Hidden Group Patterns in Democracy Developments: Bayesian Inference for Grouped Heterogeneity Online Appendix

Jaeho Kim* Le Wang†

Abstract

This online appendix provides the detailed Bayesian MCMC algorithm used in the main paper, additional empirical results for robustness checks, and Monte Carlo simulation results. In this appendix, we also further discuss about our proposed Bayesian prior and model specifications.

Appendix A. Connection to the Group Fixed-effects Model of Bonhomme and Manresa (2015)

Our Bayesian approach is closely related to Bonhomme and Manresa (2015) and indeed built on their important insights. Specifically, our model with prior distributions can be considered as one of the special cases studied in Bonhomme and Manresa (2015) under certain conditions.[1] This section elaborates on the connection between our model and the group fixed-effects model of Bonhomme and Manresa (2015). Our benchmark model is given by

$$y_{i,t} = x'_{i,t}\theta + \alpha_{g_i,t} + \nu_{i,t}, \quad \nu_{i,t} \sim N(0, \sigma^2), \tag{A.0}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$. By stacking all observations for individual i, we can rewrite Equation (A.0) as follows:

$$y_i = x_i\theta + \alpha_{g_i} + \nu_i,\ \nu_i \sim N(\mathbf{0}, \sigma^2 I_T), \tag{A.1}$$

where $y_i = [y_{i,1}, y_{i,1}, ..., y_{i,T}]'$; $x_i = [x_{i,1}, x_{i,2}, ..., x_{i,T}]'$; $\alpha_{g_i} = [\alpha_{g_i,1}, \alpha_{g_i,2}, ..., \alpha_{g_i,T}]'$; $\nu_i = [\nu_{i1}, \nu_{i1}, ..., \nu_{iT}]'$. The target posterior distribution is given by:

*Corresponds to 248 Cate Center Drive, Norman, OK 73019, U.S.A. Department of Economics, University of Oklahoma. E-mail: jaeho@ou.edu

†248 Cate Center Drive, Norman, OK 73019, U.S.A. Department of Economics, University of Oklahoma. E-mail: le.wang.econ@gmail.com

[1] The detailed expositions are provided in the online supplimentary material of Bonhomme and Manresa (2015).

$$
\begin{aligned}
&p(\mathbf{A},\mathbf{V},\mathbf{G},\theta,\beta,\sigma^2|Y) \propto p(Y|\mathbf{A},\theta,\sigma^2,\mathbf{G})p(\mathbf{G}|\mathbf{V})p(\mathbf{A},\mathbf{V},\theta,\beta,\sigma^2)\\
&\propto \Big[\prod_{i=1}^{N} p(y_i|\alpha_{g_i},\theta,\sigma^2)p(g_i|\mathbf{V})\Big]p(\mathbf{A},\mathbf{V},\theta,\beta,\sigma^2)\\
&\propto \Big[\prod_{i=1}^{N} p(y_i|\alpha_{g_i},\theta,\sigma^2)\pi_{i,g_i}\Big]p(\mathbf{A},\mathbf{V},\theta,\beta,\sigma^2) \qquad (A.2)\\
&= \Bigg[(2\pi\sigma^2)^{-\frac{TN}{2}}\exp(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\Bigg(\sum_{t=1}^{T}(y_{i,t}-x'_{i,t}\theta-\alpha_{g_i,t})^2-(2\sigma^2)ln(\pi_{i,g_i})\Bigg))\Bigg]\\
&\times p(\mathbf{A},\mathbf{V},\theta,\beta,\sigma^2),
\end{aligned}
$$

where $Y = [y'_1, y'_2, ..., y'_N]'$; $\mathbf{A} = [\alpha_1, \alpha_2, ...]$; $\mathbf{V} = [V_1, V_2, ...]$; $\mathbf{G} = [g_1, g_2, ..., g_N]$; g_i is the group index variable; $\pi_{i,g}$ is the probability that individual i belongs to the group $g \in \{1, 2, ...\}$. The Dirichlet process (DP) prior assumes that π_{i,g_i} is homogeneous for all cross-sectional units such that $\pi_{i,g_i} = \pi_{g_i}$ for $g_i \in \{1, 2, ...\}$.

For a case where prior information about the group index variable is available, Bonhomme and Manresa (2015) employ a penalized group fixed-effects (GFE) estimator that utilizes the prior dependent objective function:

$$
(\hat{\mathbf{A}},\hat{\mathbf{G}},\hat{\theta}) = \arg\min_{\mathbf{A},\mathbf{G},\theta}\sum_{i=1}^{N}\Bigg(\sum_{t=1}^{T}(y_{i,t}-x'_{i,t}\theta-\alpha_{g_i,t})^2 - Cln(\pi_{i,g_i})\Bigg). \qquad (A.3)
$$

where $C > 0$ is a penalty term. If $C = 0$, the original GFE estimator that does not consider prior information is obtained. Bonhomme and Manresa (2015) show that the penalized GFE estimator and the GFE estimator are asymptotically equivalent if $0 < \epsilon < \pi_{i,g_i} < 1 - \epsilon < 1$ for some ϵ. That means that the effect of π_{i,g_i} on the GFE estimator eventually disappears as a sample size increases if the prior information is not dogmatic. The penalized GFE estimator of $\hat{\mathbf{A}}$ and $\hat{\theta}$ is therefore given as follows:

$$
(\hat{\mathbf{A}},\hat{\theta}) = \arg\min_{\mathbf{A},\theta}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t}-x'_{i,t}\theta-\alpha_{\hat{g}_i(\mathbf{A},\theta),t})^2 \qquad (A.4)
$$

where

$$
\hat{g}_i(\mathbf{A},\theta) = \arg\min_{g}\sum_{i=1}^{N}\Bigg(\sum_{t=1}^{T}(y_{i,t}-x'_{i,t}\theta-\alpha_{g,t})^2 - Cln(\pi_{i,g})\Bigg). \qquad (A.5)
$$

for $i = 1, 2, ..., N$ given $\pi_{i,g}$ for all g. The GFE estimator of g_i is simply $\hat{g}_i(\hat{\mathbf{A}}, \hat{\theta})$. While the penalized GFE estimator and the GFE estimator are asymptotically equivalent, the finite sample properties of the two GFE estimators will differ due to the penalty term or the prior information.

The target posterior density for our model is closely related with the penalized GFE estimator. Suppose that $C = 2\sigma^2$ and diffuse priors are assumed for θ, β. Then the logarithm of the posterior kernel density associated with key model parameters becomes proportional to the objective function for the penalized GFE estimator conditional on $\pi_{i,g}$, σ^2, and G where G represents the total number of groups. That is, if we estimate the model using the maximum-likelihood estimation method based on the posterior kernel density under the aforementioned conditions, the resulting estimates will be numerically identical to the penalized GFE estimates. In the Bayesian framework, we use weakly informative priors, treat $\pi_{i,g}$, σ^2, and G as model parameters to estimate, and employ the MCMC algorithm. As discussed in Bonhomme and Manresa (2015), the Bayesian estimator that involves non-dogmatic prior information will be asymptotically identical to the GFE estimator. However, in a finite sample, weakly informative priors can produce slightly different estimates from the GFE estimates. Strong informative prior may produce very different estimates. The other Bayesian models we consider in the empirical section are also related with the penalized GFE estimator in that they all apply prior information about g_i.

Appendix B. MCMC Algorithms

Neal (2000) introduces the MCMC algorithms that draw posterior samples of the model parameters from the joint posterior distribution after integrating out the DP prior. While the algorithms have been widely used, their converge speed to the posterior distribution could be very slow because the group assignment is based on a single-move sampling approach. Here, we apply the novel slice sampling approach developed by Walker (2007) who augments the joint posterior distribution with an auxiliary variable. This method avoids the truncated approximation approach used in Ishwaran and James (2001) and achieves a much faster convergence speed of the MCMC algorithm, compared to those of Neal (2000) by sequentially sampling the group assignment variables from their joint distribution.

