이 학 박 사 학 위 논 문

Bayesian inference of moment condition model with Bayesian bootstrap and constrained Dirichlet process

베이즈 붓스트랩과 제한 디리클레 과정을 이용한 적률조건모형의 베이즈 추론

2023년 8월

서울대학교 대학원

통계학과

이 재 귀

Bayesian inference of moment condition model with Bayesian bootstrap and constrained Dirichlet process

베이즈 붓스트랩과 제한 디리클레 과정을 이용한 적률조건모형의 베이즈 추론

지도교수 이 재 용

이 논문을 이학박사 학위논문으로 제출함

2023년 4월

서울대학교 대학원

통계학과

이 재 귀

이재귀의 이학박사 학위논문을 인준함

2023년 6월

위 원 장	오 희 석
부위원장	이 재 용
위 원	임 채 영
위 원	조 성 일
위 원	이 광 민

Bayesian inference of moment condition model with Bayesian bootstrap and constrained Dirichlet process

By

Jaegui Lee

A Thesis
Submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
August, 2023

ABSTRACT

Bayesian inference of moment condition model with Bayesian bootstrap and constrained Dirichlet process

Jaegui Lee
The Department of Statistics
The Graduate School
Seoul National University

In this thesis, we consider nonparametric Bayesian methods for moment condition models. In a moment condition model, the parameter θ is defined through a moment condition $\mathbb{E}_F[g(X, \theta)] = 0$, where X is an observation and $g(X, \theta)$ is a moment function. Little research has been conducted on moment condition models using the nonparametric Bayesian methods because moment condition constrains the parameter space, making it difficult to calculate the posterior distribution. We suggest using the Bayesian bootstrap and the constrained Dirichlet process for estimating the parameter of the moment condition models. The posterior distributions under both models are defined on manifolds of the parameter space, which makes the posterior sampling complicated. We solve this problem by obtaining the posterior samples using the constrained Hamiltonian Monte Carlo. We illustrate the proposed methods with various numerical studies.

Keywords: Moment condition model, Bayesian bootstrap, constrained Dirichlet process, constrained Hamiltonian Monte Carlo, Shake algorithm

Student Number: $2016-20273$

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Moment condition models have been widely used in various fields such as econometrics and statistics. In a moment condition model, the parameter $\theta \in \Theta \subset \mathbb{R}^p$ is defined through a moment condition $\mathbb{E}_F[g(X, \theta)] = 0$, where $X \in \mathcal{X}$ is a random variable with unknown distribution F, and $g : \mathcal{X} \times \Theta \mapsto \mathbb{R}^d$ is a moment function. Suppose that the observations $X_1, \cdots, X_n$ are given and the relationship of the number of moment restrictions and the dimension of the parameter θ is $d \geq p$. In just-identified models ($d = p$), a unique solution exists that satisfies the moment condition. There are various methods to estimate such a solution, and one of the representative approaches is the Method of Moments (MM). However, in over-identified models ($d > p$), some moment restrictions may be invalid. This is because the number of moment restrictions exceeds the number of parameters, so the true data-generating process F may not satisfy the moment condition $\mathbb{E}_F[g(X, \theta)] = 0$ for all $\theta \in \Theta$.

A lot of research has been conducted to overcome these over-identified problems. Traditionally, the two-step generalized method of moments (GMM) estimator [Hansen, 1982] was suggested for the moment condition models. The two-step GMM estimator has large sample properties such as strong consistency and asymptotic normality under the stationary and ergodic variables. However, this estimator has a substantial bias [Altonji and Segal, 1996] and a poor coverage rate [Pagan and Robertson, 1997] when the degree of over-identification is high in small samples. Various one-step estimators such as GMM with continuous updating (CU) [Hansen et al., 1996], empirical likelihood (EL) [Owen, 1990, 1988, 2001], and exponential tilting (ET) [Efron, 1981] were suggested to alternate the two-step GMM estimator. These one-step estimators are invariant to the linear transformations of the moment functions [Imbens, 1997; Owen, 2001]. In addition, these estimators not only have the first-order efficiency of the two-step GMM estimator but also have higher-order asymptotic properties superior to the two-step GMM estimator [Imbens, 2002; Newey and Smith, 2004]. Numerous studies had been conducted on which estimator is most preferred among the one-step estimators. In the correctly specified models, EL is more desirable than ET and CU because EL possesses various good properties [DiCiccio et al., 1991; Hall, 1990; Imbens, 1997; Newey and Smith, 2004; Owen, 2001; Qin and Lawless, 1994]. For example, EL has a lower bias than ET and CU in finite samples, and bias-corrected EL is higher-order efficient than any other method of moment estimators. However, in the misspecified models, EL may behave worse than ET. If the moment function is unbounded, the asymptotic variance of EL may not be defined because the denominator of the influence function may be close to zero in the misspecified models. This can adversely affect the asymptotic properties of EL if there exists a misspecification in the model. On the other hand, ET is free from these problems [Imbens et al., 1995] and maintains asymptotic properties even under misspecified models. As a way to compensate for these shortcomings of EL mentioned above, the ETEL estimator was suggested [Schennach, 2004, 2007]. ETEL possesses the same $O(n^{-1})$ bias and $O(n^{-2})$ variance as EL in the correctly specified models and whose asymptotic variance can be defined in the misspecified models.

However, little research has been conducted on the moment condition models using nonparametric Bayesian methods. This is because the moment condition constrains the parameter space, making it difficult to calculate the posterior distribution. Bayesian EL, a method of using EL from a Bayesian perspective, was proposed by Lazar [Lazar, 2003], and Bayesian GMM was proposed by Shin [Shin, 2015]. In addition, Bayesian ETEL (BETEL), a method of analyzing ETEL from a Bayesian perspective, was proposed by Schennach [Schennach, 2005] and further developed by Chib [Chib et al., 2018].

We propose to use the Bayesian bootstrap [Rubin, 1981] and the constrained Dirichlet process [Ferguson, 1973], the nonparametric Bayesian methods, to estimate the parameters of the moment condition models. When estimating the parameters of the moment condition models using the constrained Dirichlet process, we propose two posterior distributions depending on the form of the data. The first case is when there are observations and covariates in the data, and the second case is when there are no covariates in the data and only observations exist. In the first case, we calculate the posterior distribution under the assumption that the distribution of the error terms is the constrained Dirichlet process. In the second case, we calculate the posterior distribution under the assumption that the distribution of observations is the constrained Dirichlet process. The posterior distributions obtained using the constrained Dirichlet process under both cases are largely affected by the base distribution. Therefore, we calculate the posterior distributions using various base distributions and then choose the base distribution as the distribution that estimates parameters well. When estimating the parameters of the moment condition models using the Bayesian bootstrap, we propose the same posterior distributions regardless of the form of data. It is known that the posterior distribution of F obtained using the Dirichlet process converges to the posterior distribution of F obtained using the Bayesian bootstrap as the concentration parameter converges to zero. Therefore, we obtain the posterior distribution of parameter θ using the Bayesian bootstrap from the equation that is modified from the posterior distribution of parameter θ using the constrained Dirichlet process when the concentration parameter converges to zero.

The posterior distributions under both proposed methods are defined on manifolds of the parameter space, which makes the posterior sampling complicated. We solve this problem by obtaining the posterior samples using the constrained Hamiltonian Monte Carlo. In the process of updating the posterior samples in the constrained Hamiltonian Monte Carlo, the Shake [Ryckaert et al., 1977], an algorithm used when there exist constraints on parameters, is used. We estimate the parameters of the moment condition models using various methods in several numerical studies. These examples show that the proposed methods outperform the competing methods by comparing the performance of each method.

The rest of this thesis is organized as follows. In Chapter 2, we suggest using the Bayesian bootstrap and the constrained Dirichlet process for estimating the parameter θ of the moment condition models. The posterior samples of both models are obtained by the constrained Hamiltonian Monte Carlo. We describe the algorithm for obtaining the posterior samples in detail in Chapter 3. In Chapter 4, we compare the proposed methods with the competing methods through various numerical studies. Finally, the conclusion is given in Chapter 5.

Chapter 2

Model

Suppose that $X \in \mathcal{X}$ is a random variable with unknown distribution F, $X_1, \cdots, X_n$ are observations of X, and $\theta \in \Theta \subset \mathbb{R}^p$ is a parameter of interest. We can express d moment restrictions about X and θ as follows.

$$\mathbb{E}_F[g(X, \theta)] = 0. \tag{2.1}$$

The equation (2.1) is called the moment condition where $g : \mathcal{X} \times \Theta \mapsto \mathbb{R}^d$ is a moment function. We want to estimate the parameter θ of the moment condition models. However, in over-identified models, some moment restrictions may be invalid. This is because the number of moment restrictions exceeds the number of parameters, so the true data-generating process F may not satisfy the equation (2.1) for all $\theta \in \Theta$. To solve this problem, we introduce the augmented moment condition which is a reformulation of the moment condition by using a nuisance parameter $\nu \in \mathcal{V} \subset \mathbb{R}^{d-p}$.

Definition 1. *(Augmented moment condition) Suppose that* $\xi := (\theta, \nu) \in (\Theta \times \mathcal{V})$ *where* $\nu \in \mathcal{V}$ *is a nuisance parameter. Then, the following equation is called the augmented moment condition.*

$$\mathbb{E}_F[g^A(X, \xi)] = \int g^A(X, \xi) dF = 0$$

where $g^A(X, \xi) := g(X, \theta) - V,\ V_i = \nu_{i-p} I(i > p)$.

Little research has been conducted on the moment condition models using the nonparametric Bayesian methods because the moment condition constrains the parameter space, making it difficult to calculate the posterior distributions. We propose to use the Bayesian bootstrap and the constrained Dirichlet process, the nonparametric Bayesian methods, to estimate the parameters of the moment condition models.

2.1 Dirichlet process

Dirichlet process (DP) [Ferguson, 1973] is a stochastic process that is a distribution over distribution. It is well known that samples drawn from the DP are discrete distributions, which cannot be described using a finite number of parameters. For this reason, DP is classified as a nonparametric Bayesian method. However, due to the problem that samples drawn from DP are discrete, the application of DP to the analysis was limited. As a solution to this problem, the Dirichlet process mixture model (DPMM) was proposed [Antoniak, 1974]. The DPMM can be widely used in nonparametric Bayesian problems thanks to the development of Markov Chain Monte Carlo (MCMC) techniques [Escobar and West, 1995; Neal, 2000]. DP is basically used for density estimation but is widely used for unsupervised learning such as clustering problems.

DP is the generalization of Dirichlet distribution, its finite-dimensional marginal distribution is the Dirichlet distribution. The definition of DP is as follows.

Definition 2. *(Dirichlet process) Suppose that α is a finite measure on $(\mathbb{R}, \mathcal{X})$. Let DP_α denote the Dirichlet process with the parameter as α. We call $\alpha(\mathcal{X})$ as the concentration parameter and $\bar{\alpha} = \alpha/\alpha(\mathcal{X})$ as the base distribution of DP. Then, the definition of the Dirichlet process is as follows.*

If $F \sim DP_\alpha$, then $(F(B_1), \cdots, F(B_n)) \sim Dir(\alpha(B_1), \cdots, \alpha(B_n))$ where $\{B_i\}_{i=1}^n$ is a measurable finite partition of $\mathcal{X}$.

DP has various properties. Representative ones are the conjugacy property and the Stick-breaking process property, and a description of them is as follows.

Theorem 2.1.1. *(Conjugacy)*

Suppose

$$F \sim DP_\alpha$$
$$X_1, \cdots, X_n | F \sim F.$$

Then

$$F|X_1, \cdots, X_n \sim DP_{\alpha + \sum_{i=1}^n \delta_{X_i}}.$$

Theorem 2.1.2. *(Stick-breaking process) [Sethuraman, 1994] Suppose that A is a positive real number and G_0 is a probability measure on $(\mathbb{R}, \mathcal{X})$. Let*

$$\begin{aligned} \theta_1, \theta_2, \cdots &\overset{i.i.d}{\sim} Beta(1, A) \\ X_1, X_2, \cdots &\overset{i.i.d}{\sim} G_0 \end{aligned}$$

and they are independent of each other. Define

$$\begin{aligned} w_1 &= \theta_1 \\ w_2 &= \theta_2(1 - \theta_1) \\ &\vdots \\ w_n &= \theta_n \prod_{i=1}^{n-1} (1 - \theta_i) \\ &\vdots \end{aligned}$$

Then

$$F = \sum_{i=1}^{\infty} w_i \delta_{X_i} \sim DP_{AG_0}.$$

In this thesis, when estimating the parameters of the moment condition models using DP, the above properties are used.

2.1.1 Constrained Dirichlet process

In this subsection, we introduce assumptions, notations, and Lemma for estimating the parameters of the moment condition models when F is distributed from the Dirichlet process constrained by the moment condition.

Suppose that $F \sim DP_\alpha$. To prevent the moment condition models from being over-identified, an augmented moment function $g^A : \mathcal{X} \times (\Theta \times \mathcal{V}) \mapsto \mathbb{R}^d$ which satisfies the below equation is used.

$$\mathbb{E}_F[g^A(X, \xi)] = 0$$

where $\xi = (\theta, \nu) \in (\Theta \times \mathcal{V}) \subset \mathbb{R}^d$. In parameter ξ, $\theta \in \Theta$ is the parameter of interest and $\nu \in \mathcal{V}$ is a nuisance parameter. For convenience, we define a function $g^* : \mathcal{M}(\mathcal{X}) \mapsto \mathbb{R}^d$ as follows

$$g^*(F) = \xi \Leftrightarrow \int g^A(x, \xi) F(dx) = 0$$

where $\mathcal{M}(\mathcal{X})$ is a collection of probability distributions defined on $\mathcal{X}$. Since the augmented moment condition uniquely determines ξ, g^* is a one-to-one function. Therefore, if g^A is a function that satisfies the condition that $\alpha \circ (g^A)^{-1}$ is not a Dirac measure, then $g^*(F)$ has a density function on μ which is a Lebesgue measure on $\mathbb{R}^d$ [Ghosal and Van der Vaart, 2017]. Therefore, we assume that the augmented moment function g^A satisfies the condition that $\alpha \circ (g^A)^{-1}$ is not a Dirac measure, and define $h(\xi : g^*, \alpha)$ as the density function of $g^*(F)$ on μ.

