Random Coefficient Logit Model with MCMC Algorithms*

Hao-Che (Howard) Hsu†

This Draft: December, 2020

Last version available here

Abstract

Numerous techniques have been developed contributing to the tools of solving the random coefficients logit model. Following the developed methods, I modify the prior distribution assumption on the aggregate demand shock and demonstrate estimating demand by sequentially updating the market share inversion process and two MCMC techniques. In particular, I present a practitioner's guide including details regarding the implementation of the algorithms.

Keywords: Random coefficient logit, Gibbs sampler, Metropolis-Hastings, MCMC

JEL Codes: C11, Q11

*I thank Colin Reinhardt for numerous suggestions. This paper also served as the course project for ECON 223A.

†Department of Economics, University of California-Irvine; haoche.hsu@uci.edu.

1 Introduction

Differentiate products depend on their heterogeneous characteristics which are the foremost factor that stimulates demand. Comparing to homogeneous products, the differentiability yields distinguishing features that lead to the forming of market power of a firm.

In general, the competition of natural resources forms homogeneous markets such as offshore oil drilling (Hendricks, Porter and Wilson, 1994) and gas leases. instead of based on oil or gas characteristics, the choice decision in such markets often hinges on how informative the competitors are. In most common markets such as grocery stores, wholesalers, or even online stores like eBay or Amazon, we can categorize products by their type and functionalities. However, it is their idiosyncratic properties that influence our final choice decisions.

Demand estimation has become an important aspect to probe consumer behavior. From choosing the brand of cereal to eat for breakfast, the movie to watch on the weekends, and the graduate programs to apply, they all involved decisions of choosing between the characteristics of differentiated products. Therefore, the sellers must explore and identify market demand and the factors that affect consumer choice.

Following the method of market-share-inversion is proposed by Berry (1994), there are various estimation techniques exploited to estimate the structural model. Besides the well known GMM method solved with a nested fixed-point algorithm (Berry, Levinsohn and Pakes, 1995), rather than solving constraint optimization for every loop, Su and Judd (2012) points out that the original problem is equivalent to formulate as a single mathematical program with equilibrium constraints (MPEC). This solution circumvents running any loops and drastically reduces the coding complexity and the total execution time. Nevertheless, this proposition relies on solvers and users need to supply the Hessian matrix[1]. Comparing the MPEC approach with the nested double loops, even though MPEC is considerably faster, sometimes incorrect solutions are estimated while the results from the iteration approach are oftentimes, more accurate, and robust.

Although this may sound like attempting to adopt the iteration procedure, some obstacles seldom make this a cumbersome choice. The most common criticized problem from the nested fixed-point algorithm is the concern of non-convergence. Under the requirements to designate a tight convergence tolerance for both iterations, oftentimes, the algorithm will fail to converge. In this case, researchers might deviate to a higher tolerance as an alternative but suffer from producing off-estimates. Besides, the nested optimization problems are handled

[1] Supplying the Hessian matrix avoids additional computation burden from numerical approximation. Moreover feeding the analytical sparse Hessian can speed up the optimization even further.

with grid searches. For every round in the outer loop, it demands the inner loop to search for a new converged candidate. This process ends when the outer loop converges which causes the entire algorithm extremely time-consuming. Researchers have experience estimations that take time from several hours to several months. This paper explores and verifies an alternative approach, using Bayesian methods to estimate the demand model.

From a frequentist perspective, Bayesian techniques are asymptotically valid. Moreover, Bayesian methods have computational advantages and sometimes, they are more straightforward to implement. In the field of Industrial Organization (IO), Bayesian approaches are utilized to dodge having to find the extremum of a badly behaved objective function. Indeed, the advantage of avoiding to compute complex integrals unlocks the path to solving problems that are once being difficult working with traditional methods. Researchers have provided Bayesian alternatives for classical IO problems. Gallant, Hong and Khwaja (2017) utilize Bayesian methods to estimate dynamic games of firms competing through investment in product quality which ultimately influences the products market shares and profitability, and a model that includes latent state variables that are serially correlated. For consumer inertia, Dubé, Hitsch and Rossi (2010) utilizes an MCMC algorithm to examine the inertia in the consumers when making brand choices. The famous dynamic binary choice problem in the literature illustrated by a bus engine replacement decision (Rust, 1987) is conducted under a different framework by Norets and Tang (2014) where they combine Bayesian inference with partial identification results.

To provide an alternative to Berry, Levinsohn and Pakes (1995), this paper follows a hybrid MCMC method proposed by Jiang, Manchanda and Rossi (2009) to estimate the random coefficient logit model that is widely used in demand estimation. Besides demonstrating the Bayesian method being a valid alternative to the classic GMM method, I additionally change the prior specification for the distribution of the aggregate demand shocks.