Our model is given by

$$y_i = x_i\theta + \alpha_{g_i} + \nu_i,\ \nu_i \sim N(\mathbf{0}, \boldsymbol{Q}_{g_i}), \tag{B.1}$$

where $y_i = [y_{i,1}, y_{i,1}, ..., y_{i,T}]'$; $x_i = [x_{i,1}, x_{i,2}, ..., x_{i,T}]'$; $\alpha_{g_i} = [\alpha_{g_i,1}, \alpha_{g_i,2}, ..., \alpha_{g_i,T}]'$; $\nu_i = [\nu_{i1}, \nu_{i1}, ..., \nu_{iT}]'$; $\boldsymbol{Q}_{g_i} = diag(\sigma^2_{g_i,1}, \sigma^2_{g_i,2}, ..., \sigma^2_{g_i,T})$. The target posterior distribution is given by:

$$\begin{aligned} p(\mathbf{A}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \theta, \beta|Y) &\propto p(Y|\mathbf{A}, \mathbf{Q}, \theta, \mathbf{G})p(\mathbf{A}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \theta, \beta) \\ &\propto \Big[\prod_{i=1}^{N} p(y_i|\alpha_{g_i}, \mathbf{Q}_{g_i}, \theta)p(g_i|\mathbf{V})\Big]\Big[\prod_{j=1}^{\infty} p(V_j|\beta)p(\alpha_j, Q_j|\phi)\Big]p(\theta)p(\beta) \end{aligned} \tag{B.2}$$

where $Y = [y_1', y_2', ..., y_N']'$; $\mathbf{A} = [\alpha_1, \alpha_2, ...]$; $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, ...]$; $\mathbf{V} = [V_1, V_2, ...]$; $\mathbf{G} = [g_1, g_2, ..., g_N]$; $p(\alpha_j, Q_j|\phi) = p(\alpha_j|\phi_\alpha)p(Q_j|\phi_Q)$. We assume independent normal and inverse gamma priors for α_j and $\sigma^2_{j,t}$, respectively. This appendix provides a Gibbs sampling algorithm by using a set of auxiliary variables $\boldsymbol{\eta} = \{\eta_1, \eta_2, ..., \eta_N\}$ that reduces the infinite dimensional parameter space to a finite space. Consider the following extended target distribution that takes into account randomness of $\boldsymbol{\eta}$:

$$p(\boldsymbol{\eta}, \mathbf{A}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \theta, \beta|Y) \propto \Big[\prod_{i=1}^{N} p(y_i|\alpha_{g_i}, \mathbf{Q}_{g_i}, \theta)\mathbb{I}(\eta_i < \pi_{g_i})\Big]\Big[\prod_{j=1}^{\infty} p(V_j|\beta)p(\alpha_j, Q_j|\phi)\Big]p(\theta)p(\beta)$$
$$= \Big[\prod_{i=1}^{N} p(y_i|\alpha_{g_i}, \mathbf{Q}_{g_i}, \theta)p(\eta_i|\pi_{g_i})\pi_{g_i}\Big]\Big[\prod_{j=1}^{\infty} p(V_j|\beta)p(\alpha_j, Q_j|\phi)\Big]p(\theta)p(\beta) \quad (B.3)$$

where $\pi_{g_i} = p(g_i|\mathbf{V}) = V_{g_i}\prod_{l=1}^{g_i-1}(1 - V_l)$; $\mathbb{I}(.)$ is the indicator function that takes one if $\eta_i < \pi_{g_i}$ and zeros otherwise; η_i follows a uniform distribution defined between 0 and 1, $\eta_i \sim \mathscr{U}(0, 1)$. The distribution of η_i conditional on π_{g_i} is a uniform distribution, $\eta_i|\pi_{g_i} \sim \mathscr{U}(0, \pi_{g_i})$. It can be easily shown that the original target posterior distribution can be recovered by integrating out η_i for $i = 1, 2, ..., N$ in equation (B.3). That is, the posterior samples of the original target distribution can be also obtained by marginalizing MCMC samples over $\boldsymbol{\eta}$ after sampling from the extended target distribution. In what follows, we provide detailed expositions for each step of our proposed Gibbs sampling algorithm based on the extended target distribution. Readers only interested in the algorithm itself can skip to the end of this section.

B.1. Posterior Simulation for $\boldsymbol{A}$, $\boldsymbol{Q}$ and $\boldsymbol{V}$

The posterior conditional density of $\boldsymbol{A}$ is proportional to the kernel of the following form:

$$p(\mathbf{A}|\boldsymbol{\eta}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \theta, \beta, Y) \propto \Big[\prod_{i=1}^{N} p(y_i|\alpha_{g_i}, \mathbf{Q}_{g_i}, \theta)\mathbb{I}(\eta_i < \pi_{g_i})\Big]\Big[\prod_{j=1}^{\infty} p(\alpha_j, \mathbf{Q}_j|\phi)\Big]. \quad (B.4)$$

Now, let C^* denote the maximum counts of groups that we decide to consider in updating model parameters. Employing a theoretically valid C^* is crucial in the proposed Gibbs sampler. We adopt the following way of computing C^* conditional on $\boldsymbol{\eta}$ and $\boldsymbol{V}$:

$$C^* = min\{c : \sum_{j=1}^{c} \pi_j > 1 - \eta^*\} \quad (B.5)$$

where $\eta^* = max\{\eta_1, \eta_2, ..., \eta_N\}$. Walker (2007) demonstrates that C^* defined by equation (B.5) has important properties such that $G^* \leq C^* < \infty$ where $G^* = max\{g_1, g_2, ..., g_N\}$, and more importantly $\eta_i > \pi_j$ for $j > C^*$.

The latter theoretical feature enables reducing the kernel of the conditional posterior density of α_g included in equation (B.4) to the density given below:

$$p(\alpha_g|\boldsymbol{\eta}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \theta, \beta, Y) \propto \Big[\prod_{i\in\mathcal{B}(g_i=g)} p(y_i|\alpha_g, \mathbf{Q}_g, \theta)\Big]p(\alpha_g|\phi_\alpha), \quad (B.6)$$

for $g = 1, 2, ..., C^*$; $\mathcal{B}(g_i = g)$ denotes the set of all i's such that $g_i = g$. The target density in equation (B.4) becomes 0 for $g = C^* + 1, C^* + 2, ..., \infty$ due to the indicator function

$\mathbb{I}(\eta_i < \pi_{g_i})$. Therefore, while it is necessary to update the model parameters of the first C^* groups, we can ignore the parameters of the infinitely many later groups. By assuming an independent normal conjugate prior for α_g, posterior sampling is carried out by the posterior normal distribution:

$$\alpha_g|\boldsymbol{\eta}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \theta, \beta, Y \sim \mathscr{N}(M_{\alpha_g}, \Sigma_{\alpha_g}) \tag{B.7}$$

where $M_{\alpha_g} = \Sigma_{\alpha_g}(\Omega_\alpha^{-1}\mu_\alpha + \sum_{i\in\mathcal{B}(g_i=g)}(I_T'\boldsymbol{Q}_g^{-1}\tilde{y}_i)$, $\Sigma_{\alpha_g} = (\Omega_\alpha^{-1} + \sum_{i\in\mathcal{B}(g_i=g)} I_T'\boldsymbol{Q}_g^{-1}I_T')^{-1}$, and $\tilde{y}_i = y_i - x_i\theta$ for $g = 1, 2, ..., C^*$. If group $g \leq C^*$ does not contain any observations, corresponding α_g is generated from the prior distribution $\mathscr{N}(\mu_\alpha, \Omega_\alpha)$.

Next, $\boldsymbol{Q}_g$ is drawn from the posterior inverse-gamma distribution using a conjugate prior. Based on equation (B.6), it is straightforward to derive the posterior density of $\boldsymbol{Q}_g$:

$$p(\boldsymbol{Q}_g|\boldsymbol{\eta}, \mathbf{A}, \mathbf{V}, \mathbf{G}, \theta, \beta, Y) \propto \Big[\prod_{i\in\mathcal{B}(g_i=g)} p(y_i|\alpha_g, \mathbf{Q}_g, \theta)\Big] p(\mathbf{Q}_g|\phi_Q). \tag{B.8}$$

Under the inverse-gamma prior $\sigma_{g,t}^2 \sim \mathscr{IG}(\frac{\nu_Q}{2}, \frac{\delta_Q}{2})$, we sample the variance parameter from the following posterior distribution:

$$\sigma_{g,t}^2|\boldsymbol{\eta}, \mathbf{A}, \mathbf{V}, \mathbf{G}, \theta, \beta, Y \sim \mathscr{IG}(\frac{\bar{\nu}_{Q_{g,t}}}{2}, \frac{\bar{\delta}_{Q_{g,t}}}{2}) \tag{B.9}$$

where $\bar{\nu}_{Q_{g,t}} = \nu_Q + N_g$; $\bar{\delta}_{Q_{g,t}} = \delta_Q + \sum_{i\in\mathcal{B}(g_i=g)} v_{i,t}'v_{i,t}$; N_g is the number of cross-sectional units in group g. The error $v_{i,t}$ is easily recovered by $v_{i,t} = (y_{i,t} - \alpha_{g,t} - x_{i,t}'\theta)$ for $g = 1, 2, ..., C^*$ and $t = 1, 2, ..., T$,

