Graduate School of
Business Administration

KOBE UNIVERSITY

ROKKO KOBE JAPAN

2016-8

BAYESIAN ESTIMATION OF BETA-TYPE DISTRIBUTION PARAMETERS BASED ON GROUPED DATA

Kazuhiko Kakamu Haruhisa Nishino

Discussion Paper Series

BAYESIAN ESTIMATION OF BETA-TYPE DISTRIBUTION PARAMETERS BASED ON GROUPED DATA

Kazuhiko Kakamu * Haruhisa Nishino †

Abstract

This study considers the estimation method of generalized beta (GB) distribution parameters based on grouped data from a Bayesian point of view. Because the GB distribution, which was proposed by McDonald and Xu (1995), includes several kinds of familiar distributions as special or limiting cases, it performs at least as well as those special or limiting distributions. Therefore, it is reasonable to estimate the parameters of the GB distribution. However, when the number of groups is small or when the number of parameters increases, it may become difficult to estimate the distribution parameters for grouped data using the existing estimation methods. This study uses a Tailored randomized block Metropolis–Hastings (TaRBMH) algorithm proposed by Chib and Ramamurthy (2010) to estimate the GB distribution parameters, and this method is applied to one simulated and two real datasets. Moreover, the Gini coefficients from the estimated parameters for the GB distribution are examined.

Key words: Generalized beta (GB) distribution; Gini coefficient; grouped data; simulated annealing; Tailored randomized block Metropolis–Hastings (TaRBMH) algorithm.

*Corresponding author. Graduate School of Business Administration, Kobe University, 2-1, Rokkodai, Nada, Kobe 657-8501, Japan. *Email*: kakamu@person.kobe-u.ac.jp

†Faculty of Law, Politics and Economics, Chiba University, Yayoi-cho 1-33, Inage-ku, Chiba 263-8522, Japan. *Email*: nishino@le.chiba-u.ac.jp

1 Introduction

The estimation of income distributions has played an important role in the measurement of inequality (e.g., the Gini coefficient), and grouped data (i.e., class frequency data) has been widely used to estimate the distribution parameters. [1] In grouped data, although x_j exists for $j = 1, 2, \ldots, n$, which is assumed to be in ascending order, only income x_i of the n_ith observation for $i = 1, 2, \ldots, k$ is observable, where strictly $k < n$. In this situation, we seek to restore the true distribution (i.e., the dotted line) from the histogram drawn using the data by estimating the parameters of the hypothetical distribution as shown in Figure 1. Furthermore, a vast body of literature has considered the estimation methods and/or hypothetical distributions.

Although the maximum likelihood estimation (MLE) has been widely used and is probably the most common method of estimating parameters for hypothetical distributions from grouped data (see McDonald and Ransom, 2008), several estimation methods have been proposed and used to estimate the parameters of the hypothetical distributions, including minimum chi-square, the scoring method, least square, method of moments, least lines, generalized method of moments, generalized least square, and Markov chain Monte Carlo (MCMC) methods (see e.g., Chotikapanich and Griffiths, 2000; Chotikapanich *et al*., 2007; McDonald and Ransom, 1979a,b; Nishino and Kakamu, 2011; van Dijk and Kloek, 1980). One reason why several estimation methods have been proposed despite that the MLE is widely used is that the estimates of the population characteristics depend on the functional form and estimation technique by McDonald and Ransom (1979a,b). Therefore, several hypothetical distributions have also been proposed and assumed as income distributions.

Among hypothetical distributions, generalized beta (GB) proposed by McDonald and Xu (1995) is the most flexible beta-type distributions, because it includes the most distributions. For example, it can be used as an income distribution and includes special or limiting cases. Comparisons of the hypothetical distributions have been examined in studies such as Atoda *et al*. (1988); Bordley *et al*. (1996); Kloek and van Dijk (1978);

[1] Studies on the estimation of income distribution from individual data exist, such as Chotikapanich and Griffiths (2008); Hasegawa and Kozumi (2003); Tachibanaki *et al*. (1997). However, we restrict our discussion to grouped data. There is another type of dataset that includes population shares and group mean incomes, and Hajargasht *et al*. (2012) consider this situation.

McDonald and Mantrala (1995); McDonald and Ransom (1979a); Majumder and Chakravarty (1990); Salem and Mount (1974); Slottje (1984); Tachibanaki *et al*. (1997). Most hypothetical distributions assumed in these papers are special or limiting GB distributions, including, for example, the GB distribution of the first and second kind (GB1 and GB2, respectively), proposed by McDonald (1984), the Singh–Maddala (SM) distribution, the beta distribution of the first kind, the beta distribution of the second kind, the gamma distribution, the chi-square distribution, and the exponential distribution (see Figure 2 in McDonald and Xu (1995) for details on the relationship of the distributions).

These papers cover many but not all studies that examine the estimation methods and hypothetical distributions from grouped data. However, as shown previously, choosing the distribution and estimation method simultaneously is critical. In addition, when the number of groups is small or when we have to estimate a large number of parameters, it sometimes becomes difficult to estimate the distribution parameters using the existing estimation methods in our experiments (see also McDonald and Mantrala, 1995). Therefore, it is reasonable to consider estimation methods that do not require the number of groups, or the hypothetical distribution. Thus, we consider the estimation procedure for the GB distribution proposed by McDonald and Xu (1995), which is the most flexible distribution in the class of beta-type distributions, as later discussed.

In this paper, we take a Bayesian approach and use the Tailored randomized block Metropolis–Hastings (TaRBMH) algorithm proposed by Chib and Ramamurthy (2010) to estimate the parameters of the GB distribution. We compare this algorithm with the existing one proposed by Chotikapanich and Griffiths (2000) using both simulated and real datasets and compare the Bayesian estimates with the MLE ones in Bordley *et al*. (1996) using a real dataset. The results show that the TaRBMH algorithm can sample MCMC draws more efficiently than the algorithm by Chotikapanich and Griffiths (2000). Moreover, the Gini coefficient from the GB distribution is examined using a real dataset, and we can show that it estimates the Gini coefficient accurately despite that the number of groups are relatively small.

The rest of this paper is organized as follows. In the next section, we introduce the features of GB distribution, including the probability density, cumulative distribution, and the likelihood functions, and we explain

the MCMC estimation procedure for this distribution. In Section 3, we examine the numerical examples of both simulated datasets and real ones that include income data from both the United States and Japan. In the Japanese dataset, the performance of the Gini coefficient is also discussed. In Section 4, we conclude the discussion and state the remaining issues.

2 The GB Distribution

2.1 The Density, Cumulative Distribution, and Likelihood Functions

While various probability distributions are used to estimate the parameters of a hypothetical income distribution, the GB distribution includes various probability distributions as special or limiting cases (see McDonald and Xu, 1995), most of which are displayed in the previous section. Therefore, it is reasonable to estimate the parameters of the GB distribution, because the GB distribution performs at least as well as the special or limiting distributions.

The GB distribution has five parameters $(a,\ b,\ c,\ p,\ q)$, and its probability density function (PDF) is written as

$$f(x) \quad = \quad \frac{|a|x^{ap-1}\left[1-(1-c)\left(\frac{x}{b}\right)^a\right]^{q-1}}{b^{ap}B(p,q)\left[1+c\left(\frac{x}{b}\right)^a\right]^{p+q}}, \quad 0<x^a<\frac{b^a}{1-c}, \tag{1}$$

where $B(p,q)$ is a beta function. For example, if we set $c=0$ or $c=1$, the distributions are reduced to GB1 and GB2, respectively. Moreover, if we set $c=1$ and $p=1$, it is reduced to an SM distribution (see Singh and Maddala, 1976), which a desirable distribution in many empirical applications. Detailed relationships among the class of distributions are summarized in McDonald and Xu (1995).