The rest of this paper is organized as follows. Section 2 describes the background for demand estimation. A summarized algorithm is provided in Section 3 and following with a detailed discussion on each algorithm. Section 4 verifies the MCMC algorithms by a Monte Carlo simulation and followed by the concluding remarks in Section 5.

2 The Demand Model

The random coefficient logit model defines the latent utility of consumer i purchasing product j in market t by

$$u_{ijt} = X_{jt}\beta_i + \xi_{jt} + \epsilon_{ijt}, \quad j = 1, ..., J \quad \text{and} \quad t = 1, ..., T. \tag{1}$$

where X_{jt} is a vector of product characteristics, ξ_{jt} is the aggregated demand shock that is common across individuals i, and ϵ_{ijt} is the idiosyncratic shock that follows a type-I extreme value distribution.

The aggregate shock is also known as the unobserved characteristics which capture the factors that relate to vertical product differentiation. Comparing to the Almost Ideal Demand System (Deaton and Muellbauer, 1980) which demand is measured based on the price and quantity of the J products including their competing commodities, economists suggest bundling the product's characteristics for consumers to choose from to resolve the curse of dimensionality problem that leads to infeasible estimation of dealing with J^2 elasticity parameters in the product space.

Turning to examine demand on the characteristics space, this permits the modeling of heterogeneous consumers as people are selecting the attributes instead of the product itself. For example, when projecting iPhone X, 11, and 12 onto the characteristic space, these three products are characterized by one variable, i.e. having two, three, and four camera sensors. Early examples of the characteristics space models include the horizontal model (Hotelling, 1929) and the vertical model Shaked and Sutton (1982).

To estimate the mean tastes of the consumers, I deviate from using the product attributes for X. Instead of directly estimate the preferences on different characteristics as demonstrated in the common estimation procedure, I use brand dummy variables to extract the taste coefficients of the brand fixed effects. Hence, $X \subset \mathbb{R}^K$ includes the brand dummies (one for each product) and the price.

2.1 Individual Heterogeneity

The random coefficients capture quality vertical differentiation. When assessing product characteristics, every consumer exerts various valuation. For instance, photography enthusiasts care about the versatility of the iPhone's camera system while the additional lenses are useless to a user who primarily uses the phone to browse the web.

By preference decomposition, the taste parameter β_i can be break up into two parts

$$\beta_i = \overline{\beta} + v_i \tag{2}$$

where $\overline{\beta}$ captures the average taste and v_i denotes every individual's preference deviation from the crowd. Combining the group's average taste with the aggregate demand shock, the mean utility can be defined as follows:

$$\delta_{jt} = X_{jt}\overline{\beta} + \xi_{jt}. \tag{3}$$

Combining Equation 2 and 3, Equation 1 can be rewritten as

$$\begin{aligned} u_{ijt} &= X_{jt}(\overline{\beta} + v_i) + \xi_{jt} + \epsilon_{ijt} \\ &= \delta_{jt} + X_{jt} v_i + \epsilon_{ijt} \end{aligned} \tag{4}$$

where the distribution of v_i can be illustrated by the demographics supplied by the econometricians or reasonable parametric assumption.

2.2 Market Share

The market share is the sales quantity of the product comparing to the sales performance of the entire market. When observing sales information that includes price and quantity, the product's true market share can be directly calculated from the data. However, market share can also be implied by the choice probabilities.

Following Equation 4, the probability for consumer i to choose product j to k in a fixed market is depicted by the following derivation:

$$\begin{aligned} p_{ij} &= p_{ij}(u_{ij} > u_{ik},\ \forall j \neq k) \\ &= p_{ij}(\delta_j + X_j v_i + \epsilon_{ij} \geq \delta_k + X_k v_i + \epsilon_{ik},\ \forall j \neq k) \\ &= p_{ij}\big[(\delta_j + X_j v_i) - (\delta_k + X_k v_i) \geq \epsilon_{ik} - \epsilon_{ij},\ \forall j \neq k\big] \\ &= \prod_{j \neq k} \big[(\delta_j + X_j v_i) - (\delta_k + X_k v_i) \geq \epsilon_{ik} - \epsilon_{ij}\big] \end{aligned}$$

The difference between two type-I extreme value shocks is distributed logistic

$$p_{ij} = \prod_{j \neq k} \frac{e^{(\delta_j + X_j v_i) - (\delta_k + X_k v_i)}}{1 + e^{(\delta_j + X_j v_i) - (\delta_k + X_k v_i)}} \tag{5}$$

then the probability of consumer i choosing product j in any fixed market takes the following closed form of multinomial logit:

$$p_{ij} = \frac{e^{\delta_j + X_j v_i}}{1 + \sum_{k=1}^{J} e^{\delta_k + X_k v_i}}. \tag{6}$$

