection model, while we have an unbiased large sample for some variables or population moments.

The proposed method can be easily generalized to consider non/semi-parametric latent variable models by incorporating the existing semi-parametric Bayesian algorithms such as the Blocked Gibbs sampler in the MCMC iteration, because the proposed quasi-Bayesian posterior distribution is proportional to the product of the likelihood and the function regarding the auxiliary information.

In this paper we focus on GLMMs, among others, but the method can be useful in various important model setups. For example, we can apply the proposed method to data combination of expetimental with observational studies to estimate population treatment effects to enhance generalizability or external validity (Shadish et.al. 2002) of the results obtained in randomized controlled trials due to biased sampling of subjects. In medical sciences, a low degree of generalizability of the results obtained in randomized controlled trials due to biased sampling of subjects has been questioned, and the proposed method will provide valid results without the assumption of selection models when the population-level moment information or a large sample dataset such as the national medical database is available (e.g., Hartman *et al.* 2015).

For another example, consider duration analysis for repeated events (Andersen and Gill 1982; Sinha and Dey 1997; Bijwaard *et al.* 2006), such as clinical trials, unemployment or interpurchase-timing. In many application settings, missing indicators that reveal the presence of missing events between two observed events (intermittent missingness) are not observed. For such cases, simple analyses without considering that some distinct true durations may be summed up to one observed duration can yield severely biased estimates especially in duration analysis. 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, large databases or other research institutes.

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.

In the next coming paper we will propose a duration model with repeated events, which has unobserved intermittent missingness using hybrid posterior incorporating population-level information regarding intermittent missingness. For such models it is not possible to obtain valid estimates to make use of auxiliary information and employ data augmentation approach by using the proposed method.

Appendix: Sketch of Proof of the Theorem 1

In this paper, we assume that the assumptions for GMM type objective function for the marginalized model (Equations (6) and (7)) holds true as in Proposition 1 of Chernozhukov and Hong (2003), which is sufficient for consistency and asymptotic normality of the quasi-Bayesian posterior mean estimator using the (practically infeasible) GMM type objective function for the marginalized model.

First, we show that the mean of the following quasi-Bayesian posterior distribution under marginal model $p(\boldsymbol{y}|\boldsymbol{\theta})$, $q(\boldsymbol{\theta}|\boldsymbol{y})_{QB*}$,

$$q(\boldsymbol{\theta}|\boldsymbol{y})_{QB*} \propto \exp\left\{R_n(\boldsymbol{\theta})\right\} \times p(\boldsymbol{\theta}), \tag{42}$$

where

$$R_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(\boldsymbol{y}|\boldsymbol{\theta}) + Q_n^*(\boldsymbol{\theta}) = l_n(\boldsymbol{\theta}) + Q_n^*(\boldsymbol{\theta}) \tag{43}$$

is constant and asymptotically normally distributed.

To avoid rigorous mathematical formulation, we use the expectation of the optimal weight matrix as $\boldsymbol{\Omega}_n$, but the following argument applies to the case with the estimated weight matrix.

The covariance between the log likelihood and the moment conditions is zero,

$$E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{y}|\boldsymbol{\theta}) \;\; m^*(\boldsymbol{y}|\boldsymbol{\theta})^T\right] = \int \frac{\partial}{\partial \boldsymbol{\theta}} p(\boldsymbol{y}|\boldsymbol{\theta}) m^*(\boldsymbol{y}|\boldsymbol{\theta})^T d\boldsymbol{y} = 0. \tag{44}$$

Then,

$$E\Big[m(\boldsymbol{y}_i|\boldsymbol{\theta}) m(\boldsymbol{y}_i|\boldsymbol{\theta})^T\Big]^{-1} = \begin{pmatrix} I(\boldsymbol{\theta})^{-1} & 0 \\ 0 & E(m^*(\boldsymbol{y}|\boldsymbol{\theta}) m^*(\boldsymbol{y}|\boldsymbol{\theta})^T)^{-1}, \end{pmatrix} \tag{45}$$

where $I(\boldsymbol{\theta})$ is the Fisher information matrix.

By using the expectation of the optimal weight matrix, the GMM objective function (Equation(6)) is expressed as

$$Q_n(\boldsymbol{\theta}) = S_n(\boldsymbol{\theta}) + Q_n^*(\boldsymbol{\theta}) = (S_n(\boldsymbol{\theta} - l_n(\boldsymbol{\theta})) + R_n(\boldsymbol{\theta}), \tag{46}$$

where

$$S_n(\boldsymbol{\theta}) = -\frac{n}{2}\Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}^T} \log p(\boldsymbol{y}_i|\boldsymbol{\theta})\Big] I(\boldsymbol{\theta})^{-1} \Big[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{y}_i|\boldsymbol{\theta})\Big]. \tag{47}$$

From similar arguments of equivalence of the asymptotic distributions of the likelihood ratio test statistics and score test statistics (e.g., Serfling, 1980), for $\boldsymbol{\theta}$ in an open neighborhood of $\boldsymbol{\theta}_0$,

$$S_n(\boldsymbol{\theta}) - \Big[l_n(\boldsymbol{\theta}) - l_n(\boldsymbol{\theta}_0)\Big] \xrightarrow{p} 0. \tag{48}$$

Then the asymptotic properties of the quasi-Bayesian posterior mean estimator using a GMM-type objective function $Q_n(\boldsymbol{\theta})$ as $L_n(\boldsymbol{\theta})$ in Equation (5) apply to the quasi-Bayesian posterior mean estimatior using $R_n(\boldsymbol{\theta})$.