To introduce the cumulative distribution function (CDF), we provide the following function:

$$I_x(p,q)=\frac{B_x(p,q)}{B(p,q)},$$

where $B_x(p,q)$ is an incomplete beta function. Then, the CDF is written as

$$F(x)=I_z(p,q),\ \text{where}\ z=\frac{\left(\frac{x}{b}\right)^a}{1+c\left(\frac{x}{b}\right)^a}. \tag{2}$$

Given the PDF and CDF, we define the likelihood function following Nishino and Kakamu (2011), which is based on the concept of selected order statistics. [2] To explain the likelihood function, let $\boldsymbol{\theta} = (a,\ b,\ c,\ p,\ q)'$ be the vector of parameters and let $\mathbf{x} = (x_1,\ x_2, \ldots,\ x_k)'$ be the vector of observations. Then, the likelihood function is defined as follows:

$$L(\mathbf{x}|\boldsymbol{\theta}) = n!\frac{F(x_1)^{n_1-1}}{(n_1-1)!}f(x_1) \times \left\{\prod_{i=2}^{k}\frac{(F(x_i)-F(x_{i-1}))^{n_i-n_{i-1}-1}}{(n_i-n_{i-1}-1)!}f(x_i)\right\}\frac{(1-F(x_k))^{n-n_k}}{(n-n_k)!}. \tag{3}$$

If we substitute (1) and (2) for (4), it becomes the likelihood function for the GB distribution. [3]

2.2 Posterior Analysis

Because we adopt a Bayesian approach, we complete the model by specifying the prior distribution over the parameters. [4] We apply the following prior:

$$\pi(\boldsymbol{\theta}) = \pi(a)\pi(b)\pi(c)\pi(p)\pi(q).$$

Given a prior density $\pi(\boldsymbol{\theta})$ and the likelihood function in (4), the joint posterior distribution can be expressed as

$$\pi(\boldsymbol{\theta}|\mathbf{x}) \propto \pi(\boldsymbol{\theta})L(\mathbf{x}|\boldsymbol{\theta}). \tag{4}$$

Finally, we assume the following prior distributions:

$$a \sim \mathcal{N}(\mu_0, \tau_0^2),\ b \sim \mathcal{G}(\alpha_0, \beta_0),\ c \sim \mathcal{B}(\gamma_0, \delta_0),\ p \sim \mathcal{G}(\epsilon_0, \zeta_0),\ q \sim \mathcal{G}(\eta_0, \nu_0),$$

[2] In the MLE, the likelihood (which is based on the multinomial distribution) is widely used. Nishino and Kakamu (2011) applied a likelihood based on selected order statistics to the log-normal distribution, which is more exact than that based on the multinomial distribution. Therefore, we use this likelihood and the concept of selected order statistics as summarized in, for example, David and Nagaraja (2003). However, which likelihood we use is not critical.

[3] If the PDF and CDF for the concerning distribution are available, we can apply this likelihood function to the distribution. The PDFs and CDFs for the beta-type distributions are summarized in, for example, Hajargasht *et al*. (2012); Kleiber and Kotz (2003).

[4] A Bayesian approach was first proposed by Chotikapanich and Griffiths (2000) using a random walk Metropolis–Hastings (RWMH) algorithm. In this paper, we compare the performance of their algorithm with our proposed algorithm in the numerical examples. Therefore, their algorithm is introduced in Appendix A.

where $\mathcal{G}(a,b)$ and $\mathcal{B}(a,b)$ denote the gamma and beta distribution, respectively.

To obtain the posterior estimates, we implement the TaRMBH algorithm proposed by Chib and Ramamurthy (2010) as follows.

1. Separate $\boldsymbol{\theta}$ into a 3×1 vector $\boldsymbol{\theta}_1$ and a 2×1 vector $\boldsymbol{\theta}_2$ randomly.

2. For $j = 1,\ 2$, implement the following Metropolis–Hastings steps.

   (a) Generate $\boldsymbol{\theta}_j^{new}$ from a multivariate t distribution, $t(\hat{\boldsymbol{\theta}}_j, \boldsymbol{\Sigma}_j, \nu)$, with mean $\hat{\boldsymbol{\theta}}_j$, covariance $\boldsymbol{\Sigma}_j$, and ν degrees of freedom. [5] Here,

   $$\hat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta}_j} \log\left\{\pi(\boldsymbol{\theta})L(\mathbf{x}|\boldsymbol{\theta})\right\},$$

   $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2^{(m-1)\prime})'$ for $j = 1$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{(m)\prime}, \boldsymbol{\theta}_2')'$ for $j = 2$ using simulated annealing by Goffe *et al*. (1994). Here,

   $$\boldsymbol{\Sigma}_j = \left(-\frac{\partial^2 \log\left\{\pi(\boldsymbol{\theta})L(\mathbf{x}|\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\theta}_j \partial\boldsymbol{\theta}_j'}\right)^{-1}\Bigg|_{\boldsymbol{\theta}_j=\hat{\boldsymbol{\theta}}_j}.$$

   (b) If $j = 1$, compute

   $$\alpha_1(\boldsymbol{\theta}_1^{m-1}, \boldsymbol{\theta}_1^{new}) = \min\left\{\frac{\pi(\boldsymbol{\theta}_1^{new}|\boldsymbol{\theta}_2^{(m-1)}, \mathbf{x})q(\boldsymbol{\theta}_1^{(m-1)}|\hat{\boldsymbol{\theta}}_1, \boldsymbol{\Sigma}_1)}{\pi(\boldsymbol{\theta}_1^{(m-1)}|\boldsymbol{\theta}_2^{(m-1)}, \mathbf{x})q(\boldsymbol{\theta}_1^{new}|\hat{\boldsymbol{\theta}}_1, \boldsymbol{\Sigma}_1)}, 1\right\},$$

   and if $j = 2$, compute

   $$\alpha_2(\boldsymbol{\theta}_2^{m-1}, \boldsymbol{\theta}_2^{new}) = \min\left\{\frac{\pi(\boldsymbol{\theta}_2^{new}|\boldsymbol{\theta}_1^{(m)}, \mathbf{x})q(\boldsymbol{\theta}_2^{(m-1)}|\hat{\boldsymbol{\theta}}_2, \boldsymbol{\Sigma}_2)}{\pi(\boldsymbol{\theta}_2^{(m-1)}|\boldsymbol{\theta}_1^{(m)}, \mathbf{x})q(\boldsymbol{\theta}_2^{new}|\hat{\boldsymbol{\theta}}_2, \boldsymbol{\Sigma}_2)}, 1\right\},$$

   where $q(\boldsymbol{\theta}_j^{new}|\hat{\boldsymbol{\theta}}_j, \boldsymbol{\Sigma}_j)$ is a multivariate t distribution given in (a).

   (c) Generate a value u_j from $\mathcal{U}(0,1)$, where $\mathcal{U}(a,b)$ is an uniform distribution on the interval (a,b).

   (d) If $u_j \leq \alpha_j\left(\boldsymbol{\theta}_j^{(m-1)}, \boldsymbol{\theta}_j^{new}\right)$, set $\boldsymbol{\theta}_j^{(m)} = \boldsymbol{\theta}_j^{new}$, otherwise $\boldsymbol{\theta}_j^{(m)} = \boldsymbol{\theta}_j^{(m-1)}$.

3. Return to step 1, and set m to $m + 1$.

[5] In the numerical examples discussed below, we set $\nu = 15$ as recommended by Chib and Ramamurthy (2010). In addition, as in Chib and Ramamurthy (2010), the inverse of the negative Hessian, which is a covariance matrix $\boldsymbol{\Sigma}_j$, may not be positive definite. Thus, we also compute a modified Cholesky algorithm by Nocedal and Wright (2000).

In all the numerical examples discussed in the next section, we set the hyper parameters as $\mu_0 = 0$, $\tau_0^2 = 100$, $\gamma_0 = \delta_0 = \epsilon_0 = \zeta_0 = \eta_0 = \nu_0 = 1.0$. Results reported in the next section are generated using Ox version 7.00 (OS_X_64/U) (Doornik, 2009).