The choice of outside option[2] when $k = 0$ is normalized to zero utility and the aggregation of individual choice probability across all consumers yields the product market share for

[2] The outside option is to purchase the alternative which is excluded from the specified product (characteristics) set.

product j in market t:

$$s_{jt} = \int \frac{e^{\delta_{jt}+X_{jt}v_i}}{1+\sum_{k=1}^{J} e^{\delta_{kt}+X_{kt}v_i}} \mathrm{d}G(v_i) \tag{7}$$

where G is the demographic distribution. Moreover, Equation 7 presents the relationship between the implied market share s_{jt} and the aggregate demand shock ξ_{jt} which is denoted as follows:

$$s_{jt} = h\big(\xi_{jt}\big|X_{jt}, \overline{\beta}, \Sigma\big) \quad \forall j = 1, ..., J. \tag{8}$$

3 The Sampler

To begin with, I start with a more in-depth discussion about the randomness in the random coefficient. In the demand literature, the random part of the consumer's preference v_i consists of two components. The demographics D_i reveals the observed consumer characteristics and some unobserved additional characteristics ν_i such as whether the consumer has a disabled family member when considering to buy a car or the consumer owns a dog which will affect their decision when buying a house. The preference formulation is presented as follows:

$$\beta_i = \overline{\beta} + \Pi D_i + \Sigma \nu_i \tag{9}$$

where D_i is a vector of demographic variables whose distribution needs to be further identified by the data, and ν_i follows a multivariate normal distribution. However, rather than bringing additional demographics information into the model, the individual Heterogeneity can be assumed normally distributed with valid reasons. For example, demographics represented by income distribution captures the preference discrepancy between wealthy and poor people.

Indeed, different income levels exhibit disparate valuations to product characteristics. However, when the scope of the market is considered far more broadly than a group of different zip code areas or a city, a particular type of demographic structure can no longer describe the average taste of the group. Hence, the distribution of the randomness in the individuals' preferences can be assumed following a normal distribution

$$\beta_i = \overline{\beta} + v_i \quad \text{s.t.} \quad v_i \sim N\big(0, \Sigma\big) \tag{10}$$

which can be specified as

$$\beta_i \sim N\big(\overline{\beta}, \Sigma\big)$$

and the covariance matrix Σ can be specified by the Cholesky factor

$$\Sigma = C'C \quad \text{s.t.} \quad C = \begin{bmatrix} e^{r_{11}} & r_{12} & \cdots & r_{1k} \\ 0 & e^{r_{22}} & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{r_{kk}}. \end{bmatrix} \tag{11}$$

where diagonal elements are exponentiated to enforce positive-definiteness. The prior distribution for r is considered as follows:

$$\begin{aligned} r_{jj} &\sim N\big(0, \sigma^2_{r_{jj}}\big) \quad \text{for} \;\; j = 1, ..., K \\ r_{jk} &\sim N\big(0, \sigma^2_{r_{\text{off}}}\big) \quad \text{for} \; j, k = 1, ..., K, \;\; j < k \end{aligned} \tag{12}$$

where $\sigma^2_{r_{jj}}$ and $\sigma^2_{r_{\text{off}}}$ are assumed with large variance and no correlation to avoid likely getting negative values (steps) when moving along the random walk chain.

Contrasting to the nested fixed-point algorithm, this step effectively lowers the non-convergence probability. During the Bayesian estimation in Algorithm 4, this reparameterization mitigates the issue of the possibility of a large proportion of draws in the algorithm are rejected simply due to the failure on positive definiteness criterion.

Algorithm 1: ROAD-MAP Bayesian estimation of random coefficients model

```
Specify the priors for β̄ (β̄₀, V_β̄), τ² (ω₀, κ₀), and r (elements in Σ; r_jj, r_jk)
Initiate δ₀, τ², R (tolerance), ℓ (random walk step size)
for iteration ← 1 to 10000 (large number) do
    Execute Algorithm 2 (Contraction)
    Execute Algorithm 3 (Gibbs)
    Execute Algorithm 4 (Metropolis-Hastings)
    Lists ← record β̄, δ_jt, Σ, τ² sequences
    Discard lists' items recorded during the burn-in period (first 3000 iterations)
    Compute the second dimensional mean of the lists to obtain taste coefficients,
      products mean utility, and individuals distribution's covariance matrix
end
return δ_jt, p_ijt, β̄, τ², Σ
```

Finally, as discussed in Jiang, Manchanda and Rossi (2009), comparing to the structural demand setup proposed by Berry (1994), one additional assumption is made to specify the likelihood. The aggregate demand shocks are independently distributed across all products j with identical variances

$$\xi_{jt} \sim N\big(0, \tau^2\big). \tag{13}$$

The road-map of estimating random coefficients model by Bayesian analysis is presented in Algorithm 1. Each step is going to be discussed in detail in the subsequent sections.