Similar to posterior sampling of $\mathbf{A}$ and $\mathbf{Q}$, we derive the posterior conditional distribution of $\mathbf{V}$ by truncating the infinite dimensional parameter space with conditioning $\boldsymbol{\eta}$. The posterior density of $\mathbf{V}$ is proportional to the density:

$$\begin{aligned} p(\mathbf{V}|\boldsymbol{\eta}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \theta, \beta, Y) &\propto \Big[\prod_{i=1}^{N} p(\eta_i|\pi_{g_i})\pi_{g_i}\Big]\Big[\prod_{j=1}^{\infty} p(V_j|\beta)\Big] \\ &= \left[\prod_{i=1}^{N} p(\eta_i|\pi_{g_i})\left(V_{g_i}\prod_{j=1}^{g_i-1}(1-V_j)\right)\right]\Big[\prod_{j=1}^{\infty} p(V_j|\beta)\Big] \end{aligned} \tag{B.10}$$

where $p(V_j|\beta)$ is a Beta distribution, $\mathscr{B}e(1, \beta)$. Based on the partially collapsed Gibbs sampler that employs the density resulted from integrating $\boldsymbol{\eta}$ out of equation (B.8), the posterior distribution of V_g is given by:

$$p(V_g|\boldsymbol{\eta}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \theta, \beta, Y) \propto \left[\prod_{i\in\mathcal{B}(g_i=g)}\left(V_{g_i}\prod_{j=1}^{g_i-1}(1-V_j)\right)\right] p(V_g|\beta). \tag{B.11}$$

Therefore, posterior sampling is implemented by using the posterior beta distribution:

$$V_g|\boldsymbol{\eta}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \theta, \beta, Y \sim \mathscr{B}e(1 + \sum_{i=1}^{N}\mathbb{I}(g_i = g), \beta + \sum_{i=1}^{N}\mathbb{I}(g_i > g)) \tag{B.12}$$

for $g = 1, 2, ..., C^*$. Whenever a group is empty, V_g is sampled from the prior distribution. For $g > C^*$, it is unnecessary to update V_g because the extended target density is always 0.

B.2. Posterior Simulation for $\boldsymbol{G}$ and $\boldsymbol{\eta}$

We generate g_i conditional on $\boldsymbol{G}_{/i}$ for $i = 1, 2, ..., N$. Here, $\boldsymbol{G}_{/i}$ denotes $\boldsymbol{G}$ except g_i. The posterior density of g_i is proportional to the kernel of the following form:

$$p(g_i = g|\boldsymbol{\eta}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{V}, \boldsymbol{G}_{/i}, \theta, \beta, Y) \propto p(y_i|\alpha_g, \boldsymbol{Q}_g, \theta)\mathbb{I}(\eta_i < \pi_g) \quad (B.13)$$

for $g \leq C^*$, and

$$p(g_i = g|\boldsymbol{\eta}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{V}, \boldsymbol{G}_{/i}, \theta, \beta, Y) = 0 \quad (A.14)$$

for $c > C^*$. The above result shows how to sample g_i conditional on C^*. We compute the posterior probability of $g_i = g$ through self-normalization:

$$Pr(g_i = g|\boldsymbol{\eta}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{V}, \boldsymbol{G}_{/i}, \theta, \beta, Y) = \frac{p(y_i|\alpha_g, \boldsymbol{Q}_g, \theta)\mathbb{I}(\eta_i < \pi_g)}{\sum_{j=1}^{C^*} p(y_i|\alpha_j, \boldsymbol{Q}_j, \theta)\mathbb{I}(\eta_i < \pi_j)}. \quad (B.15)$$

Conditional on $\mathbf{V}$ and $\mathbf{G}$, it is straightforward to sample η_i. As shown in equation (B.3), η_i is simulated from the uniform distribution $\mathscr{U}(.)$ defined between 0 and π_{g_i}:

$$\eta_i|\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{V}, \boldsymbol{G}, \theta, \beta, Y \sim \mathscr{U}(0, \pi_{g_i}) \quad (B.16)$$

It is important to recognize that values of C^* and η^* should be re-computed after this Gibbs sampling step.

B.3. Label Switching

A conventional MCMC algorithm for estimating mixture models often does not fully explore the parameter space because the posterior distribution of model parameters can have mutiple peaks. This means that the Markov-chain for a conventional MCMC algorithm can get stuck in a local mode. To address such problem, our MCMC algorithm adopts the label switching step developed by Papaspiliopoulos and Roberts (2008).

The label switching move is based on a Metropolis-Hastings algorithm that utilizes two proposal distributions. The first proposal distribution begins by randomly selecting two non-empty groups. Suppose that j-th and and l-th groups are selected, and the two groups contain N_j and N_l observations respectively. Then the prior group probabilities for the two groups are given by $\pi_j^{N_j}$ for the j-th group and $\pi_l^{N_l}$ for the l-th group. After switching labels between the two groups (i.e., after swapping all observations between the two groups), the prior group probabilities change to $\pi_j^{N_l}$ and $\pi_l^{N_j}$. Accordingly, the acceptance probability for the proposed move is give by:

$$\min\left(1,\ \frac{\pi_j^{N_l}\pi_l^{N_j}}{\pi_j^{N_j}\pi_l^{N_l}} = (\frac{\pi_l}{\pi_j})^{N_j - N_l}\right). \quad (B.17)$$

All other densities for the likelihood, the priors of the not selected groups, and the candidate generating density are canceled in the above minimum acceptance probability.

The second proposal further facilitates the convergence of the MCMC algorithm. This proposal distribution randomly select two consecutive groups, j-th and $j+1$-th groups from all non-empty groups. Then the labels of the two groups switch along with V_j and V_{j+1}. Switching V_j and V_{j+1} greatly simplifies the acceptance probability. Similar to the first proposal distribution, we have the following acceptance probability for the second proposed move:

$$\min\left(1,\ \frac{\pi_j^{*N_{j+1}}\pi_{j+1}^{*N_j}}{\pi_j^{N_j}\pi_{j+1}^{N_{j+1}}}\right). \tag{B.18}$$

where π_j^* and π_{j+1}^* are the new group probabilities after switching V_j and V_{j+1}. The minimum acceptance probability is further simplified as:

$$\begin{aligned}\frac{\pi_j^{*N_{j+1}}\pi_{j+1}^{*N_j}}{\pi_j^{N_j}\pi_{j+1}^{N_{j+1}}} &= \frac{[V_{j+1}(1-V_{j-1})(1-V_{j-2})...(1-V_1)]^{N_{j+1}}[V_j(1-V_{j+1})(1-V_{j-1})...(1-V_1)]^{N_j}}{[V_j(1-V_{j-1})(1-V_{j-2})...(1-V_1)]^{N_j}[V_{j+1}(1-V_j)(1-V_{j-1})...(1-V_1)]^{N_{j+1}}} \\ &= \frac{(1-V_{j+1})^{N_j}}{(1-V_j)^{N_{j+1}}}\end{aligned} \tag{B.19}$$

The two proposal distributions that generate label switching are also randomly selected with probability 0.5. If a new label switching proposal is accepted, all group-dependent model parameters should be properly assigned according to the new group label.

B.4. Posterior Simulation for β

Escobar and West (1995) have shown that the number of estimated groups under a DP prior could be sensitive to β. Therefore, it has been recommended to include β in Gibbs sampling procedures as an unknown model parameter instead of choosing an arbitrary value. In this section, we briefly review the Gibb sampling approach for β, which is developed by Escobar and West (1995).

The parameter β has the posterior density of the following form:

$$p(\beta|\boldsymbol{\eta}, \mathbf{A}, \mathbf{Q}, \mathbf{G}, \theta, Y) \propto p(\mathbf{G}|\beta)p(\beta). \tag{B.20}$$

The prior for β is assumed to be $\beta \sim Gamma(a_0, b_0)$. Escobar and West (1995) propose a two-step approach to generate β by introducing a latent variable κ. Following their approach, we first generate κ conditional on β from a Beta distribution given below:

$$\kappa|\beta, \boldsymbol{G} \sim \mathscr{B}(\beta+1, N). \tag{B.21}$$

In the second step, we sample β conditional on κ according to the mixed Gamma distribution of the following form:

$$p(\beta|\kappa, \boldsymbol{G}) = \frac{a_0 + G - 1}{a_0 + G - 1 + N(b_0 - log(\kappa))}\mathscr{G}(a_0 + G, b_0 - log(\kappa)) + \frac{N(b_0 - log(\eta)}{a_0 + G - 1 + N(b_0 - log(\kappa))}\mathscr{G}(a_0 + G - 1, b_0 - log(\kappa) \quad (B.22)$$

where G is the tatal number of existing groups. Escobar and West (1995) prove that the MCMC samples of β generated from the above procedure converge to samples from the true target distribution in equation (B.2).

