Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Semiparametric Quasi-Bayesian Inference with Dirichlet Process Priors: Application to Nonignorable Missing Responses

Igari Ryosuke, Takahiro Hoshino

26 June, 2017
DP2017-020
https://ies.keio.ac.jp/en/publications/7905/

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

Semiparametric Quasi-Bayesian Inference with Dirichlet Process Priors: Application to Nonignorable Missing Responses

Igari Ryosuke, Takahiro Hoshino

Keio-IES DP2017-020
26 June, 2017
JEL classification: C11; C14; C15
Keywords: Dirichlet Process Mixture Model; Blocked Gibbs Sampler; GMM; Auxiliary Information; Selection Model

Abstract

Quasi-Bayesian inference, in which we can use an objective function such as generalized method of moments (GMM), M-estimators, or empirical likelihoods instead of log-likelihood functions, has been studied in Bayesian statistics.However, existing quasi-Bayesian estimation methods do not incorporate Bayesian semiparametric modeling such as Dirichlet process mixtures. In this study, we propose a semiparametric quasi-Bayesian inference with Dirichlet process priors based on the method proposed by Hoshino and Igari (2017) and Igari and Hoshino (2017), which divide the objective function into likelihood function and objective function of GMM.In the proposed method, auxiliary information such as population information can be incorporated in a GMM-type function, whereas the likelihood function is expressed as infinite mixtures.In the resulting Markov chain Monte Carlo (MCMC) algorithm, the GMM-type objective function is considered in the Metropolis Hastings algorithm in the blocked Gibbs sampler.For illustrative purposes, we apply the proposed estimation method to the missing data analysis with nonignorable responses, in which the missingness depends on the dependent variable.We show the performance of our model using a simulation study.

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

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

Acknowledgement: This work was partly supported by JSPS KAKENHI Grant Numbers JP 15J03947, JP 26285151.

Semiparametric Quasi-Bayesian Inference with Dirichlet Process Priors: Application to Nonignorable Missing Responses

June 26, 2017

Ryosuke Igari and Takahiro Hoshino

Graduate School of Economics, Keio University

Abstract

Quasi-Bayesian inference, in which we can use an objective function such as generalized method of moments (GMM), M-estimators, or empirical likelihoods instead of log-likelihood functions, has been studied in Bayesian statistics. However, existing quasi-Bayesian estimation methods do not incorporate Bayesian semiparametric modeling such as Dirichlet process mixtures. In this study, we propose a semiparametric quasi-Bayesian inference with Dirichlet process priors based on the method proposed by Hoshino and Igari (2017) and Igari and Hoshino (2017), which divide the objective function into likelihood function and objective function of GMM. In the proposed method, auxiliary information such as population information can be incorporated in a GMM-type function, whereas the likelihood function is expressed as infinite mixtures. In the resulting Markov chain Monte Carlo (MCMC) algorithm, the GMM-type objective function is considered in the Metropolis Hastings algorithm in the blocked Gibbs sampler. For illustrative purposes, we apply the proposed estimation method to the missing data analysis with nonignorable responses, in which the missingness depends on the dependent variable. We show the performance of our model using a simulation study.

Keyword: *Dirichlet Process Mixture Model; Blocked Gibbs Sampler; GMM; Auxiliary Information; Selection Model; Misspecified Missing Mechanism*

1 Introduction

Recently, quasi-Bayesian inference methods or the Bayesian generalized method of moments (GMM) have been developed and applied in various studies (Chernozhukov and Hong, 2003; Hoshino, 2008; Yin, 2009; Yang and He, 2012). In quasi-Bayesian inference, we can use an objective function such as GMM, M-estimators, or empirical likelihoods instead of log-likelihood functions. Most existing applications of the quasi-Bayesian estimation method emphasize the robustness of the estimation results in that the correct specification of the entire model is not required (e.g., Li and Jiang, 2016). The classical (non-Bayes) GMM methods (or empirical likelihood methods) can easily include more restrictions than the number of parameters (Hansen, 1982). Additionally, vast literature exists that proposes non-Bayesian methods for making an inference that incorporates auxiliary information using classical GMM methods (Imbens and Lancaster, 1994; Hellerstein and Imbens, 1999; Nevo, 2003) and the empirical likelihood method (Qin, 2000; Qin and Zhang, 2007; Chaudhuri *et al.*, 2008). However, numerical optimization (and integration) is required in classical (non-Bayes) GMM methods, which is often difficult in complex models. Even when the optimization of an objective function is difficult, a quasi-Bayesian inference such as Bayesian GMM can be also available by using Markov chain Monte Carlo (MCMC) without numerical optimization. Additionally, the Bayesian GMM can incorporate external information into an objective function, such as the classical GMM.

Moreover, vast literature exists on non or semiparametric Bayesian inference, such as the Dirichlet process mixtures model (DPM) (e.g., Ferguson, 1973; Ishwaran and James, 2001; Hjort *et al.*, 2010). In the DPM, the distribution is expressed using an infinite mixture distribution that can reproduce various distributions, and blocked Gibbs sampling algorithms are available to easily estimate parameters (Ishwaran and James, 2001) in a manner similar to that of Bayesian finite mixture models. Additionally, semiparametric Bayesian estimation methods have been proposed and applied to weaken the parametric assumptions in various fields (e.g., Lee and Berger, 2001; Hoshino, 2013). If semiparametric Bayesian methods are used in a quasi-Bayesian inference, parameters can be estimated using blocked Gibbs sampling without numerical optimization with constraining the restriction from external information such as the GMM. However, the quasi-Bayesian method with latent variable models or semiparametric inference has not been developed. In this circumstance, Hoshino and Igari (2017) and Igari and Hoshino (2017) proposed a new quasi-Bayesian inference that divides the objective function into two components: a likelihood function and moment restrictions. In their method the quasi-posterior mean estimator is shown consistent and asymptotically normally distributed. By dividing the objective functions in a quasi-Bayesian inference, we can incorporate latent variables and moment restrictions from population-level information through data augmentation (e.g., Tanner and Wong, 1987; Albert and Chib, 1993) into the models in an MCMC implementation.