The objective function $R_n(\boldsymbol{\theta})$ satisfies the assumptions required for Theorem 2 (consistency and asymptotic normality of the quasi-Bayesian estimator) in Chernozhukov and Hong (2003) to hold true, and the limiting distribution is equivalent to that of the corresponding extremum estimator.

Next, by integrating the latent variables, we obtain the quasi-Bayesian posterior distribution of $\boldsymbol{\theta}$ as

$$\int q(\boldsymbol{\theta}, \boldsymbol{f}|\boldsymbol{y})_{QB*} d\boldsymbol{f} = \frac{\{\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta})\} \times \exp\left[Q_n^*(\boldsymbol{\theta})\right] \times p(\boldsymbol{\theta})}{\int \{\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta})\} \times \exp\left[Q_n^*(\boldsymbol{\theta})\right] \times p(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \tag{49}$$

which is equivalent to the quasi-Bayesian posterior distribution under the marginal model, $p(\boldsymbol{y}|\boldsymbol{\theta})$, $q(\boldsymbol{\theta}|\boldsymbol{y})_{QB*}$. This shows the consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}_{QB*}$.

References

AGRESTI, A. & KATERI, M. (2011). *Categorical data analysis 3rd*. Springer Berlin Heidelberg.

ALBERT, J. H. & CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. R. Statist. Soc. A.* **88**, 669–679.

ANDERSEN, P.K. & 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. & PAAP, R. (2006). Modeling Purchases as Repeated Eventsa. *J. Bus. Econ. Stat* **24**, 487-502.

CHAUDHURI, S. , HANDCOCK, M. & 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. & HONG, H. (2003). An MCMC approach to classical estimation. *J. Econom.* **115**, 293–346.

CHUNG, Y. & DUNSON, D. B. (2009). Nonparametric Bayes conditional distribution modeling with variable selection. *J. Am. Statist. Ass.* **104**, 1646–1660.

COLE, S. R. & STUART, E. A. (2010). Generalizing evidence from randomized clinical trials to target populations the actg 320 trial.. *Am. J. Epidemiol.* **172**, 107–115.

DOBSON, A. J. & BARNETT, A. (2008). *An introduction to generalized linear models*. CRC press.

DUNSON, D.B. (2000). Bayesian latent variable models for clustered mixed outcomes. *J. R. Statist. Soc. B.* **62**, 355–366.

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.)*. Oxford University Press.

GOLDSTEIN, H. (2010). *Multilevel Statistical Models, 4th ed.*. Wiley.

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

HARTMAN, E. , GRIEVE, R. , RAMSAHAI, R. & SEKHON, J.S. (2015). From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. *J. R. Statist. Soc. A.* **178**, 757–778.

HECKMAN, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–161.

HJORT, N. L., HOLMES, C., MÜLLER, P. & WALKER, S. G. (2010). *Bayesian nonparametrics*. Cambridge University Press.

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.

IMBENS, G. & LANCASTER, T. (1994). Combining micro and macro data in microeconometric models. *Rev. Econ. Stud.* **61**, 655–680.

ISHWARAN, H. & JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Ass.* **96**, 161–173.

KIM, J. Y. (2002). Limited information likelihood and Bayesian analysis. *J. Econom.* **107**, 175–193.

KLEINMAN, K. P. & IBRAHIM, J. G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics* **54**, 921–938.

KUNDU, S. & DUNSON, D.B. (2014). Latent factor models for density estimation. *Biometrika.* **101**, 641–654.

KYUNG, M. , GILL, J. & CASELLA, G. (2010). Are randomized controlled trials the (G) old Standard? From clinical intelligence to prescriptive analytics.. *Ann. Statist.* **38**, 979–1009.

LEE, J. & BERGER, J. O. (2001). Semiparametric Bayesian Analysis of Selection Models. *J. Am. Statist. Ass.* **96**, 1397–1409.

LI, C. & JIANG, W. (2016). On oracle property and asymptotic validity of Bayesian generalized method of moments. *J. Multivar. Anal.* **52**, 132–147.

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

QIN, J. & ZHANG, B. (2007). Empiricallikelihoodbased inference in missing response problems and its application in observational studies. *J. R. Statist. Soc. B.* **69**, 101–122.

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics* John WIley & Sons, New York, NY.

SHADISH, W.R., COOK, T.D. & CAMPBELL, D.T. (2002). *Experimental and Quasi-Experimental Design for Generalized Causal Inference* Houghton Mifflin, Boston, MA.

SINHA, D. & DEY, D. K. (1997). Semiparametric Bayesian analysis of survival dat. *J. Am. Statist. Ass.* **92**, 1195–1212.

TANNER, M. A. & WONG, W. H. (1987). The calculation of posterior distributions by data augmentation.. *J. Am. Statist. Ass.* **82**, 528–540.

VAN POUCKE, S. , THOMEER, M. , HEATH, J. & VUKICEVIC, M. (2016). Are randomized controlled trials the (G) old Standard? From clinical intelligence to prescriptive analytics.. *J Med Internet Res* **18**, 4–14.

WALKER, J. & BEN-AKIVA, M. (2002). Generalized Random Ulitlity Model. *Mathematical Social Sciences.* **43**, 303–343.

YANG, Y. & 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.