Bayesian Estimation of Fixed Effects Models with Large Datasets

Hang Qian*
The MathWorks, Inc.

August 2, 2024

Abstract

In hierarchical prior longitudinal models, random effects are estimated by the Gibbs sampler. We show that fixed effects can be handled by a similar Gibbs sampler under a diffuse prior on the unobserved heterogeneity. The dummy variable approach for fixed effects is computationally intensive and has the out-of-memory risk, while the Gibbs sampler can reproduce the dummy variable estimator without creating dummy variables, and therefore avoids the memory burden. Compared to alternating projections and other classical approaches, our method simplifies both inference and estimation of the limited dependent variable models with fixed effects. The proposed method is applied to a real-world mortgage dataset for classification with three-way fixed effects on banks, regions, and loan purposes.

Keywords: Panel data, Unobserved effects, Big data, Markov Chain Monte Carlo
MSC subject classifications: 62F10, 62F15, 65Y04

*55 Centre Street, Natick, MA, U.S., 01760. Email: matlabist@gmail.com. The author would like to thank the editor and two anonymous reviewers for the comments and insights that greatly improve the paper.

1 Introduction

Random and fixed effects are essential to econometric analysis of panel data. The model of random effects assumes that the unobserved heterogeneity is a random variable uncorrelated with regressors, while fixed effects allow for arbitrary dependence between unobserved effects and observed regressors (Wooldridge, 2010, p.286). In classical econometrics, random effects are assimilated into error terms and estimated by the Generalized Least Squares (GLS) methods, while fixed effects are treated as parameters and estimated by the Least Squares Dummy Variable (LSDV) regressions. In Bayesian statistics, random effects are studied in the hierarchical (multilevel) models. The Gibbs samplers of Gelfand and Smith (1990) and Chib and Carlin (1999) address random effects estimation.

Our research is motivated by a natural question: what is the Bayesian counterpart of fixed effects estimation by dummy variables? We show that the Gibbs sampler of Gelfand and Smith (1990) can be adjusted to estimate fixed effects under a diffuse prior on the unobserved heterogeneity. Compared to the classical LSDV estimator for fixed effects, the advantage of Bayesian estimation is scalability to multi-way and high-dimensional fixed effects models, where it is not feasible to create multitudinous dummy variables in the computer memory, which is a common issue in the labor economics literature (e.g., Abowd et al. (1999), Carneiro et al. (2012), Mittag (2019), among others).

Least squares estimation is closely related to linear projections. Feasible dummy variable estimators using alternating projections are proposed by Guimaraes and Portugal (2010), Gaure (2013), Somaini and Wolak (2016), Correia et al. (2020) and others. Compared to alternating projections and other iterative algorithms, our method has three advantages. First, as a full Bayesian approach, the Gibbs sampler simplifies inference by incorporating parameter uncertainty into posterior samples, which provide a Bayesian counterpart of the LSDV point estimator and the standard error. Second, our method handles random and fixed effects in the unified framework under suitable prior specifications. The Gibbs sampler never explicitly creates dummy variables and is ideal for in-memory computing. Third, the advantage of our approach is more prominent in the limited dependent variable models, in which computing the classical standard errors by the bootstrap with resampling data can be complex and computationally expensive (Guimaraes and Portugal, 2010). In contrast, data augmentation makes the Gibbs sampler a powerful tool for posterior simulation of random and fixed effects. Parameter uncertainty is characterized by posterior samples, which simplify inference both conceptually and computationally.

The remainder of the paper is organized as follows. Section 2 proposes an efficient Gibbs sampling method for estimating fixed effects. Section 3 extends fixed effects in linear regressions to the limited dependent variable models. Section 4 is devoted to a Monte Carlo exercise that compares the Gibbs sampler to non-stochastic iterative algorithms. Section 5 provides an application on a real-world mortgage dataset. Section 6 concludes the paper.