In this paper, we propose a semiparametric quasi-Bayesian inference with Dirichlet process priors using the method proposed by Hoshino and Igari (2017) and Igari and Hoshino (2017). In the proposed method, auxiliary information—such as population information—can be incorporated in the GMM-type function, whereas the likelihood function is expressed as infinite mixtures. In the resulting MCMC algorithm, the GMM-type objective function is considered in the Metropolis Hastings algorithm in the blocked Gibbs sampler. For illustrative purposes, we apply the proposed estimation method to missing data analysis (e.g., Little and Rubin, 2002) with nonignorable missing responses in which the missingness depends on the response variable. Several models have been proposed, such as the Tobit type II model (Heckman, 1979; Amemiya, 1984) for nonignorable missingness; however, the selection mechanism must be correctly specified. Other methods employ non or semiparametric model formulation for the selection model (e.g., Lee and Berger, 2001; van Hasselt, 2011; Hoshino, 2013); however, these models are weakly identified and the results obtained are instable. Instead, in this study, we use marginal population-level information to avoid model misspecification.

2 Semiparametric Quasi-Bayesian Inference

2.1 Dirichlet Process Mixture Model

First, we define the semiparametric Bayesian model with the DPM. In the Dirichlet process priors $G \sim DP(\alpha, G_0)$, G is expressed as follows:

$$G(\boldsymbol{\theta}) = \sum_{k=1}^{\infty} \pi_k \delta_{\boldsymbol{\theta}_k}, \ \ \delta_{\boldsymbol{\theta}_k} \sim G_0, \tag{1}$$

where $\delta_{\boldsymbol{\theta}_k}$ denotes a discrete measure concentrated at $\boldsymbol{\theta}_k$ and $\sum_{k=1}^{\infty} \pi_k = 1$.

In practice, let $\boldsymbol{y} = (y_1, ..., y_n)$ be a dependent vector; the model is:

$$p(\boldsymbol{y}|\cdot) = \sum_{k=1}^{\infty} \pi_k p(\boldsymbol{y}|\boldsymbol{\theta}_k). \tag{2}$$

The stick-breaking process is:

$$\pi_1 = V_1,\ \pi_k = V_k \prod_{h=1}^{k-1}(1 - V_h),\ V_k \sim Be(1, \alpha), \tag{3}$$

where $\alpha \sim Ga(a_\alpha, b_\alpha)$.

To estimate the parameters, the blocked Gibbs sampler (Ishwaran and James, 2001) is widely used. In implementing the blocked Gibbs sampler, we generate the multinomial indicator q_i that indicates that subject i is allocated to each component in each MCMC iteration:

$$p(q_i = k|\cdot) = \frac{\pi_k p(y_i|\boldsymbol{\theta}_k, q_i = k)}{\sum_{l=1}^{M} \pi_l p(y_i|\boldsymbol{\theta}_l, q_i = l)}, \tag{4}$$

where M is the maximum number of components.

2.2 Quasi-Bayesian Inference and Bayesian GMM

Next, we define the quasi-Bayesian inference (Chernozhukov and Hong, 2003) with a GMM-type objective function. Let $\boldsymbol{\theta}$ be a parameter with a r-dimensional vector. Then, the quasi-Bayesian posterior is:

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

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 such as GMM, M-estimators, or empirical likelihoods instead of log-likelihood functions (Chernozhukov and Hong, 2003; Hoshino, 2008; Yin, 2009; Yang and He, 2012).

The quasi-Bayesian posterior means are represented as follows:

$$\hat{\boldsymbol{\theta}} = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} q(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = \int_{\boldsymbol{\Theta}} \boldsymbol{\theta} \left(\frac{exp\{L_n(\boldsymbol{\theta})\}p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} exp\{L_n(\boldsymbol{\theta})\}p(\boldsymbol{\theta})d\boldsymbol{\theta}} \right) d\boldsymbol{\theta}. \tag{6}$$

It is shown that under mild regularity conditions, the quasi-Bayesian posterior means are consistent and asymptotically normally distributed (Kim, 2002; Chernozhukov and Hong, 2003; Yin, 2009; Yang and He, 2012).

The GMM-type objective function is defined as follows:

$$L_n(\boldsymbol{\theta}) = -\frac{n}{2}\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}(\boldsymbol{y}_i|\boldsymbol{\theta})\Big)^T \boldsymbol{\Omega}_n^{-1}(\boldsymbol{\theta})\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}(\boldsymbol{y}_i|\boldsymbol{\theta})\Big), \tag{7}$$

where $\boldsymbol{m}(\boldsymbol{y}_i|\boldsymbol{\theta})$ is a moment restriction that is $E\big[\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})\big] = 0$, and $\boldsymbol{\Omega}_n(\boldsymbol{\theta})$ is the optimal weight matrix:

$$\boldsymbol{\Omega}_n(\boldsymbol{\theta}) = E\Big[\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta})^T\Big]. \tag{8}$$

Yin (2009) and Li and Jiang (2016) used the generalized estimation equation (GEE) for moment restriction $\boldsymbol{m}(\boldsymbol{y}_i|\boldsymbol{\theta})$. Their method can be applied to the longitudinal data in which the observations in the same subject are correlated. However, if additional restrictions exist, such as theory constraints, except for the restrictions of each parameter, their method using GEE must simultaneously incorporate both the restrictions of each parameter and additional restrictions in $\boldsymbol{m}(\boldsymbol{y}_i|\boldsymbol{\theta})$, in which flexible modeling, such as latent variable modeling or non/semiparametric Bayesian modeling, can not be admitted.

In this study, to enable flexible modeling, we divide moment restriction $\boldsymbol{m}(\boldsymbol{y}_i|\boldsymbol{\theta})$ into two parts: (1) a likelihood function and (2) additional moment restrictions. That is to say:

$$\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial\boldsymbol{\theta}^T}\log p(\boldsymbol{y}|\boldsymbol{\theta}) \quad \boldsymbol{m}^{*T}(\boldsymbol{y}|\boldsymbol{\theta})\Big)^T. \tag{9}$$

From this formulation, we have some computational flexibility and can construct flexible models, such as random effects or semiparametric models. When we draw other parameters unrelated to additional moment restriction $\boldsymbol{m}^{*T}(\boldsymbol{y}|\boldsymbol{\theta})$ in an MCMC implementation, we can easily draw samples from only the likelihood function (and prior distribution). Additionally, we can easily include latent variables in Equation (9) (see Hoshino and Igari (2017) for details).