It is known that DP_α has a (g^*, μ)-disintegration $(DP_{\alpha,\xi}, \xi \in \mathbb{R}^d)$ where $DP_{\alpha,\xi}$ is a finite measure defined on $\mathcal{M}(\mathcal{X})$ [Chang and Pollard, 1997]. $DP_{\alpha,\xi}$ has a fiber $(g^*)^{-1}(\xi)$ as a support, and satisfies the equation (2.3) for all measurable function $f : \mathcal{X} \mapsto \mathbb{R}^+$.

$$(g^*)^{-1}(\xi) \coloneqq \{F \in \mathcal{M}(\mathcal{X}) : g^*(F) = \xi\},\ \xi \in \mathbb{R}^d$$

$$\int_{\mathcal{M}} f(F) DP_\alpha(dF) = \int_{\mathbb{R}^d} \int_{(g^*)^{-1}(\xi)} f(F) DP_{\alpha,\xi}(dF)\mu(d\xi) \tag{2.2}$$

$$= \int_{\mathbb{R}^d} \int_{(g^*)^{-1}(\xi)} f(F) \frac{DP_{\alpha,\xi}(dF)}{DP_{\alpha,\xi}(\mathcal{M})} DP_{\alpha,\xi}(\mathcal{M})\mu(d\xi)$$

$$= \int_{\mathbb{R}^d} \int_{(g^*)^{-1}(\xi)} f(F) DP_\alpha(dF|g^*(F) = \xi) h(\xi : g^*, \alpha)\mu(d\xi) \tag{2.3}$$

where

$$DP_\alpha(dF|g^*(F) = \xi) = \frac{DP_{\alpha,\xi}(dF)}{DP_{\alpha,\xi}(\mathcal{M})}$$

$$h(\xi : g^*, \alpha) = DP_{\alpha,\xi}(\mathcal{M}) = DP_{\alpha,\xi}((g^*)^{-1}(\xi)).$$

The equation (2.2) is satisfied by the Disintegration theorem. We can express more simply the equation (2.3) as

$$\begin{aligned} DP_\alpha(dF) &= DP_{\alpha,\xi}(dF)\mu(d\xi) \\ &= \frac{DP_{\alpha,\xi}(dF)}{DP_{\alpha,\xi}(\mathcal{M})} DP_{\alpha,\xi}(\mathcal{M})\mu(d\xi) \\ &= DP_\alpha(dF|g^*(F) = \xi) h(\xi : g^*, \alpha)\mu(d\xi) \end{aligned}$$

Suppose that $F \sim DP_\alpha$ and $X_1, \cdots, X_n|F \overset{i.i.d.}{\sim} F$. Then Q which is the joint distribution of F and $\mathbf{X}_n = (X_1, \cdots, X_n)$ can be expressed as

$$\begin{aligned} Q(d\mathbf{x}_n, dF) &= \prod_{i=1}^n F(dx_i) DP_\alpha(dF) \\ &= DP_{\alpha + \sum_{i=1}^n \delta_{x_i}}(dF) Polya_\alpha(d\mathbf{x}_n). \end{aligned}$$

where $\mathbf{x}_n = (x_1, \cdots, x_n)$ and $Polya_\alpha(d\mathbf{x}_n)$ is the distribution of Pȯlya sequence. Using the above equation, we can establish the following Lemma.

Lemma 2.1.3. *Suppose that $F \sim DP_\alpha$ and $X_1, \cdots, X_n | F \sim F$. Then the following equation is satisfied.*

$$\prod_{i=1}^{n} F(dx_i) DP_{\alpha,\xi}(dF) = DP_{\alpha+\sum_{i=1}^{n}\delta_{x_i},\xi}(dF) Polya_\alpha(d\mathbf{x}_n), \ \mu\text{−}a.a.\ \xi.$$

Proof. To prove this Lemma, we calculate the (g^*, μ)-disintegration of measure Q. Since the equation (2.4) holds, Q has a (g^*, μ)-disintegration $(\prod_{i=1}^{n} F(dx_i) DP_{\alpha,\xi}(dF), \ \xi \in \mathbb{R}^d)$.

$$\begin{aligned} Q(d\mathbf{x}_n, dF) &= \prod_{i=1}^{n} F(dx_i) DP_\alpha(dF) \\ &= \prod_{i=1}^{n} F(dx_i) DP_{\alpha,\xi}(dF)\mu(d\xi) \end{aligned} \tag{2.4}$$

On the one hand, since the equation (2.5) holds, Q has a (g^*, μ)-disintegration $(DP_{\alpha+\sum_{i=1}^{n}\delta_{x_i},\xi}(dF) Polya_\alpha(d\mathbf{x}_n), \ \xi \in \mathbb{R}^d)$.

$$\begin{aligned} Q(d\mathbf{x}_n, dF) &= DP_{\alpha+\sum_{i=1}^{n}\delta_{x_i}}(dF) Polya_\alpha(d\mathbf{x}_n) \\ &= DP_{\alpha+\sum_{i=1}^{n}\delta_{x_i},\xi}(dF)\mu(d\xi) Polya_\alpha(d\mathbf{x}_n) \\ &= DP_{\alpha+\sum_{i=1}^{n}\delta_{x_i},\xi}(dF) Polya_\alpha(d\mathbf{x}_n)\mu(d\xi) \end{aligned} \tag{2.5}$$

Because of the uniqueness of the (g^*, μ)-disintegration, the equation (2.4) and (2.5) are equivalent. Therefore, the equation (2.6) is satisfied.

$$\prod_{i=1}^{n} F(dx_i) DP_{\alpha,\xi}(dF) = DP_{\alpha+\sum_{i=1}^{n}\delta_{x_i},\xi}(dF) Polya_\alpha(d\mathbf{x}_n), \ \mu - a.a\ \xi \tag{2.6}$$

2.1.2 Dirichlet process posterior

In this subsection, we estimate the parameter of the moment condition models using the constrained Dirichlet process. We propose two posterior distributions depending on the form of the data. The first case is when there are observations and covariates in the data, and the second case is when there are no covariates in the data and only observations exist.

First, we calculate the posterior distribution of θ when there are observations and covariates in the data.

Model

Suppose that $y_i \in \mathbb{R}^{d_y}$ is an observation, $x_i \in \mathbb{R}^{d_x}$ is a covariate for $i = 1, \cdots, n$, and $\theta \in \Theta$ is a parameter of interest. Define the error terms as

$$\epsilon_i \coloneqq t(y_i, x_i, \theta) \overset{i.i.d.}{\sim} F,\ i = 1, \cdots, n.$$

Since the expectation of the error term is zero, the distribution F and the parameter θ satisfy the moment condition $\mathbb{E}_F[g(\epsilon, \theta)] = 0$. As mentioned in the previous subsection, we use an augmented moment function g^A to prevent the moment condition models from being over-identified. Therefore, we redefine the error term as

$$\epsilon_i \coloneqq t(y_i, x_i, \xi) \overset{i.i.d.}{\sim} F,\ i = 1, \cdots, n.$$

and the distribution F and the parameter $\xi = (\theta, \nu)$ satisfy the below augmented moment condition.

$$\mathbb{E}_F[g^A(\epsilon, \xi)] = \int g^A(\epsilon, \xi) dF(\epsilon) = 0.$$

If $\nu = 0$, the augmented moment condition is equivalent to the moment condition. Therefore, the form of the final model is as given in equation (2.7).

$$\begin{aligned}
\epsilon_i = t(y_i, x_i, \xi) &\overset{i.i.d.}{\sim} F, \quad i = 1, 2, \cdots, n \\
\int g^A(\epsilon, \xi) dF(\epsilon) &= 0 \\
\nu &= 0.
\end{aligned} \tag{2.7}$$

We define an assumption to calculate the posterior distribution of ξ using the constrained Dirichlet process. Before describing the assumption, we denote some notations. Define the error term $\boldsymbol{\epsilon}_n$ for given $\mathbf{x}_n$, $\mathbf{y}_n$, and ξ as

$$\boldsymbol{\epsilon}_n = t(\mathbf{y}_n, \mathbf{x}_n, \xi) = (t(y_1, x_1, \xi), \cdots, t(y_n, x_n, \xi))^T.$$

Also, define $\epsilon^* = (\epsilon_1^*, \ldots, \epsilon_k^*)$ as distinct values of $\boldsymbol{\epsilon}_n$, and define $\Pi(\boldsymbol{\epsilon}_n)$ as a partition of $\boldsymbol{\epsilon}_n$ on the $[n] = \{1, 2, \cdots, n\}$. Then, assumption **A1** is as follows.

A1 Suppose that $\Pi(\boldsymbol{\epsilon}_n)$ is uniquely determined by the given $\mathbf{x}_n$ and $\mathbf{y}_n$. Therefore, $\Pi(\boldsymbol{\epsilon}_n)$s are equivalent for all ξ.

If the function t does not depend on the parameter ξ and has the inverse function for given $\mathbf{x}_n$ and $\mathbf{y}_n$, assumption **A1** is satisfied.

Prior

Suppose that α, a finite measure defined on $\mathcal{X}$, satisfies the below function

$$\alpha(dx) = AG_0(x)dx$$

where A is a positive real number and $G_0(\cdot)$ is a density function on $\mathcal{X}$. We set the prior of (F, ξ) as follows.

$$\begin{aligned}
\pi(dF, d\xi) &= DP_\alpha(dF|g^*(F) = (\theta, \nu))\pi(\theta)d\theta\delta_0(d\nu) \\
&= \frac{DP_{\alpha,(\theta,\nu)}(dF)}{DP_{\alpha,(\theta,\nu)}(\mathcal{M})}\pi(\theta)d\theta\delta_0(d\nu) \\
&= \frac{DP_{\alpha,(\theta,\nu)}(dF)}{h(\theta, \nu : g^*, \alpha)}\pi(\theta)d\theta\delta_0(d\nu).
\end{aligned}$$

where $h(\xi : g^*, \alpha)$ is a density function of $g^*(F)$ when $F \sim DP_\alpha$.

Posterior

We would like to calculate the posterior distribution of (F, ξ). Suppose that $b(y : x, \theta)$ is the density function of y when $\epsilon = t(y, x, \theta) \sim G_0(\cdot)$. Then, the density function b can be calculated as follows.

$$b(y : x, \theta) = G_0(t(y, x, \theta))|t'(y, x, \theta)|.$$

If the function t does not depend on θ, the density function b does not depend on θ, either.

Theorem 2.1.4. *(Posterior distribution) Suppose that the model is equation* (2.7) *and the assumption* **A1** *holds. Then, the posterior distributions of F and θ are*

$$\pi(dF, d\theta|\mathbf{y}_n) \propto DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta, 0)) \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta) \frac{h(\theta, 0 : g^*, \alpha_n(\theta))}{h(\theta, 0 : g^*, \alpha)} \pi(\theta) d\theta$$
$$= DP_{\alpha_n(\theta),(\theta,0)}(dF) \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta) \frac{\pi(\theta)}{h(\theta, 0 : g^*, \alpha)} d\theta$$
$$\pi(d\theta|\mathbf{y}_n) \propto \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta) \frac{h(\theta, 0 : g^*, \alpha_n(\theta))}{h(\theta, 0 : g^*, \alpha)} \pi(\theta) d\theta.$$

where $\alpha_n(\theta) \coloneqq \alpha + \sum_{i=1}^{n} \delta_{\epsilon_i}$.

Proof. First, $\pi(dF, d\xi, d\boldsymbol{\epsilon}_n)$ is calculated as follows.

$$\begin{aligned}
&\pi(dF, d\xi, d\boldsymbol{\epsilon}_n) \\
&= \prod_{i=1}^{n} F(d\epsilon_i) \pi(dF, d\xi) \\
&= \prod_{i=1}^{n} F(d\epsilon_i) \frac{DP_{\alpha,(\theta,\nu)}(dF)}{h(\theta, \nu : g^*, \alpha)} \pi(\theta) d\theta \delta_0(d\nu) \\
&= DP_{\alpha_n(\theta),(\theta,\nu)}(dF) Polya_\alpha(d\boldsymbol{\epsilon}_n) \frac{\pi(\theta)}{h(\theta, \nu : g^*, \alpha)} d\theta \delta_0(d\nu) \qquad (2.8) \\
&= \frac{DP_{\alpha_n(\theta),(\theta,\nu)}(dF)}{h(\theta, \nu : g^*, \alpha_n(\theta))} h(\theta, \nu : g^*, \alpha_n(\theta)) Polya_\alpha(d\boldsymbol{\epsilon}_n) \frac{\pi(\theta)}{h(\theta, \nu : g^*, \alpha)} d\theta \delta_0(d\nu) \\
&= DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta, \nu)) Polya_\alpha(d\boldsymbol{\epsilon}_n) \frac{h(\theta, \nu : g^*, \alpha_n(\theta))}{h(\theta, \nu : g^*, \alpha)} \pi(\theta) d\theta \delta_0(d\nu) \\
&= DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta, \nu)) \Big[\sum_{k=1}^{n} \prod_{i=1}^{k} G_0(\epsilon_i^*) d\epsilon_i^* p_A(\Pi(\boldsymbol{\epsilon}_n)) \Big] \\
&\quad \times \frac{h(\theta, \nu : g^*, \alpha_n(\theta))}{h(\theta, \nu : g^*, \alpha)} \pi(\theta) d\theta \delta_0(d\nu).
\end{aligned}$$

The equation (2.8) holds by the Lemma 2.1.3. Using the above equation, we can calculate the following equation using the variable transformation.

$$\begin{aligned}
&\pi(dF, d\xi, d\mathbf{y}_n)\\
&= DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta,\nu))\Big[\sum_{k=1}^{n}\prod_{i=1}^{k} G_0(t(y_i^*, x_i^*, \theta))|t'(y_i^*, x_i^*, \theta)|dy_i^* p_A(\Pi(\mathbf{y}_n))\Big]\\
&\quad\times \frac{h(\theta,\nu : g^*, \alpha_n(\theta))}{h(\theta,\nu : g^*, \alpha)}\pi(\theta)d\theta\delta_0(d\nu)\\
&= DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta,\nu))\Big[\sum_{k=1}^{n}\prod_{i=1}^{k} b(y_i^* : x_i^*, \theta)dy_i^* p_A(\Pi(\mathbf{y}_n))\Big]\\
&\quad\times \frac{h(\theta,\nu : g^*, \alpha_n(\theta))}{h(\theta,\nu : g^*, \alpha)}\pi(\theta)d\theta\delta_0(d\nu) \qquad (2.9)
\end{aligned}$$

From the equation (2.9), we can get the posterior distribution of F and ξ as

$$\begin{aligned}
\pi(dF, d\xi|\mathbf{y}_n) \propto\ & DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta,\nu))\prod_{i=1}^{k} b(y_i^* : x_i^*, \theta)\\
&\times \frac{h(\theta,\nu : g^*, \alpha_n(\theta))}{h(\theta,\nu : g^*, \alpha)}\pi(\theta)d\theta\delta_0(d\nu). \qquad (2.10)
\end{aligned}$$