3 Numerical Examples

3.1 Simulated Data

To illustrate the Bayesian approach discussed in the previous section, we compare the algorithm with that by Chotikapanich and Griffiths (2000) using a simulated dataset. We set the number of observations to $n = 100,000$ and assume a decile number of groups ($k = 9$). Given n and k, we assume that the true data-generating process (DGP) is a GB distribution,[6] where the parameters are $a = 5$, $b = 30$, $c = 0.95$, $p = 0.5$, and $q = 0.8$. Furthermore, generate x_j for $j = 1, 2, \ldots, n$. The generated random numbers are sorted in ascending order, and x_i corresponds to the n_ith observation. Then, $i = 1, 2, \ldots, k$ is picked up and $\mathbf{x} = (x_1, x_2, \ldots, x_k)'$ are collected. Given the dataset, we run a RWMH algorithm using $800,000$ iterations and discarding the first $300,000$, while we run a TaRBMH algorithm using $11,000$ iterations and discarding the first $1,000$.

Table 1 shows the posterior estimates of the parameters using RWMH and TaRBMH. From Table 1, we first observe that the posterior means are close to each other, while the 95% credible intervals from TaRBMH are slightly wider than those from RWMH. In addition, all parameters include true values in the 95% credible intervals. Therefore, we conclude that both Bayesian approaches work well in the parameter estimation of the GB distribution. However, a 30-times difference appears in the inefficiency factors (IF), which approximate the ratio of the numerical variance of the estimate from the MCMC chain relative to that from hypothetical i.i.d. draws.

To illustrate the differences in the mixing of the two algorithms over the parameter space, we plot the sample draws from the posterior distribution for the parameters together with auto-correlation functions (ACFs)

[6] To generate a random number from GB with parameters a, b, c, p, and q, first generate Z from a beta distribution with parameters p and q. Then, $X = b\left(\frac{Z}{1-cZ}\right)^{\frac{1}{a}}$ becomes the random number from the GB distribution.

in Figure 2. The top panel corresponds to the draws from the TaRBMH algorithm, and the bottom panel corresponds to those from the RWMH algorithm. As these plots show, the RWMH chain is highly persistent with the ACFs retaining significant mass even at lag length 300. On the other hand, the TaRBMH ACFs decay quickly (within lag length 200 for all parameters).

Finally, we discuss the reason why the convergence of the MCMC chain is slow. Figure 3 shows the counter plots of the marginal joint posterior distributions of each parameter. Figure 3 implies that correlations between a and p, between a and q, between b and q, and between p and q are very high. In the GB distribution, the shape parameters are a, p, and q, and several combinations of these three parameters might lead to similar distribution shapes. Therefore, such high correlations lead to the identification problem and make the convergence of the MCMC chain slow. However, because TaRBMH speeds up convergence, we can conclude that our algorithm is superior to the RWMH algorithm in terms of mixing.

3.2 U.S. Family Income Data

As far as we know, the GB distribution is estimated for grouped data only by McDonald and Xu (1995) and Bordley *et al*. (1996). While McDonald and Xu (1995) did not seem to successfully estimate parameter c and estimated the GB2 distribution instead, Bordley *et al*. (1996) estimated all parameters including c every 5 years from 1970 to 1990. Therefore, it is reasonable to compare the results from Bordley *et al*. (1996) with the Bayesian approaches. We examine U.S. income data, which is estimated in Bordley *et al*. (1996). To estimate the parameters of the GB distribution from these data, we run an RWMH algorithm using $200,000$ iterations and discarding the first $100,000$, while we run a TaRBMH algorithm using $11,000$ iterations and discarding the first $1,000$ for every year.

Table 2 shows the estimation results from Bordley *et al*. (1996) and the posterior estimates of the parameters using the RWMH and TaRBMH algorithms. First, if we compare the posterior estimates of both algorithms, we see that the 95% credible intervals from the TaRBMH algorithm are wider than those from the RWMH algorithm. However, the posterior means are generally close to each other and the IF from the TaRBMH

algorithm is much smaller than that from the RWMH algorithm except for the year 1975. Therefore, the TaRBMH algorithm searches through wider parameter spaces efficiently, and we focus on the results from the TaRBMH thereafter.

Next, we compare the estimation results from Bordley *et al*. (1996) with the posterior estimates. Focusing on the results for 1970 shows that all MLE estimates are close to the posterior means and are included in the 95% credible intervals. The same situation occurs for the results for 1975. However, for the results for 1980, the MLE estimates of the parameters b, c, and q out of the five total parameters are not within the 95% credible intervals of the Bayesian estimates. In addition, four and five parameters respectively in 1985 and 1990 are not within the 95% credible intervals. Therefore, these results imply that the MLE estimates may not achieve the optima.

To confirm the effects of parameter c, its marginal posterior distributions are displayed in Figure 4. The figure shows a clear truncation of the posterior distribution at 1.0 in 1975. However, the truncation is not obvious in 1970, 1980, 1985, and 1990. Therefore, the estimates of c in 1970, 1980, 1985, and 1990 are not 1.0. The differences in the estimates may be caused by the failure of the global maximization in the MLE. Moreover, parameter c plays an important role in the estimation of the income distribution in the United States because parameter c is estimated to be different from 1.0.

Once the parameters of the GB distribution are estimated, we can calculate the Gini coefficient from the parameters. [7] Table 3 and Figure 5 show the Gini coefficients of the U.S. family income data. The Gini coefficients are calculated from the MCMC draws of TaRBMH. From the results, we can observe that the inequalities have increasing trends. To confirm this tendency, Figure 6 shows the mobility of income distributions. From the figure, we can confirm that the modes slightly move to higher income and becomes lower. These might be the cause of the increasing trend of the Gini coefficients.

[7] There is no analytical expression of the Gini coefficient for the GB distribution that differs from other beta-type distributions. Therefore, we calculate the Gini coefficient from the GB distribution using numerical integration.

3.3 Family Income and Expenditure Survey in Japan

Finally, we examine data from the Family Income and Expenditure Survey (FIES) in Japan prepared by the Statistics Bureau, Ministry of Internal Affairs and Communications. We use workers' households data from the 2009 FIES. Quintile and decile data exist, and the sample size in both datasets is $n = 10,000$. To estimate the parameters of the GB distribution from the FIES data in Japan, we run a TaRBMH algorithm using $21,000$ iterations and discarding the first $1,000$.

Table 4 shows the posterior estimates from the FIES data in Japan. Table 4 implies that the estimates of c differ from 1.0. Therefore, parameter c also plays an important role in income distribution in Japan. We find that the posterior means and the 95% credible intervals differ slightly. This is especially the case in parameter p, where only the number of groups differs. Therefore, the posterior means in one result are included in the 95% credible intervals in the other result. To visually see the differences, GB distributions fitted to the representative sample are displayed in Figure 7. The figure implies that the mode and the shape in the distribution are slightly different. However, both distributions seem to fit to the histogram drawn from the dataset. Therefore, our proposed algorithm works well even when the number of groups is small.

Finally, we examine the estimated Gini coefficient from the GB distribution as same as the case of U.S. family income data. Table 5 shows the posterior estimates of the Gini coefficients. Although the posterior estimates of the original parameters slightly differ, those of the Gini coefficients are similar to each other. However, the 95% credible interval of the quintile data is wider than that of the decile data. Figure 8 shows the posterior distributions of the Gini coefficients, and we confirm that the posterior mode is different and the distribution from the quintile data is skewed. As shown by Kakamu (2015), [8] the effect of the number of groups may also appear in the distribution variance of the Gini coefficient in the GB distribution. To discuss the accuracy of the Gini coefficient, we calculated the lower and upper bounds of the Gini coefficient as proposed by

[8] Kakamu (2015) examined the performance of the Gini coefficients assuming Singh–Maddala and Dagum distributions using Monte Carlo experiments and showed that the effects of not only the number of observations but also the number of groups appears in the Gini coefficients in terms of the root mean square errors (RMSEs).