In this study, we apply the following quasi-Bayesian joint posterior distribution for parameter vector $\boldsymbol{\theta}$

$$q(\boldsymbol{\theta}|\boldsymbol{y})_{QB*} = \frac{\big\{\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{\theta})\big\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})}{\int \big\{\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{\theta})\big\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{10}$$

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

$$Q_n^*(\boldsymbol{\theta}) = -\frac{n}{2}\Big[\frac{1}{n}\sum_{i=1}^n \boldsymbol{m}^*(\boldsymbol{y}_i|\boldsymbol{\theta})\Big]^T \boldsymbol{\Omega}_n^{*-1}(\boldsymbol{\theta})\Big[\frac{1}{n}\sum_{i=1}^n \boldsymbol{m}^*(\boldsymbol{y}_i|\boldsymbol{\theta})\Big], \tag{11}$$

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

Note that the quasi-Bayesian posterior distribution (Equation (10)) is proportional to the likelihood $\prod_{i=1}^n p(\boldsymbol{y}_i|\boldsymbol{\theta})$ times the following quasi-Bayesian posterior distribution, conditional on the external information of the moment $\boldsymbol{m}^*(\boldsymbol{y}|\boldsymbol{\theta})$:

$$q(\boldsymbol{\theta}|\boldsymbol{m}^*)_{QB*} = \frac{\exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})}{\int\{\exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{12}$$

See Hoshino and Igari (2017) for the proof of the consistency and asymptotic properties of the estimator.

2.3 Semiparametric Quasi-Bayesian Inference and Algorithm

In a semiparametric quasi-Bayesian inference, we replace Equation (10) with the DPM format,

$$\boldsymbol{m}(\boldsymbol{y}|\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial\boldsymbol{\theta}^T}\log\sum_{k=1}^{\infty}p(\boldsymbol{y}|\boldsymbol{\theta}_k, \boldsymbol{q}=k)p(\boldsymbol{q}=k) \quad \boldsymbol{m}^{*T}(\boldsymbol{y}|\boldsymbol{\theta})\Big)^T. \tag{13}$$

where $\boldsymbol{m}^*(\boldsymbol{y}|\boldsymbol{\theta})$ are the moment restriction with the DPM format.

We consider that the S-dimensional auxiliary information $\boldsymbol{y}^* = (y_1^*, \ ..., \ y_S^*)^T$ are available. The moment restrictions $\boldsymbol{m}^*(y_i|\boldsymbol{\theta})$ are determined by letting:

$$\boldsymbol{m}^*(y_i|\boldsymbol{\theta}) = \begin{bmatrix} I_i^1\left[y_1^* - E[y_i|\boldsymbol{\theta}]\right] \\ \cdots \\ I_i^S\left[y_S^* - E[y_i|\boldsymbol{\theta}]\right] \end{bmatrix}, \tag{14}$$

where $I_i^s = 1$ when subject i belongs to group s (e.g., gender or range of age), and the expected value is $E[y_i|\boldsymbol{\theta}] = \sum_{k=1}^{\infty}\pi_k E[y_i|\boldsymbol{\theta}_k]$.

The semiparametric quasi-Bayesian posterior distribution is:

$$q(\boldsymbol{\theta}, \boldsymbol{q}|\boldsymbol{y})_{SQB*} = \frac{\{\prod_{i=1}^n\sum_{k=1}^{\infty}p(\boldsymbol{y}_i|\boldsymbol{\theta}_k, q_i=k)p(q_i=k)\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})}{\int\{\prod_{i=1}^n\sum_{k=1}^{\infty}p(\boldsymbol{y}_i|\boldsymbol{\theta}_k, q_i=k)p(q_i=k)\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{15}$$

Blocked Gibbs Sampling Algorithm

In our augmentation approach, the algorithm for drawing samples of $\boldsymbol{\theta}$ from Equation (10) is very straightforward.

Sampling $\boldsymbol{\theta}_k$

We use the blocked Gibbs sampler and draw $\boldsymbol{\theta}_k$ for each other, which corresponds to Equation

(10).

$$q(\boldsymbol{\theta}_k, \boldsymbol{q}|\boldsymbol{y})_{SQB*} = \frac{\{\prod_{i=1\in\{q_i=k\}}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta}_k, q_i = k)p(q_i = k)\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta}_k)}{\int\{\prod_{i=1\in\{q_i=k\}}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta}_k, q_i = k)p(q_i = k)\} \times \exp\Big[Q_n^*(\boldsymbol{\theta})\Big] \times p(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k}. \tag{16}$$

In this setup, because it is difficult to draw samples $\boldsymbol{\theta}_k$ directly from this distribution, we use the Metropolis-Hastings algorithm by drawing the candidate of $\boldsymbol{\theta}_k$, $\boldsymbol{\theta}_k^{(cand)}$ from the candidate distribution $c(\boldsymbol{\theta}_k)$ and accept the value with the following probability:

$$p(\boldsymbol{\theta}_k^{(old)} \to \boldsymbol{\theta}_k^{(cand)}) = min\left\{\frac{q(\boldsymbol{\theta}_k^{(cand)}, \boldsymbol{q}|\boldsymbol{y})_{SQB*}c(\boldsymbol{\theta}_k^{(old)}|\boldsymbol{\theta}_k^{(cand)})}{q(\boldsymbol{\theta}_k^{(old)}, \boldsymbol{q}|\boldsymbol{y})_{SQB*}c(\boldsymbol{\theta}_k^{(cand)}|\boldsymbol{\theta}_k^{(old)})}, 1\right\}, \tag{17}$$

where $\boldsymbol{\theta}_k^{(old)}$ is the value obtained in the previous MCMC iteration.

Sampling q_i

The probability that subject i belongs to component k given all other parameters is:

$$p(q_i = k|\cdot) = \frac{\pi_k p(\boldsymbol{y}_i|\boldsymbol{\theta}_k, q_i = k)}{\sum_{l=1}^{M} \pi_l p(\boldsymbol{y}_i|\boldsymbol{\theta}_l, q_i = l)}, \tag{18}$$

where M is a the finite number of components.

Similarly, we draw other parameters $V_1, ..., V_{M-1}, r$ using the usual blocked Gibbs sampler (Ishwaran and James, 2001; Gelman *et al.*, 2013) and show them in Section 3.5.