2 From Random to Fixed Effects

Consider the multi-way random effects regression:

$$Y_{ijt} = \alpha_i + \gamma_j + \delta_t + x_{ijt}\beta + \sigma\varepsilon_{ijt}, \tag{1}$$

where the unobserved effects α_i, γ_j, δ_t follow normal distributions $N\left(0, \Lambda_\alpha^{-1}\right)$, $N\left(0, \Lambda_\gamma^{-1}\right)$ and $N\left(0, \Lambda_\delta^{-1}\right)$, respectively. The parameters $\Lambda_\alpha, \Lambda_\gamma, \Lambda_\delta$ are scalars. The unobserved effects are uncorrelated with the regressors x_{ijt}. Equation (1) shows three-way random effects over the dimensions $i = 1, \ldots, n$, $j = 1, \ldots, J$ and $t = 1, \ldots, T$, and our method can be extended to higher dimensions and unbalanced panel data. Denote the random vectors as $\alpha = (\alpha_1, \ldots, \alpha_n)'$, $\gamma = (\gamma_1, \ldots, \gamma_J)'$ and $\delta = (\delta_1, \ldots, \delta_T)'$. Also, we specify a normal prior $N\left(0, \Lambda_\beta^{-1}\right)$ for β, and an inverse gamma prior $IG\left(a, b\right)$ for σ^2.

The Gibbs sampler of Gelfand and Smith (1990) cycles through the full posterior conditional distributions of $\alpha, \beta, \gamma, \delta, \sigma^2$ by a sequence of Bayesian linear regressions. First, to update β and σ^2 conditional on α, γ, δ, we use a pooled regression with the response variable $Y_{ijt} - \alpha_i - \gamma_j - \delta_t$ and regressors x_{ijt}. Second, to update α conditional on $\beta, \gamma, \delta, \sigma^2$, we run n separate (parallel) linear regressions, in which the response variable is $Y_{ijt} - \gamma_j - \delta_t - x_{ijt}\beta$ and the only regressor is an intercept (i.e., a column of ones). Third, the unobserved effects γ are estimated by J intercept-only regressions with the response variable $Y_{ijt} - \alpha_i - \delta_t - x_{ijt}\beta$. Fourth, the unobserved effects δ are updated by T intercept-only regressions with the response variable $Y_{ijt} - \alpha_i - \gamma_j - x_{ijt}\beta$. In random effects models, the variance terms Λ_α^{-1}, Λ_γ^{-1}, Λ_δ^{-1} are usually specified in a hierarchical structure with the conjugate inverse gamma priors, and the Gibbs sampler also updates the variance components.

Chib and Carlin (1999) improve the Gibbs sampler of Gelfand and Smith (1990) by marginalization over random effects, so that $\alpha, \beta, \gamma, \delta$ can be updated simultaneously in the same block. Marginalization is achieved by combining random effects α, γ, δ with the disturbance term ε_{ijt} (as in the error components model). The combined disturbances are correlated, with a non-diagonal covariance matrix. Sampling β from the marginalized posterior is analogous to the classical GLS regression for the random effects model.

The key assumption of random effects is that the random variables $\alpha_i, \gamma_j, \delta_t$ are uncorrelated with x_{ijt}, which implies that observing x_{ijt} does not change the distributions $N\left(0, \Lambda_k^{-1}\right)$, $k \in \{\alpha, \gamma, \delta\}$. Fixed effects relax the assumption by allowing for arbitrary dependence between unobserved effects and observed regressors. To clarify, in econometric parlance, fixed effects do not refer to the parameter β. Instead, fixed effects refer to the vectors α, γ, δ as unknown parameters associated with dummy variable matrices $D_\alpha, D_\gamma, D_\delta$ in the vector/matrix form of Equation (1):

$$Y = D_\alpha \alpha + D_\gamma \gamma + D_\delta \delta + x\beta + \sigma\varepsilon, \tag{2}$$