Gastwirth (1972), which are 0.238 and 0.261, respectively, for the quintile data. [9] The result indicates that both posterior means are included in the lower and upper bounds of the Gini coefficient. The 95% credible interval for the decile data approaches the lower and upper bounds, whereas that for the quintile data is slightly wider than these bounds. Then, the GB distribution can estimate the Gini coefficient accurately, and the accuracy increases as the number of groups increases.

4 Conclusions

This paper considered the estimation of the GB distribution parameters for grouped data from a Bayesian point of view. To estimate the parameters of the distribution, we utilized the TaRBMH algorithm proposed by Chib and Ramamurthy (2010). We examined numerical examples with one simulated and two real datasets. In the numerical examples, we compared a TaRBMH algorithm with the RWMH algorithm proposed by Chotikapanich and Griffiths (2000). From the results, we confirmed that TaRBMH is more efficient than RWMH in terms of mixing. In addition, using empirical results of U.S. income data, we showed that the estimated parameters of our Bayesian approach may differ from those of the MLE. Finally, our Bayesian approach could estimate the parameters of the GB distribution even if the number of groups are relatively small, and the Gini coefficient from the GB distribution could be calculated accurately.

Finally, $10,000$ iterations of TaRBMH takes approximately 6 hours, while $2,000,000$ iterations of RWMH takes about 10 minutes. The computation time is critical, but the shorter iterations of the RWMH algorithm might lead to the failure of an accurate estimation of the true parameters. We observed that the 95% credible intervals from RWMH are narrower than those from TaRBMH in the simulated and real datasets. Although it is not reported in this paper, more than $50,000,000$ iterations were required to make the MCMC chain converge using FIES data with RWMH, and this takes more than 6 hours. Overcoming efficiency issues of the MCMC

[9] To calculate the nonparametric lower and upper bounds by Gastwirth (1972), class income means are required in addition to income classes and the frequencies. In the income data from FIES, the class income means are also available. Thus, we can compare the results in this datasets.

chain and time-consuming problems simultaneously is left to future work.

A MCMC Schemes by Chotikapanich and Griffiths (2000)

In this appendix, we briefly explain the MCMC procedures proposed by Chotikapanich and Griffiths (2000). Chotikapanich and Griffiths (2000) first proposed an MCMC method to estimate the distribution parameters from grouped data using an RWMH algorithm. Their algorithm is as follows.

1. Generate a candidate value $\boldsymbol{\theta}^{new}$ from $\mathcal{N}(\boldsymbol{\theta}^{(m-1)}, c^2\boldsymbol{\Sigma})$, where c is a tuning parameter and $\boldsymbol{\Sigma}$ is the maximum likelihood covariance estimate. [10]

2. Compute

$$\alpha\left(\boldsymbol{\theta}^{(m-1)}, \boldsymbol{\theta}^{new}\right) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^{new}|\mathbf{x})}{\pi(\boldsymbol{\theta}^{(m-1)}|\mathbf{x})}\right\}.$$

   If any of the elements of $\boldsymbol{\theta}^{new}$ fall outside the feasible parameter region, $\alpha\left(\boldsymbol{\theta}^{(m-1)}, \boldsymbol{\theta}^{new}\right) = 0$.

3. Generate a value u from $\mathcal{U}(0,1)$.

4. If $u \leq \alpha\left(\boldsymbol{\theta}^{(m-1)}, \boldsymbol{\theta}^{new}\right)$, set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{new}$, otherwise set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{(m-1)}$.

5. Return to step 1, with m set to $m+1$.

Acknowledgements

Previous versions of this paper were presented at the European Seminar on Bayesian Econometrics (ESOBE 2012) in Vienna, Computational and Financial Econometrics (CFE 2012) in Oviedo, Econometric Society of Australian Meeting (ESAM 2013) in Sydney, International Society for Bayesian Analysis World Meeting

[10] It may be difficult to estimate using maximum likelihood methods, and such an estimation may be sensitive to the choice of initial values. Therefore, we implement simulated annealing by Goffe *et al*. (1994) to find the mode. The setup for simulated annealing is the same as that used in the TaRBMH.

(ISBA 2014) in Cancun, International Workshop on Bayesian Econometrics and Computation at Kobe University, and at a work-in-progress seminar at Tokyo Institute of Technology. We would like to thank the seminar/conference participants, especially Sungbae An, Gholamreza Hajargasht, Hideo Kozumi, Yasuhiro Omori, Scott Sisson, and Herman van Dijk for their valuable comments and suggestions. This work is partially supported by KAKENHI #26380266, #24530222, and #25245035.

References

[1] Atoda, N., T. Suruga, and T. Tachibanaki, 1988, Statistical inference of functional forms for income distribution. The Economic Studies Quarterly 39, 14–40.

[2] Bordley, R.F., J.B. McDonald, and A. Mantrala, 1996, Something new, something old: Parametric models for the size distribution of income. Journal of Income Distribution 6, 91–103.

[3] Chib, S. and S. Ramamurthy, 2010, Tailored randomized block MCMC methods with applications to DSGE models. Journal of Econometrics 155, 19–38.

[4] Chotikapanich, D. and W.E. Griffiths, 2000, Posterior distributions for the Gini coefficient using grouped data. Australian and New Zealand Journal of Statistics 42, 383–392.

[5] Chotikapanich, D. and W.E. Griffiths, 2008, Estimating income distributions using a mixture of gamma densities. (Chotikapanich, D. ed) Modeling Income Distributions and Lorenz Curves. Springer, New York, 285–302.

[6] Chotikapanich, D., D.S.P. Rao, and K.K. Tang, 2007, Estimating income inequality in China using grouped data and the generalized beta distribution. Review of Income and Wealth 53, 127–146.

[7] David, H.A. and H.N. Nagaraja, 2003, Order Statistics, 3rd ed., Wiley, New York.

[8] Doornik, J.A., 2009, Ox: An Object Oriented Matrix Programming Language, Timberlake, London.

[9] Gastwirth, J.L. 1972, The estimation of the Lorenz curve and Gini index. Review of Economics and Statistics 54, 306–316.

[10] Goffe, W.L., G.D. Ferrier, and J. Rogers, 1994, Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60, 65–99.

[11] Hajargasht, G., W,E. Griffiths, J. Brice, D.S.P. Rao, and D. Chotikapanich, 2012, Inference for income distributions using grouped data. Journal of Business and Economic Statistics 30, 563–575.

[12] Hasegawa, H. and H. Kozumi, 2003, Estimation of Lorenz curves: A Bayesian nonparametric approach. Journal of Econometrics 115, 277–292.

[13] Kakamu, K., 2015, Simulation studies comparing Dagum and Singh–Maddala income distributions. forthcoming in Computational Economics.

[14] Kleiber, C. and S. Kotz, 2003, Statistical Size Distributions in Economics and Actuarial Science, Wiley, New York.

[15] Kloek, T. and H.K. van Dijk, 1978, Efficient estimation of income distribution parameters. Journal of Econometrics 8, 61–74.

[16] McDonald, J.B., 1984, Some generalized functions for the size distribution of income. Econometrica 52, 647–663.

[17] McDonald, J.B. and A. Mantrala, 1995, The distribution of personal income: Revisited. Journal of Applied Econometrics 10, 201–204.

[18] McDonald, J.B. and M.R. Ransom, 1979a, Functional forms, estimation techniques and the distribution of income. Econometrica 47, 1513–1525.

[19] McDonald, J.B. and M.R. Ransom, 1979b, Alternative parameter estimators based upon grouped data. Communications in Statistics A8, 899–917.