3 Application to Missing Data Analysis

3.1 Semiparametric Selection Model

Let $\boldsymbol{y}$ be a dependent variable vector, $\boldsymbol{x}$ be an independent variable vector, and $\boldsymbol{z}$ be a missing indicator vector. Then, the dependent variable vector $\boldsymbol{y}$ contains nonignorable responses.

The selection model we considered here is:

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\lambda})p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma}), \tag{19}$$

where $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma})$ is a selection model that represents the missing mechanisms, and $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\lambda})$ is a parametric model such as a linear regression, a proportional hazard model, or a Poisson regression model.

We consider a selection model for which y_i is missing, corresponding to the values of its own variable. We define y_i as observed when $z_i = 1$ and y_i is missing when $z_i = 0$. In the selection model, including the sample selection model, it is widely known that if the missing mechanism is misspecified, the estimated results are severely biased. Then, we use the DPM in the selection mechanism $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma})$ to avoid misspecification:

$$p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma}) = \sum_{k=1}^{\infty} \pi_k p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma}_k). \tag{20}$$

Although we focus on the nonignorable missing responses in this study, our models can be easily extended to the missing covariates problem (e.g., Ibrahim *et al.*, 2001; Kato and Hoshino, submitted).

Although the DPM is flexible, the practical identification of the model is often weak and we use moment restrictions from the external information.

3.2 Likelihood Function

The likelihood function of the semiparametric selection model is:

$$\begin{aligned} L &= \prod_{i=1}^{n} \Big\{ p(y_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) p(z_i|y_i, \boldsymbol{\gamma}) \Big\}^{z_i} \Big\{ \int \{ p(y_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) p(z_i|y_i, \boldsymbol{\gamma}) \} dy_i \Big\}^{1-z_i} \\ &= \prod_{i=1}^{n} \Big\{ p(y_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) \sum_{k=1}^{\infty} \pi_k p(z_i|y_i, \boldsymbol{\gamma}_k) \Big\}^{z_i} \Big\{ \int \{ p(y_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) \sum_{k=1}^{\infty} \pi_k p(z_i|y_i, \boldsymbol{\gamma}_k) \} dy_i \Big\}^{1-z_i}. \end{aligned} \tag{21}$$

3.3 Restriction from Population-level Information

The S-dimensional population-level information $\boldsymbol{y}^* = (y_1^*, \ ..., \ y_S^*)^T$ is set to identify the parameters of the true distribution. The moment restrictions from the population-level information $\boldsymbol{m}^*(y_i|\boldsymbol{\lambda})$ are determined by letting:

$$\boldsymbol{m}^*(y_i|\boldsymbol{\lambda}) = \begin{bmatrix} I_i^1 \left[y_1^* - E[y_i|\boldsymbol{x}_i, \boldsymbol{\lambda}] \right] \\ \cdots \\ I_i^S \left[y_S^* - E[y_i|\boldsymbol{x}_i, \boldsymbol{\lambda}] \right] \end{bmatrix}, \tag{22}$$

where $I_i^s = 1$ when subject i belongs to group s (e.g., gender or range of age). The latent variables y_i^{miss} or components of the DPM q_i are not included in the expected value $E[y_i|\boldsymbol{x}_i, \boldsymbol{\lambda}]$, the blocked Gibbs sampler, or Monte Carlo integration, and are not required in the calculation of the moment restriction.

The objective function of the moment restriction is:

$$Q_n^*(\boldsymbol{\lambda}) = -\frac{n}{2} \Big(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}^*(y_i|\boldsymbol{\lambda}) \Big)^T \boldsymbol{\Omega}_n^{*-1}(\boldsymbol{\lambda}) \Big(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}^*(y_i|\boldsymbol{\lambda}) \Big), \tag{23}$$

where $\boldsymbol{\Omega}_n^*(\boldsymbol{\lambda}) = E\Big[\boldsymbol{m}^*(\boldsymbol{y}|\boldsymbol{\lambda})\boldsymbol{m}^*(\boldsymbol{y}|\boldsymbol{\lambda})^T\Big]$.

3.4 Quasi-Bayesian Posterior

$$
\begin{aligned}
q(\boldsymbol{\lambda}, \gamma|\boldsymbol{y}, \boldsymbol{z}) &\propto L \times exp\big\{Q_n^*(\boldsymbol{\lambda})\big\} \times p(\boldsymbol{\lambda}) \times \prod_{k=1}^{M} p(\boldsymbol{\gamma}_k) \\
&= \prod_{i=1}^{n} \Bigg\{p(y_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) \sum_{k=1}^{M} \pi_k p(z_i|y_i, \boldsymbol{\gamma}_k)\Bigg\}^{z_i} \Bigg\{\int \{p(y_i|\boldsymbol{x_i}, \boldsymbol{\lambda}) \sum_{k=1}^{M} \pi_k p(z_i|y_i, \boldsymbol{\gamma}_k)\} dy_i\Bigg\}^{1-z_i} \\
&\quad \times exp\Bigg\{-\frac{n}{2}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}^*(y_i|\boldsymbol{\lambda})\Big)^T \boldsymbol{\Omega}_n^{*-1}(\boldsymbol{\lambda})\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}^*(y_i|\boldsymbol{\lambda})\Big)\Bigg\} \\
&\quad \times p(\boldsymbol{\lambda}) \times \prod_{k=1}^{M} p(\boldsymbol{\gamma}_k),
\end{aligned}
\tag{24}
$$

where $p(\boldsymbol{\lambda})$ and $p(\boldsymbol{\gamma}_k)$ are prior distributions for $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}_k$.

3.5 Estimation

We use the MCMC method to estimate the parameters of the proposed model. In addition to the algorithms in Section 2.3, we introduce the algorithm corresponding to the selection model using data augmentation.

3.5.1 Blocked Gibbs Sampler for DPM

Sampling q_i

The probability that subject i belongs to component k given all other parameters is:

$$
p(q_i = k|\cdot) = \frac{\pi_k p(z_i|y_i, \boldsymbol{\gamma}_k)}{\sum_{l=1}^{M} \pi_l p(z_i|y_i, \boldsymbol{\gamma}_l)}. \tag{25}
$$

Sampling V_k

The stick-breaking weight V_k is generated from the Beta distribution:

$$
V_k \sim Beta\Bigg(1 + n_k, \alpha + \sum_{h=k+1}^{M} n_h\Bigg), \tag{26}
$$

where n_k means the number of subjects that belong to component k.