where the data Y and x are formulated by stacking observations along dimensions i, j and t. Denote $X = (D_\alpha, D_\gamma, D_\delta, x)$ and $\theta = (\alpha', \gamma', \delta', \beta')'$. The classical fixed effects estimator $\hat{\theta}$ is defined by the ordinary least squares with dummy variables (i.e., the LSDV estimator):

$$\hat{\theta} = (X'X)^{-1} X'Y. \tag{3}$$

However, the computational challenge is that it is not feasible to compute $\hat{\theta}$ directly by Equation (3), as the matrix X can be too large to be loaded into the computer memory, let alone the computationally intensive matrix multiplication $X'X$ and inversion. It is possible to sweep out one-way fixed effects by within transformation. Alternating projections are feasible and work well for estimating multi-way fixed effects models, but inference remains a challenging task.

The Gibbs sampler is a form of the Markov Chain Monte Carlo (MCMC) simulation that has the target distribution as the invariant distribution. Our aim is to reproduce the classical LSDV estimator in the Bayesian framework, by setting $\hat{\theta}$ as the mean of the target distribution of the Gibbs sampler.

We adjust the random effects Gibbs sampler of Gelfand and Smith (1990) in the following ways: the sampling procedures for updating $\alpha, \beta, \gamma, \delta, \sigma^2$ remain unchanged, but $\Lambda_\alpha, \Lambda_\gamma, \Lambda_\delta$ are specified by researchers instead of estimated by data. The target distribution of the Gibbs sampler is given by

$$p\left(\alpha, \beta, \gamma, \delta, \sigma^2 \left| Y \right.\right) \propto \left(\sigma^2\right)^{-\left(a+\frac{nJT}{2}+1\right)} e^{-\sigma^{-2}b-\frac{1}{2}\sigma^{-2}(Y-X\theta)'(Y-X\theta)-\frac{1}{2}\theta'\Lambda\theta}, \tag{4}$$

where $\Lambda = diag\left(\Lambda_\alpha I_n, \Lambda_\gamma I_J, \Lambda_\delta I_T, \Lambda_\beta\right)$, and I_n, I_j, I_T are identity matrices.

Proposition 1 shows that the random effects model reduces to the fixed effects model if we specify a diffuse prior: $\Lambda = 0$. The implication is that the Gibbs sampler can reproduce the LSDV estimator without constructing large dummy variables. The posterior means of $\alpha, \beta, \gamma, \delta$ computed from the Gibbs samples correspond to the LSDV results. The posterior standard deviation is a Bayesian counterpart of the classical standard error, as Proposition 1 indicates that $Var\left(\theta \left| Y, \sigma^2 \right.\right) = \sigma^2 \left(X'X\right)^{-1}$. Our intuition for the diffuse prior is that arbitrary dependence between fixed effects and regressors induces a noninformative distribution of unobserved effects conditional on observed regressors. Proposition 1 also suggests that specifying a non-zero Λ in the prior leads to a regularized fixed effects estimator as in the ridge regression.

Proposition 1 *The posterior conditional distribution $p\left(\theta \left| Y, \sigma^2 \right.\right)$ obtained from Equation (4) is normal $N\left(\bar{\mu}, \bar{\Lambda}^{-1}\right)$, where*

$$\bar{\mu} = \left(\Lambda + \sigma^{-2}X'X\right)^{-1}\left(\sigma^{-2}X'Y\right),$$

$$\bar{\Lambda} = \Lambda + \sigma^{-2}X'X.$$

Under a diffuse prior ($\Lambda = 0$), the posterior mean of θ is identical to the LSDV estimator:

$$E\left(\theta \left| Y \right.\right) = \hat{\theta}.$$

Proof: By completing the squares with respect to θ in Equation (4), we obtain the posterior conditional distribution:

$$p\left(\theta \left| Y, \sigma^2 \right.\right) \propto e^{-\frac{1}{2}(\theta-\bar{\mu})'\bar{\Lambda}(\theta-\bar{\mu})},$$

which is the normal distribution $N\left(\bar{\mu}, \bar{\Lambda}^{-1}\right)$. If we specify a diffuse prior by taking $\Lambda = 0$, the posterior conditional mean matches the LSDV estimator:

$$E\left(\theta \left| Y, \sigma^2\right.\right) = \left(X'X\right)^{-1} X'Y.$$

Note that the posterior conditional mean does not depend on σ^2, so the posterior (unconditional) mean is still the LSDV estimator. ■