3.1 Estimate Mean Utility

The mean utility is specified by the product dummies Xs, the taste coefficients, and the aggregate demand shock. The product taste coefficients and therefore the mean utility δ are identified by minimizing the observed market share from the data and the shares implied by choice probability.

To compute the implied market share, the integral in Equation 7 is calculated by simulated integration. I simulate the market share by drawing from R individuals

$$s^R_{jt}\big(\delta_{jt}, \Sigma\big) = \frac{1}{R}\sum_{r=1}^{R} \frac{e^{\delta_{jt}+\sum_r \left(X_{jt} v_i^r\right)}}{1+\sum\limits_{k=1}^{J} e^{\delta_{kt}+\sum_r \left(X_{kt} v_i^r\right)}} \tag{14}$$

where v_i^r is drawn from $N(0,1)$. Recall from Equation 7 and 10, given Σ, s_{jt} is only a function of δ_{jt}. Matching the observed shares and the market shares implied by the data, Berry (1994) propose an inversion

$$\delta_{jt} = s^R_{jt}\big(s_{jt}, \Sigma\big)^{-1} \tag{15}$$

to recover the mean utility. The inevitability of s^R_{jt} is shown under general condition (Berry, Gandhi and Haile, 2013). This inversion is done numerically by iterating over

$$\delta^n_{jt} = \delta^{n-1}_{jt} + \log(s_{jt}) - \log\left(s^R_{jt}\right) \tag{16}$$

until $|\delta^n_{jt} - \delta^{n-1}_{jt}| < 10^{-14}$. As discussed in Dubé, Fox and Su (2012), a high tolerance can propagate and prevent convergence of the outer loop in the nested fixed-point algorithm. However, in my simulation, it turns out that the numerical error from the estimation of mean utility can also impact the outcome of Bayesian estimation.

The steps for the contraction process are summarized in Algorithm 2. The covariance matrix of the demographic distribution Σ is initialized with the diagonals multiplied by 100 or 200 to avoid a dogmatic prior. The mean utility and the consumer choice probability obtained from the contraction algorithm are passed to Algorithm 4 to estimate the entries in Equation 11 and the elements of the Jacobian matrix respectively. When the covariance matrix is updated through the Metropolis-Hastings sampling (Section 3.3), the re-run of Algorithm 2 depends on the accept-reject decision in Algorithm 4.

Algorithm 2: CONTRACTION: estimates the mean utility δ_{jt}

Input: s_{jt} (observed market shares), X (brand dummies and price), Σ (variance of demographic distribution), R (number of simulated individuals), M (number of markets), Tolerance

Output: δ_{jt} (mean utility), p_{ijt} (consumer choice probability)

$Tolerance \leftarrow 10^{-14}$
for $t \leftarrow 1$ ***to*** T **do**
    while *mean utility sequence difference* $>$ *Tolerance* **do**
        Compute $\delta_{jt} + X_{jt} v_i$
        Calculate consumer choice probability p_{ijt} by Equation 5
        Calculate the implied market share by Equation 7
        Numerically invert the market share by Equation 16 to obtain δ_{jt}^{n} and δ_{jt}^{n-1}
        mean utility sequence difference $\leftarrow |\delta_{jt}^{n} - \delta_{jt}^{n-1}|$
    end
end
return δ_{jt}, p_{ijt}

3.2 Gibbs Sampling

The average taste $\overline{\beta}$ and the covariance matrix of the aggregate demand shocks τ^2 is estimated by a Gibbs sampler by performing a Bayes regression analysis on Equation 3 with the conjugate priors specified as the following[3]:

$$\begin{aligned} \xi_{jt} &\sim N\left(0, \tau^2\right) \\ \overline{\beta} &\sim MVN\left(\overline{\beta_0}, V_{\overline{\beta}}\right) \\ \tau^2 &\sim IW\left(\omega_0, \kappa_0\right) \end{aligned} \tag{17}$$

which I deviate from the prior specifications assumed by Jiang, Manchanda and Rossi (2009). The sampling rounds for the average taste $\overline{\beta}$ and covariance of the aggregate demand shocks τ^2 are drawn from the conditional posterior distributions:

$$\pi\left(\overline{\beta}|\delta_{jt}, \tau^2\right) \sim N\left[\left(V_{\overline{\beta}}^{-1} + \frac{X_{jt}' X_{jt}}{\tau^2}\right)^{-1}\left(V_{\overline{\beta}}^{-1}\overline{\beta_0} + \frac{X_{jt}'\delta_{jt}}{\tau^2}\right), \left(V_{\overline{\beta}}^{-1} + \frac{X_{jt}' X_{jt}}{\tau^2}\right)^{-1}\right] \tag{18}$$

$$\pi\left(\tau^2|\delta_{jt}, \overline{\beta}\right) \sim IW\left[\omega_0 + n, \kappa_0 + \left(\delta_{jt} - X_{jt}'\overline{\beta}\right)'\left(\delta_{jt} - X_{jt}'\overline{\beta}\right)\right] \tag{19}$$

where n is the number of observations JT. The estimated $\overline{\beta}$ and τ^2 are passed on to Algorithm 4 for the accept-reject sampler.