Sampling α

The precision parameter α plays a role in controlling the prior on the number of clusters

(Gelman *et al.*, 2013). The conditional distribution of α follows the Gamma distribution:

$$\alpha \sim Ga\Bigg(a_\alpha + M - 1, b_\alpha - \sum_{k=1}^{M-1} log(1 - V_k)\Bigg), \tag{27}$$

where a_α and b_α are the hyper parameters of the prior distribution of α.

Sampling $\boldsymbol{\gamma}_k$

Sampling $\boldsymbol{\gamma}_k$ is the same as the Bayesian finite mixture model.

$$p(\boldsymbol{\gamma}_k|\cdot) \propto p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma_k})p(\boldsymbol{\gamma}_k) = \Bigg\{\prod_{i\in\{q_i=k\}}^{n} p(z_i|y_i, q_i = k, \boldsymbol{\gamma_k})\Bigg\}p(\boldsymbol{\gamma}_k) \tag{28}$$

Because the posterior distribution of the logistic regression model is not any probability distribution, we generate samples using the Metropolis-Hastings (MH) algorithm.

3.5.2 Sampling y_i^{miss}

We generate the $y_i^{miss}(z_i = 0)$ through MH data augmentation (Lee and Berger, 2001). The posterior distribution is:

$$p(y_i^{miss}|z_i = 0, \boldsymbol{x}_i, \boldsymbol{\lambda}, \boldsymbol{\gamma}_k, q_i = k) = \frac{p(y_i^{miss}|z_i = 0, \boldsymbol{x}_i, \boldsymbol{\lambda})p(z_i = 0|y_i^{miss}, \boldsymbol{\gamma}_k, q_i = k)}{\int p(y_i^{miss}|z_i = 0, \boldsymbol{x}_i, \boldsymbol{\lambda})p(z_i = 0|y_i^{miss}, \boldsymbol{\gamma}_k, q_i = k)dy_i^{miss}}. \tag{29}$$

We set the candidate distribution, $c(y_i^{miss}|\cdot)$. Then, the probability of accepting a new candidate sample $y^{miss(cand)}$ is as follows:

$$p(y^{miss(old)} \to y^{miss(cand)}) = min\Bigg\{1, \frac{p(z_i = 0|y_i^{miss(cand)}, \boldsymbol{\gamma}_k, q_i = k)c(y_i^{miss(old)}|y_i^{miss(cand)})}{p(z_i = 0|y_i^{miss(old)}, \boldsymbol{\gamma}_k, q_i = k)c(y_i^{miss(cand)}|y_i^{miss(old)})}\Bigg\}. \tag{30}$$

Then, we set $\boldsymbol{y}^{comp} = \big((\boldsymbol{y}^{obs})^T, (\boldsymbol{y}^{miss})^T\big)^T$ and draw $\boldsymbol{\lambda}$ using them.

3.5.3 Sampling $\boldsymbol{\lambda}$

The $\boldsymbol{\lambda}$ is drawn from the quasi-Bayesian posterior that incorporates a moment restriction from auxiliary information. When we assume the multivariate normal distribution for $p(\boldsymbol{\lambda})$, the quasi-Bayesian posterior is:

$$\begin{aligned}q(\boldsymbol{\lambda}|\boldsymbol{y}^{comp}) &\propto p(\boldsymbol{y}^{comp}|\boldsymbol{x},\boldsymbol{\lambda}) \times exp\big\{Q_n^*(\boldsymbol{\lambda})\big\} \times p(\boldsymbol{\lambda}) \\ &\propto \prod_{i=1}^{n}\{p(\boldsymbol{y}_i|\boldsymbol{x}_i,\boldsymbol{\lambda})\} \times exp\bigg\{-\frac{n}{2}\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}^*(y_i|\boldsymbol{\lambda})\Big)^T\boldsymbol{\Omega}_n^{*-1}(\boldsymbol{\lambda})\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{m}^*(y_i|\boldsymbol{\lambda})\Big)\bigg\} \\ &\quad \times exp\bigg\{-\frac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_0)^T\boldsymbol{\Lambda}_0^{-1}(\boldsymbol{\lambda}-\boldsymbol{\lambda}_0)\bigg\}\end{aligned} \tag{31}$$

where $\boldsymbol{\lambda}_0$ and $\boldsymbol{\Lambda}_0$ are the mean vector and variance-covariance matrix for $p(\boldsymbol{\lambda})$.

We draw a new candidate sample using random walk MH:

$$\boldsymbol{\lambda}^{cand} \sim N(\boldsymbol{\lambda}^{old}, \Psi) \tag{32}$$

where $\boldsymbol{\lambda}^{old}$ is a sample in a previous MCMC iteration and Ψ is a diagonal variance parameter for random walk MH.

Then, the probability of accepting a new candidate sample is:

$$p(\boldsymbol{\lambda}^{old} \to \boldsymbol{\lambda}^{cand}) = min\bigg\{1, \frac{q(\boldsymbol{\lambda}^{cand}|\boldsymbol{y}^{comp},\boldsymbol{x})}{q(\boldsymbol{\lambda}^{old}|\boldsymbol{y}^{comp},\boldsymbol{x})}\bigg\}. \tag{33}$$

4 Simulation Study

4.1 Model

In the simulation study, we assume a linear regression model for $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\lambda})$:

$$y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i, \epsilon_i \sim N(0,\sigma^2), \tag{34}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\beta}^T, \sigma^2)^T$.

Then, the probability of $z_i = 1$ is modeled using a logistic regression model:

$$logit[p(z_i = 1|y_i, \boldsymbol{\gamma})] = \gamma_0 + y_i\gamma_1. \tag{35}$$

However, if the missing mechanisms are misspecified, the estimated results are severely biased.

We consider that the true missing mechanism is expressed by the quadratic function of y_i:

$$logit[p^*(z_i = 1|y_i, \boldsymbol{\gamma}^*)] = \gamma_0^* + y_i\gamma_1^* + y_i^2\gamma_2^*, \tag{36}$$

where $\boldsymbol{\gamma}^* = (\gamma_0^*, \gamma_1^*, \gamma_2^*,)^T$ are true parameters and $p^*(z_i = 1|y_i, \boldsymbol{\gamma}^*)$ is a true missing mechanism.