The Gibbs sampler allows estimation of fixed and random effects in a unified framework, where unobserved effects differ in the prior specification: fixed effects have a diffuse prior $N\left(0, \Lambda_k^{-1}\right)$ with $\Lambda_k = 0$, $k \in \{\alpha, \gamma, \delta\}$, while random effects have a hierarchical prior (Λ_k follows a gamma distribution, that is, Λ_k^{-1} follows an inverse gamma distribution):

$$p\left(\Lambda_k\right) \propto \Lambda_k^{a_k - 1} e^{-b_k \Lambda_k},$$

and the posterior conditional distribution takes a conjugate form:

$$p\left(\Lambda_k \left| Y, \theta\right.\right) \propto \Lambda_k^{\bar{a}_k - 1} e^{-\bar{b}_k \Lambda_k},$$

where $\bar{a}_k$ and $\bar{b}_k$ are given by

$$\bar{a}_\alpha = a_\alpha + \frac{n}{2}, \bar{a}_\gamma = a_\gamma + \frac{J}{2}, \bar{a}_\delta = a_\delta + \frac{T}{2},$$
$$\bar{b}_\alpha = b_\alpha + \frac{\alpha'\alpha}{2}, \bar{b}_\gamma = b_\gamma + \frac{\gamma'\gamma}{2}, \bar{b}_\delta = b_\delta + \frac{\delta'\delta}{2}.$$

Conditional on Λ_k, there is no difference between fixed and random effects in terms of updating model parameters.

Regardless of estimating fixed or random effects, the Gibbs sampler never constructs dummy variable matrices and avoids the out-of-memory risk, as it sequentially updates the scalars $\alpha_i, \gamma_j, \delta_t$, for each $i = 1, \ldots, n$, $j = 1, \ldots, J$ and $t = 1, \ldots, T$. The computational and memory efficiency makes the Gibbs sampler scalable to big data applications.

The Gibbs sampler simplifies inference on both the slope coefficients β and (functions of) the fixed effects α, γ, δ. The posterior distribution of a function like $f\left(\alpha, \beta, \gamma, \delta\right)$ can be approximated by plugging each posterior draw into the function.

3 Extension to Limited Dependent Variable Models

The advantage of Bayesian computing is more prominent in the limited dependent variable models, in which data augmentation makes the Gibbs sampler a powerful tool for posterior simulation with random and fixed effects. The Probit and Tobit are popular models for Bayesian computing by the Gibbs sampler of Chib (1992) and Albert and Chib (1993).

We extend Equation (2) to the Probit model with a binary response variable:

$$Y^* = D_\alpha \alpha + D_\gamma \gamma + D_\delta \delta + x\beta + \sigma\varepsilon, \tag{5}$$

$$Y = I\left(Y^* > 0\right),$$

where Y^* is a vector of latent variables, and observations are coded as one if $Y^* > 0$, and zero otherwise. We normalize $\sigma = 1$ for parameter identification.

From a classical viewpoint, the fixed effects Probit model is fraught with computational and statistical difficulties due to an incidental parameters problem that causes estimation inconsistency (Greene, 2004). Whether an estimator is consistent depends on which index tends to infinity. If we consider $n \to \infty$ and a finite sample to estimate each α_i, $i = 1, \ldots, n$, then inconsistency of α_i estimate causes inconsistency of β estimate in a nonlinear model. However, the incidental parameters problem can be avoided if the sample size increases without changing the number of unknown parameters. The mortgage application in Section 5 illustrates that point. For each bank, state, and loan purpose, we have hundreds of loan records (e.g., many home buyers obtain mortgages from a large bank in a state), and the number of loans grows over time. All parameters can be consistently estimated, as the number of parameters remains unchanged with the increasing sample size.