B.5. Posterior Sampling for θ

This section explains how to sample group-invariant model parameters θ. For the purpose, we transform the model to eliminate the group fixed effects as follows:

$$\bar{y}_i = x_i\theta + v_i,\ v_i \sim \mathscr{N}(\mathbf{0}, \boldsymbol{Q}_{g_i}), \quad (B.23)$$

where $\bar{y}_i = y_i - \alpha_{g_i}$. Equation (B.23) is a simple panel data model with a known form of heteroskedasticity. Therefore, the remaining Gibbs sampling steps are standard. Using a normal conjugate prior $\theta \sim \mathscr{N}(\mu_\theta, \Omega_\theta)$, the conditional posterior distribution of θ is given by:

$$\theta|\boldsymbol{\eta}, \mathbf{A}, \mathbf{Q}, \mathbf{V}, \mathbf{G}, \beta, Y \sim \mathscr{N}(M_\theta, \Sigma_\theta) \quad (B.24)$$

where

$$M_\theta = \Sigma_\theta(\Omega_\theta^{-1}\mu_\theta + \sum_{i=1}^{N} x_i'\boldsymbol{Q}_{g_i}^{-1}\bar{y}_i)$$

$$\Sigma_\theta = (\Omega_\theta^{-1} + \sum_{i=1}^{N} x_i'\boldsymbol{Q}_{g_i}^{-1}x_i)^{-1}.$$

This step completes the proposed MCMC sampling algorithm. A brief summary of our Bayesian algorithm is given below.

Proposed MCMC Algorithm

- Start the MCMC sampler by assigning initial values of model parameters and group indicator variables. To choose good initial values, we generate K random samples of α from $\mathscr{N}(\hat{\alpha}_{OLS}, 100 \times \hat{C}ov(\hat{\alpha}_{OLS}))$ where $\hat{\alpha}_{OLS}$ is the OLS estimate of α and $\hat{C}ov(\hat{\alpha}_{OLS})$ is its covariance estimate when $G = 1$. Also set $\{\theta = \hat{\theta}_{OLS}, \sigma^2_{g_i,t} = \hat{\sigma}^2_{OLS}\}$ as their initial values. Then apply the algorithm in the section B.2 to the generated samples of α, $\{\alpha^{(1)}, \alpha^{(2)}, ..., \alpha^{(K)}\}$ after excluding $\mathbb{I}(\eta_i < \pi_g)$ in equation (B.15). This step assigns each cross-sectional unit to one element of $\{\alpha^{(1)}, \alpha^{(2)}, ..., \alpha^{(K)}\}$. Eliminate all empty elements in $\{\alpha^{(1)}, \alpha^{(2)}, ..., \alpha^{(K)}\}$ that do not contain any observations. Set new labels for the survived elements such that $\{\alpha_1, \alpha_2, ..., \alpha_G\}$ where α_g indicates the g-th survived element. Set $\boldsymbol{G} = \{g_1, g_2, ..., g_N\}$ according to the new labels. Compute C^* for the next iteration. Repeat below Steps 1 to 9 until convergence is achieved.

1) Draw α_g from $\mathscr{N}(M_{\alpha_g}, \Sigma_{\alpha_g})$ for $g = 1, 2, ..., C^*$.
2) Draw $\sigma^2_{g,t}$ from $\mathscr{IG}(\frac{\bar{\nu}_{Q_{g,t}}}{2}, \frac{\bar{\delta}_{Q_{g,t}}}{2})$ for $t = 1, 2, ..., T$ and $g = 1, 2, ..., C^*$.
3) Draw V_g from $\mathscr{B}e(1 + \sum_{i=1}^{N} \mathbb{I}(g_i = g), \beta + \sum_{i=1}^{N} \mathbb{I}(g_i > g))$ for $c = 1, 2, ..., C^*$.
4) Draw η_i from $\mathscr{U}(0, \pi_{g_i})$ where $\pi_{g_i} = V_{g_i} \prod_{l=1}^{g_i-1}(1 - V_l)$ for $i = 1, 2, ..., N$.
5) Perform the label switching algorithm of Papaspiliopoulos and Roberts (2008).
6) Compute η^* and C^* and draw more α_g's, Q_g's and V_g's from the prior distributions if the new value of C^* is larger than that of C^* in the previous MCMC iteration.
7) Draw g_i by using the posterior probability:

$$\frac{p(y_i|\alpha_g, Q_g, \theta)\mathbb{I}(\eta_i < \pi_g)}{\sum_{j=1}^{C^*} p(y_i|\alpha_j, Q_j, \theta)\mathbb{I}(\eta_i < \pi_j)},$$

   for $i = 1, 2, ..., N$.
8) Draw κ from $\mathscr{B}(\beta + 1, N)$ first and then β from the mixed Gamma distribution in equation (B.22).
9) Draw θ from $\mathscr{N}(M_\theta, \Sigma_\theta)$.

Appendix C. Convergence Check

A panel data model with time-varying heterogeneity involves high-dimensional unobserved states. Thus, the mixing property of the proposed algorithm should be carefully examined. We closely follow the idea of Gelman et al. (1992) to examine the convergence property of our proposed Bayesian algorithm. Specifically, we repeat a target MCMC algorithm several times but with different initial parameter values. If the posterior distributions estimated with different initial values are similar, it can be interpreted as evidence of satisfactory convergence. For a simple model that involves a DP prior, Hastie et al. (2015) has implemented the same method.

We run the proposed algorithm with randomly selected G in the initial MCMC iteration. The initial values of the other group-dependent parameters are also randomly drawn in each estimation trial. The procedure to generate G and the group-dependent parameters in the initial MCMC iteration is explained in detail in *Appendix B*. For convergence check, we also include the result obtained using deterministic initial values. The deterministic initial values are set to be same as the OLS estimates and the group count is assumed $G = 1$.

For the empirical models, we monitor the group-independent parameters, β and θ_2, to check convergence. Note that it is not trivial to monitor the group-dependent parameters over MCMC iterations because of label switching. Figure A.1 shows the posterior distributions of the selected parameters that are generated by 5 estimation trials. It is evident via visual inspection that the proposed sampler produces almost identical posterior distributions, regardless of different initial values. This result indicates that all MCMC chains coverage well to the correct target posterior distribution.

Appendix D. Deriving a Marginal t Distribution from a Scale Mixture of Normals Distribution

Consider the densities of $\nu_{i,t}$ and λ_i given by:

$$p(\nu_{i,t}|\lambda_i, \sigma^2) = (\frac{\lambda_i}{2\pi\sigma^2})^{\frac{1}{2}} \exp(-\frac{\lambda_i}{2\sigma^2}\nu_{i,t}^2), \quad (D.1)$$

$$p(\lambda_i) = \frac{\frac{\tau}{2}^{\frac{\tau}{2}}}{\Gamma(\frac{\tau}{2})}\lambda_i^{\frac{\tau}{2}-1} \exp(-\frac{\tau}{2}\lambda_i). \quad (D.2)$$

To obtain the marginal density of $\nu_{i,t}$, we integrate out λ_i from the joint density of $\nu_{i,t}$ and λ_i as follows:

$$\begin{aligned}
p(\nu_{i,t}|\sigma^2) &= \int p(\nu_{i,t}|\lambda_i, \sigma^2)p(\lambda_i)d\lambda_i \\
&= (\frac{1}{2\pi\sigma^2})^{\frac{1}{2}} \frac{\frac{\tau}{2}^{\frac{\tau}{2}}}{\Gamma(\frac{\tau}{2})} \int \lambda_i^{\frac{\tau+1}{2}-1} \exp(-\frac{1}{2}(\frac{\nu_{i,t}^2}{\sigma^2} + \tau)\lambda_i)d\lambda_i \\
&= \frac{\Gamma(\frac{\tau+1}{2})}{\Gamma(\frac{\tau}{2})}(\frac{1}{2\pi\sigma^2})^{\frac{1}{2}}\left(\frac{\tau}{2}\right)^{\frac{\tau}{2}}\left(\frac{(\frac{\nu_{i,t}^2}{\sigma^2} + \tau)}{2}\right)^{-(\frac{\tau+1}{2})} \quad (D.3) \\
&= \frac{\Gamma(\frac{\tau+1}{2})}{\Gamma(\frac{\tau}{2})\sqrt{\pi\sigma^2}}\tau^{\frac{\tau+1}{2}-\frac{1}{2}}(\frac{\nu_{i,t}^2}{\sigma^2} + \tau)^{-(\frac{\tau+1}{2})} \\
&= \frac{\Gamma(\frac{\tau+1}{2})}{\Gamma(\frac{\tau}{2})\sqrt{\pi\sigma^2\tau}}(\frac{\nu_{i,t}^2}{\tau\sigma^2} + 1)^{-(\frac{\tau+1}{2})}
\end{aligned}$$

The third equality in (D.3) follows the fact that the term inside the integral is the kernel of the gamma distribution with $a = (\tau + 1)/2$ shape parameter and $b = (\frac{\nu_{i,t}^2}{\sigma^2} + \tau)/2$ rate parameter. Thus, we obtain the density of a non-standardized t distribution in the last equation in (D.3). Equation (D.3) explains why the scale mixture of normals distributions becomes a t distribution by integrating out λ_i.