However, the estimated model (Equation (35)) is misspecified. Then, we use the DPM in the selection mechanism $p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\gamma})$ to avoid misspecification:

$$p(z|y_i, \boldsymbol{\gamma}) = \sum_{k=1}^{\infty} \pi_k p(z|y_i, \boldsymbol{\gamma}_k). \tag{37}$$

4.2 Simulation Condition

Here, we show the performance of the proposed model using a simulation study for the NMAR data using a semiparametric quasi-Bayesian selection model. Now, we explain the method of generating the data. First, we generate independent variable x_i and dependent variable y_i on the regression model using true parameters. Second, we create a missing indicator z_i using a logistic regression model based on the quadratic function of y_i. To be more concrete, we set the probability in which the dependent variable y_i is missing as a function of the value y_i and y_i^2. Third, we let y_i be missing when $z_i = 0$.

We set the proportion of missing indicators to about $20 \sim 30\%$. We compose the number of moment restrictions ($NMR = 3, 5, 7$) from auxiliary information on the dependent variable y among the range of the covariate x. We set the sample size $n = 1000$ and generate 1,000 datasets.

For the model comparison, we estimate six models:

(1) Ignoring Missingness (list-wise case deletion) without Moment

(2) Selection Model without Moment

(3) DPM Selection Model without Moment (Lee and Berger, 2001)

(4) Ignoring Missingness (list-wise case deletion) with Moment (Bayesian alternative to Imbens and Lancastor, 1994)

(5) Selection Model with Moment

(6) DPM Selection Model with Moment (Proposed Model)

In the MCMC procedure, we draw 4,000 MCMC samples after 2,000 burn-in phases and confirm the convergence of the parameters using the method in Geweke (1992).

4.3 Results

We show in Table.1 the MSE ($\times 10^2$) and the ratio of the MSE in which the MSEs of model.6 are fixed to one, and in Table.2 coverage from a 99% Bayesian credible interval. From Table.1, results of model.6 (=proposed model) perform the best for all parameters and conditions. In contrast, from Table.2, the coverages of model.6 (=proposed model) also perform the best for all parameters and conditions. In particular, the coverages of variance parameter σ^2 are bad in all models except for those of the proposed model.

Next, we show the boxplot of each parameter in the cases $1,000$ when $NMR = 5$ in Figure.3. From them, we understand that the proposed models can appropriately reproduce the true parameters; however, those from the other models have large biases.

Table 1: Simulation Result (MSE)

	MSE×10^2						**MSE Ratio**					
	Model.1	Model.2	Model.3	Model.4	Model.5	Model.6	Model.1	Model.2	Model.3	Model.4	Model.5	Model.6
NMR=3												
β_0	16.160	1.503	1.376	0.192	1.042	0.018	906.3	84.3	77.1	10.7	58.4	-
β_1	37.533	25.624	11.062	0.162	4.897	0.032	1176.7	803.3	346.8	5.1	153.5	-
σ^2	66.272	44.221	33.045	23.376	32.061	0.802	82.6	55.1	41.2	29.2	40.0	-
NMR=5												
β_0	16.187	1.482	1.370	0.023	0.303	0.009	1892.8	173.3	160.2	2.7	35.4	-
β_1	37.511	25.556	11.062	0.104	1.055	0.026	1445.5	984.8	426.3	4.0	40.6	-
σ^2	66.100	43.887	32.852	23.071	22.033	0.907	72.9	48.4	36.2	25.4	24.3	-
NMR=7												
β_0	16.126	1.427	1.308	0.058	9.752	0.027	605.0	53.5	49.1	2.2	365.9	-
β_1	37.545	25.503	10.955	0.151	24.874	0.064	586.7	398.5	171.2	2.4	388.7	-
σ^2	66.379	43.997	32.896	23.919	88.300	1.090	60.9	40.4	30.2	21.9	81.0	-

Table 2: Simulation Result (Coverage)

	Model.1	Model.2	Model.3	Model.4	Model.5	Model.6
NMR=3						
β_0	0.000	0.745	0.735	0.902	0.986	1.000
β_1	0.000	0.000	0.001	0.990	0.001	1.000
σ^2	0.000	0.008	0.007	0.021	0.040	0.989
NMR=5						
β_0	0.000	0.767	0.750	0.998	0.997	1.000
β_1	0.000	0.000	0.000	0.984	0.071	1.000
σ^2	0.000	0.007	0.004	0.015	0.083	0.989
NMR=7						
β_0	0.000	0.774	0.768	1.000	0.894	1.000
β_1	0.000	0.000	0.004	0.999	0.861	1.000
σ^2	0.000	0.008	0.004	0.012	0.143	0.992

Next, we show the trace plot of the proposed model in Figure.4. The trace plot of the