[20] McDonald, J.B. and M.R. Ransom, 2008, The generalized beta distribution as a model for the distribution of income: Estimation of related measures of inequality. (Chotikapanich, D. ed) Modeling Income Distributions and Lorenz Curves, Springer, New York, 147–166.

[21] McDonald, J.B. and Y.J. Xu, 1995, A generalization of the beta distribution with applications. Journal of Econometrics 66, 133–152.

[22] Majumder, A. and S.R. Chakravarty, 1990, Distribution of personal income: Development of a new model and its application to U.S. income data. Journal of Applied Econometrics 5, 189–196.

[23] Nishino, H. and K. Kakamu, 2011, Grouped data estimation and testing of Gini coefficients using lognormal distributions. Sankhya Series B 73, 193–210.

[24] Nocedal, J. and S.J. Wright, 2000, Numerical Optimization. Second Edition, Springer, New York.

[25] Salem, A.B.Z. and T.D. Mount, 1974, A convenient descriptive model of income distribution: The gamma density. Econometrica 42, 1115–1127.

[26] Singh, S.K. and G.S. Maddala, 1976, A function for size distribution of income. Econometrica 47, 1513–1525.

[27] Slottje, D.J., 1984, A measure of income inequality in the U.S. for the years 1952–1980 based on the beta distribution of the second kind. Economics Letters 15, 369–375.

[28] Tachibanaki, T., T. Suruga, and N. Atoda, 1997, Estimations of income distribution parameters for individual observations by maximum likelihood method. Journal of the Japan Statistical Society 27, 191–203.

[29] van Dijk, H.K. and T. Kloek, 1980, Inferential procedures in stable distributions for class frequency data on incomes. Econometrica 48, 1139–1148.

Table 1: Simulated data

		RWMH				TaRBMH			
	true values	Mean	95%CI		IF	Mean	95%CI		IF
a	5.00	5.171	4.258	6.080	11961.408	5.554	4.058	7.451	178.217
b	30.00	30.417	28.318	33.305	6585.192	30.169	28.131	33.413	93.938
c	0.95	0.945	0.913	0.972	8297.133	0.950	0.905	0.980	85.169
p	0.50	0.470	0.382	0.589	11591.862	0.440	0.304	0.618	151.172
q	0.80	0.784	0.554	1.195	7977.529	0.726	0.439	1.185	148.400

Note: Posterior means (Mean), 95% credible intervals (95%CI), and inefficiency factors (IF) are displayed. The acceptance rates are as follows: around 20% in the RWMH algorithm, and around 90% in the TaRBMH algorithm.

Table 2: Real data: U.S. family income data

U.S. Family Income Data in 1970									
	MLE	RWMH				TaRBMH			
		Mean	95%CI		IF	Mean	95%CI		IF
a	4.797	4.734	4.673	4.794	6449.905	4.818	4.391	5.259	67.085
b	44.29	44.121	43.641	44.632	6065.901	44.213	42.992	45.607	49.977
c	0.997	0.996	0.994	0.997	5985.317	0.997	0.995	0.999	3.951
p	0.316	0.322	0.315	0.329	4989.927	0.316	0.286	0.350	65.848
q	0.695	0.701	0.685	0.715	4411.731	0.695	0.603	0.805	66.966
U.S. Family Income Data in 1975									
	MLE	RWMH				TaRBMH			
		Mean	95%CI		IF	Mean	95%CI		IF
a	2.887	2.935	2.715	3.161	81.357	2.944	2.689	3.185	85.644
b	54.87	54.767	52.260	57.567	89.930	54.686	52.120	57.898	76.089
c	1.000	0.996	0.988	1.000	11.589	0.996	0.988	1.000	5.753
p	0.561	0.551	0.501	0.605	87.320	0.549	0.497	0.612	82.158
q	1.587	1.573	1.356	1.823	92.656	1.564	1.338	1.860	85.611
U.S. Family Income Data in 1980									
	MLE	RWMH				TaRBMH			
		Mean	95%CI		IF	Mean	95%CI		IF
a	2.587	2.776	2.558	3.016	276.391	2.807	2.559	3.051	93.655
b	64.48	59.148	54.687	64.175	153.673	58.643	54.202	64.140	81.448
c	1.000	0.977	0.958	0.997	25.167	0.977	0.958	0.997	17.206
p	0.599	0.554	0.501	0.610	247.684	0.547	0.494	0.610	90.075
q	1.961	1.589	1.285	1.948	185.164	1.551	1.253	1.943	91.084
U.S. Family Income Data in 1985									
	MLE	RWMH				TaRBMH			
		Mean	95%CI		IF	Mean	95%CI		IF
a	2.498	2.753	2.524	3.009	2432.152	2.747	2.503	3.009	86.002
b	66.06	58.317	54.192	63.100	1292.948	58.421	53.942	63.615	80.575
c	1.000	0.978	0.961	0.997	177.366	0.978	0.961	0.997	20.545
p	0.578	0.520	0.466	0.576	2199.231	0.522	0.468	0.581	79.589
q	1.793	1.330	1.088	1.618	1656.529	1.337	1.070	1.657	88.151
U.S. Family Income Data in 1990									
	MLE	RWMH				TaRBMH			
		Mean	95%CI		IF	Mean	95%CI		IF
a	2.731	3.055	2.799	3.324	1945.881	3.086	2.797	3.373	103.874
b	62.19	54.885	51.629	58.954	1116.695	54.524	51.176	58.917	92.802
c	1.000	0.978	0.967	0.993	122.056	0.977	0.966	0.992	13.945
p	0.519	0.460	0.416	0.508	1803.985	0.455	0.411	0.508	103.749
q	1.358	0.980	0.817	1.192	1458.683	0.960	0.794	1.192	111.004

Note: MLE estimates are based on the results from Bordley *et al*. (1996). The acceptance rates are as follows: around 50% in the RWMH algorithm every year, around 70% in the TaRBMH in 1975, and around 90% in the TaRBMH algorithm in other years.

Table 3: Gini coefficients for U.S. data

Year	Mean	95%CI	
1970	0.341	0.337	0.345
1975	0.349	0.346	0.352
1980	0.357	0.354	0.360
1985	0.379	0.375	0.384
1990	0.382	0.376	0.389

Table 4: Real data: Family Income and Expenditure Survey in Japan in 2009

	Quintile				Decile			
	Mean	95%CI		IF	Mean	95%CI		IF
a	2.794	1.584	5.370	448.561	2.136	1.347	3.122	401.557
b	0.635	0.454	0.817	63.685	0.509	0.350	0.669	110.153
c	0.951	0.889	0.996	146.477	0.933	0.894	0.960	214.113
p	1.701	0.553	3.608	534.857	2.895	1.359	5.884	409.815
q	1.569	0.466	3.863	409.743	1.720	0.855	3.647	278.706

Note: The acceptance rates are as follows: around 20% in the RWMH algorithm, and around 80% in the TaRBMH one.