[3]For computational purposes, inverse-Gamma distribution is substituted by inverse-Wishart distribution.

Algorithm 3: GIBBS SAMPLING: estimate $\overline{\beta}$ and τ^2

Input: δ_0 (initiate with mean utility from Algorithm 2), X (brand dummies and price), τ^2 (variance of N), $\overline{\beta_0}$ (mean of MVN), $V_{\overline{\beta}}$ (variance of MVN), ω_0 (mean of IW), κ_0 (variance of IW)

Output: $\overline{\beta}$ (average taste) and τ^2 (covariance of aggregate demand shock)

In the first loop, compute Equation 18 and 19 with parameters' initial value

Update Equation 18 with new τ^2 and draw $\overline{\beta}$

Update Equation 19 with new $\overline{\beta}$ and draw τ^2

return $\overline{\beta}$, τ^2

3.3 Metropolis-Hastings Sampling

In the last part, we calculate the covariance matrix of the demographic distribution presented in Equation 11. This describes the correlation between individuals. Comparing to Section 3.2, the full conditionals presented in Equation 24 is not analytically tractable. Therefore, the estimation of r employs a random walk Metropolis chain.

The joint density of shares at time t can be obtained by adopting the Change-of-Variable theorem by using the relationship between s_{jt} and ξ_{jt} specified in Equation 8:

$$\pi\left(s_{jt}|X_{jt},\overline{\beta},\Sigma,\tau^2\right)=\phi\Big(\underbrace{h^{-1}\left(s_{jt}|X_{jt},\overline{\beta},\Sigma\right)}_{\xi_{jt}}\Big|\tau^2\Big)\left(J_{s_{jt}\to\xi_{jt}}\right)^{-1}\quad\text{for } j=1,...,J \tag{20}$$

where following from Equation 3

$$\xi_{jt}=\delta_{jt}-X_{jt}\overline{\beta} \tag{21}$$

which is updated by passing in δ_{jt} from Algorithm 2 and $\overline{\beta}$ is passed in from Algorithm 3. The $J\times J$ Jacobian matrix at time t, $J_{s_{jt}\to\xi_{jt}}$, is given by

$$\left\|\begin{bmatrix}\frac{\partial s_{1t}}{\partial\xi_{1t}} & \frac{\partial s_{1t}}{\partial\xi_{2t}} & \cdots & \frac{\partial s_{1t}}{\partial\xi_{Jt}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial s_{Jt}}{\partial\xi_{1t}} & \frac{\partial s_{Jt}}{\partial\xi_{2t}} & \cdots & \frac{\partial s_{Jt}}{\partial\xi_{Jt}}\end{bmatrix}\right\| \tag{22}$$

where the elements are specified below:

$$\frac{\partial s_{jt}}{\partial\xi_{kt}}=\begin{cases}\displaystyle\int -p_{ijt}(1-p_{ikt})\phi\left(\theta^i|\overline{\theta},\Sigma\right)\mathrm{d}\theta^i, & \text{if } \ k=j\\ \displaystyle\int -p_{ijt}p_{ikt}\phi\left(\theta^i|\overline{\theta},\Sigma\right)\mathrm{d}\theta^i, & \text{if } \ k\neq j.\end{cases} \tag{23}$$

where p_{ijt} is the consumers' choice probabilities specified previously with fixed market t in

Equation 5. The elements r of the Σ matrix are updated by evaluating the following joint log-posterior of all parameters

$$\pi\left(\overline{\beta}, r, \tau^2 \middle| \{s_{jt}, X_{jt}\}_{t=1}^{T}\right) \tag{24}$$
$$\propto \underbrace{\prod_{t=1}^{T}\left(J^{-1}(s_{jt}, X_{jt}, r)\prod_{j=1}^{J} h^{-1}(s_{jt}|X_{jt}, \overline{\beta}, r)\right)}_{\text{part-I}} \times \underbrace{\prod_{j=1}^{K}\exp\left\{-\frac{r_{jj}^2}{2\sigma_{r_{jj}}^2}\right\}}_{\text{part-II}} \times \underbrace{\prod_{j=1}^{K-1}\prod_{k=j+1}^{K}\exp\left\{-\frac{r_{jk}^2}{2\sigma_{r_{\text{off}}}^2}\right\}}_{\text{part-III}}$$

where only the terms that relate to r are considered.