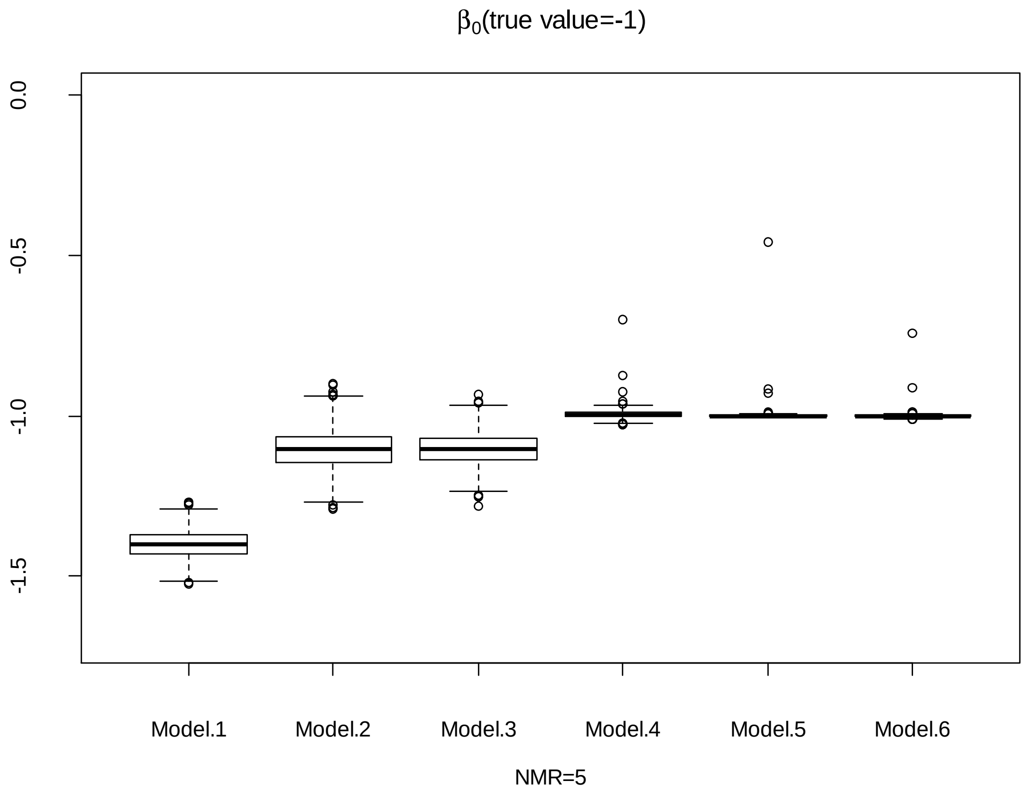

Figure 1: Boxplots of β_0 (NMR=5)

MCMC is stable and convergence is confirmed using the method by Geweke (1992).

5 Conclusion

In this paper, we propose a semiparametric quasi-Bayesian inference with Dirichlet process priors using Hoshino and Igari (2017) and Igari and Hoshino (2017). The proposed method can estimate parameters using a blocked Gibbs sampler, which is one of the major algorithms in the DPM, with incorporating external information into the objective function. For illustrative purposes, we apply the proposed estimation method to missing data analysis with not missing random (NMAR) data. We show the performance of the proposed models using a simulation study. From these studies, the existing models such as the selection model or the semiparametric Bayesian selection model cannot work appropriately when the missing mechanisms are misspecified.

Although we apply our method to the selection models in which moment restrictions from external information are composed only on the $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\lambda})$, the proposed method can be easily generalized to consider the internal function of the Dirichlet process mixture (Equation. (1)). Then, the method and algorithms introduced in Section 2.3 can be applicable. We can easily

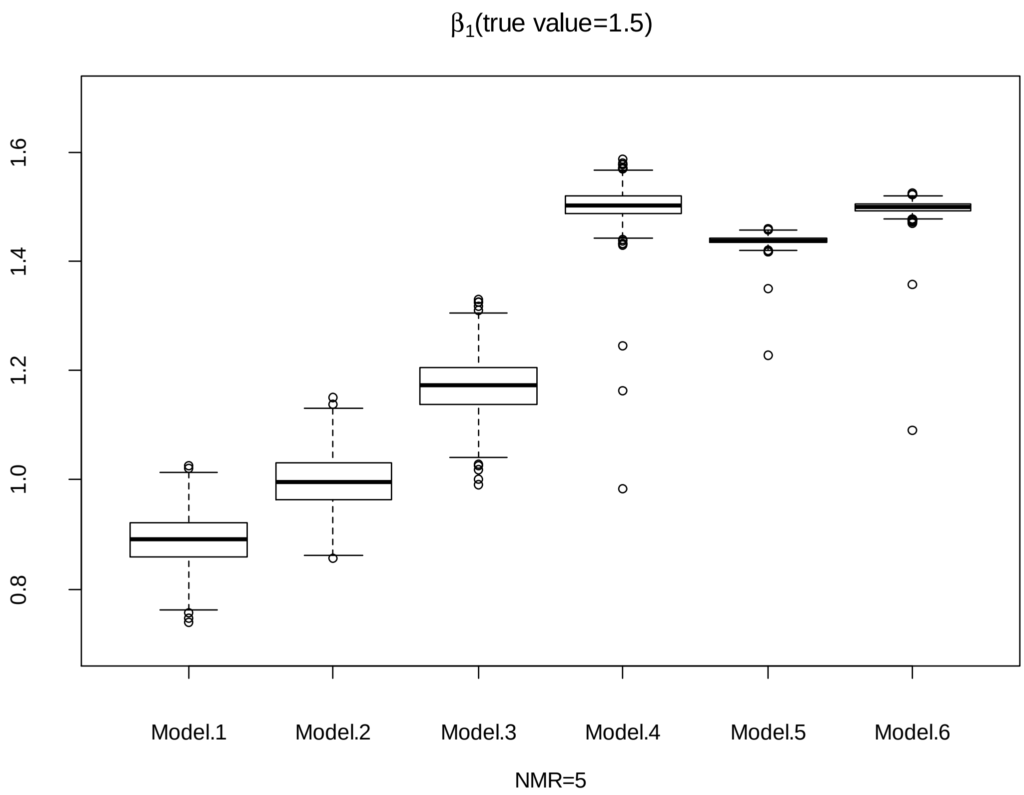

Figure 2: Boxplots of β_1 (NMR=5)

apply our approaches to propensity score adjustments, shared parameter models, pattern mixture models, and sample selection models in a similar manner.

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.

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics.* **24**, 3-61.

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.

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

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics.* **1**, 209-230.

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.