Bayesian inference focuses on the posterior distribution of parameters and latent variables conditional on available observations (i.e., data are fixed). The Gibbs sampler of Albert and Chib (1993) simulates latent variables from truncated normal distributions, and the Gibbs sampler of Gelfand and Smith (1990) cycles through $\alpha, \beta, \gamma, \delta$. Under a diffuse prior discussed in Section 2, we combine two samplers for estimating the Probit model with fixed effects. See Section 5 for an application.

As Guimaraes and Portugal (2010, p.640) commented: "[w]ith nonlinear models, it is possible to estimate correctly the vector β, but there is no easy solution for estimation of the associated standard errors. While it would be possible to bootstrap the standard errors, this solution is likely to be computationally very expensive." The difficulty of classical inference can be avoided by the Gibbs sampler, which generates posterior draws that facilitate Bayesian inference.

For linear models, the fixed effects LSDV estimator $\hat{\theta}$ in Equation (3) is also the maximum likelihood estimator (MLE). For limited dependent variable models, the classical fixed effects estimator $\hat{\theta}$ is defined by the MLE with dummy variables:

$$\hat{\theta} = \arg\max_{\theta} \ln p\left(Y \left| \theta \right.\right), \tag{6}$$

and the log likelihood of the Probit model in Equation (5) takes the form

$$\ln p\left(Y \left| \theta \right.\right) = \sum \ln \Phi\left[\left(2Y - 1\right) X\theta\right],$$

where $X = (D_\alpha, D_\gamma, D_\delta, x)$ and $\theta = (\alpha', \gamma', \delta', \beta')'$. The summation is taken over all elements in the likelihood vector by the normal c.d.f. Φ.

Bayesian estimation of random effects is implemented by the Gibbs sampler with data augmentation. As an extension of Proposition 1 to the limited dependent variable models, Proposition 2 indicates that the random effects estimator reduces to the fixed effects estimator if we specify a diffuse prior: $\Lambda = 0$. The posterior mode/mean matches the MLE, and the posterior standard deviation provides a Bayesian counterpart of the MLE asymptotic standard error.

Proposition 2 *Under a diffuse prior ($\Lambda = 0$), the posterior mode of θ is identical to the MLE $\hat{\theta}$ in Equation (6), and the posterior covariance matrix $Var\left(\theta \left| Y \right.\right)$ tends to the inverse Hessian of the log likelihood (as the MLE asymptotic covariance matrix).*

Proof: Bayes' rule indicates that

$$p\left(\theta \left| Y \right.\right) = \frac{p\left(\theta\right) p\left(Y \left| \theta \right.\right)}{p\left(Y\right)}.$$

Under a diffuse prior $p(\theta) \propto 1$, the maximizer of the posterior density (i.e., the posterior mode) is identical to the maximizer of the likelihood function (i.e., the MLE):

$$\hat{\theta} = \arg\max_{\theta} \ln p\left(\theta \left| Y \right.\right).$$

Consider the second-order Taylor expansion of the log likelihood around $\hat{\theta}$:

$$p(\theta|Y) = \exp\left[\ln p\left(Y\left|\hat{\theta}\right.\right) + \frac{1}{2}\left(\theta-\hat{\theta}\right)'\frac{\partial \ln p(Y|\theta_b)}{\partial\theta_b\partial\theta_b'}\left(\theta-\hat{\theta}\right)\right]/p(Y),$$

where the first-order term equals zero because of $\frac{\partial \ln p\left(Y|\hat{\theta}\right)}{\partial\hat{\theta}} = 0$, and θ_b is between θ and $\hat{\theta}$. As the sample size increases, $\hat{\theta}$ converges to the true parameters denoted as θ_0, and $p(\theta|Y)$ collapses to a point probability mass on θ_0. It follows that θ_b converges to θ_0. Asymptotically, we have