Algorithm 4: METROPOLIS-HASTINGS SAMPLING: estimate Σ

Input: δ_{jt} (mean utility obtained from Algorithm 2), p_{ijt} (consumer choice probability obtained from Algorithm 2), $\overline{\beta}$ and τ^2 (average taste and variance of N are obtained from Algorithm 3), X (brand dummies and price), Σ (variance of demographic distribution), R (number of simulated individuals), $\sigma^2_{r_{jj}}$ (variance of r_{jj}), $\sigma^2_{r_{\text{off}}}$ (variance of r_{jk})

Output: Σ (variance of demographic distribution)

In the first loop, compute Equation 24 with r_{ii}, r_{jk}, initial values
Compute ξ_{jt} by Equation 21
Compute the updated r by Equation 25

`// evaluate posterior densities with` r^{current} (D_c) `and` r^{update} (D_u)
for $t \leftarrow 1$ ***to*** T **do**
  Compute the Jacobian matrix by Equation 22 and 23
  $D \leftarrow$ Compute the part-I log-posterior ++
end
Compute $D \leftarrow D +$ part-II log-posterior + part-III log-posterior

`// accept-reject criteria`
if $\exp\{D_u - D_c\} \geq$ *draw from* $U[0, 1]$ **then**
  $\Sigma \leftarrow r^{\text{update}}$
  Do not execute Algorithm 2 in next loop
end
return Σ, *Algorithm 2 skipping decision*

Follow from Equation 20, $h^{-1}(\cdot)$ in the above part-I is drawn from $MVN(0, \tau^2)$. The multivariate normal pdf is evaluated at ξ_{jt} which is calculated by Equation 21 for a particular market t.

The proposal distribution is the multivariate normal distribution with a hyperparameter "step size" (ℓ) multiplied by the covariance matrix and τ^2 is passed from Algorithm 3. The updating of the elements r follows the equation

$$r^{\text{update}} = r^{\text{current}} + MVN\big(0, I_K \cdot \ell\big). \tag{25}$$

For part-II and part-III[4] in Equation 24, the two densities are evaluated from the multivariate normal distribution assumed in Equation 12. The acceptance decision is decided by comparing a number η drawn from $U[0,1]$ and the probability

$$\alpha = \min\left\{1, \frac{\pi\left(\overline{\beta}, r^{\text{update}}, \tau^2 \middle| \{s_{jt}, X_{jt}\}_{t=1}^T\right)}{\pi\left(\overline{\beta}, r^{\text{current}}, \tau^2 \middle| \{s_{jt}, X_{jt}\}_{t=1}^T\right)}\right\}.$$

The elements in Σ is updated if $\alpha \geq \eta$. However, if the decision results in a rejection, then r^{current} is used to evaluate the joint posterior and at the same time, Algorithm 2 is skipped in the next loop.

4 Simulation

In this section, I demonstrate a Monte Carlo simulation of this hybrid Bayesian approach. Panel data is often used to study questions related to the filed of Industrial Organization. For the considered product set, the demand for each brand is observed across different areas (e.g. census regions, states, cities) and across different time periods (e.g. across 10 years). For each particular time-region states, they are defined as different markets.

The data generating process involves three products and covers 300 markets. An example for this scenario is the demand for iPad mini, iPad Air and iPad Pro of people in the fifty states from 2014 to 2020. For each market, the demand collects the behavior of 10000 consumers.

The average preferences of these three iPad models and their prices are [0.1, 0.5, 0.9, -0.2]. The random utility specified in Equation 10 is assumed without any correlation and a variance of 0.3. The variance of the aggregate demand shock in Equation 13 is 0.2.