The posterior distribution of F and θ is calculated as follows by integrating out the nuisance parameter ν in equation (2.10)

$$\begin{aligned}
\pi(dF, d\theta|\mathbf{y}_n) \propto\ & DP_{\alpha_n(\theta)}(dF|g^*(F) = (\theta,0))\prod_{i=1}^{k} b(y_i^* : x_i^*, \theta)\\
&\times \frac{h(\theta,0 : g^*, \alpha_n(\theta))}{h(\theta,0 : g^*, \alpha)}\pi(\theta)d\theta. \qquad (2.11)
\end{aligned}$$

and the posterior distribution of θ is calculated as follows by integrating out the distribution F in equation (2.11)

$$\pi(dF, d\theta|\mathbf{y}_n) \propto \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta)\frac{h(\theta,0 : g^*, \alpha_n(\theta))}{h(\theta,0 : g^*, \alpha)}\pi(\theta)d\theta. \qquad (2.12)$$

Corollary 2.1.5. *Suppose that the model is equation* (2.7) *and assumption* **A1** *holds. If* $\xi = \theta$, *and the error term* ϵ *does not depend on* θ, *then the posterior distributions of* F *and* θ *are*

$$
\begin{aligned}
\pi(dF, d\theta|\mathbf{y}_n) &\propto DP_{\alpha_n}(dF|g^*(F) = \theta)\frac{h(\theta : g^*, \alpha_n)}{h(\theta : g^*, \alpha)}\pi(\theta)d\theta \\
&= DP_{\alpha_n,\theta}(dF)\frac{\pi(\theta)}{h(\theta : g^*, \alpha)}d\theta && (2.13) \\
\pi(d\theta|\mathbf{y}_n) &\propto \frac{h(\theta : g^*, \alpha_n)}{h(\theta : g^*, \alpha)}\pi(\theta)d\theta && (2.14) \\
\pi(dF|\mathbf{y}_n) &\propto \frac{\pi(g^*(F))}{h(g^*(F) : g^*, \alpha)}DP_{\alpha_n}(dF). && (2.15)
\end{aligned}
$$

where $\alpha_n := \alpha + \sum_{i=1}^n \delta_{\epsilon_i}$ *does not depend on* θ.

Proof. First, we derive the equation (2.13) and (2.14). According to the assumptions, $\xi = \theta$, and each error term ϵ_i does not depend on θ. Hence, $\alpha_n(\theta)$ does not depend on θ and we denote $\alpha_n(\theta) = \alpha_n$. Also, the distribution of $\mathbf{y}_n$ does not depend on θ. Therefore, we can derive the equation (2.13) and (2.14).

Next, we derive the equation (2.15). Suppose that $r(F)$ is a bounded and continuous real-valued function. Then, the following equation holds.

$$
\begin{aligned}
\int r(F)\pi(dF|\mathbf{y}_n) &= \int\int r(F)\pi(dF, d\xi|\mathbf{y}_n) \\
&= \int\int r(F)DP_{\alpha_n}(dF|g^*(F) = \xi)\frac{h(\xi : g^*, \alpha_n)}{h(\xi : g^*, \alpha)}\pi(\xi)d\xi \\
&= \int\int r(F)\frac{\pi(\xi)}{h(\xi : g^*, \alpha)}DP_{\alpha_n}(dF|g^*(F) = \xi)h(\xi : g^*, \alpha_n)d\xi \\
&= \int r(F)\frac{\pi(g^*(F))}{h(g^*(F) : g^*, \alpha)}DP_{\alpha_n}(dF). && (2.16)
\end{aligned}
$$

The equation (2.16) is satisfied by the equation (2.3). Therefore, we can derive the equation (2.15).

Calculation of the posterior distribution

For the calculation of the posterior distribution, the equation (2.17) obtained in Theorem 2.1.4 is used.

$$\pi(dF, d\theta|y_n) \propto DP_{\alpha_n(\theta),(\theta,0)}(dF) \prod_{i=1}^{k} b(y_i^* : \theta, x_i^*) \frac{\pi(\theta)}{h(\theta, 0 : g^*, \alpha)} d\theta. \tag{2.17}$$

However, generating posterior samples using the equation (2.17) is a complex issue. Hence, we use the Stick-breaking process [Sethuraman, 1994] to simplify the posterior distribution of F and θ. $F \sim DP_{\alpha_n(\theta)}$ in the equation (2.17) can be decomposed as

$$F = u \sum_{i=1}^{\infty} [\tau_i \prod_{j<i} (1 - \tau_i)] \delta_{Z_i} + (1 - u) \sum_{i=1}^{n} w_i \delta_{X_i}$$

by using the Stick-breaking process [Sethuraman, 1994] where

$$\begin{aligned} u &\sim Beta(A, n) \\ \tau_1, \tau_2, \cdots &\overset{i.i.d.}{\sim} Beta(1, A) \\ (w_1, \cdots, w_n) &\sim Dir(1, \cdots, 1) \\ Z_1, Z_2, \cdots &\overset{i.i.d.}{\sim} G_0 \end{aligned}$$

Therefore, if we express the posterior distributions of F and θ as the posterior distribution of (θ, u, τ, Z, w), then we can get the below equation.

$$\pi(\theta, u, \tau, Z, w|\mathbf{y}_n) \propto Beta(u|A, n) \prod_{i=1}^{m-1} Beta(\tau_i|1, A) \prod_{i=1}^{m} G_0(Z_i)$$
$$\times Dir(w|1, 1, \ldots, 1) \times \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta) \times \frac{\pi(\theta)}{h(\theta, 0 : g^*, \alpha)}$$

for sufficiently large $m \in \mathbb{N}$. We use the posterior distribution of (θ, u, τ, Z, w) instead of the posterior distribution of (F, θ) for computational convenience.

Second, we calculate the posterior distribution of θ when there are no covariates in the data and only observations exist.

Model

Suppose that $y_i \in \mathbb{R}^{d_y}$ is an observation for $i = 1, \cdots, n$, and $\theta \in \Theta$ is a parameter of interest. We define the distribution of the observations as

$$y_i \overset{i.i.d.}{\sim} F, \; i = 1, \cdots, n.$$

The distribution F and the parameter θ satisfy the moment condition $\mathbb{E}_F[g(y, \theta)] = 0$, and we use an augmented moment function g^A as in the case where there are observations and covariates in the data. Hence, the distribution F and the parameter $\xi = (\theta, \nu)$ satisfy the below augmented moment condition.

$$E_F[g^A(y, \xi)] = \int g^A(y, \xi) dF(y) = 0.$$

For the augmented moment condition to be equivalent to the moment condition, it must be $\nu = 0$. Therefore, the form of the final model is as given in equation (2.18)

$$
\begin{aligned}
y_i &\overset{i.i.d.}{\sim} F, \quad i = 1, 2, \cdots, n \\
\int g^A(y, \xi) dF(y) &= 0 \\
\nu &= 0.
\end{aligned}
\tag{2.18}
$$

Prior

We set the prior of (F, ξ) as follows.

$$
\begin{aligned}
\pi(dF, d\xi) &= DP_\alpha(dF | g^*(F) = (\theta, \nu)) \pi(\theta) d\theta \delta_0(d\nu) \\
&= \frac{DP_{\alpha,(\theta,\nu)}(dF)}{DP_{\alpha,(\theta,\nu)}(\mathcal{M})} \pi(\theta) d\theta \delta_0(d\nu) \\
&= \frac{DP_{\alpha,(\theta,\nu)}(dF)}{h(\theta, \nu : g^*, \alpha)} \pi(\theta) d\theta \delta_0(d\nu).
\end{aligned}
$$

where $h(\xi : g^*, \alpha)$ is a density function of $g^*(F)$ when $F \sim DP_\alpha$.

Posterior

The posterior distributions of F and θ can be obtained using Theorem 2.1.4. However, since observations $\mathbf{y}_n$ do not depend on θ, we can apply Corollary 2.1.5 to calculate the posterior distributions. Therefore, the posterior distributions of F and θ are calculated as follows

$$
\begin{aligned}
\pi(dF, d\theta | \mathbf{y}_n) &\propto DP_{\alpha_n,(\theta,0)}(dF) \frac{\pi(\theta)}{h(\theta, 0 : g^*, \alpha)} d\theta \\
\pi(d\theta | \mathbf{y}_n) &\propto \frac{h(\theta, 0 : g^*, \alpha_n)}{h(\theta, 0 : g^*, \alpha)} \pi(\theta) d\theta
\end{aligned}
$$

where $\alpha_n \coloneqq \alpha + \sum_{i=1}^n \delta_{y_i}$.

2.2 Bayesian bootstrap

The Bayesian bootstrap (BB) [Rubin, 1981] is a Bayesian alternative to the Bootstrap, which is a technique used for estimating uncertainty in a model or parameter of interest. In Bootstrap, the sampling distribution is estimated by resampling the original data sets. On the other hand, in the Bayesian bootstrap, the probability of each original sample being generated is set as a parameter, and then the posterior distribution is calculated.

A lot of research on BB has been done, including studies on various models such as the finite population model [Lo, 1988], right censored data [Hjort, 1991; Lo, 1993], and the proportional hazard model [Kim and Lee, 2003]. As in Bootstrap, BB does not require distributional assumptions for random variables. In addition, BB has several advantages over Bootstrap. First, the posterior distribution obtained using BB has a much smoother shape than the sampling distribution obtained using Bootstrap. Second, BB can infer the parameter of interest, whereas Bootstrap cannot.

Various theoretical studies about BB also have been conducted. BB and Bootstrap are first-order asymptotically equivalent [Lo, 1987]. Furthermore, BB approximation to the posterior distribution of the population mean is more accurate than Bootstrap and standard normal approximations in the second-order sense [Weng, 1989].

In this section, we estimate the parameter of the moment condition models using BB. Unlike when calculating the posterior distributions using DP in the previous section, we propose the same posterior distribution regardless of the form of the data. When proposing the posterior distribution using BB, we use the posterior distributions obtained using DP. This is because there is a relationship between the posterior distribution of F obtained using DP and BB. Before proposing the posterior distribution of the parameters of the moment condition model, we introduce BB briefly.

2.2.1 Bayesian bootstrap

Suppose that $w = (w_1, \cdots, w_n)$ is a probability of observations being generated as follows.

$$\mathbb{P}(X = X_i|w) = w_i, \ i = 1, \cdots, n.$$

The prior of w is set as

$$\pi(w) \propto \prod_{i=1}^{n} w_i^{l_i}$$

and the likelihood function is

$$L(X_1, \cdots, X_n|w) \propto \prod_{i=1}^{n} w_i.$$

Therefore, the posterior distribution of w is calculated as

$$\pi(w|X_1, \cdots, X_n) \propto \prod_{i=1}^{n} w_i^{l_i+1} \sim Dir(l_1 + 2, \cdots, l_n + 2).$$

The hyperparameters of the prior of w can get various values, but commonly, $l_1 = \cdots = l_n = -1$ is used. There are two main reasons for using the hyperparameters this way: the consistency of the posterior distribution and the relationship between BB and DP. The first reason is related to the consistency of the posterior distribution. It is well known that if $\pi(w)$ is a proper prior, then the posterior distribution of w is inconsistent. Therefore, in order for the posterior distribution of w to be consistent, all hyperparameters, $l_1, \cdots, l_n$, should be negative. The second reason is related to the relationship between the posterior distribution of BB and DP. If $l_1 = \cdots = l_n = -1$, then the posterior distribution of w is

$$\pi(w|X_1, \cdots, X_n) \sim Dir(1, \cdots, 1).$$

Hence, the posterior distribution of F in BB is

$$F|X_1, \cdots, X_n = \sum_{i=1}^{n} w_i \delta_{X_i}, \ w \sim Dir(1, \cdots, 1).$$

The above equation is related to the posterior distribution of F in DP. When $F \sim DP_\alpha$, the asymptotic posterior distribution of F in DP is calculated as follows

$$F|X_1, \cdots, X_n \sim DP_{\alpha + \sum_{i=1}^{n} \delta_{X_i}} \overset{\alpha(\mathcal{X}) \to 0}{\rightsquigarrow} DP_{\sum_{i=1}^{n} \delta_{X_i}}.$$

Therefore, the posterior distribution of F in DP converges to the posterior distribution of F in BB as $\alpha(\mathcal{X})$ converges to 0.

For these reasons, the prior and the posterior distribution of w in BB are as follows

$$\pi(w) \propto \prod_{i=1}^{n} w_i^{l_i}$$

$$\pi(w|X_1, \cdots, X_n) \sim Dir(1, \cdots, 1).$$

2.2.2 Asymptotic posterior when $\alpha(\mathcal{X}) \to 0$

We would like to calculate the asymptotic posterior distribution of θ obtained in Theorem 2.1.4 as the concentration parameter of DP converges to 0. Suppose that

$$F_A \sim DP_\alpha$$

We denote the base measure of DP as

$$\alpha(dx) = AG_0(x)dx$$

where A is a positive real number and $G_0(x)$ is a density function on $\mathcal{X}$. To calculate the asymptotic posterior distribution obtained using DP, we assume two conditions.

B1 (Existence of the density function)
Define as

$$F_{\epsilon,\theta} \coloneqq \frac{1}{n}\sum_{i=1}^{n} \delta_{t(y_i,x_i,\theta)}$$

$$\alpha_n(\theta) \coloneqq \alpha + nF_{\epsilon,\theta}$$

and suppose that

$$F_A \sim DP_{\alpha=AG_0(\cdot)}$$

$$F_0 = \delta_Y,\ Y \sim G_0(\cdot)$$

$$F_{A,n} \sim DP_{\alpha_n(\theta)}$$

$$F_{0,n} \sim DP_{nF_{n,\theta}}.$$

Then, $h(\xi : g^*, A)$, $h(\xi : g^*, 0)$, $h(\xi : g^*, A, n)$, and $h(\xi : g^*, 0, n)$, which are density functions of $g^*(F_A)$, $g^*(F_0)$, $g^*(F_{A,n})$, and $g^*(F_{0,n})$, exist.

B2 (Uniformly convergence condition)

There exists $A_0 = A_0(\xi) > 0$ which satisfies

$$\lim_{r\to 0} \sup_{0<A<A_0} \left| \frac{\mathbb{P}(g^*(F_A) \in B(\xi, r))}{\lambda(B(\xi, r))} - h(\xi : g^*, A) \right| = 0$$

$$\lim_{r\to 0} \sup_{0<A<A_0} \left| \frac{\mathbb{P}(g^*(F_{A,n}) \in B(\xi, r))}{\lambda(B(\xi, r))} - h(\xi : g^*, A, n) \right| = 0$$

for a.a-ξ where λ is a Lebesgue measure.