$$p(\theta|Y) \propto \exp\left[\frac{1}{2}\left(\theta-\hat{\theta}\right)'\frac{\partial \ln p(Y|\theta_0)}{\partial\theta_0\partial\theta_0'}\left(\theta-\hat{\theta}\right)\right],$$

which indicates that the asymptotic posterior mean $E(\theta|Y)$ equals the MLE $\hat{\theta}$, and the asymptotic posterior covariance matrix $Var(\theta|Y)$ is the inverse Hessian of the log likelihood, namely $\left[-\frac{\partial \ln p(Y|\theta_0)}{\partial\theta_0\partial\theta_0'}\right]^{-1}$. ■

4 Monte Carlo Study

The commonalities and differences between the Gibbs sampler and alternative methods for estimating multi-way fixed effects models can be visualized in a Monte Carlo study. The data generating process is Equation (1), in which the fixed effects $\alpha_i, \gamma_j, \delta_t$ are random samples from standard normal distributions, and three covariates in x_{ijt} are generated by adding standard normal variates to $\alpha_i, \gamma_j, \delta_t$, so that covariates are correlated with fixed effects. To benchmark against alternative estimators, it is desirable to compute the ground-truth LSDV estimator directly by Equation (3), so we consider a moderate sample size of 10000 ($n = 100$, $T = 10$ and $J = 10$) in this simulation study.

Table 1 shows estimation results of five methods: 1) the ground-truth LSDV, 2) the alternating projections method of Gaure (2013), 3) the zigzag Gauss-Seidel iterative algorithm of Guimaraes and Portugal (2010), 4) our Gibbs sampler for fixed effects, and 5) the Gibbs sampler of Gelfand and Smith (1990) for random effects. Proposition 1 indicates that our Gibbs sampler under the diffuse prior collapses to the Guimaraes and Portugal (2010) algorithm if $\sigma^2 \to 0$. In practice, specifying σ^2 as small as the machine epsilon (about 2×10^{-16}) can replicate non-stochastic iterations using our Gibbs sampler program.

Table 1 indicates that the random effects estimator is inconsistent, while other estimators are sufficiently close to the ground-truth LSDV estimator.

Figure 1 highlights different behaviors of the Gibbs sampler and non-stochastic iterative algorithms. The estimators of Gaure (2013) and Guimaraes and Portugal (2010) monotonically converge to the ground truth, and the convergence speed of alternating projections is superior. In contrast, the convergence of the Gibbs sampler refers to approximating the target distribution specified in Equation (4), but the Gibbs samples per se are volatile. Such volatility provides valuable information on parameter uncertainty: the sample standard deviation is 0.0323, which provides Bayesian estimation of the classical LSDV standard error 0.0322. Also, the sample mean 0.9602 is a Bayesian counterpart of the classical LSDV point estimator 0.9607. Figure 1 indicates that the Gibbs sampler has burn-in periods before reaching the stationary distribution of the Markov chain. The key point of the paper stems from the fact that letting the Gibbs sampler run on provides the variation we need for a Bayesian counterpart of the classical standard errors.

	LSDV	AP	Zigzag	Gibbs	Random
β_1	1.0430	1.0430	1.0430	1.0427	1.1635
	(0.0326)	(0.0326)		(0.0326)	(0.0352)
β_2	0.9385	0.9385	0.9385	0.9386	0.9498
	(0.0326)	(0.0326)		(0.0325)	(0.0325)
β_3	0.9607	0.9607	0.9607	0.9602	0.9716
	(0.0322)	(0.0322)		(0.0323)	(0.0324)

Table 1: Parameter estimation by LSDV, alternating projections (AP) of Gaure (2013), Zigzag iterative algorithm of Guimaraes and Portugal (2010), Gibbs samplers for fixed and random effects. Standard errors/deviations are in parentheses.