Table 5: Gini coefficients for Japan data

Quintile			Decile		
Mean	95%CI		Mean	95%CI	
0.247	0.224	0.265	0.243	0.235	0.251

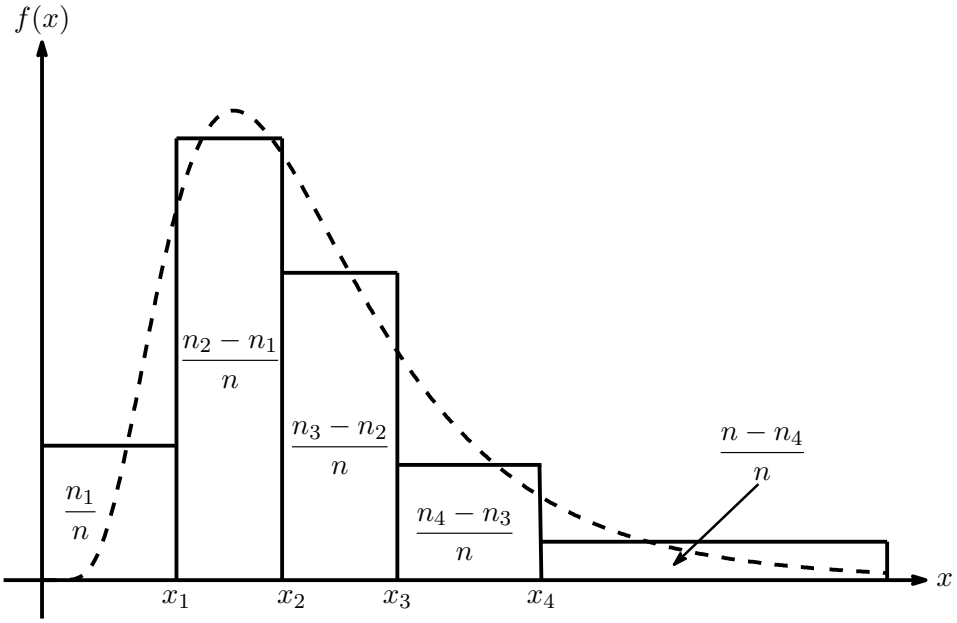

Figure 1: Quintile Data

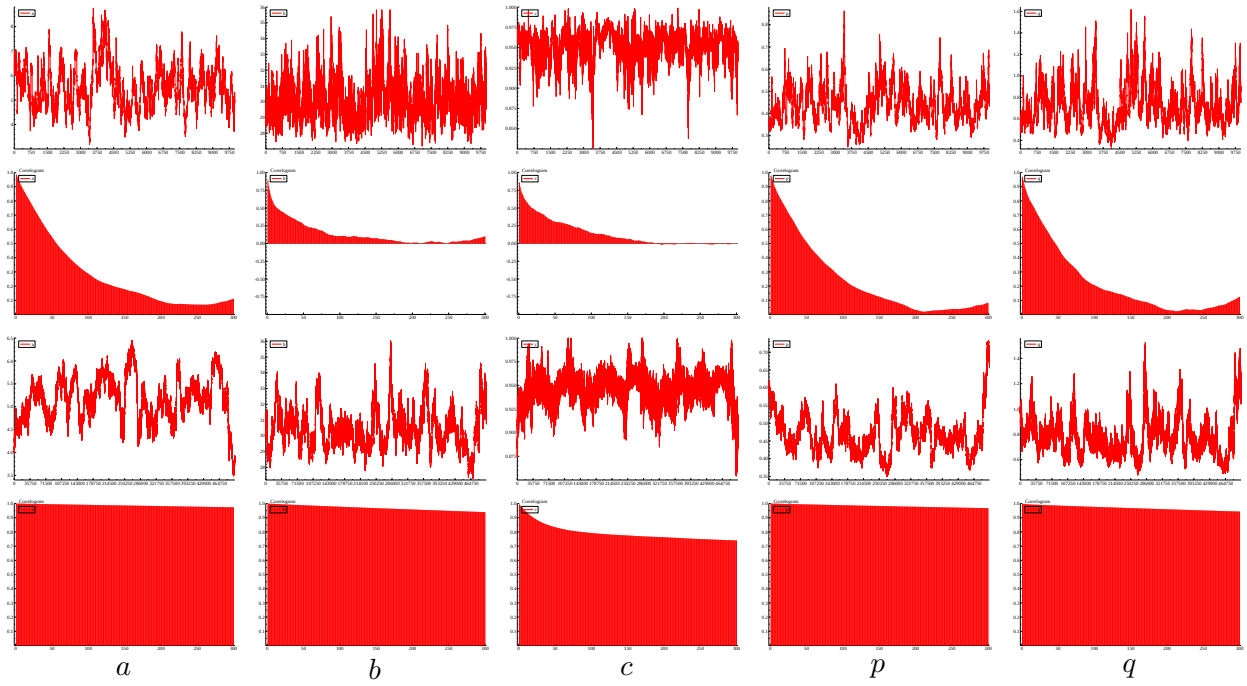

Figure 2: Sampling results from simulated data: A time-series plots of draws from the posterior and corresponding auto-correlation functions using the TaRBMH (top panel) and the RWMH (bottom panel)