Appendix E. Monte Carlo Simulation Evidence of Finite-Sample Performances of Proposed Bayesian Algorithm

Here we conduct three sets of Monte Carlo simulation exercises corresponding to various data generating processes (DGPs). These DGPs differ in how the group membership and the group fixed effects are correlated with the covariates, $x_{i,t}$. Such designs allow us to examine how our approach performs for a typical DGP, but also how sensitive the performances are to the correlation between the group fixed effects and the covariate. For all the exercises, we generate 600 artificial datasets with the model parameters set to be similar to the estimates obtained using our real data. The sample size of each artificial data set is set to the sample size of our replication data, and log-GDP per capita is used as the explanatory variable $x_{i,t}$ in all data generating processes. In sum, our approach performs well in these exercises.

Data Generating Process 1

$$y_{i,t} = \alpha_{g_i,t} + \theta_1 y_{i,t-1} + \theta_2 x_{i,t} + \nu_{i,t},\ \nu_{it} \sim N(0,\sigma^2) \tag{E.1}$$
$$\{\theta_1 = 0.4, \theta_2 = 0.06, \sigma^2 = 0.01\}$$

Individual units are assigned to groups in order such that each group contains the same number of individual units. And the true values of $\alpha_{g,t}$ are set to resemble the estimated pattern of $\alpha_{g,t}$ in our application. For each data set, our estimation method is applied to obtain the Bayesian posterior means of the model parameters. We use non-informative priors in this Monte-Carlo simulation. The reported results are the sample averages and standard deviations of those 600 different estimates. We also report the frequentist coverage of the true parameter based on the Bayesian 95% credible interval. We consider different numbers of groups ranging from 3 to 5 for this DGP.

The results in Table A1 show that the posterior distributions of the common parameters are close to the true ones. The coverage probability is well controlled when $G = 3$. However, it slightly differ from the nominal coverage 0.95 as G increases from 3 to 5. It is because for a given sample size, adding one more group to the DGP substantially increases estimation uncertainty. The sample means of the 95 % credible intervals are similar across all cases. However, their standard deviations are different, and the difference seems to affect the coverage probability. In each simulation, we store the posterior probability of group counts. For DGP 1, our algorithm always yields the highest posterior probability mass at the true value of G. By using a representative sample, Figure A.2 shows that regardless of how many groups exist in the assumed DGP, the posterior estimates of the time-varying group fixed effects track the true values of the group-dependent parameters well. The posterior estimates are obtained by using the post-processing approach as in Molitor et al. (2010) and Liverani et al. (2015). Refer to Appendix F for more details about the post-processing approach.

Note that in DGP 1, the true group assignment and the true values of the group fixed effects are exogenously chosen, independent of the covariates, $x_{i,t}$. We now turn to two additional DGPs to investigate how sensitive the performance of the proposed MCMC algorithm is to this assumption by allowing $\alpha_{gi,t}$ and g_i to be determined by $x_{i,t}$. For DGP 2, we allow for a strong correlation between the group fixed effects and the covariates. In DGP 3, we further allow that the group membership is nonparametrically determined by the covariates.

Data Generating Process 2

$$y_{i,t} = \alpha_{g_i,t} + \theta_1 y_{i,t-1} + \theta_2 x_{i,t} + \nu_{i,t},\ \nu_{it} \sim N(0,\sigma^2) \tag{E.2}$$
$$\alpha_{g_i,t} = \gamma_{g_i} \bar{x}_t, \tag{E.3}$$
$$\{\theta_1 = 0.4, \theta_2 = 0.06, \sigma^2 = 0.01\}$$

where $\bar{x}_t = \frac{1}{N}\sum_{i=1}^{N}(x_{i,t}) - \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}(x_{i,t})$. In DGP 2, we assume that $G = 3$, $\gamma_1 = 4$, $\gamma_2 = 2$, and $\gamma_3 = 0$. Individual units are assigned to the three groups in order such that each group contains the same number of individual units.

Data Generating Process 3

$$y_{i,t} = \alpha_{g_i,t} + \theta_1 y_{i,t-1} + \theta_2 x_{i,t} + \nu_{i,t},\ \ \nu_{it} \sim N(0, \sigma^2) \qquad (E.4)$$

$$\alpha_{g_i,t} = \gamma_{g_i}\bar{x}_t,\ \ g_i = m(x_{i,t}) \qquad (E.5)$$

$$\{\theta_1 = 0.4, \theta_2 = 0.06, \sigma^2 = 0.01\}$$

In DGP 3, individual units are assigned to three groups according to the function $m(.)$ of $x_{i,t}$. We take the same true values of $\{G, \gamma_1, \gamma_2, \gamma_3\}$ as DGP 2. For true group assignment in DGP 3, we compute the long-run growth rate of $x_{i,t}$ as $\Delta x_{it}\% = \frac{x_{i,T} - x_{i,1}}{x_{i,1}} \times 100$ and estimate high, medium, and low growth groups using the *kmeans* clustering method. For all individual units in the high growth group, g_i is set to 1. For all individual units in the medium growth group, g_i is set to 2. All remaining units in the low group growth takes $g_i = 3$.

We apply the proposed Bayesian algorithm to each data set generated by DGP 2 and DGP 3. The simulation results are reported in Table A.2. In line with with the theoretical prediction of Bonhomme and Manresa (2015), the results confirm that the model parameters are well estimated even when the relationship between $\alpha_{g_i,t}$ and $x_{i,t}$ is not explicitly modeled in estimation. Not only the point estimates are close to the true values, but the 95% Bayesian credible intervals tend to produce accurate coverage probabilities for the true parameters. In Figure A.3, we further confirm that the estimated group fixed effects are close to the true values using representative samples. (As before, we use the post-processing approach to obtain the posterior estimates of the group fixed effects. Refer to Appendix F for more details about the post-processing approach.)

Appendix F. Posterior Estimates of the Group Fixed Effects and the Group Membership

To facilitate interpretations of our Bayesian results and comparisons to the classical approach, we employ a modified version of the post-processing approach explained in Molitor et al. (2010) and Liverani et al. (2015). This approach is robust to changing group membership and "label switching", typical for the Bayesian group-based models.[2]

First, we obtain the posterior modes of $\alpha_{g_i} = [\alpha_{g_i,1}, \alpha_{g_i,2}, .., \alpha_{g_i,T}]'$ for all individual units, which is robust to changing group membership and the label switching problem. Then, we apply *kmeans* clustering to the set of the posterior modes of $\alpha_{g_i,t}$.[3] The *kmeans* clustering operates given a selected number of groups. For example, in the paper, for *Model 1*, *Model*

[2] Because labels associated with each group change while running the MCMC algorithm, the meaning of being included in group g ($g_i = g$) could be completely different across MCMC iterations. That is, $g_i = 1$ could mean that individual i belongs to a high democracy group at some MCMC iterations but at other MCMC iterations, $g_i = 1$ could indicates that the same individual i belongs to a low democracy group.

[3] We use the mode of each posterior marginal distribution to report the results in Figure A2 and A3. The results do not change when the posterior mean is used instead. This represents another fundamental difference between the Bayesian and Classical approaches. The GFE estimator is obtained by minimizing the objective function in equation (A.3) which is proportional to the joint density of the model parameters and the group assignments. However, any difference in using the marginal or joint distribution will quickly disappear when the sample size increases because both will be concentrated around the true values. Our empirical application appears be an example of this case.

2, and *Model 3*, we choose $G = 4$ to facilitate direct comparisons of the Bayesian estimation results with those of Bonhomme and Manresa (2015). For *Model 4* and *Model 5*, we select $G = 5$ and $G = 4$, respectively, because the posterior probabilities of G are highest at those values. After the optimal partition is chosen by the *k-means* clustering, we collect the posterior modes of $\alpha_{g_i,t}$ for countries classified in the same group and compute the sample average of their individual posterior modes. Figure 2 depicts the resulting sample averages of the posterior modes using the groups identified by the *kmeans* clustering. In the Monte-Carlo simulation, we also apply the same post-processing approach to obtain the posterior estimates of the group fixed effects. The results are depicted in Figure A.2 and A.3.

Finally, once the optimal partition is obtained using the *kmeans* clustering, the posterior probabilities of group membership is defined by the probability that a country is classified together with the benchmark country in the group. The benchmark country in each group is the country that is most often included in the same group and is least often included in the other groups during the MCMC iterations, again identified by the *kmeans* clustering.

G. Results Incorporating More Non-informative Priors

In this section, we experiment with more non-informative priors to assess the robustness of our results (the five models in Section 3). Recall that our dependent variable varies from 0 to 1. In our main empirical analysis, the weakly informative priors already cover the wide range of the values for the group effects (approximately from -2 to 1 with 95 %). Here, we consider even more diffuse priors with the prior variance of the base distribution for the group effects being 100 times the estimated variance-covariance matrix of pooled OLS estimates. Overall, the more non-informative priors do not impact the main conclusion of Bonhomme and Manresa (2015) that the long-run and the short-run effects of income on democracy is very limited once the group heterogeneity is controlled for.