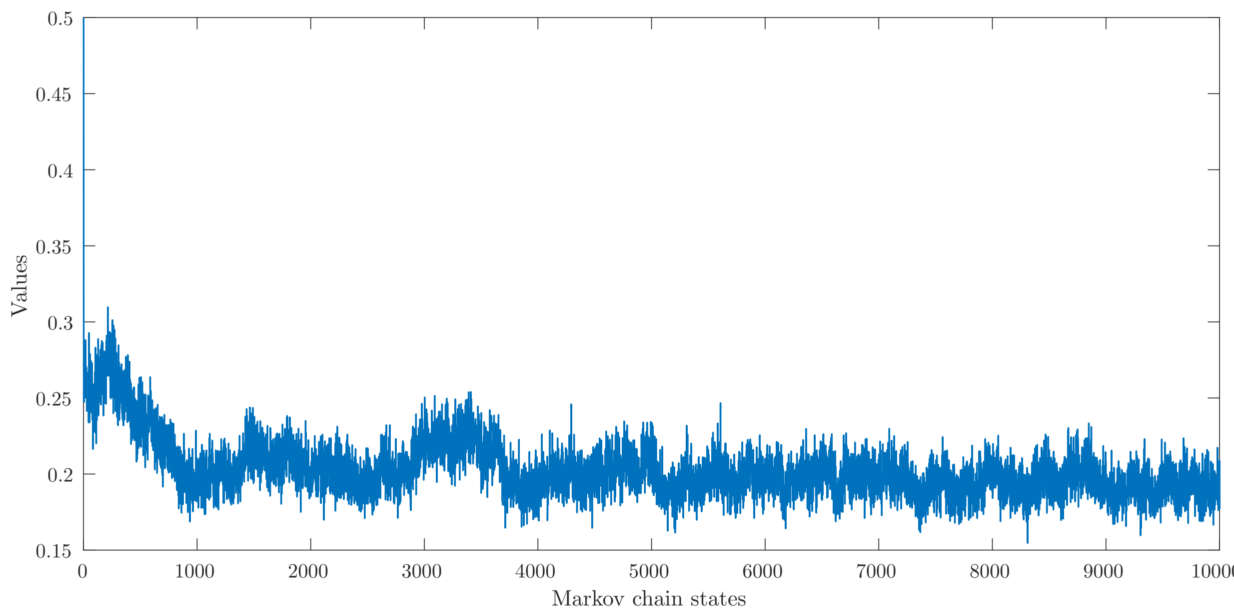
To initiate Algorithm 1, the mean utilities δ_0 for each observations across all states and periods are set to their own observed market shares observed from the data. The aggregate shock γ^2 is set to 0.5. The brand dummies and price's coefficients are set to 0. For the hyper parameters, there are 1,000 simulated individuals (R). Following the literature, the contraction tolerance R is set to 10^{-14} and the random walk step size ℓ is set to 0.002.

There are two popular specifications for the random taste covariance matrix σ: the unre-

[4] The joint posterior only includes the kernels of the prior distributions specified in Equation 12.

stricted covariance matrix and a restricted diagonal matrix. The restricted matrix benefits the estimation due to sparsity which greatly improves the run-time. However, not knowing the generating process, I will use an unrestricted covariance matrix in this estimation.

Figure 1: The τ^2 updating process in the Markov chain



Note: The variance of the aggregate demand shock is updated by Algorithm 3. The true value is 0.2 and the parameter is initiated to 0.5.

As presented in Equation 12, both $\sigma_{r^2_{jj}}$ and $\sigma_{r^2_{\text{off}}}$ are initialized with an identity matrix and multiplied by 100 to set up as a weak prior. The Bayesian updating process goes through 10,000 states and the burn-in period is set to 5,000.

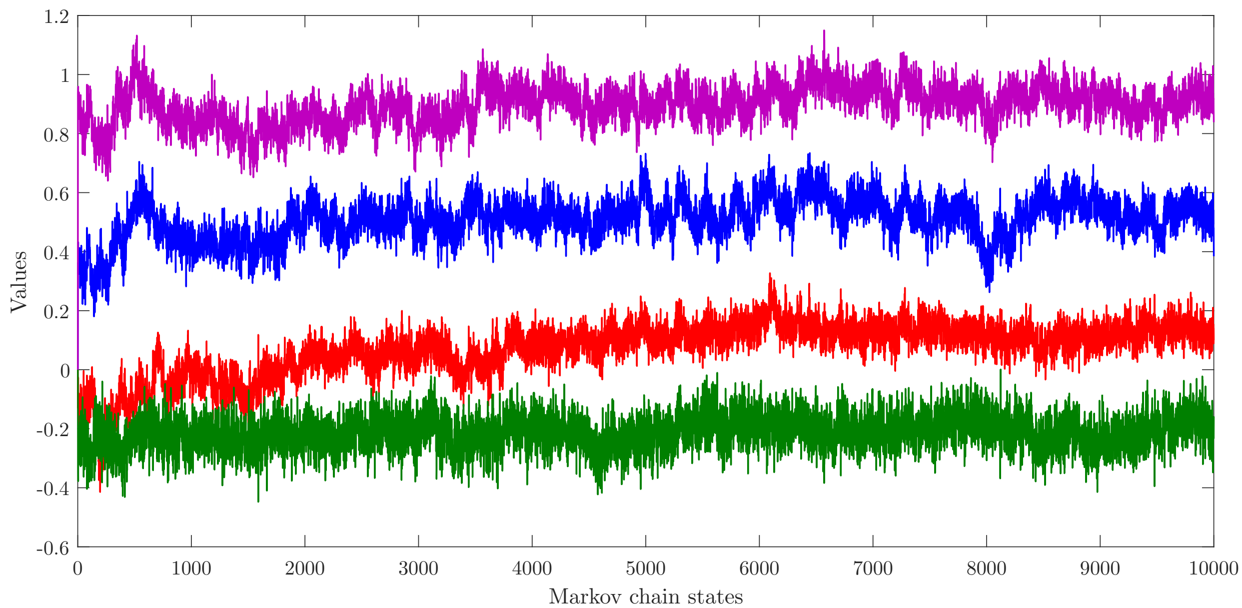
For Algorithm 2, the simulated integral is averaged across 1,000 draws. The updating process of τ is illustrated in Figure 1. Starting from the initial value 0.5, the chain converges toward 0.2 in 1,000 cycles. After discarding the burn-in values, the average of the rest of the states is 0.1948.