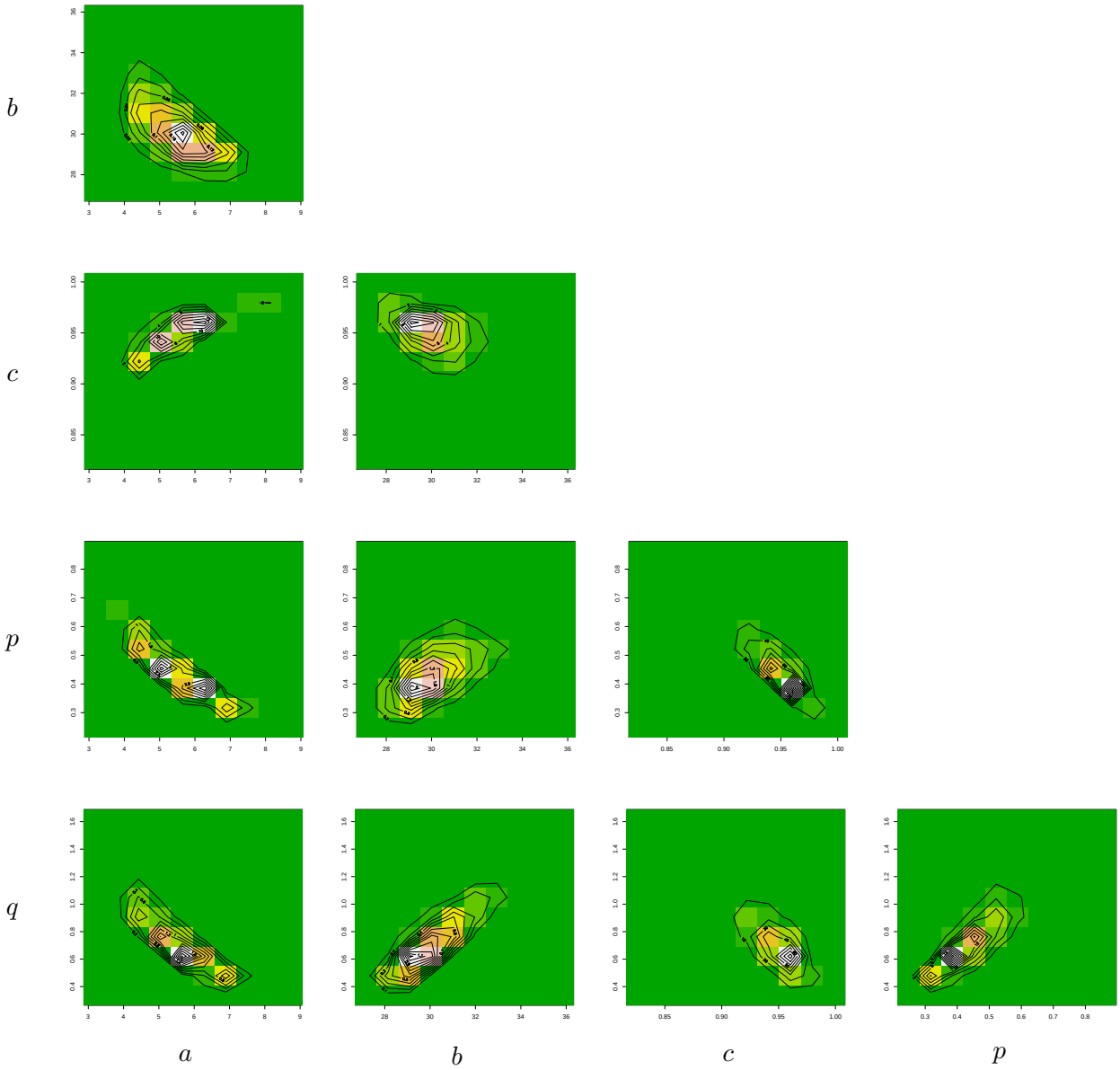

Figure 3: Joint posterior distributions from simulated data

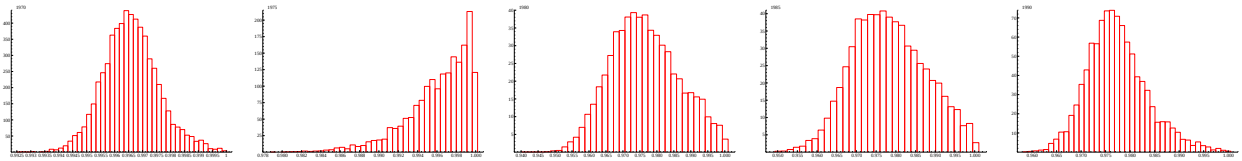

Figure 4: Marginal posterior distributions of c from U.S. income data

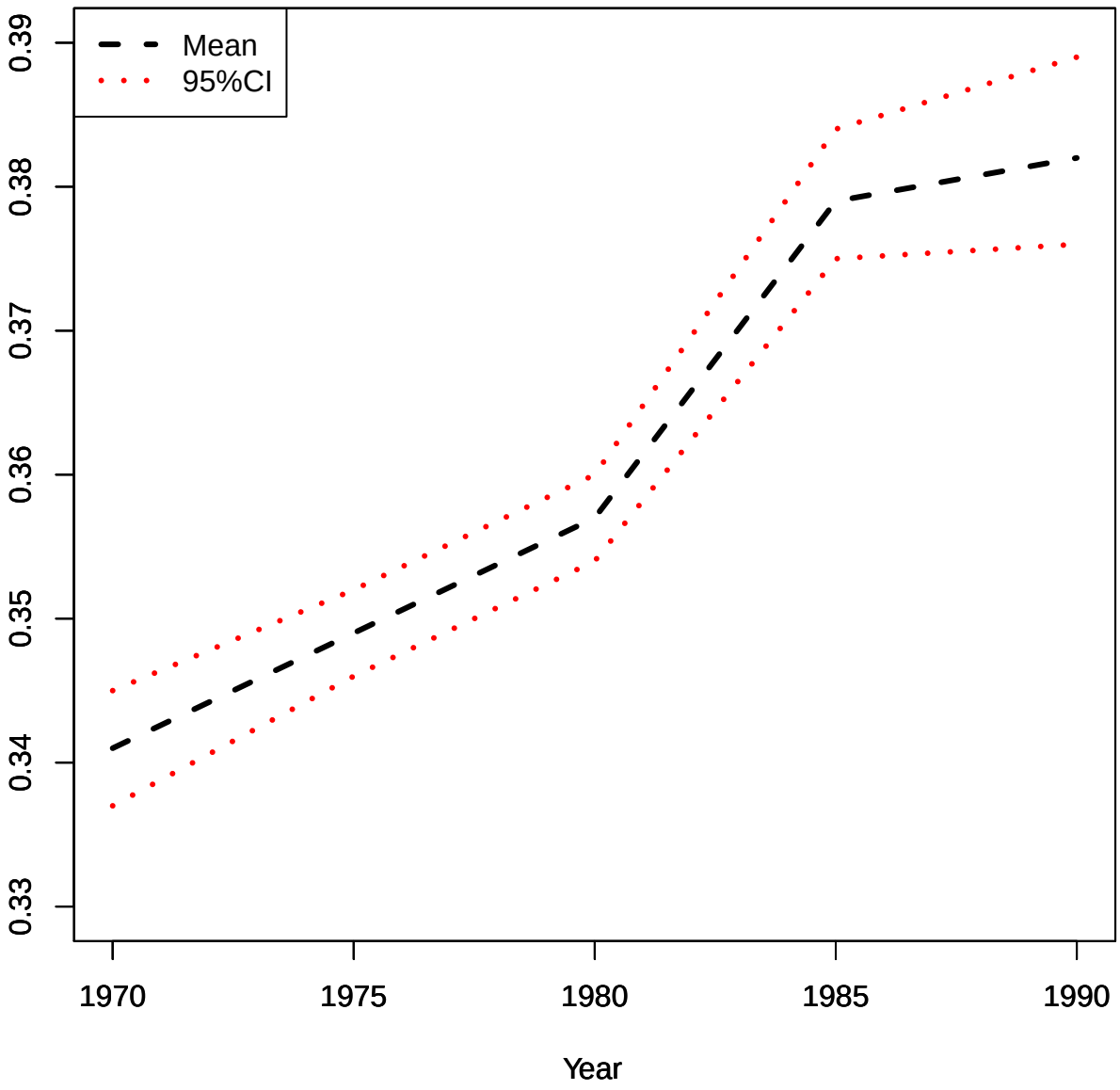

Figure 5: The trends of the U.S. Gini coefficients