If $h(\xi : g^*, A)$, the density function of $g^*(F_A)$, exists, then the following equation holds [Folland, 1999].

$$\lim_{r\to 0} \left| \frac{\mathbb{P}(g^*(F_A) \in B(\xi, r))}{\lambda(B(\xi, r))} - h(\xi : g^*, A) \right| = 0, \ \forall - a.a\ \xi, \forall A > 0. \tag{2.19}$$

The condition **B2** assumes that equation (2.19) holds uniformly for A.

Lemma 2.2.1. *Assume that **B1** and **B2** hold. Then, the following equations are satisfied.*

$$\lim_{A\to 0} h(\xi : g^*, A) = h(\xi : g^*, 0), \ \forall \xi - a.e.$$

$$\lim_{A\to 0} h(\xi : g^*, A, n) = h(\xi : g^*, 0, n), \ \forall \xi - a.e.$$

Proof. First, we derive the first equation. Set a sequence (A_n) which satisfies $A_n \to 0$. Then, the equation (2.20) holds for $\lambda - a.a\ \xi$ [Folland, 1999]

$$\lim_{r\to 0} \left| \frac{\mathbb{P}(g^*(F_{A_n}) \in B(\xi, r))}{\lambda(B(\xi, r))} - h(\xi : g^*, A_n) \right| = 0, \ \forall - a.a\ \xi, \forall n \tag{2.20}$$

Assume that ξ satisfies the equation (2.20). Then the following equations are satisfied.

$$\begin{aligned}
\lim_{A\to 0} h(\xi : g^*, A) &= \lim_{A\to 0}\lim_{r\to 0} \frac{\mathbb{P}(g^*(F_A) \in B(\xi, r))}{\lambda(B(\xi, r))} \\
&= \lim_{r\to 0}\lim_{A\to 0} \frac{\mathbb{P}(g^*(F_A) \in B(\xi, r))}{\lambda(B(\xi, r))} \\
&= \lim_{r\to 0} \frac{\mathbb{P}(g^*(F_0) \in B(\xi, r))}{\lambda(B(\xi, r))} \\
&= h(\xi : g^*, 0).
\end{aligned}$$

The first and last equations in the above equation hold by the Theorem 3.22 of Folland [Folland, 1999] and the second equation in the above equation holds by the assumption **B2**. Therefore, the first equation is proved.

The second equation can be derived in the same way.

Therefore, when $A \to 0$, the asymptotic posterior distribution of θ is

$$\pi_0(\theta|\mathbf{y}_n) = \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta) \frac{h(\theta, 0 : g^*, 0, n)}{h(\theta, 0 : g^*, 0)} \pi(\theta).$$

Theorem 2.2.2. *Assume that **B1** and **B2** hold. Then, the following equation is satisfied.*

$$\pi(\theta|\mathbf{y}_n) \to \pi_0(\theta|\mathbf{y}_n), \ \forall - a.a.\theta.$$

Proof. The posterior distribution of θ is

$$\pi(\theta|\mathbf{y}_n) = \prod_{i=1}^{k} b(y_i^* : x_i^*, \theta) \frac{h(\theta, 0 : g^*, \alpha_n(\theta))}{h(\theta, 0 : g^*, \alpha)} \pi(\theta).$$

By the Lemma 2.2.1, the following equations are satisfied as $A \to 0$

$$h(\theta, 0 : g^*, A, n) \to h(\theta, 0 : g^*, 0, n),\ \forall - a.e\ \theta$$

$$h(\theta, 0 : g^*, A) \to h(\theta, 0 : g^*, 0),\ \forall - a.e\ \theta$$

Therefore, the following equation is satisfied as $A \to 0$.

$$\pi(\theta|\mathbf{y}_n) \to \pi_0(\theta|\mathbf{y}_n),\ \forall - a.a.\ \theta.$$

2.2.3 Bayesian bootstrap posterior

As mentioned above, when $F \sim DP_\alpha$, the posterior distribution of F converges to the posterior distribution obtained using BB as $\alpha(\mathcal{X})$ converges to zero. Therefore, we use the posterior distributions obtained in Theorem 2.1.4 and Corollary 2.1.5 to deduce the posterior distribution obtained using BB. We propose the posterior distributions obtained using BB according to the form of the data as in Section 2.1. When suggesting the posterior distributions obtained using BB, we remove the influence of the prior of parameter θ.

First, we calculate the posterior distribution of θ when there are observations and covariates in the data. Suppose that $y_i \in \mathbb{R}^{d_y}$ is an observation, $x_i \in \mathbb{R}^{d_x}$ is a covariate for $i = 1, \cdots, n$. Define the error term as

$$\epsilon_i \coloneqq t(y_i, x_i, \theta) \overset{i.i.d.}{\sim} F,\ i = 1, \cdots, n.$$

By the Theorem 2.1.4, the posterior distribution obtained using DP is

$$\pi(dF, d\theta|\mathbf{y}_n) \propto DP_{\alpha_n(\theta),(\theta,0)}(dF)\prod_{i=1}^{k} b(y_i^* : x_i^*, \theta)\frac{\pi(\theta)}{h(\theta, 0 : g^*, \alpha)}d\theta \tag{2.21}$$

where $\alpha_n(\theta) := \alpha + \sum_{i=1}^{n} \delta_{\epsilon_i}$. As $\alpha(\mathcal{X}) \to 0$, the base measure of the posterior distribution converges as follows

$$\alpha_n(\theta) \to \sum_{i=1}^{n} \delta_{\epsilon_i}.$$

However, the density function $h(\theta, 0 : g^*, \alpha)$ in equation (2.21) depends on the base measure α. In addition, $\prod_{i=1}^{k} b(y_i^* : x_i^*, \theta)$ is influenced by the prior of θ. Since we would like to remove the influence of the base measure of DP and the prior of θ, we modify equation (2.21) as follows.

$$\pi(dF, d\theta|\mathbf{y}_n) \propto DP_{\alpha_n(\theta),(\theta,0)}(dF)d\theta \tag{2.22}$$

where $\alpha_n(\theta) = \sum_{i=1}^{n} \delta_{\epsilon_i}$.

Second, we calculate the posterior distribution of θ when there are no covariates in the data and only observations exist. Suppose that $y_i \in \mathbb{R}^{d_y}$ is an observation for $i = 1, \cdots, n$. Then, the distribution of observations is as follows.

$$y_i \overset{i.i.d.}{\sim} F, \; i = 1, \cdots, n.$$

By the Corollary 2.1.5, the posterior distribution obtained using DP is

$$\pi(dF, d\theta|\mathbf{y}_n) \propto DP_{\alpha_n,(\theta,0)}(dF)\frac{\pi(\theta)}{h(\theta, 0 : g^*, \alpha)}d\theta \tag{2.23}$$

where $\alpha_n := \alpha + \sum_{i=1}^{n} \delta_{y_i}$. In the same way as when there are covariates and observations in the data, we would like to remove the influence of the base measure of DP and the prior of θ. Hence, the modified equation is

$$\pi(dF, d\theta|\mathbf{y}_n) \propto DP_{\alpha_n,(\theta,0)}(dF)d\theta \tag{2.24}$$

where $\alpha_n = \sum_{i=1}^{n} \delta_{y_i}$.

Since the form of equation (2.22) and (2.24) are equivalent, the posterior distribution is proposed as follows regardless of the form of the data.

$$\pi(dF, d\theta|\mathbf{y}_n) \propto DP_{\alpha_n,(\theta,0)}(dF)d\theta.$$

where $\alpha_n = \sum_{i=1}^{n} \delta_{y_i}$.

Chapter 3

Algorithm

We suggested estimating the parameter θ of the moment condition models using the Bayesian bootstrap and the constrained Dirichlet process. The posterior distributions of θ obtained using both models are calculated in Chapter 2. However, since the parameter space of θ is constrained by the moment condition, it is difficult to obtain the posterior samples of both models. We solve this problem by generating the posterior samples of both models using the constrained Hamiltonian Monte Carlo.

The process of obtaining samples using the Hamiltonian Monte Carlo consists of two main steps. In the first step, auxiliary parameters are generated from a normal distribution, and in the second step, parameters are updated using Hamiltonian dynamics. In the second step of the Hamiltonian Monte Carlo, we use the Shake algorithm to update the posterior samples of both models. In this Chapter, we introduce the Hamiltonian Monte Carlo, a method used to obtain the posterior samples, and the Shake, an algorithm that updates the posterior samples in the Hamiltonian Monte Carlo when the parameter space is constrained.

3.1 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC), a type of MCMC, is an algorithm for obtaining a sequence of random samples that converges to the posterior distribution. Hamiltonian Monte Carlo was devised by applying Hamiltonian dynamics to Monte Carlo, which was first proposed under the name Hybrid Monte Carlo to tackle calculations in Lattice Quantum Chromodynamics [Duane et al., 1987]. As time went by, Hybrid Monte Carlo started appearing in various textbooks under the name Hamiltonian Monte Carlo [Bishop and Nasrabadi, 2006; MacKay and Mac Kay, 2003]. It began to be used extensively in statistical computing thanks to Neal's influential review [Neal et al., 2011]. In the mid-2010s, NUTS (No-U-Turn Sampler) [Hoffman et al., 2014], an algorithm that updates posterior samples more efficiently in HMC, was developed, which could significantly reduce the time to generate posterior samples.

A sequence of random samples obtained using HMC converges to the posterior distribution much faster than that obtained through the traditional MCMC methods like Metropolis-Hastings and Gibbs sampler. This occurs because of the way the HMC updates samples. In the process of updating samples using HMC, auxiliary parameters are randomly generated from a normal distribution at each iteration. This makes samples obtained using HMC have a lower correlation between the successive samples than that obtained using traditional MCMC methods while maintaining a high probability of acceptance rate. Thanks to this feature, HMC has recently received considerable attention in Bayesian analysis. We describe the process of obtaining a sequence of random samples using HMC in detail in the following paragraphs.

Suppose that q is a parameter of interest. The target density we want to estimate is $\pi(q)$, which is the posterior distribution of q. To apply Hamiltonian dynamics, we introduce an auxiliary parameter p which has the same dimension as q and is distributed from the normal distribution with zero mean vector. The parameters q and p serve as the state and momentum of the particle in Hamiltonian dynamics, respectively. Since Hamiltonian dynamics is used in the process of updating samples in HMC, the target density $\pi(q)$ and the conditional density $\pi(p|q)$ are defined as follows to apply Hamiltonian dynamics.

$$\pi(q) = \exp\left(-U(q)\right)$$
$$\pi(p|q) = \exp\left(-K(p,q)\right)$$

where $U(q)$ and $K(p,q)$ are the potential energy and the kinetic energy, respectively. By classical mechanics, the Hamiltonian $H(p,q)$ can be expressed as the sum of the potential energy $U(q)$ and the kinetic energy $K(p,q)$. Therefore, the Hamiltonian $H(p,q)$ can be expressed in terms of the target density $\pi(q)$ and the conditional density $\pi(p|q)$ as follows.

$$
\begin{aligned}
H(p,q) &= K(p,q) + U(q) \\
&= -\log \pi(p|q) - \log \pi(q) \\
&= \frac{1}{2} p^T M(q)^{-1} p - \log \pi(q).
\end{aligned}
$$

where $M(q)$ is a mass matrix that is symmetric and positive definite.

The algorithm for obtaining a sequence of random samples using HMC consists of two main steps. In the first step, we move the contour of $\pi(p,q)$ by generating p from the normal distribution $N(0, M(q))$. In the second step, samples are updated over one contour of $\pi(p,q)$ using Hamiltonian dynamics. Hamiltonian dynamics is a differential equation that expresses the rate of change in the state and momentum of the particle over time, expressed as follows.

$$
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial q} H(p,q) \tag{3.1}
$$

$$
\frac{\partial q}{\partial t} = \frac{\partial}{\partial p} H(p,q). \tag{3.2}
$$

By the law of conservation of energy, the samples before and after the update using the equation (3.1) and (3.2) have the same Hamiltonian. Therefore, both samples lie on the same contour of the density $\pi(p,q)$. Through this process, we generate a Markov chain $((p^{(t)}, q^{(t)}),\ t = 1, \cdots, N)$ having equation (3.3) as the invariant distribution.

$$
\pi(p,q) = \exp(-H(p,q)) \tag{3.3}
$$

The target density $\pi(q)$ can be estimated using the Markov chain $((q^{(t)}),\ t = 1, \cdots, N)$ obtained from the Markov chain $((p^{(t)}, q^{(t)}),\ t = 1, \cdots, N)$.

3.1.1 Constrained Hamiltonian Monte Carlo: Shake

There are several algorithms used in the second step of HMC for updating samples. As the parameters we want to estimate should satisfy the moment conditions, samples obtained using HMC need to be updated within a constrained parameter space. Therefore, in the process of obtaining the posterior samples of the moment condition models, an algorithm that can update samples on non-linear manifolds should be used. We suggest using the Shake algorithm, one of the methods used in the Hamiltonian Monte Carlo when the parameters of interest are constrained.

The Shake [Ryckaert et al., 1977] is an algorithm of HMC that can be used to update samples when there exist constraints on parameters. When updating samples using the Shake algorithm, the Lagrange multiplier method is used to calculate the rate of change in the state and momentum of the particle over time. In this subsection, the moment condition is expressed as $c(q) = 0$ for convenience. By the Lagrange multiplier method, the Hamiltonian $H^*(p, q)$ to be used in the Shake algorithm can be decomposed as follows.

$$\begin{aligned} H^*(p, q) &= H(p, q) + \lambda^T c(q) \\ &= K(p, q) + U(q) + \lambda^T c(q). \end{aligned}$$

where λ is a Lagrange multiplier. Using this decomposition of Hamiltonian $H^*(p,q)$, we can calculate the rate of change in the state and momentum of the particle over time as follows.

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial q}H^*(p,q) = -\big(\frac{\partial}{\partial q}H(p,q) + C(q)^T\lambda\big) \tag{3.4}$$
$$\frac{\partial q}{\partial t} = \frac{\partial}{\partial p}H^*(p,q) = \frac{\partial}{\partial p}H(p,q) \tag{3.5}$$

where $C(q) = \frac{\partial}{\partial q}c(q)$. In the process of updating samples using the Shake algorithm, the Leapfrog algorithm which is a method of numerically solving ordinary differential equations is used. The process of obtaining a sequence of random samples that converge to the joint distribution $\pi(p,q)$ using the Shake algorithm based on the two equations (3.4) and (3.5) is as follows.