Table A.7 displays the non-informative priors for the base distributions of the DP prior and the DPM prior, as well as the posterior estimates. As we can see, the results are rather robust to the alternative priors used. The posterior means for the coefficient on log-GDP per capita, θ_1, are very similar to what we obtained in the main empirical section. For *Model 1, Model 2,* and *Model 3*, the posterior means change only slightly by 0.01 for both the short-run and the long-run effects. For *Model 4* and *Model 5*, the posterior means of the same parameter are nearly identical to the results of the main empirical section.

In addition to θ_1, we also find similar results for θ_2, the auto-regressive parameter. For *Model 1* and *Model 2,* the posterior means of θ_2 are slightly higher than the main empirical section, perhaps because group counts are estimated to be slightly smaller. *Model 3*, *Model 4* and *Model 5* produce again, nearly identical estimates for θ_2. WAIC is higher for all models except *Model 5.*

References

Bonhomme, Stéphane and Elena Manresa. 2015. "Grouped patterns of heterogeneity in panel data." *Econometrica* 83:1147–1184.

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

Gelman, Andrew, Donald B Rubin, et al. 1992. "Inference from iterative simulation using multiple sequences." *Statistical science* 7:457–472.

Hastie, David I, Silvia Liverani, and Sylvia Richardson. 2015. "Sampling from Dirichlet process mixture models with unknown concentration parameter: mixing issues in large data implementations." *Statistics and computing* 25:1023–1037.

Ishwaran, Hemant and Lancelot F James. 2001. "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association* 96:161–173.

Liverani, Silvia, David I Hastie, Lamiae Azizi, Michail Papathomas, and Sylvia Richardson. 2015. "PReMiuM: An R package for profile regression mixture models using Dirichlet processes." *Journal of statistical software* 64:1.

Molitor, John, Michail Papathomas, Michael Jerrett, and Sylvia Richardson. 2010. "Bayesian profile regression with an application to the National Survey of Children's Health." *Biostatistics* 11:484–498.

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

Papaspiliopoulos, Omiros and Gareth O Roberts. 2008. "Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models." *Biometrika* 95:169–186.

Walker, Stephen G. 2007. "Sampling the Dirichlet mixture model with slices." *Communications in StatisticsSimulation and Computation®* 36:45–54.

Table A1: Sampling Distribution of the Bayesian Posterior Estimator

		DGP1 ($G = 3$)		DGP1 ($G = 4$)		DGP1 ($G = 5$)	
		Model Parameters					
	True	Mean (SD)	95 % CS (FC)	Mean (SD)	95 % CS (FC)	Mean (SD)	95 % CS (FC)
θ_1	0.4	0.417 (0.035)	[0.345,0.490] (0.926)	0.424 (0.035)	[0.351,0.497] (0.911)	0.427 (0.035)	[0.354,0.502] (0.905)
θ_2	0.07	0.068 (0.004)	[0.058,0.078] (0.951)	0.067 (0.004)	[0.057,0.077] (0.928)	0.067 (0.005)	[0.057,0.076] (0.896))
$\frac{\theta_2}{1-\theta_1}$	0.117	0.117 (0.006)	[0.104,0.131] (0.970)	0.117 (0.006)	[0.104,0.131] (0.950)	0.117 (0.006)	[0.103,0.131] (0.963)

	Probability for Group Counts		
	Mean (SD)	Mean (SD)	Mean (SD)
$Pr(G \leq 2)$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$Pr(G = 3)$	**0.963** (0.086)	0.000 (0.000)	0.000 (0.000)
$Pr(G = 4)$	0.035 (0.085)	**0.941** (0.101)	0.000 (0.000)
$Pr(G = 5)$	0.000 (0.001)	0.056 (0.097)	**0.887** (0.135)
$Pr(G = 6)$	0.000 (0.000)	0.001 (0.006)	0.105 (0.120)
$Pr(G = 7)$	0.000 (0.000)	0.000 (0.000)	0.006 (0.019)
$Pr(G = 8)$	0.000 (0.000)	0.000 (0.000)	0.000 (0.001)
$Pr(G \geq 9)$	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

Note: Simulation number = 600; Total Iteration/Burn-in Iteration = 50,000/10,000; G represents the number of non-empty groups; 'CS' represents the Bayesian credible set; 'FC' represents the frequentist coverage of the true parameter based on the Bayesian credible set; G represents the number of non-empty groups. The base distribution of the DP prior is $\mathcal{N}(\hat{\alpha}_{OLS}, 100 \times \hat{Cov}(\hat{\alpha}_{OLS}))$ where $\hat{\alpha}_{OLS}$ is the OLS estimate of α and $\hat{Cov}(\hat{\alpha}_{OLS})$ is the OLS estimate of $Cov(\hat{\alpha}_{OLS})$. We assume non-informative priors for the other parameters.

Table A2: Sampling Distribution of the Bayesian Posterior Estimator

		DGP2 ($G=3$)		DGP3 ($G=3$)	
		Model Parameters			
	True	Mean (SD)	95 % CS (FC)	Mean (SD)	95 % CS (FC)
θ_1	0.4	0.405	[0.333,0.479]	0.419	[0.347,0.492]
		(0.035)	(0.961)	(0.034)	(0.925)
θ_2	0.07	0.069	[0.059,0.079]	0.068	[0.057,0.078]
		(0.004)	(0.953)	(0.005)	(0.945)
$\frac{\theta_2}{1-\theta_1}$	0.117	0.116	[0.103,0.130]	0.117	[0.102,0.133]
		(0.006)	(0.950)	(0.007)	(0.958)
		Probability for Group Counts			
		Mean (SD)		Mean (SD)	
$Pr(G \leq 2)$		0.000		0.000	
		(0.000)		(0.000)	
$Pr(G=3)$		**0.992**		**0.991**	
		(0.041)		(0.030)	
$Pr(G=4)$		0.007		0.008	
		(0.040)		(0.030)	
$Pr(G=5)$		0.000		0.000	
		(0.001)		(0.000)	
$Pr(G=6)$		0.000		0.000	
		(0.001)		(0.000)	
$Pr(G=7)$		0.000		0.000	
		(0.000)		(0.000)	
$Pr(G=8)$		0.000		0.000	
		(0.000)		(0.000)	
$Pr(G \geq 9)$		0.000		0.000	
		(0.000)		(0.000)	

Note: Simulation number = 600; Total Iteration/Burn-in Iteration = 50,000/10,000; G represents the number of non-empty groups; 'CS' represents the Bayesian credible set; 'FC' represents the frequentist coverage of the true parameter based on the Bayesian credible set; G represents the number of non-empty groups. The base distribution of the DP prior is $\mathcal{N}(\hat{\alpha}_{OLS}, 100 \times \hat{Cov}(\hat{\alpha}_{OLS}))$ where $\hat{\alpha}_{OLS}$ is the OLS estimate of α and $\hat{Cov}(\hat{\alpha}_{OLS})$ is the OLS estimate of $Cov(\hat{\alpha}_{OLS})$. We assume non-informative priors for the other parameters.

Table A3: Posterior Probability of Group Membership for Model 1

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Argentina	0.000	1.000	0.000	0.000
Australia	0.997	0.000	0.000	0.000
Austria	0.998	0.000	0.000	0.000
Burundi	0.000	0.000	0.002	0.972
Belgium	0.997	0.000	0.000	0.000
Benin	0.000	0.000	1.000	0.000
Burkina Faso	0.001	0.000	0.000	0.000
Bolivia	0.000	0.573	0.001	0.000
Brazil	0.057	0.272	0.022	0.001
Central African Republic	0.000	0.000	0.999	0.000
Canada	0.997	0.000	0.000	0.000
Switzerland	0.997	0.000	0.000	0.000
Chile	0.000	0.000	0.004	0.000
China	0.000	0.000	0.000	0.989
Cote d'Ivoire	0.000	0.000	0.000	0.987
Cameroon	0.000	0.000	0.000	1.000
Congo, Rep.	0.000	0.000	0.042	0.834
Colombia	0.979	0.000	0.000	0.000
Costa Rica	1.000	0.000	0.000	0.000
Cyprus	0.944	0.004	0.000	0.000
Denmark	0.997	0.000	0.000	0.000

Table A3: Posterior Probability of Group Membership for Model 1(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Dominican Republic	0.993	0.000	0.000	0.000
Algeria	0.000	0.000	0.000	0.984
Ecuador	0.000	0.000	0.000	0.000
Egypt, Arab Rep.	0.000	0.000	0.000	0.942
Spain	0.159	0.001	0.000	0.000
Finland	0.978	0.003	0.000	0.000
France	0.997	0.000	0.000	0.000
Gabon	0.000	0.000	0.000	0.982
United Kingdom	0.997	0.000	0.000	0.000
Ghana	0.000	0.000	0.003	0.000
Guinea	0.000	0.000	0.000	0.980
Greece	0.001	0.317	0.000	0.000
Guatemala	0.045	0.013	0.000	0.000
Honduras	0.002	0.008	0.000	0.000
Indonesia	0.003	0.001	0.000	0.006
India	0.997	0.000	0.000	0.000
Ireland	0.997	0.000	0.000	0.000
Iran	0.000	0.000	0.000	0.983
Iceland	0.997	0.000	0.000	0.000
Israel	0.989	0.001	0.000	0.000
Italy	0.997	0.000	0.000	0.000