5 An Application

Big data come from various sources, one of which is large administrative data. Under the U.S. Home Mortgage Disclosure Act (HMDA), financial institutions must disclose loan-

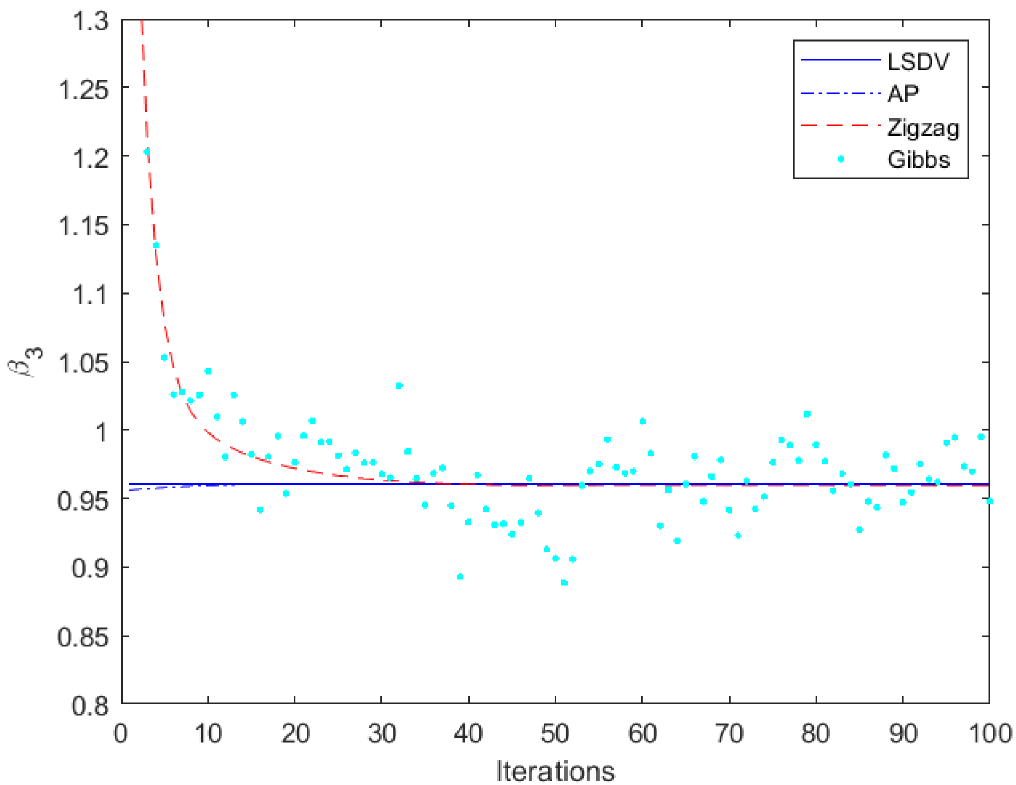

Figure 1: Convergence modes in stochastic and non-stochastic iterative algorithms. Estimators/draws for β_3 are plotted at each iteration of alternating projections of Gaure (2013), the zigzag algorithm of Guimaraes and Portugal (2010) and the Gibbs sampler for fixed effects.

level data to the public. Millions of loan records have been added to the HMDA database, in which hundreds of variables on the characteristics of loans, properties, applicants, and lenders are publicly available. In this application, we consider loan actions (i.e., loan originated or denied). The descriptive statistics of the data in the year 2018 indicate that 51% of loans were originated, 17% were denied, and the remaining were withdrawn or purchased on the secondary market. The loan purposes included home purchase (51%), home improvement (8%), refinancing (15%), cash-out refinancing (17%) and others.