1. Set initial value $q^{(0)}$
2. Repeat $n = 0, 1, \cdots, N-1$
   (a) Generate $p^{(n)} \sim N(0, M(q^{(n)}))$
   (b) Define as $(p_0^{(n)}, q_0^{(n)}) := (p^{(n)}, q^{(n)})$
   (c) Repeat $j = 0, \cdots, L-1$
      i. $p_{j+\frac{1}{2}}^{(n)} = p_j^{(n)} - \frac{\epsilon}{2}\Big(\frac{\partial}{\partial q}H(p_j^{(n)}, q_j^{(n)}) + C(q_j^{(n)})^T\lambda\Big)$
      ii. $q_{j+1}^{(n)} = q_j^{(n)} + \epsilon\frac{\partial}{\partial p}H(p_{j+\frac{1}{2}}^{(n)}, q_j^{(n)})$
      iii. $c(q_{j+1}^{(n)}) = 0$
      iv. $p_{j+1}^{(n)} = p_{j+\frac{1}{2}}^{(n)} - \frac{\epsilon}{2}\Big(\frac{\partial}{\partial q}H(p_{j+\frac{1}{2}}^{(n)}, q_{j+1}^{(n)}) + C(q_{j+1}^{(n)})^T\lambda\Big)$
   (d) Calculate acceptance rate

$$r = \min\Big\{1, \exp\big(H(p^{(n)}, q^{(n)}) - H(p_L^{(n)}, q_L^{(n)})\big)\Big\}$$

(e) Generate $u \sim Unif(0, 1)$

(f) update $q^{(n+1)} = \begin{cases} q_L^{(n)} & \text{if } \ u < r \\ \\ q^{(n)} & \text{if } \quad \text{o.w.} \end{cases}$

where ϵ is a step size of updating, which is a positive small value. Theoretically, in HMC, there is no need to execute the acceptance-rejection step performed in (d) to (f) of Step 2. However, due to numerical errors occurring during the process of updating samples in (c) of Step 2, $(p_0^{(n)}, q_0^{(n)})$ and $(p_L^{(n)}, q_L^{(n)})$ may not be on the same contour of $\pi(p, q)$. This problem is solved by using the acceptance-rejection step which is used in the Metropolis-Hastings.

Chapter 4

Numerical Studies

In this Chapter, we estimate the parameters of the moment condition models using various methods in several examples. Two examples are used for the analysis: in the first example, moments of the nonparametric model are inferred in the just-identified moment condition model, and in the second example, the coefficients of the instrumental variable regression model are estimated in the over-identified moment condition model. In each example, the proposed methods, BB and DP, and various competing methods are used to estimate the parameters of the moment condition models. As competing methods, GMM, EL, and Bootstrap, which are Frequentist methods, and BETEL, which is a nonparametric Bayesian method are used. The performance of each method is compared through various figures such as MSE, bias, and so on.

4.1 Example 1: Moments of nonparametric model

In this section, we would like to estimate the moments of one-dimensional data in the nonparametric model. The data to be used in this example is the garden earthworms data surveyed by Perl & Fuller [Pearl and Fuller, 1905]. Perl & Fuller investigated various characteristics of garden earthworms to examine the variations and correlations of those characteristics. The characteristics used in the survey were the number of somites and the position of the clitellum and so on, and 487 garden earthworms were investigated to confirm this. In this example, the number of somites of garden earthworms is used as a variable.

Before proceeding with the analysis, exploratory data analysis (EDA) was conducted to figure out the data. Since there is only one variable, we draw a histogram to visualize the distribution of the data. The histogram of the number of garden earthworms according to the number of somites is presented in Figure 4.1. Most of the data points are located close to the mean. However, there are several data points that are far away from the mean, and the number of such data points is not small. For this reason, we anticipate that the variance of the data will not be very small. Additionally, the shape of the histogram is not symmetric; it appears to be skewed to the left. Thus, the skewness of the number of somites is expected to be negative.

We would like to estimate the moments of the numbers of

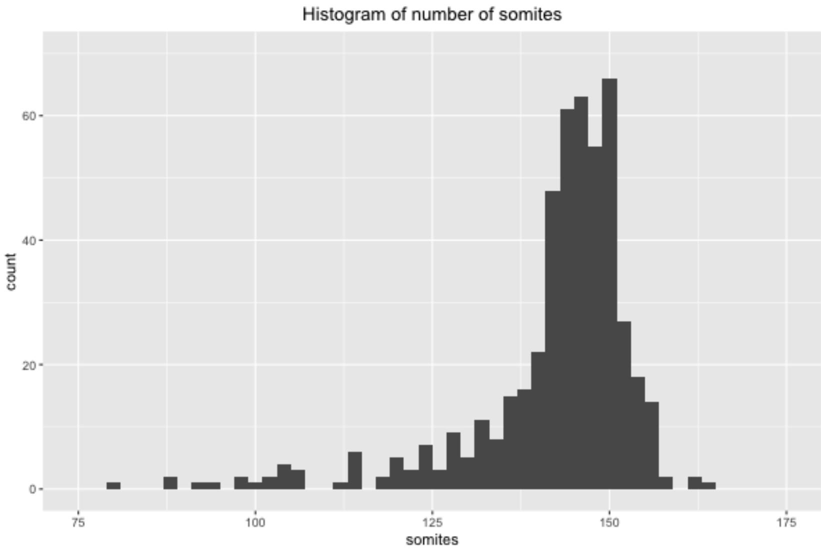

Figure 4.1: Histogram of the number of somites of garden earthworms

somites of garden earthworms using this data. The moments we want to estimate are the mean (μ), standard deviation (σ), skewness (γ), and kurtosis (κ) of the number of somites of the garden earthworm. By definition, the moments should satisfy the restrictions (4.1) to (4.4).

$$\mathbb{E}_F[X - \mu] = 0 \tag{4.1}$$

$$\mathbb{E}_F[(\frac{X - \mu}{\sigma})^2 - 1] = 0 \tag{4.2}$$

$$\mathbb{E}_F[(\frac{X - \mu}{\sigma})^3 - \gamma] = 0 \tag{4.3}$$

$$\mathbb{E}_F[(\frac{X - \mu}{\sigma})^4 - \kappa - 3] = 0. \tag{4.4}$$

The first method that can be thought of for estimating each moment is the Method of Moments. However, the Method of Moments often provides biased estimators as well as fails to estimate moments in over-identified models. Therefore, we estimate moments using various methods for estimating the parameters of the moment condition model instead of the Method of Moments. Through the moment restrictions (4.1) to (4.4), we can derive moment functions g_1 and g_2 to be used in the analysis as follows.

$$g_1(X,\theta) = \begin{pmatrix} X-\theta_1 \\ e^{-2\theta_2}(X-\theta_1)^2-1 \\ e^{-3\theta_2}(X-\theta_1)^3-\theta_3 \\ e^{-4\theta_2}(X-\theta_1)^4-e^{\theta_4} \end{pmatrix}, \quad \text{where } \theta = (\mu, \log(\sigma), \gamma, \log(\kappa+3))$$

$$g_2(X,\theta) = \begin{pmatrix} X-\theta_1 \\ \big((X-\theta_1)/\theta_2\big)^2-1 \\ \big((X-\theta_1)/\theta_2\big)^3-\theta_3 \\ \big((X-\theta_1)/\theta_2\big)^4-(\theta_4+3) \end{pmatrix}, \quad \text{where } \theta = (\mu, \sigma, \gamma, \kappa)$$

In each moment function, θ_i means the ith element of parameter θ. The moment function g_2 is obtained by the restrictions (4.1) to (4.4), and the moment function g_1 is obtained by transforming the σ and κ of g_2. We estimate μ, σ, γ, and κ using these moment functions.

For estimating the parameters of the moment condition models, various methods such as BB and DP, which are proposed methods, and GMM, EL, and Bootstrap, which are Frequentist methods are used. When estimating the parameter of the moment condition models using DP, we set the concentration parameter and the base distribution of DP as $A = 0.5$ and $G_0 = t_{(5)}$, respectively,

		Parameter			
Function	Method	μ	σ	γ	κ
-	MM	142.715	11.853	-2.179	5.857
g_1	GMM	142.715	11.853	-2.179	5.857
g_1	EL	142.715	11.853	-2.179	5.857
g_1	Bootstrap	142.756	11.825	-2.149	5.527
g_1	BB	142.726	11.802	-2.105	5.432
g_1	DP	142.661	12.109	-2.228	6.481
g_2	GMM	142.715	11.853	-2.179	5.857
g_2	EL	142.715	11.853	-2.179	5.857
g_2	Bootstrap	142.756	11.825	-2.149	5.527
g_2	BB	142.716	11.766	-2.151	5.528
g_2	DP	142.674	12.234	-2.288	6.442

Table 4.1: Estimators of $\mu, \sigma, \gamma, \kappa$ for garden earthworm data

and set the prior of θ to a non-informative prior. In the process of estimating the parameters using BB, DP, and Bootstrap, 90,000 samples are used. When obtaining posterior samples of BB and DP, 100,000 samples are generated using CHMC, then the first 10,000 samples are removed through burn-in.

The estimators obtained using each method are given in Table 4.1. Additionally, the estimators obtained through the Method of Moments are also included in Table 4.1. We compare the differences in estimators according to the variable transformation of the moment function and the methods used for parameter estimation.

By comparing the estimators when g_1 and g_2 are used as the moment functions, the difference in estimators according to variable transformation is confirmed. There is little difference in the estimators according to the moment functions. Therefore, it can be concluded that the variable transformation of the moment function does not significantly affect the parameter estimation. Next, we compare the estimators of each method. Upon comparing the estimators of moments, it appears that there is little difference among the methods. In particular, GMM and EL provide estimators that are exactly equivalent to the Method of Moment estimators (MME). The estimators obtained using BB, DP, and Bootstrap have differences from the MME, which seems to be occurred due to the randomness of the samples. Among them, the estimators obtained using BB and Bootstrap are similar to estimators obtained by other methods, whereas estimators obtained using DP exhibits noticeable differences, particularly prominent in the parameter κ.

In order to better figure out the characteristics of each method, the distributions of the estimators obtained using each method are compared. In Figure 4.2 and 4.3, the distributions of μ, σ, γ, and κ are given. Figure 4.2 and 4.3 list the distributions of estimators when g_1 and g_2 are used as the moment functions, respectively.

In each figure, the posterior distributions of parameters were drawn for BB and DP, the empirical distribution for Bootstrap, and the asymptotic distributions for GMM and EL. Each distribution is compared with MME which was indicated by the red vertical line.

At first, we compare the shape of the distributions according to the moment function. As in the previous comparison of estimators, there is little difference in the distributions according to the moment functions. Therefore, it can be concluded that the variable transformation of the moment function does not significantly affect the distributions of the estimators. Next, we compare the distributions of estimators according to the methods. It can be seen that the distributions of estimators obtained using BB, Bootstrap, GMM, and EL are symmetric, while that of DP are not. The degree of asymmetry of the posterior distributions of estimators obtained using DP increase as the order of the moment increases. Specifically, the distribution of μ shows little difference between DP and the other methods, and its shape remains close to symmetry. On the other hand, the distributions of γ and κ exhibit substantial differences between DP and the other methods, with the posterior distributions of estimators obtained using DP being highly skewed and having long tails. Among the symmetric distributions, the distributions obtained using BB, Bootstrap, and EL have similar shapes. Whereas, the estimators obtained using GMM have larger variances than the estimators obtained using other methods. For this reason, the distributions of the estimators using GMM have different shapes from that of the other methods.

We thought that the difference between the posterior distributions obtained using DP and the distribution obtained using the other methods was caused by the wrong choice of the base measure

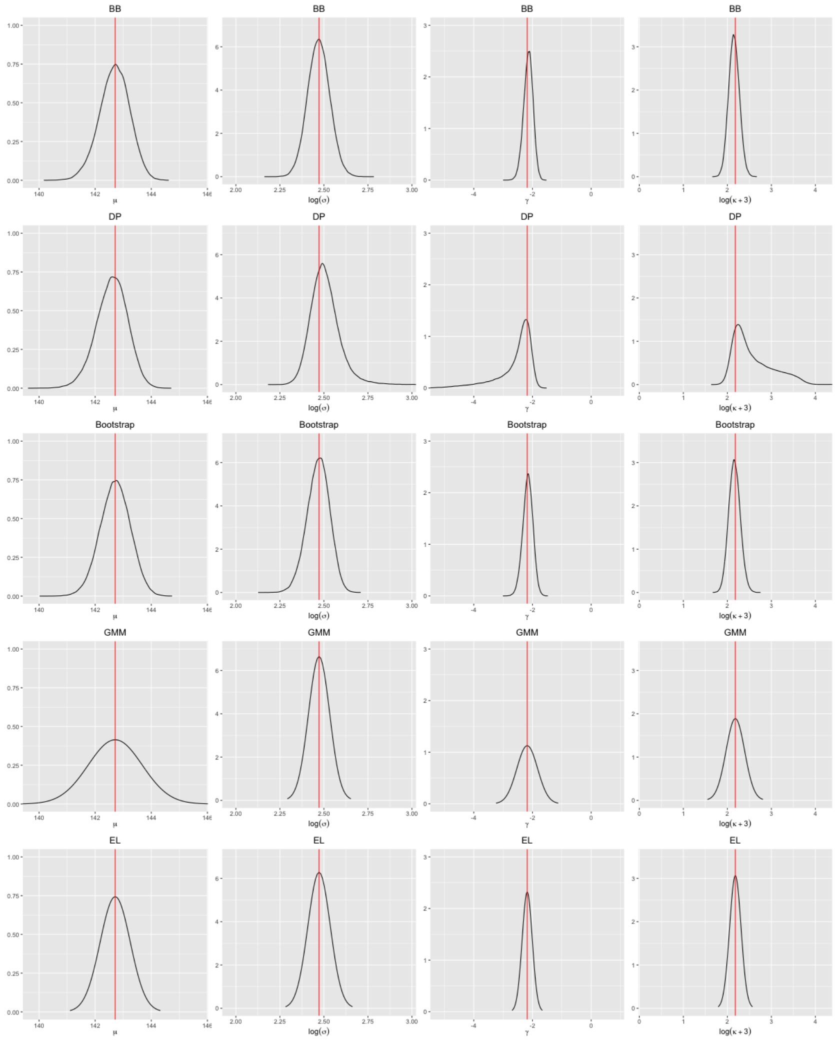

Figure 4.2: Distributions of μ, σ, γ, κ for garden earthworm data using g_1 as a moment function.