Table A3: Posterior Probability of Group Membership for Model 1(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Jamaica	0.997	0.000	0.000	0.000
Jordan	0.000	0.000	0.214	0.622
Japan	0.994	0.000	0.000	0.000
Kenya	0.000	0.000	0.000	0.713
Korea, Rep.	0.017	0.221	0.061	0.001
Sri Lanka	0.994	0.000	0.000	0.000
Luxembourg	0.986	0.000	0.000	0.000
Morocco	0.006	0.000	0.000	0.270
Madagascar	0.000	0.107	0.633	0.000
Mexico	0.000	0.003	0.050	0.016
Mali	0.000	0.000	1.000	0.000
Mauritania	0.000	0.000	0.000	0.972
Malawi	0.000	0.000	0.999	0.000
Malaysia	0.186	0.000	0.000	0.021
Niger	0.000	0.000	0.995	0.000
Nigeria	0.000	0.000	0.000	0.000
Nicaragua	0.001	0.036	0.015	0.001
Netherlands	0.997	0.000	0.000	0.000
Norway	0.997	0.000	0.000	0.000
Nepal	0.068	0.000	0.000	0.000
New Zealand	0.997	0.000	0.000	0.000

Table A3: Posterior Probability of Group Membership for Model 1(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Panama	0.000	0.000	0.679	0.000
Peru	0.000	0.000	0.000	0.000
Philippines	0.004	0.138	0.332	0.000
Portugal	0.061	0.000	0.000	0.000
Paraguay	0.001	0.001	0.006	0.087
Romania	0.000	0.000	0.965	0.011
Rwanda	0.000	0.000	0.000	0.988
Singapore	0.000	0.000	0.000	0.986
Sierra Leone	0.000	0.000	0.000	0.029
El Salvador	0.018	0.677	0.000	0.000
Sweden	0.996	0.000	0.000	0.000
Syrian Arab Republic	0.000	0.000	0.000	0.985
Chad	0.000	0.000	0.002	0.975
Togo	0.000	0.000	0.004	0.954
Thailand	0.022	0.338	0.000	0.000
Trinidad and Tobago	0.996	0.000	0.000	0.000
Tunisia	0.000	0.000	0.000	0.989
Turkey	0.011	0.625	0.000	0.000
Taiwan	0.001	0.018	0.261	0.002
Tanzania	0.000	0.005	0.682	0.014
Uganda	0.000	0.000	0.019	0.665

Table A3: Posterior Probability of Group Membership for Model 1(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Uruguay	0.003	0.457	0.001	0.000
United States	0.997	0.000	0.000	0.000
Venezuela, RB	0.996	0.000	0.000	0.000
South Africa	0.000	0.000	0.622	0.000
Congo, Dem. Rep.	0.000	0.000	0.000	0.986
Zambia	0.000	0.001	0.828	0.005

Table A3: Posterior Probability of Group Membership for Model 2

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Argentina	0.000	1.000	0.000	0.000
Australia	0.999	0.000	0.000	0.000
Austria	0.999	0.000	0.000	0.000
Burundi	0.000	0.000	0.000	0.966
Belgium	0.999	0.000	0.000	0.000
Benin	0.000	0.000	0.000	0.000
Burkina Faso	0.003	0.000	0.000	0.000
Bolivia	0.000	0.599	0.000	0.000
Brazil	0.068	0.275	0.000	0.002
Central African Republic	0.000	0.000	0.000	0.000
Canada	0.999	0.000	0.000	0.000
Switzerland	0.999	0.000	0.000	0.000
Chile	0.000	0.000	0.000	0.000
China	0.000	0.000	0.000	0.985
Cote d'Ivoire	0.000	0.000	0.000	0.982
Cameroon	0.000	0.000	0.000	0.986
Congo, Rep.	0.000	0.000	0.000	0.839
Colombia	0.986	0.000	0.000	0.000
Costa Rica	0.998	0.000	0.000	0.000
Cyprus	0.941	0.005	0.000	0.000
Denmark	0.999	0.000	0.000	0.000

Table A3: Posterior Probability of Group Membership for Model 2(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Dominican Republic	0.995	0.000	0.000	0.000
Algeria	0.000	0.000	0.000	0.975
Ecuador	0.000	0.000	0.020	0.000
Egypt, Arab Rep.	0.000	0.000	0.000	0.941
Spain	0.197	0.001	0.000	0.000
Finland	0.979	0.003	0.000	0.000
France	0.999	0.000	0.000	0.000
Gabon	0.000	0.000	0.000	0.977
United Kingdom	0.999	0.000	0.000	0.000
Ghana	0.000	0.000	0.994	0.000
Guinea	0.000	0.000	0.000	0.975
Greece	0.002	0.307	0.000	0.000
Guatemala	0.058	0.013	0.000	0.000
Honduras	0.001	0.008	0.001	0.000
Indonesia	0.008	0.000	0.042	0.009
India	0.998	0.000	0.000	0.000
Ireland	1.000	0.000	0.000	0.000
Iran	0.000	0.000	0.000	0.978
Iceland	0.999	0.000	0.000	0.000
Israel	0.991	0.001	0.000	0.000
Italy	0.999	0.000	0.000	0.000

Table A3: Posterior Probability of Group Membership for Model 2(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Jamaica	0.999	0.000	0.000	0.000
Jordan	0.000	0.000	0.000	0.633
Japan	0.995	0.000	0.000	0.000
Kenya	0.000	0.000	0.003	0.730
Korea, Rep.	0.026	0.222	0.001	0.001
Sri Lanka	0.997	0.000	0.000	0.000
Luxembourg	0.988	0.000	0.000	0.000
Morocco	0.013	0.001	0.008	0.306
Madagascar	0.000	0.097	0.000	0.000
Mexico	0.001	0.004	0.344	0.024
Mali	0.000	0.000	0.000	0.000
Mauritania	0.000	0.000	0.000	0.967
Malawi	0.000	0.000	0.000	0.000
Malaysia	0.242	0.000	0.002	0.030
Niger	0.000	0.000	0.000	0.000
Nigeria	0.000	0.000	1.000	0.000
Nicaragua	0.005	0.033	0.005	0.001
Netherlands	0.999	0.000	0.000	0.000
Norway	0.999	0.000	0.000	0.000
Nepal	0.083	0.000	0.004	0.000
New Zealand	0.999	0.000	0.000	0.000

Table A3: Posterior Probability of Group Membership for Model 2(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Panama	0.000	0.000	0.312	0.000
Peru	0.000	0.000	0.083	0.000
Philippines	0.005	0.133	0.000	0.000
Portugal	0.078	0.001	0.001	0.000
Paraguay	0.004	0.001	0.003	0.101
Romania	0.000	0.000	0.001	0.011
Rwanda	0.000	0.000	0.000	0.983
Singapore	0.000	0.000	0.000	0.981
Sierra Leone	0.001	0.000	0.049	0.032
El Salvador	0.023	0.700	0.000	0.000
Sweden	0.998	0.000	0.000	0.000
Syrian Arab Republic	0.000	0.000	0.000	1.000
Chad	0.000	0.000	0.000	0.967
Togo	0.000	0.000	0.000	0.948
Thailand	0.038	0.333	0.000	0.000
Trinidad and Tobago	0.998	0.000	0.000	0.000
Tunisia	0.000	0.000	0.000	0.985
Turkey	0.017	0.660	0.000	0.000
Taiwan	0.003	0.015	0.071	0.002
Tanzania	0.001	0.005	0.012	0.015
Uganda	0.000	0.000	0.006	0.656

Table A3: Posterior Probability of Group Membership for Model 2(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Uruguay	0.002	0.479	0.000	0.000
United States	0.999	0.000	0.000	0.000
Venezuela, RB	0.997	0.000	0.000	0.000
South Africa	0.000	0.000	0.000	0.000
Congo, Dem. Rep.	0.000	0.000	0.000	0.979
Zambia	0.000	0.001	0.000	0.006