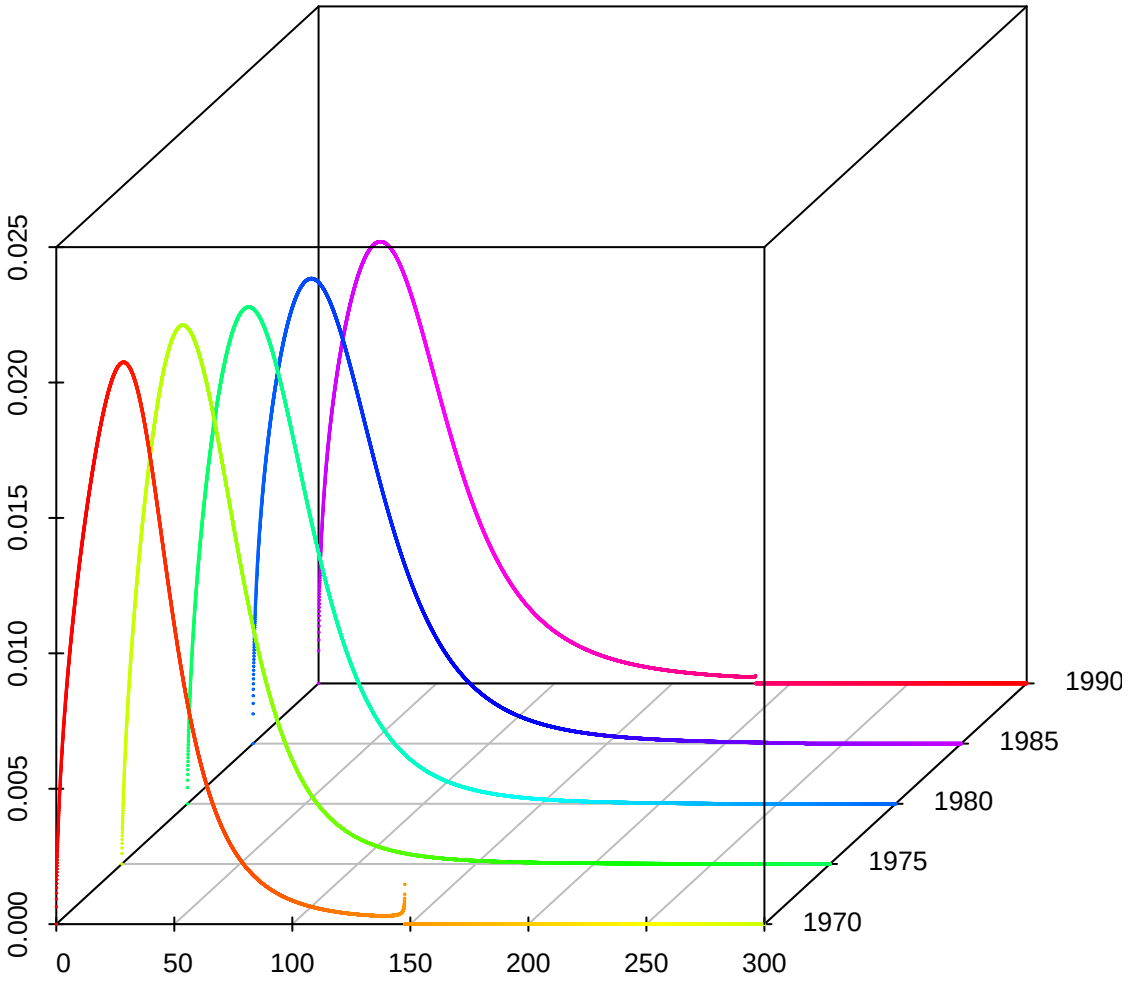

Figure 6: The mobility of U.S. income distributions

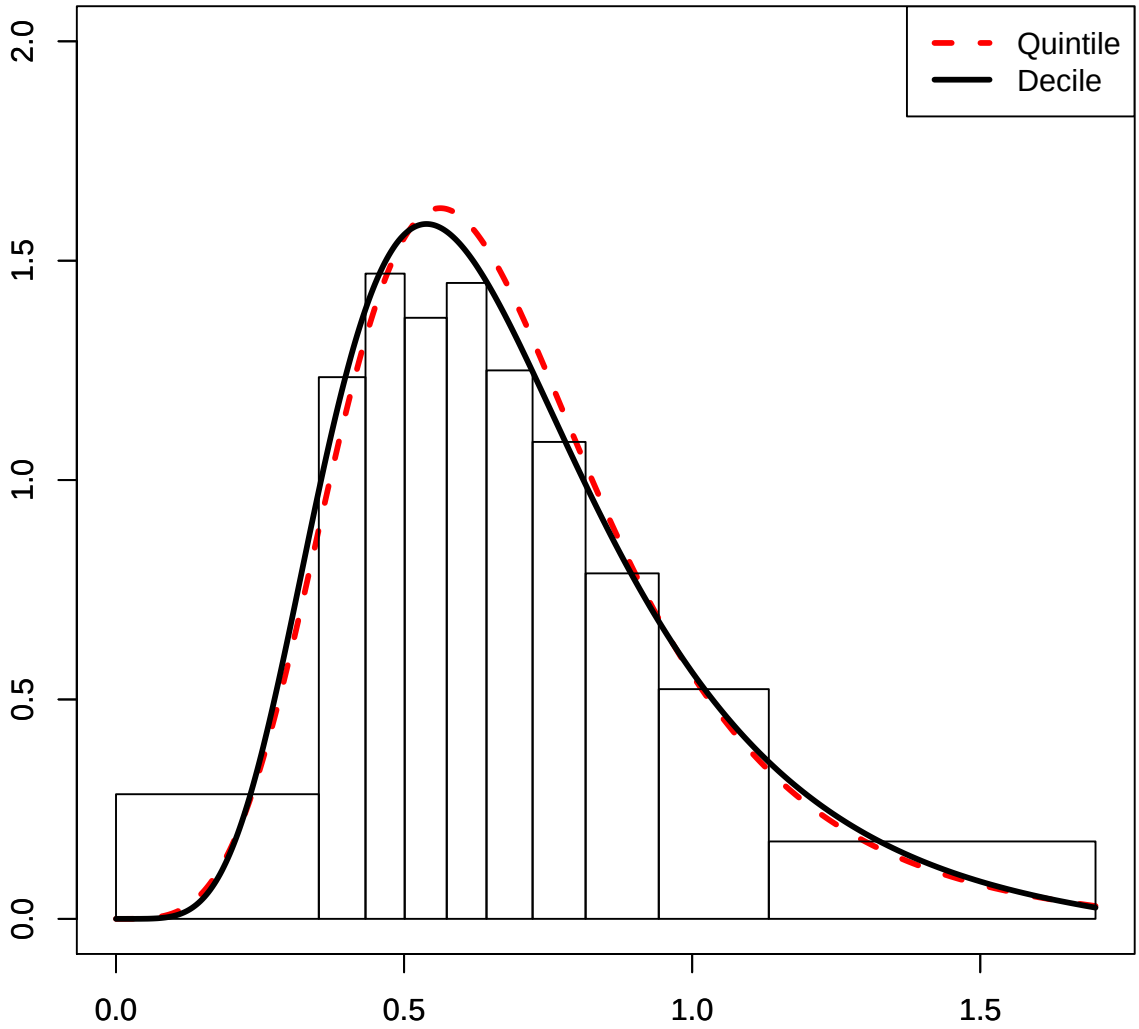

Figure 7: GB distributions fitted to the representative sample

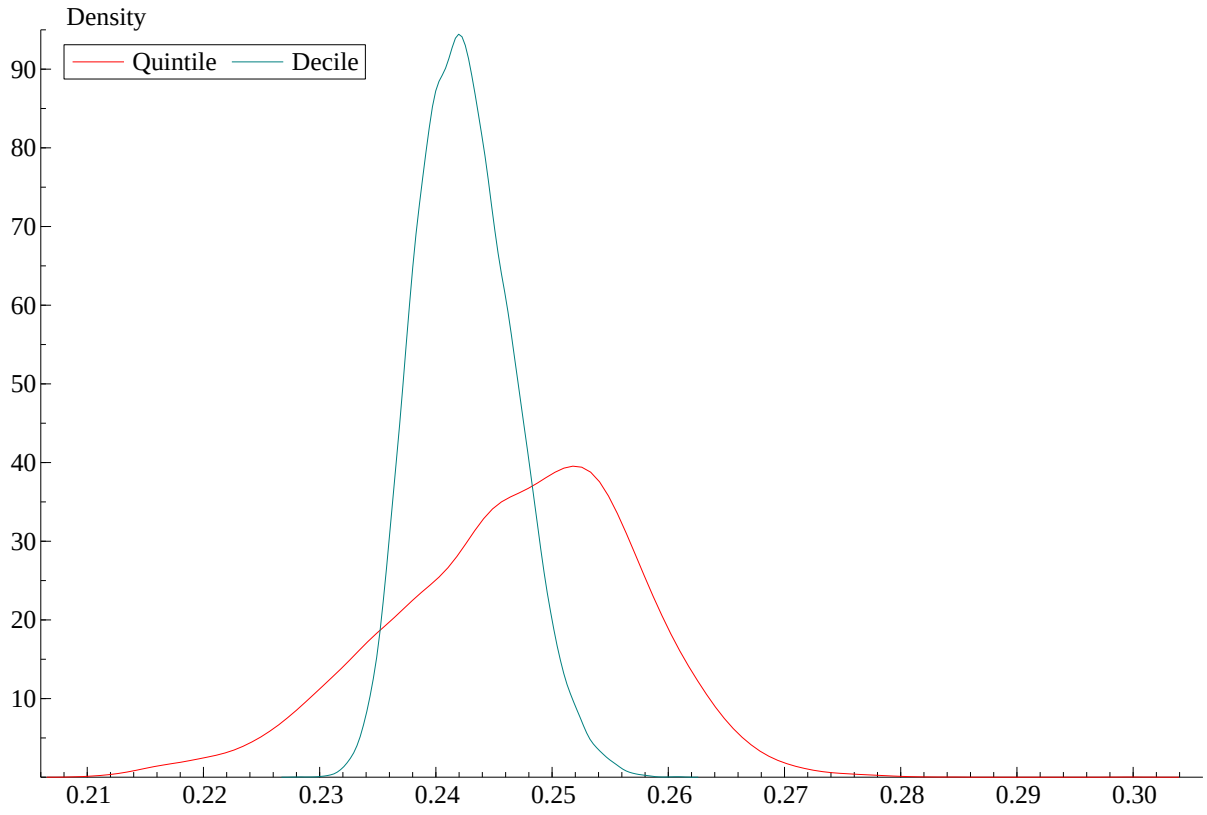

Figure 8: Posterior distributions of the Gini coefficients

[2016.3.31 1220]