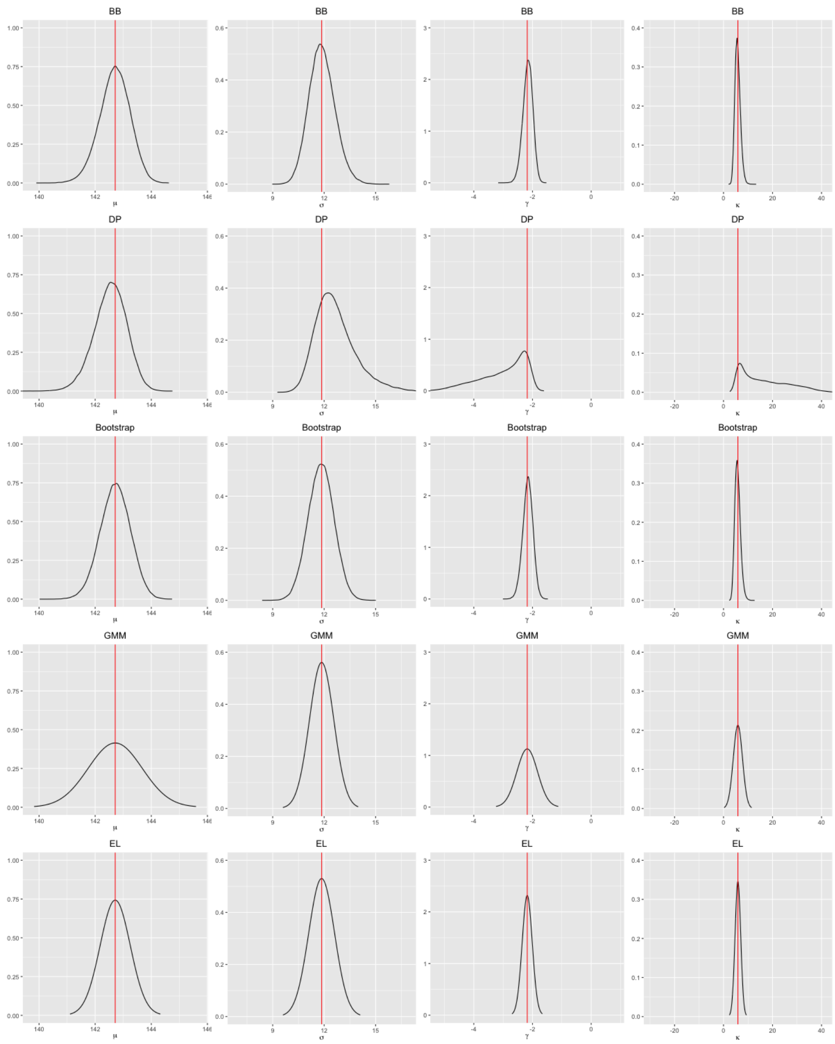

Figure 4.3: Distributions of μ, σ, γ, κ for garden earthworm data using g_2 as a moment function.

of DP. In order to confirm this thought, we compare the posterior distributions of the estimators obtained using DP by setting the base distributions to various distributions. As the base distribution of DP, the Location-scale t-distribution with degrees of freedom as 5 is used. In this setup, we use the scale parameter as 1 and the location parameter as $\frac{i}{10}$ times $(i = 0, 1, \cdots, 20)$ the sample mean of the data.

The posterior distributions of μ, σ, γ, and κ obtained using DP are given in Figure 4.4 and 4.5. For convenience, only the posterior distributions where the location parameters of DP are $\frac{i}{2}$ times $(i = 0, 1, 2, 3, 4)$ the sample mean of the data is compared. From these figures, it can be seen that the posterior distributions of parameters obtained using DP are heavily influenced by the center of the base distribution. As the center of the base distribution approaches the sample mean of the data, the posterior distributions obtained using DP become more symmetric, with shorter tails. Further, the MAP estimators obtained using DP get closer to the MME. Based on these results, it can be inferred that setting the center of the base distribution of DP as the sample mean of the data would lead to better results. Therefore, when estimating the parameters of the moment condition models using DP, we suggest choosing the base distribution of DP as $t_{(r)}(\hat{\mu}, \hat{\Sigma})$ where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance of the data, respectively.

We compare the estimators and the posterior distributions of the estimators obtained using each method. In this case, the base

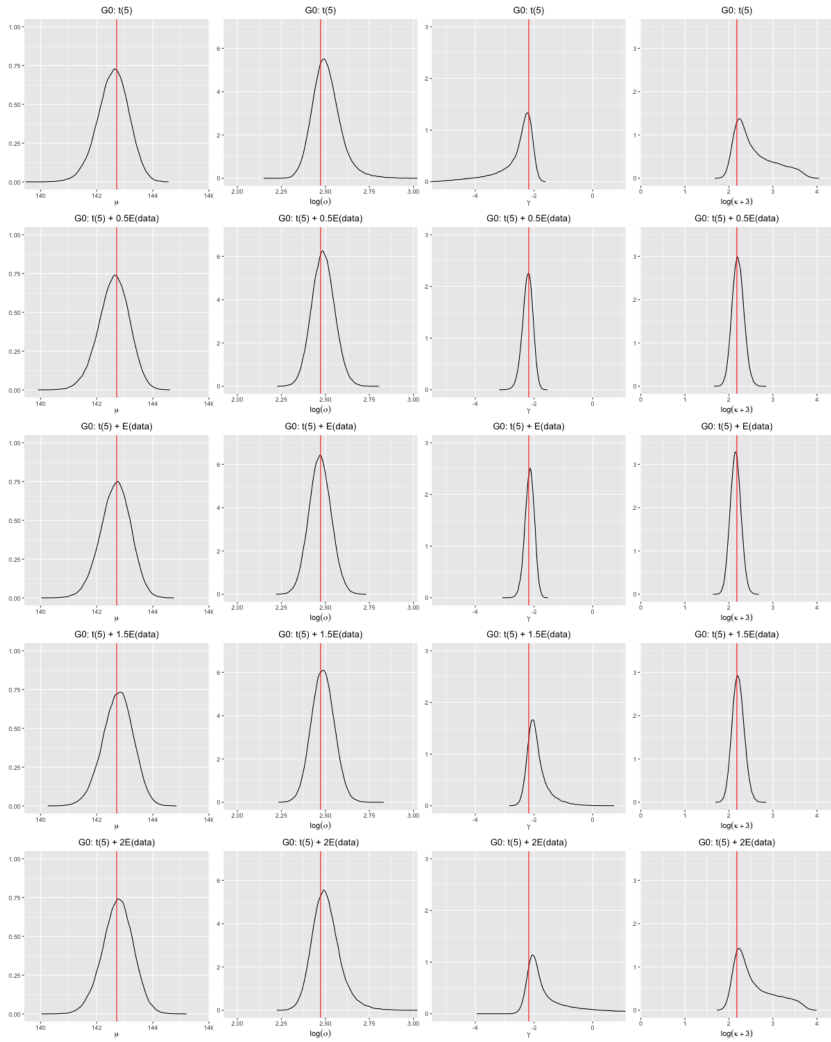

Figure 4.4: Posterior distributions of $\mu, \sigma, \gamma, \kappa$ obtained through DP for garden earthworm data using g_1 as a moment function.

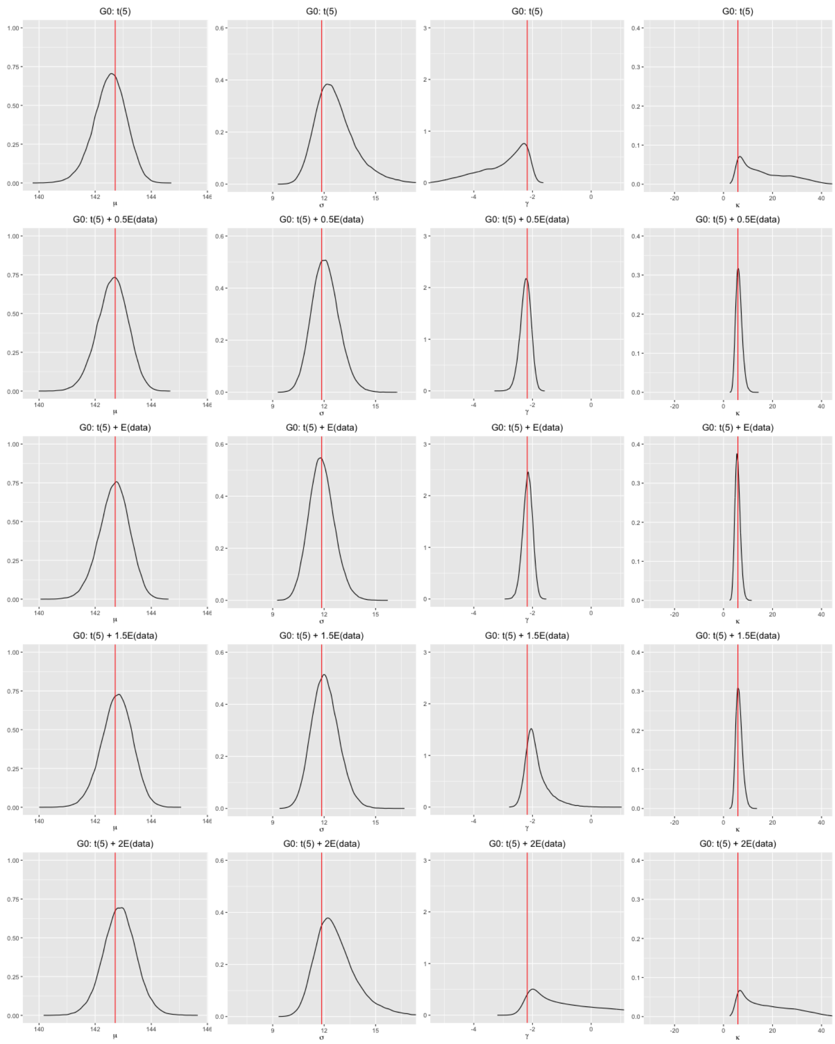

Figure 4.5: Posterior distributions of $\mu, \sigma, \gamma, \kappa$ obtained through DP for garden earthworm data using g_2 as a moment function.

distribution of DP is $t_{(5)}(\hat{\mu}, \hat{\Sigma})$ where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance of the data, respectively. At first, we compare the estimators obtained using each method. The estimators are given in Table 4.2. It can be seen that the difference between the estimators obtained using DP and those of other methods is greatly reduced. Second, we compare the distributions of the estimators obtained using each method. The posterior distributions obtained using DP are similar to the distributions obtained using BB, Bootstrap, and EL which are unskewed and have short tails.

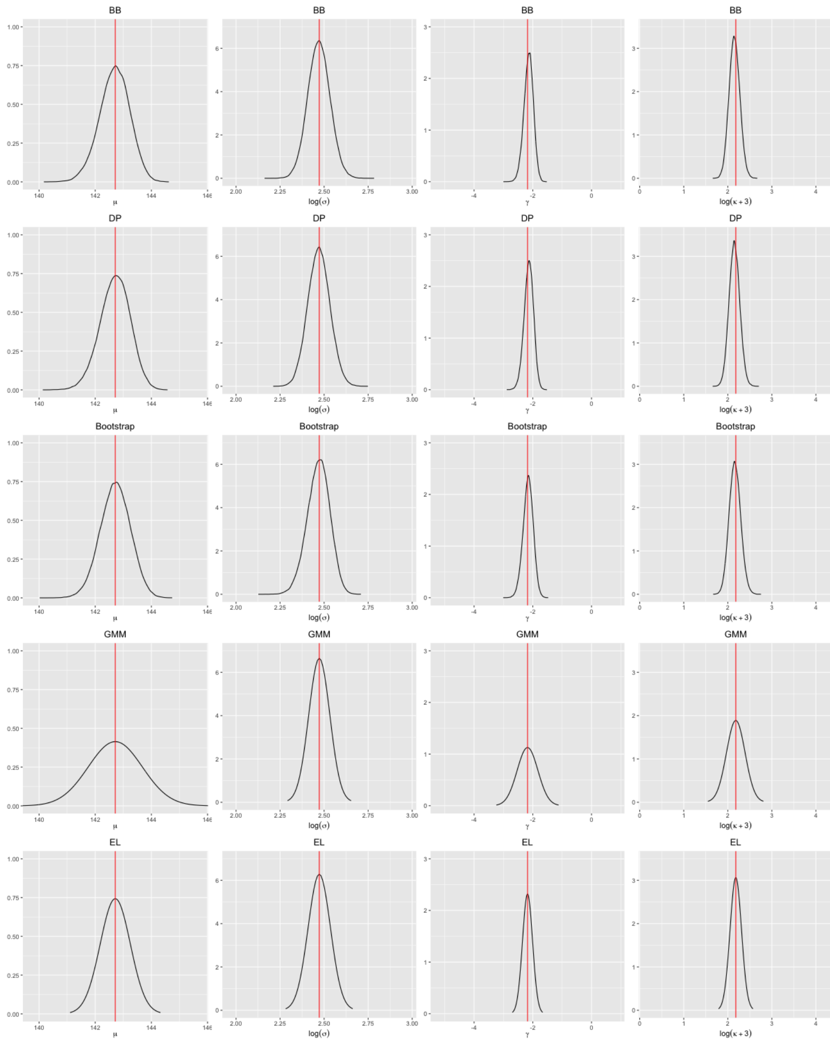

Figure 4.6: Distributions of μ, σ, γ, κ for garden earthworm data using g_1 as a moment function with base distribution of DP as $t_{(5)}(\hat{\mu}, \hat{\Sigma})$.