Next, the average product taste is estimated by Algorithm 3. Starting from the initial values of $[0, 0, 0, 0]$, the chain moves close to the true values also at around 1,000 iterations.

Presented in Figure 2, the red, blue, and purple lines corresponds to the average taste for products one, two, and three. The green sequence depicted the average taste on the products' price. The updating process has 10,000 iterations. After the burn-in, the mean preference values are $[0.1279, 0.5346, 0.9209, -0.2006]$.

It is worth mentioning that in the general settings, the aggregate shock is endogenous. The counter price endogeneity, a simultaneous equation estimation can be added to Algorithm 3 where a vector of instruments $\widetilde{X_{jt}}$ will be added into the sampler.

Figure 2: The estimation of $\overline{\beta}$



Note: The updating process of $\overline{\beta}$ is responsible by Algorithm 3. The true average preference for the three products and their prices (green line) is $[0.1, 0.5, 0.9, -0.2]$ and the chain starts updating from $[0, 0, 0, 0]$ for 10,000 iterations.

Finally, I also report the estimated covariance matrix of the demographic distribution in Equation 26. Σ is estimated by Algorithm 4 and used as inputs in Algorithm 2.

$$\Sigma = \begin{bmatrix} 0.0147 & 0.0271 & 0.0538 & -0.0053 \\ 0.0271 & 0.1709 & 0.1273 & -0.1023 \\ 0.0538 & 0.1273 & 0.6328 & 0.1914 \\ -0.0050 & -0.1023 & 0.1914 & 0.5547 \end{bmatrix} \tag{26}$$

The estimation uses an unrestricted matrix with all the diagonal elements being positive values which is an advantage of the reparametization illustrated in Equation 11.

5 Conclusion

This article demonstrates demand estimation with a Bayesian approach and provides a practitioner's guide regarding the details when implementing the algorithm. Under the Bayesian framework with modified aggregate demand shock prior, the random coefficients can be estimated with two MCMC methods. Comparing to the nested fixed-point algorithm, this approach avoids the dreadful non-convergence predicament. Furthermore, without optimizing two nested loops, the Bayesian approach considerably appears to be more efficient.

References

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. "Connected Substitutes and Invertibility of Demand." *Econometrica*, 81(5): 2087–2111.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, 25(2): 242–262.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Deaton, Angus, and John Muellbauer. 1980. "An Almost Ideal Demand System." *Competition policy and antitrust*, 70(3): 312–326.

Dubé, Jean-Pierre, Günter J. Hitsch, and Peter E. Rossi. 2010. "State dependence and alternative explanations for consumer inertia." *RAND Journal of Economics*, 41(53): 417–445.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su. 2012. "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation." *Econometrica*, 80(5): 2231–2267.

Gallant, A. Ronald, Han Hong, and Ahmed Khwaja. 2017. "A Bayesian Approach to Estimation of Dynamic Models with Small and Large Number of Heterogeneous Players and Latent Serially Correlated States." *Working paper*.

Hendricks, Kenneth, Robert H. Porter, and Charles A. Wilson. 1994. "Relaxing Price Competition Through Product Differentiation." *Econometrica*, 62(6): 1415–1444.

Hotelling, Harold. 1929. "Stability in Competition." *Economic Journal*, 39(153): 41–57.

Jiang, Renna, Puneet Manchanda, and Peter E. Rossi. 2009. "Bayesian analysis of random coefficient logit models using aggregate data." *Journal of Econometrics*, 149(2): 136–148.

Norets, A., and X. Tang. 2014. "Semiparametric Inference in Dynamic Binary Choice Models." *Review of Economic Studies*, 81(3): 1229–1262.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 55(5): 999–1033.

Shaked, Avner, and John Sutton. 1982. "Relaxing Price Competition Through Product Differentiation." *Review of Economic Studies*, 49(1): 3–13.

Su, Che-Lin, and Kenneth L. Judd. 2012. "Constrained Optimization Approaches to Estimation of Structural Models." *Econometrica*, 80(5): 2213–2230.