We consider the Probit model in Equation (5), where loans are originated if $Y^* > 0$, and denied otherwise. The regressors x include debt-to-income ratio, loan-to-value ratio, loan size, indicators for the fixed rate and subordinate lien, the number of units, applicant age, gender, and joint filing status. The model has three-way fixed effects: 500 largest banks, 50 U.S. states and 5 loan purposes. The sample size is about 7.5 million, which implies that it would cost over 30GB memory space if we created a dummy variable matrix.

The model is fitted by combining the Gibbs sampler discussed in Section 2 with that of Albert and Chib (1993). The summary statistics of the posterior distributions are shown in Table 2. All coefficients have the expected signs, and parameter uncertainty is small. The debt-to-income ratio and the loan-to-value ratio are the most important factors for loan actions. They are negatively related to the loan probability. The empirical results suggest useful strategies to increase the chances of successful loan applications: consider a fixed rate large-size loan on a single-family house, avoid the subordinate lien, find a co-applicant for joint filing, etc.

6 Conclusion

Although our Gibbs sampler is designed for Bayesian estimation of fixed effects, non-Bayesian users can interpret Bayesian computing as a vehicle for reproducing a classical estimator that cannot be implemented due to the out-of-memory problem. The idea is to specify the classical LSDV estimator for fixed effects as the mean of the target distribution, and then construct a Gibbs sampler that has the target distribution as its invariant distribution. The fixed effects Gibbs sampler does not create any dummy variables, and

	mean	std	quantile05	quantile95
debt/Income	-2.821	0.004	-2.828	-2.815
loan/Value	-0.711	0.004	-0.717	-0.704
loanSize	2.822	0.010	2.805	2.839
fixedRate	0.136	0.002	0.133	0.140
subordinate	-0.093	0.002	-0.097	-0.089
units	-0.205	0.002	-0.209	-0.201
age	0.035	0.004	0.028	0.042
female	0.071	0.002	0.068	0.073
jointFilers	0.180	0.001	0.178	0.182

Table 2: Summary of posterior distributions of loan action Probit regression with fixed effects estimated by the Gibbs sampler.

therefore is suitable for in-memory computing with large datasets.

The Gibbs sampler is more versatile than reproducing the LSDV estimator. First, the posterior samples provide a point estimator (e.g., the posterior mean) as well as parameter uncertainty (e.g., the posterior standard deviation). Second, the Gibbs sampler can handle models with both fixed and random effects. Fixed effects have diffuse priors, while the variances of random effects are estimated by data. Third, the Gibbs sampler supports fixed effects in nonlinear models such as binary response variables.

References

Abowd, J., F. Kramarz, and D. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*, 251–333. 2

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*, 669–679. 7, 12

Carneiro, A., P. Guimaraes, and P. Portugal (2012). Real wages and the business cycle:

accounting for worker and firm heterogeneity. *American Economic Journal: Macroeconomics 4*, 133–152. 2

Chib, S. (1992). Bayesian inference in the Tobit censored regression model. *Journal of Econometrics 51*, 79–99. 7

Chib, S. and B. P. Carlin (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing 9*, 17–26. 2, 4

Correia, S., P. Guimaraes, and T. Zylkin (2020). Fast Poisson estimation with high-dimensional fixed effects. *Stata Journal 20*, 95–115. 2

Gaure, S. (2013). OLS with multiple high dimensional category variables. *Computational Statistics and Data Analysis 66*, 8–18. 2, 9, 10, 11

Gelfand, A. E. and A. F. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409. 2, 3, 4, 5, 7, 9

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal 7*, 98–119. 7

Guimaraes, P. and P. Portugal (2010). A simple feasible alternative procedure to estimate models with high-dimensional fixed effects. *Stata Journal 10*, 628–649. 2, 3, 7, 9, 10, 11

Mittag, N. (2019). A simple method to estimate large fixed effects models applied to wage determinants. *Labour Economics 61*, 1–10. 2

Somaini, P. and F. Wolak (2016). An algorithm to estimate the two-way fixed effects model. *Journal of Econometric Methods 5*, 143–152. 2

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data, Second Edition*. Cambridge: The MIT Press. 2