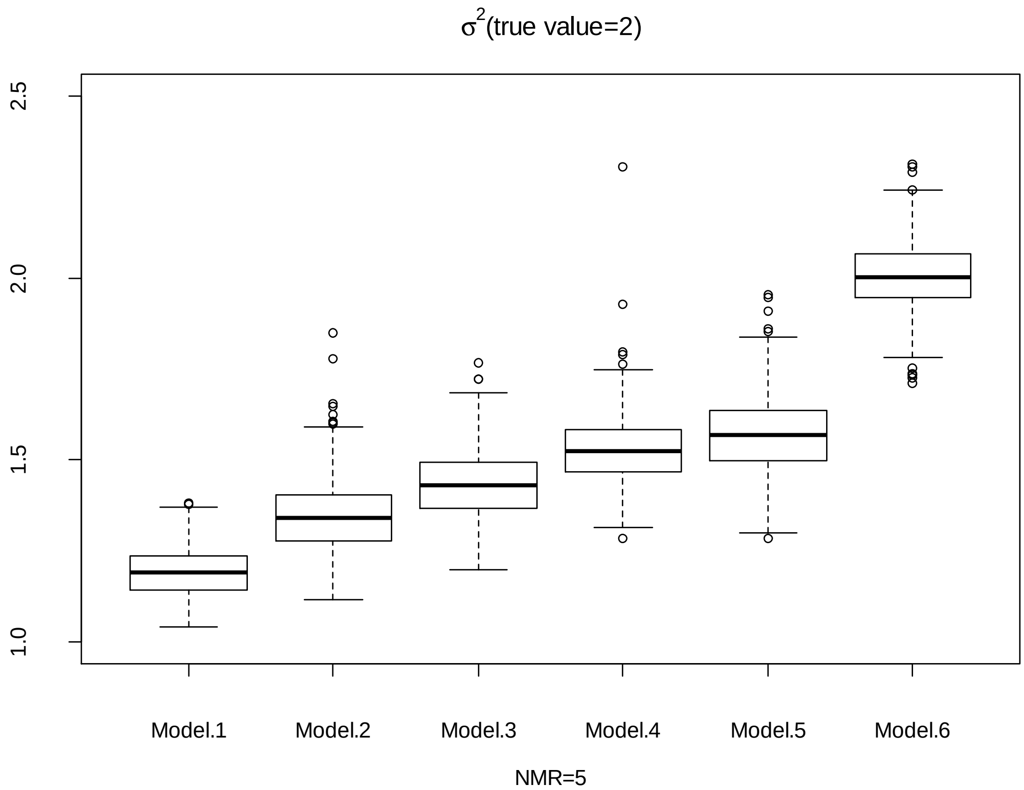

Figure 3: Boxplots of σ^2 (NMR=5)

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.

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

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.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian nonparametrics.* Cambridge University Press.

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

Hoshino, T. (2013). Semiparametric Bayesian estimation for marginal parametric potential outcome modeling: Application to causal inference. *Journal of the American Statistical Association.* **108**, 1189-1204.

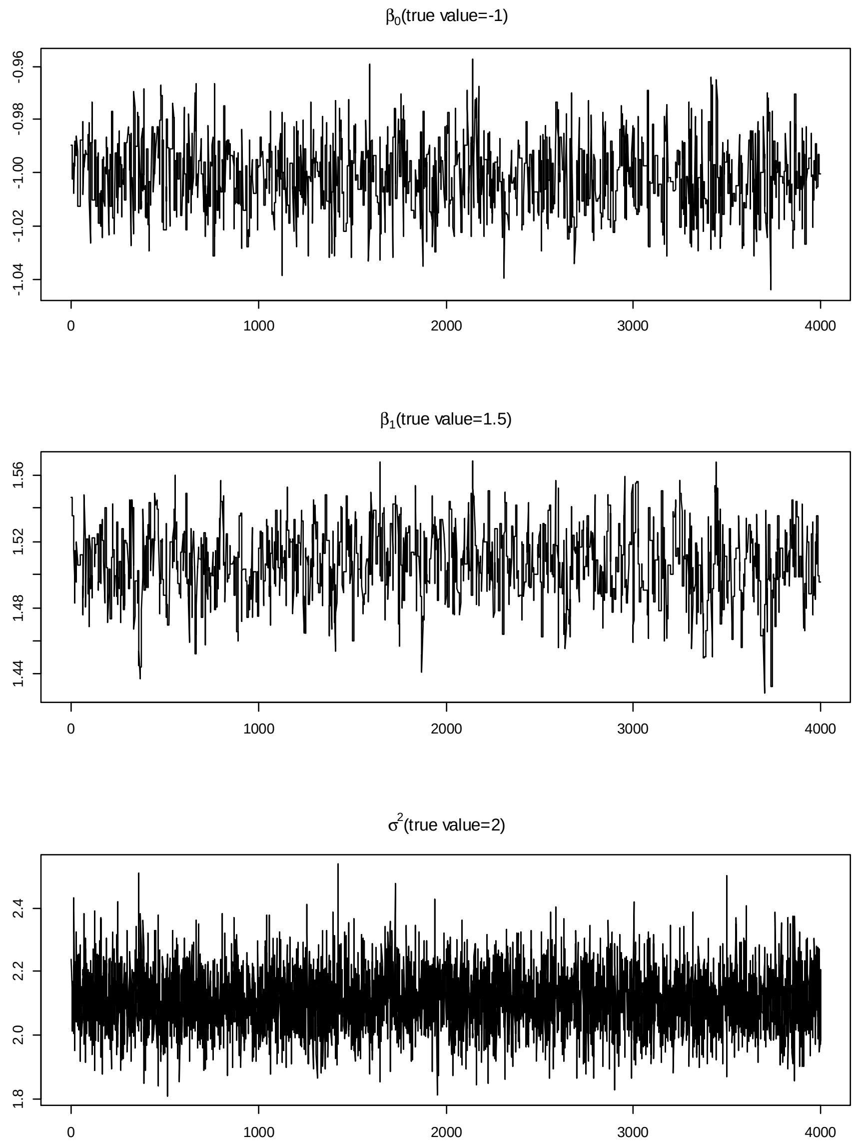

Figure 4: Trace Plot of Proposed Model (NMR=5)

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.*

Ibrahim, J. G. Chen, M. H., and Lipsitz, S. (2001). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika.* **88**, 551-564.

Igari, R., and Hoshino, T. (2017). Bayesian data combination approach for repeated durations under unobserved missing indicators: Application to interpurchase-timing in marketing. *Keio-IES- discussion paper series.*

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

Ishwaran, H., and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*. **96**, 161-173.

Kato, R., and Hoshino, T. (submitted). Semiparametric Bayes multiple imputation for regression models with missing mixed continuous-discrete covariates.

Kim, J. Y. (2002). Limited information likelihood and Bayesian analysis. *Journal of Econometrics*. **107**, 175-193.

Lee, J., and Berger, J. O. (2001). Semiparametric Bayesian analysis of selection models. *Journal of the American Statistical Association*. **96**, 1397-1409.

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.

Little, R. J., and Rubin, D. B. (2002). *Statistical analysis with missing data*. J Wiley & Sons.

Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *Journal of Business & Economic Statistics*. **21**, 43-52.

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

Qin, J., and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B*. **69**, 101-122.

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.

van Hasselt, M. (2011). Bayesian inference in a sample selection model. *Journal of Econometrics*. **165**, 221-232.

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

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