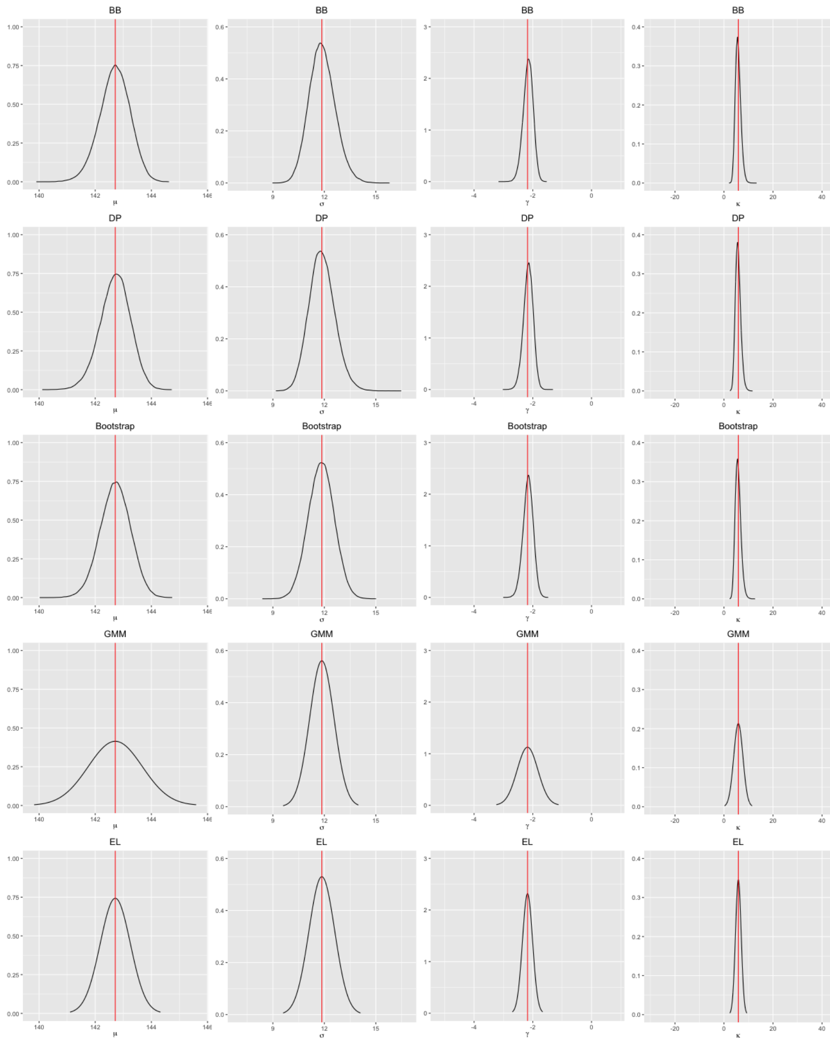

Figure 4.7: Distributions of μ, σ, γ, κ for garden earthworm data using g_2 as a moment function with base distribution of DP as $t_{(5)}(\hat{\mu}, \hat{\Sigma})$.

		Parameter			
Function	Method	μ	σ	γ	κ
	Sample	142.715	11.853	-2.179	5.857
g_1	GMM	142.715	11.853	-2.179	5.857
g_1	EL	142.715	11.853	-2.179	5.857
g_1	Bootstrap	142.756	11.825	-2.149	5.527
g_1	BB	142.726	11.802	-2.105	5.432
g_1	DP	142.739	11.808	-2.126	5.486
g_2	GMM	142.715	11.853	-2.179	5.857
g_2	EL	142.715	11.853	-2.179	5.857
g_2	Bootstrap	142.756	11.825	-2.149	5.527
g_2	BB	142.716	11.766	-2.151	5.528
g_2	DP	142.752	11.768	-2.138	5.564

Table 4.2: Estimators of $\mu, \sigma, \gamma, \kappa$ for garden earthworm data with the base distribution of DP as $t_{(5)}(\hat{\mu}, \hat{\Sigma})$.

We estimated the parameters of the moment condition models using various methods and compared their values and distributions. The estimators obtained using each method had similar values and distributions, but it was not possible to determine which method estimated the parameters better because the true values of the parameters were not known. Therefore, we conduct a simulation study to compare the performance of each method.

The shape of the distribution of the data used in this simulation study is set to be as similar as possible to the histogram of the garden earthworm data given in Figure 4.1. The true model of the simulation data is as follows.

$$X_i \overset{i.i.d.}{\sim} \sum_{j=1}^{4} \phi_j N(\mu_j, \sigma_j^2),\ i = 1, \cdots, 487$$

where $\phi = (\frac{3}{118}, \frac{7}{118}, \frac{15}{118}, \frac{93}{118}), \mu = (105, 115, 136, 146), \sigma = (20, 15, 10, 5)$. We generate 100 simulation data sets from the true model. Each simulation data set is one-dimensional data with 487 observations. As in the analysis of garden earthworm data, we use BB, DP, GMM, EL, and Bootstrap to estimate the parameters of the moment condition models and compare their performance using Mean Squared Error (MSE). MSE of each method is given in Table 4.3.

Depending on the moments, there is a difference in the performance of the estimator. For μ, σ, and γ, GMM and EL showed the lowest MSE, while for κ, BB and DP have the lowest MSE. Among these, for μ and γ, the MSE of each method did not differ significantly. On the other hand, for σ and κ, there were substantial differences in MSE between the competing methods, GMM and EL, and the proposed methods, BB and DP. This seems to have occurred due to the randomness that arose in the process of generating the posterior samples of BB and DP, which are non-parametric Bayesian methods. Therefore, when estimating the parameters of the moment condition models in just-identified cases, it is difficult to say which of the proposed methods, BB and DP, and the competing methods, GMM, EL, and Bootstrap, show better performance.

		Parameter			
Function	Method	μ	σ	γ	κ
	Sample	0.298	0.891	0.062	3.761
g_1	GMM	**0.298**	**0.891**	**0.062**	3.761
g_1	EL	**0.298**	**0.891**	**0.062**	3.761
g_1	Bootstrap	0.304	0.911	0.063	3.857
g_1	BB	0.310	0.957	0.067	**3.530**
g_1	DP	0.306	0.961	0.067	3.541
g_2	GMM	**0.298**	**0.891**	**0.062**	3.761
g_2	EL	**0.298**	**0.891**	**0.062**	3.761
g_2	Bootstrap	0.304	0.911	0.063	3.857
g_2	BB	0.303	0.929	0.063	3.435
g_2	DP	0.317	0.929	0.069	**3.379**

Table 4.3: MSE of $\mu, \sigma, \gamma, \kappa$ for simulation data set

4.2 Example 2: IV regression

Instrumental Variables (IV) regression used in various fields such as statistics, econometrics, and epidemiology is a statistical technique to estimate causal relationships between variables when there is a possibility of endogeneity or omitted variable bias. Endogeneity occurs when one or more independent variables are correlated with the error term in a regression model, which causes coefficient estimators such as Ordinary Least Squares (OLS) estimators to be inconsistent and biased. IV regression solves this problem by introducing instrumental variables which are uncorrelated with the error term and correlated with the endogenous variable which is correlated with the error term.

In this section, we would like to estimate the coefficient of IV regression using the moment condition model. Consider the linear model

$$y = \alpha + \beta x + \delta s + \epsilon, \quad \mathbb{E}[\epsilon] = 0. \tag{4.5}$$

Our goal is to estimate the regression coefficient $\theta = (\alpha, \beta, \delta)$. However, in this model, the independent variable x is correlated with the error term ϵ, unlike the independent variable s. For this reason, we need to use a method other than OLS to estimate the parameter θ. To solve this problem, we introduce instrument variables z_1 and z_2 which are correlated with the independent variable x and uncorrelated with the error term ϵ. Therefore, the regression coefficient θ should satisfy the moment restrictions (4.6) to (4.9)

$$\mathbb{E}[y - \alpha - \beta x - \delta s] = 0 \tag{4.6}$$

$$\mathbb{E}[(y - \alpha - \beta x - \delta s)z_1] = 0 \tag{4.7}$$

$$\mathbb{E}[(y - \alpha - \beta x - \delta s)z_2] = 0 \tag{4.8}$$

$$\mathbb{E}[(y - \alpha - \beta x - \delta s)s] = 0. \tag{4.9}$$

Through the moment restrictions (4.6) to (4.9), we can derive moment function g to be used in this example as follows.

$$g(X, \theta) = \begin{pmatrix} y - \alpha - \beta x - \delta s \\ (y - \alpha - \beta x - \delta s)z_1 \\ (y - \alpha - \beta x - \delta s)z_2 \\ (y - \alpha - \beta x - \delta s)s \end{pmatrix}$$

The simulation data set $X = (x, y, s, z_1, z_2)$ is randomly generated from the true model. The true model of the model (4.5) is as follows.

$$\begin{aligned} &y = 1 + 0.5x + 0.7s + \epsilon \\ &x = z_1 + z_2 + s + u \\ &z_j \overset{iid}{\sim} N(0.5, 1^2), \quad j = 1, 2 \\ &s \sim Unif(0, 1) \end{aligned}$$

The error terms (ϵ, u) are generated from a Gaussian copula whose diagonal and off-diagonal entries of the covariance matrix are 1 and 0.7, respectively. The marginal distribution of ϵ is $\frac{1}{2}N(0.5, 0.5^2) + \frac{1}{2}N(-0.5, 1.118^2)$ and the marginal distribution of u is $N(0, 1^2)$. We set the number of observations to various values to estimate parameter θ under various conditions. The number of observations of simulation data sets is $n = 10, 25, 50, 100, 250, 500, 1000, 2000$, and 100 simulation data sets are generated from the true model for each n. In addition, 100 prediction data sets are generated which have 500 observations. We estimate θ for each simulation data set. Using these estimators, we calculate MSE for each simulation data set and predict the response variable y of the prediction data sets using the estimated IV regression model.

For estimating the regression coefficient θ, various methods are used, including the proposed methods BB and DP, Frequentist methods GMM and EL, and the nonparametric Bayesian method BETEL. MM and Bootstrap, the methods used in the previous section, cannot be used in the over-identified moment condition models, so they could not be used in this example. We use the Location-scale t-distribution and a non-informative prior as the prior of θ. When estimating θ using BETEL, $t_{(2.5)}(0, 5^2)$ is used as the prior of θ, and when estimating θ using DP, $t_{(2.5)}(0, 5^2)$ and a non-informative prior are used as the prior of θ. Furthermore, we set the concentration parameter and the base distribution of DP as $A = 0.5$ and $G_0 = t_{(5)}(\hat{\mu}, \hat{\Sigma})$ where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance of the data, respectively. In the process of estimating θ using nonparametric Bayesian methods BB, DP, and BETEL, we use MCMC to generate posterior samples. The number of posterior samples to be generated was set differently according to the number of observations n in the data. Specifically, when $n = 10$, we generate 1 million samples; when

$n = 25$, we generate 100,000 samples, and when n is 50 or more, we generate 10,000 samples. In each case, we remove the first 10% of the posterior samples through burn-in.

We compare the performance of the methods used to estimate θ using MSE and bias. First, the performance of each method is compared through MSE. The MSEs calculated using the estimators obtained by each method are given in Table 4.4 and 4.5. The MSEs for data with a small number of observations ($n = 10, 25, 50, 100$) are given in Table 4.4, and the MSEs for data with a large number of observations ($n = 250, 500, 1000, 2000$) are given in Table 4.5. In each table, the prior of θ used in each method is written in parentheses right after BETEL and DP. That is when $t_{(2.5)}(0, 5^2)$ is used as the prior, it is expressed as (t), and when the non-informative prior is used, it is expressed as (non).

The comparison of the MSEs with small n is given in Table 4.4. Before comparing the MSEs of each method, it is important to note that BETEL failed to estimate parameters in some data sets. Specifically, for $n = 10$, it could not estimate the parameters of 25 data sets, and for $n = 25$, it failed to estimate the parameters of 1 data set. Therefore, we only used estimators of parameters using the data set which can estimate the parameters to calculate the MSEs of BETEL.

When $n = 10$, the difference in MSEs according to the estimating methods is clearly visible. The competing methods, GMM, EL, and BETEL, have large MSEs, while the proposed methods, BB and DP, have relatively small MSEs. Among the competing meth-

		Method					
n	Parameter	GMM	EL	BETEL(t)	BB	DP(non)	DP(t)
10	α	1.776	7×10^{14}	1.894	0.939	0.759	**0.369**
	β	0.407	2×10^{14}	1.169	0.176	0.111	**0.076**
	δ	6.410	6×10^{15}	3.182	2.633	2.127	**0.644**
	predict	2.999	1×10^{15}	6.433	2.210	1.694	**1.362**
10(2)	α	1.598	1.251	1.894	0.829	0.767	**0.372**
	β	0.365	0.235	1.169	0.128	0.110	**0.076**
	δ	4.254	3.096	3.182	2.456	2.137	**0.651**
	predict	2.729	2.237	6.433	1.822	1.692	**1.363**
25	α	0.283	0.297	0.368	0.284	0.280	**0.202**
	β	0.023	0.023	0.065	0.019	0.020	**0.018**
	δ	0.850	0.779	0.959	0.741	0.723	**0.473**
	predict	1.158	1.189	1.470	1.148	1.149	**1.102**
50	α	0.107	0.103	0.108	0.106	0.102	**0.089**
	β	0.013	0.014	0.016	0.013	0.013	**0.012**
	δ	0.341	0.317	0.337	0.309	0.308	**0.244**
	predict	1.069	1.079	1.108	1.065	1.067	**1.052**
100	α	0.044	0.045	0.049	0.044	0.044	**0.040**
	β	0.007	0.007	0.007	0.006	0.007	**0.006**
	δ	0.121	0.125	0.122	0.123	0.121	**0.105**
	predict	1.028	1.033	1.048	1.027	1.031	**1.022**

Table 4.4: MSE of each method for $n = 10, 25, 50, 100$

		Method					
n	Parameter	GMM	EL	BETEL(t)	BB	DP(non)	DP(t)
250	α	0.027	0.026	0.029	0.026	0.027	**0.025**
	β	0.003	0.003	0.003	**0.002**	0.003	0.003
	δ	0.058	0.058	0.061	0.058	0.058	**0.056**
	predict	0.988	0.991	0.995	0.988	0.988	**0.987**
500	α	0.009	0.009	0.009	**0.009**	0.009	0.009
	β	0.001	0.001	0.001	**0.001**	0.001	0.001
	δ	0.027	0.027	0.027	0.027	0.027	**0.025**
	predict	0.983	0.984	0.983	0.982	0.983	**0.981**
1000	α	0.005	0.005	0.005	0.005	0.005	**0.005**
	β	0.001	0.001	0.001	0.001	**0.001**	0.001
	δ	0.013	0.013	0.013	0.013	0.013	**0.012**
	predict	0.977	0.977	**0.975**	0.977	0.977	0.976
2000	α	0.002	0.002	0.002	0.002	0.002	**0.002**
	β	0.000	0.000	0.000	0.000	**0.000**	0.000
	δ	0.006	0.006	**0.006**	0.006	0.006	0.006
	predict	0.977	0.977	**0.976**	0.977	0.977	0.976

Table 4.5: MSE of each method for $n = 250, 500, 1000, 2000$

ods, the MSEs of EL are extremely large because the estimators of the two data sets differ too much from the true values of the parameters. Therefore, we excluded these two data sets and compared the MSEs using the estimators obtained from the remaining 98 data sets, denoted in row $n = 10(2)$. In this case, BETEL had no differences in MSE, as it was unable to estimate parameters from those two data sets from the beginning. Even with the remaining 98 data sets, the MSEs of BB and DP are still lower than the MSEs of GMM, EL, and BETEL. An interesting observation is that the MSEs of EL are greatly reduced and that of GMM and BB are slightly decreased. In contrast, the MSEs of DP hardly changed and are still smaller than that of the other methods. Therefore, it can be said that DP estimates parameters relatively well compared to other methods even for data with fewer observations and large error terms.

When $n = 25$, the MSEs of all methods decreased. In this case, the proposed methods, BB and DP, still outperform the competing methods, GMM, EL, and BETEL, in estimating θ. However, compared to $n = 10$, the MSEs between the methods are substantially reduced. When $n = 50$ and $n = 100$, the differences diminish further, and no significant differences are observed among the methods. Additionally, comparing the MSEs of DP with different prior, since the true values of α, β, and δ are close to zero, estimators obtained using $t_{(2.5)}(0, 5^2)$ as the prior outperforms the estimators obtained using a non-informative prior.

The comparison of the MSEs with large n is given in Table

4.5. In these cases, all methods have similar and low MSEs. Consequently, if n is sufficiently large, there is no significant difference in the estimators according to the methods, and all methods estimate parameters well.

Based on the comparison of MSE for each method, it can be concluded that the proposed methods, BB and DP, outperformed the competing methods, GMM, EL, and BETEL, in parameter estimation. When estimating parameters for data with a small number of observations, there were notable performance differences among the methods, but as the number of observations increased, these differences diminished rapidly. After reaching a certain level of sample size, there were no significant differences in performance among the estimation methods. Among the proposed methods, the MSEs of DP are lower than that of BB since the base distribution of DP reflected the information of the data. Additionally, estimators obtained using $t_{(2.5)}(0, 5^2)$ as the prior resulted in lower MSE compared to using a non-informative prior, since the true values of the parameters are close to zero in each case.

Next, we compare the performance of each method in more detail using the biases for each data set. Biases are compared using the estimators of the data sets with $n = 10$ because the difference in MSEs between the methods was the largest when $n = 10$. Since comparing the biases of all methods simultaneously may not be visually clear, we proceed by iteratively comparing the biases of the two methods. Specifically, DP with a non-informative prior, one of the proposed methods, and the other methods, GMM, EL, BETEL, and BB, are sequentially compared. The results of comparing the biases of each method are given in Figures 4.8 to 4.12. In each figure, the x-axis represents the index of the simulation data set, and the y-axis represents the absolute values of the biases. The red dots indicate the biases of DP, and the blue dots represent the biases of the other methods.

A comparison of the biases of DP and GMM is given in Figure 4.8. It can be seen that the biases of GMM are greater than those of DP in most data sets. The biases of GMM are large in some data sets, while the biases of DP are close to zero in most data sets. The frequency of the large biases of GMM is much higher than that of DP.

A comparison of the biases of DP and EL is given in Figure 4.9 and 4.10. At first, we tried to compare the biases of DP and EL using all data sets. However, as shown in Figure 4.9, the biases of EL are extremely large in some data sets, making it impossible to compare the biases. Therefore, we compare the biases of DP and EL using only 98 data sets, except for the two data sets where the biases of EL are extremely large. From Figure 4.10, we can conclude that the result of comparing the biases of DP and EL is similar to the comparison of the biases of DP and GMM. The biases of EL are greater than that of DP in most data sets, and the biases of EL are large in some data sets, while the biases of DP are close to zero in most data sets. The frequency of the large biases of EL is much higher than that of DP.

A comparison of the biases of DP and BETEL is given in Figure

4.11. We compare the biases of 75 data sets in which BETEL can estimate the parameter θ. The biases of BETEL are larger than those of DP in most data sets. The biases of BETEL are large in many data sets, while the biases of DP are close to zero in most data sets. The frequency of the large biases of BETEL is much higher than that of GMM and EL as well as DP.

In addition, we compare the two proposed methods, BB and DP. A comparison of the biases of DP and BB is given in Figure 4.12. The biases of the two proposed methods are similar in most data sets except for some data sets. In these data sets, the biases of BB are greater than the biases of DP, but this frequency is low.

Based on the comparison of biases, it can be seen that the proposed methods, DP and BB, perform better in estimating the parameters compared to the competing methods, GMM, EL, and BETEL. The biases of DP are lower than GMM, EL, and BETEL in most data sets. Furthermore, the frequency of large biases of DP is much lower than that of GMM, EL, and BETEL. As a result of comparing the biases of DP and BB, the two methods have similar biases in most of the data.

In this section, we estimated the coefficients of the IV regression models using various moment condition model estimation methods. The proposed methods, BB and DP, performed better than the competing methods, GMM, EL, and BETEL. While there were no significant differences in performance among the methods when n was large, substantial differences emerged when n was small. Therefore, we can conclude that when estimating the parameters of the moment condition models in over-identified cases, the proposed methods show better performance than the competing methods. Among the proposed methods, DP showed slightly better performance than BB. This is considered to be because the base distribution of DP reflected the information of the data.

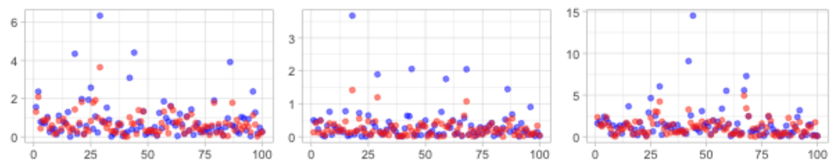

Figure 4.8: Bias of DP (red) and GMM (blue) $n = 10$: the bias of α (left), the bias of β (center), the bias of δ (right).

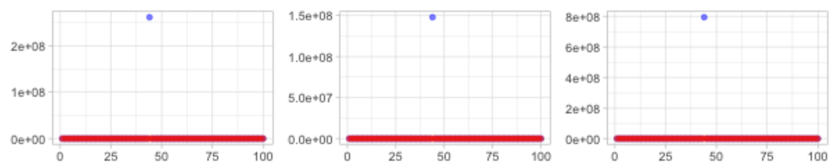

Figure 4.9: Bias of DP (red) and EL (blue) with $n = 10$ (using 100 data sets): the bias of α (left), the bias of β (center), the bias of δ (right).

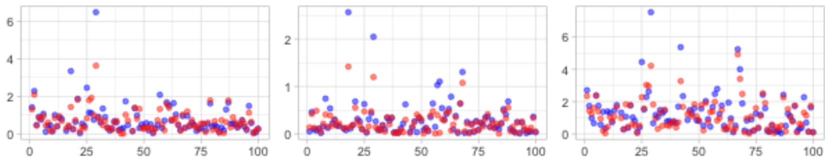

Figure 4.10: Bias of DP (red) and EL (blue) with $n = 10$ (using 98 data sets): the bias of α (left), the bias of β (center), the bias of δ (right).

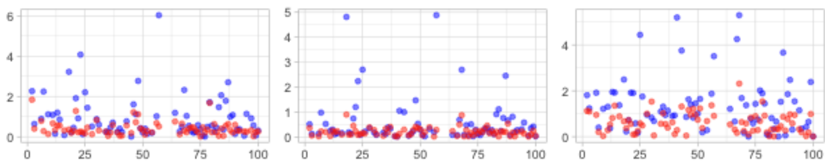

Figure 4.11: Bias of DP (red) and BETEL (blue) with $n = 10$: the bias of α (left), the bias of β (center), the bias of δ (right).

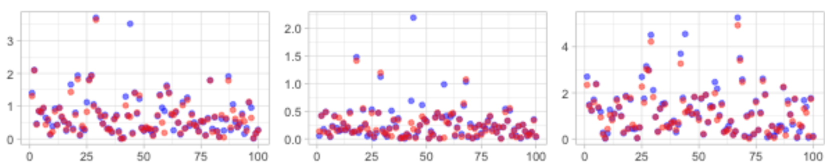

Figure 4.12: Bias of DP (red) and BB (blue) with $n = 10$: the bias of α (left), the bias of β (center), the bias of δ (right).

Chapter 5

Conclusion

In this thesis, we propose two nonparametric Bayesian methods, the Bayesian bootstrap and the constrained Dirichlet process, to estimate the parameters of the moment condition models. Several Frequentist methods such as GMM, EL, ET, and ETEL have been proposed as methods for estimating the moment condition models. However, little research has been conducted on the moment condition models using the nonparametric Bayesian methods because the moment condition constrains the parameter space, making it difficult to calculate the posterior distribution. We solve this problem by obtaining the posterior samples using constrained Hamiltonian Monte Carlo. When updating the posterior samples in the second step of the constrained Hamiltonian Monte Carlo, we used the Shake algorithm.

Through various numerical studies, we estimated the moment condition models using the proposed methods and compared them with the competing methods such as GMM, EL, and Bootstrap which are the Frequentist methods, and BETEL which is a non-parametric Bayesian method. In just-identified models, there was no significant difference in performance between the proposed methods and the competing methods. It is judged that the small amount of difference that exists in estimators is caused by the randomness of the posterior samples of the proposed methods obtained using the constrained Hamiltonian Monte Carlo. In over-identified models, the performance of the proposed methods is much better than that of the competing methods. While there were no significant differences when n was large, substantial differences emerged when n was small.

However, the estimators and the posterior distributions obtained using DP were affected by the form of the base distribution. As the center of the base distribution approaches the sample mean of the data, the posterior distributions obtained using DP become more symmetric, with shorter tails. Further, the MAP estimators obtained using DP get closer to the true values of the parameters. Therefore, we suggest using the base distribution of DP as a Location-scale t-distribution whose location parameter and scale parameter are the sample mean and the sample covariance of the data, respectively. In this case, DP estimates parameters slightly better than BB because the base distribution of DP reflects the information of the data.

We suggest estimating the parameters of the moment condition models in two ways depending on the situation. If we have prior information about the parameters, we suggest using DP for estimating the parameters of the moment condition models because the posterior distribution obtained using BB cannot reflect the prior of parameters. If we do not have prior information about the parameters, we suggest estimating parameters using either BB or DP, depending on the subjectivity of the analyst.

Bibliography

Joseph G Altonji and Lewis M Segal. Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366, 1996.

Charles E Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pages 1152–1174, 1974.

Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Joseph T Chang and David Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317, 1997.

Siddhartha Chib, Minchul Shin, and Anna Simoni. Bayesian estimation and comparison of moment condition models. *Journal of the American Statistical Association*, 113(524):1656–1668, 2018.

Thomas DiCiccio, Peter Hall, and Joseph Romano. Empirical likelihood is bartlett-correctable. *the Annals of Statistics*, pages 1053–1061, 1991.

Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2): 216–222, 1987.

Bradley Efron. Nonparametric standard errors and confidence intervals. *canadian Journal of Statistics*, 9(2):139–158, 1981.

Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588, 1995.

Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

Subhashis Ghosal and Aad Van der Vaart. *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press, 2017.

Peter Hall. Pseudo-likelihood theory for empirical likelihood. *The Annals of Statistics*, pages 121–140, 1990.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054, 1982.

Lars Peter Hansen, John Heaton, and Amir Yaron. Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280, 1996.

Nils Lid Hjort. Bayesian and empirical bayesian bootstrapping. *Preprint series. Statistical Research Report http://urn.nb.no/URN: NBN: no-23420*, 1991.

Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

Guido Imbens, Phillip M Johnson, and Richard H Spady. Information theoretic approaches to inference in moment condition models, 1995.

Guido W Imbens. One-step estimators for over-identified generalized method of moments models. *The Review of Economic Studies*, 64(3):359–383, 1997.

Guido W Imbens. Generalized method of moments and empirical likelihood. *Journal of Business & Economic Statistics*, 20(4):493–506, 2002.

Yongdai Kim and Jaeyong Lee. Bayesian bootstrap for proportional hazards models. *The Annals of Statistics*, 31(6):1905–1922, 2003.

Nicole A Lazar. Bayesian empirical likelihood. *Biometrika*, 90(2):319–326, 2003.

Albert Y Lo. A large sample study of the bayesian bootstrap. *The Annals of Statistics*, pages 360–375, 1987.

Albert Y Lo. A bayesian bootstrap for a finite population. *The annals of statistics*, pages 1684–1695, 1988.

Albert Y Lo. A bayesian bootstrap for censored data. *The Annals of Statistics*, pages 100–123, 1993.

David JC MacKay and David JC Mac Kay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

Radford M Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

Whitney K Newey and Richard J Smith. Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255, 2004.

Art Owen. Empirical likelihood ratio confidence regions. *The annals of statistics*, 18(1):90–120, 1990.

Art B Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

Art B Owen. *Empirical likelihood*. Chapman and Hall/CRC, 2001.

Adrian R Pagan and John Robertson. Gmm and its problems. *Manuscript, Australia National University*, 1997.

Raymond Pearl and Wilbur N Fuller. Variation and correlation in the earthworm. *Biometrika*, 4(1-2):213–229, 1905.

Jin Qin and Jerry Lawless. Empirical likelihood and general estimating equations. *the Annals of Statistics*, 22(1):300–325, 1994.

Donald B Rubin. The bayesian bootstrap. *The annals of statistics*, pages 130–134, 1981.

Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

Susanne M Schennach. Exponentially tilted empirical likelihood. *Working Pa*, 2004.

Susanne M Schennach. Bayesian exponentially tilted empirical likelihood. *Biometrika*, 92(1):31–46, 2005.

Susanne M Schennach. Point estimation with exponentially tilted empirical likelihood. 2007.

Jayaram Sethuraman. A constructive definition of dirichlet priors. *Statistica sinica*, pages 639–650, 1994.

Min Chul Shin. *Bayesian Gmm*. University of Pennsylvania, 2015.

Chung-Sing Weng. On a second-order asymptotic property of the bayesian bootstrap mean. *The Annals of Statistics*, pages 705–710, 1989.

국문초록

본 논문은 비모수 베이즈 방법인 베이즈 붓스트랩과 제한 디리클레 과정을 사용해 적률조건모형을 추정하는 것을 제안한다. 적률조건모형에서는 관측치 X와 적률식 $g(X, \theta)$에 대해 관심모수 θ가 적률조건 $\mathbb{E}_F[g(X, \theta)] = 0$에 의해 결정된다. 하지만 적률조건은 모수공간을 제한하여 사후분포의 계산을 어렵게 하기 때문에 그동안 많은 연구가 진행되지 않았다. 본 논문에서는 제약이 있는 해밀턴 동역학 모의 실험 방법을 이용해 제약이 있는 모수공간에서 사후표본을 생성하는 아이디어를 제안함으로써 이러한 문제를 해결한다. 제안된 아이디어는 비모수 모형에서의 적률의 추론, 도구변수가 있는 회귀모형 등의 예제를 통해 적률조건모형을 추정하고, 기존 모형과 성능을 비교한다.

주요어: 적률조건 모형, 베이지안 붓스트랩, 제한된 디리클레 과정, 제약이 있는 해밀턴 동역학 모의 실험 방법, Shake 알고리즘

학 번: 2016-20273