Table A4: Posterior Probability of Group Membership for Model 3

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Argentina	0.003	0.004	0.001	0.000
Australia	1.000	0.000	0.000	0.000
Austria	1.000	0.000	0.000	0.000
Burundi	0.000	0.000	0.000	0.951
Belgium	1.000	0.000	0.000	0.000
Benin	0.000	0.000	1.000	0.000
Burkina Faso	0.001	0.000	0.010	0.002
Bolivia	0.001	0.013	0.002	0.000
Brazil	0.005	0.065	0.001	0.000
Central African Republic	0.000	0.000	1.000	0.000
Canada	1.000	0.000	0.000	0.000
Switzerland	1.000	0.000	0.000	0.000
Chile	0.001	0.001	0.059	0.002
China	0.000	0.000	0.000	1.000
Cote d'Ivoire	0.000	0.000	0.000	0.976
Cameroon	0.000	0.000	0.000	0.988
Congo, Rep.	0.000	0.000	0.034	0.831
Colombia	0.944	0.000	0.000	0.000
Costa Rica	0.998	0.000	0.000	0.000
Cyprus	0.955	0.016	0.000	0.000
Denmark	1.000	0.000	0.000	0.000

Table A4: Posterior Probability of Group Membership for Model 3(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Dominican Republic	0.989	0.004	0.000	0.000
Algeria	0.000	0.000	0.000	0.974
Ecuador	0.000	1.000	0.000	0.000
Egypt, Arab Rep.	0.000	0.000	0.000	0.915
Spain	0.002	0.986	0.000	0.000
Finland	0.993	0.000	0.000	0.000
France	1.000	0.000	0.000	0.000
Gabon	0.000	0.000	0.000	0.961
United Kingdom	1.000	0.000	0.000	0.000
Ghana	0.001	0.031	0.270	0.002
Guinea	0.000	0.000	0.000	0.956
Greece	0.006	0.080	0.000	0.000
Guatemala	0.029	0.000	0.000	0.000
Honduras	0.000	0.975	0.000	0.000
Indonesia	0.000	0.000	0.000	0.009
India	1.000	0.000	0.000	0.000
Ireland	1.000	0.000	0.000	0.000
Iran	0.000	0.000	0.000	0.983
Iceland	1.000	0.000	0.000	0.000
Israel	0.997	0.000	0.000	0.000
Italy	1.000	0.000	0.000	0.000

Table A4: Posterior Probability of Group Membership for Model 3(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Jamaica	1.000	0.000	0.000	0.000
Jordan	0.000	0.000	0.232	0.421
Japan	0.999	0.000	0.000	0.000
Kenya	0.000	0.000	0.000	0.610
Korea, Rep.	0.000	0.004	0.002	0.000
Sri Lanka	0.979	0.000	0.000	0.000
Luxembourg	0.998	0.001	0.000	0.000
Morocco	0.000	0.001	0.000	0.054
Madagascar	0.000	0.003	0.112	0.000
Mexico	0.000	0.009	0.016	0.002
Mali	0.000	0.000	1.000	0.000
Mauritania	0.000	0.000	0.000	0.952
Malawi	0.000	0.000	0.998	0.000
Malaysia	0.072	0.001	0.000	0.007
Niger	0.000	0.000	0.997	0.000
Nigeria	0.000	0.006	0.004	0.030
Nicaragua	0.000	0.001	0.005	0.000
Netherlands	1.000	0.000	0.000	0.000
Norway	1.000	0.000	0.000	0.000
Nepal	0.002	0.990	0.000	0.000
New Zealand	1.000	0.000	0.000	0.000

Table A4: Posterior Probability of Group Membership for Model 3(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Panama	0.000	0.002	0.953	0.000
Peru	0.000	0.833	0.000	0.002
Philippines	0.000	0.006	0.009	0.000
Portugal	0.001	0.990	0.000	0.000
Paraguay	0.000	0.000	0.002	0.013
Romania	0.000	0.000	0.980	0.006
Rwanda	0.000	0.000	0.000	0.986
Singapore	0.000	0.000	0.000	0.969
Sierra Leone	0.000	0.000	0.000	0.018
El Salvador	0.268	0.000	0.000	0.000
Sweden	1.000	0.000	0.000	0.000
Syrian Arab Republic	0.000	0.000	0.000	0.982
Chad	0.000	0.000	0.000	0.959
Togo	0.000	0.000	0.002	0.932
Thailand	0.041	0.028	0.000	0.000
Trinidad and Tobago	1.000	0.000	0.000	0.000
Tunisia	0.000	0.000	0.000	0.986
Turkey	0.079	0.001	0.000	0.002
Taiwan	0.000	0.003	0.220	0.000
Tanzania	0.000	0.000	0.400	0.007
Uganda	0.000	0.012	0.015	0.591

Table A4: Posterior Probability of Group Membership for Model 3(cont.)

Country	High	Transition (Early,Rapid)	Transition (Late,Rapid)	Low
Uruguay	0.001	0.015	0.000	0.000
United States	1.000	0.000	0.000	0.000
Venezuela, RB	0.996	0.000	0.000	0.000
South Africa	0.002	0.000	0.489	0.000
Congo, Dem. Rep.	0.000	0.000	0.000	0.977
Zambia	0.000	0.002	0.481	0.011

Table A5: Posterior Probability of Group Membership for Model 4

Country	High (Stable)	High (Unstable)	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Argentina	0.000	0.004	1.000	0.205	0.000
Australia	1.000	0.000	0.000	0.000	0.000
Austria	1.000	0.000	0.000	0.000	0.000
Burundi	0.000	0.000	0.007	0.014	0.962
Belgium	1.000	0.000	0.000	0.000	0.000
Benin	0.000	0.001	0.205	1.000	0.000
Burkina Faso	0.000	0.004	0.991	0.207	0.001
Bolivia	0.000	0.004	0.995	0.205	0.000
Brazil	0.000	0.045	0.892	0.223	0.001
Central African Republic	0.000	0.001	0.189	0.969	0.000
Canada	1.000	0.000	0.000	0.000	0.000
Switzerland	1.000	0.000	0.000	0.000	0.000
Chile	0.000	0.004	0.995	0.206	0.000
China	0.000	0.000	0.001	0.001	0.994
Cote d'Ivoire	0.000	0.000	0.001	0.001	0.993
Cameroon	0.000	0.000	0.000	0.000	1.000
Congo, Rep.	0.000	0.003	0.739	0.431	0.012
Colombia	0.000	0.275	0.671	0.108	0.000
Costa Rica	0.992	0.007	0.000	0.000	0.000
Cyprus	0.000	0.581	0.387	0.062	0.000
Denmark	1.000	0.000	0.000	0.000	0.000

Table A5: Posterior Probability of Group Membership for Model 4(cont.)

Country	High (Stable)	High (Unstable)	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Dominican Republic	0.000	0.208	0.756	0.125	0.000
Algeria	0.000	0.001	0.299	0.072	0.681
Ecuador	0.000	0.004	0.993	0.206	0.000
Egypt, Arab Rep.	0.000	0.000	0.039	0.010	0.951
Spain	0.000	0.763	0.200	0.025	0.000
Finland	0.000	0.976	0.014	0.003	0.000
France	1.000	0.000	0.000	0.000	0.000
Gabon	0.000	0.001	0.087	0.029	0.893
United Kingdom	1.000	0.000	0.000	0.000	0.000
Ghana	0.000	0.004	0.995	0.206	0.000
Guinea	0.000	0.000	0.001	0.001	0.984
Greece	0.000	0.004	0.995	0.206	0.000
Guatemala	0.000	0.009	0.981	0.204	0.000
Honduras	0.000	0.005	0.987	0.208	0.000
Indonesia	0.000	0.004	0.992	0.206	0.001
India	0.000	0.367	0.579	0.082	0.000
Ireland	1.000	0.000	0.000	0.000	0.000
Iran	0.000	0.000	0.018	0.006	0.973
Iceland	1.000	0.000	0.000	0.000	0.000
Israel	0.000	0.964	0.025	0.005	0.000
Italy	1.000	0.000	0.000	0.000	0.000

Table A5: Posterior Probability of Group Membership for Model 4(cont.)

Country	High (Stable)	High (Unstable)	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Jamaica	0.023	0.767	0.169	0.022	0.000
Jordan	0.000	0.002	0.270	0.548	0.258
Japan	0.000	0.988	0.004	0.002	0.000
Kenya	0.000	0.000	0.018	0.005	0.976
Korea, Rep.	0.000	0.025	0.887	0.241	0.000
Sri Lanka	0.000	0.271	0.682	0.120	0.000
Luxembourg	0.000	1.000	0.004	0.001	0.000
Morocco	0.000	0.013	0.740	0.163	0.198
Madagascar	0.000	0.001	0.359	0.808	0.000
Mexico	0.000	0.006	0.963	0.225	0.001
Mali	0.000	0.001	0.210	0.955	0.000
Mauritania	0.000	0.000	0.009	0.015	0.959
Malawi	0.000	0.002	0.375	0.802	0.000
Malaysia	0.000	0.007	0.978	0.209	0.001
Niger	0.000	0.001	0.238	0.924	0.000
Nigeria	0.000	0.004	0.995	0.206	0.000
Nicaragua	0.000	0.011	0.943	0.211	0.001
Netherlands	1.000	0.000	0.000	0.000	0.000
Norway	1.000	0.000	0.000	0.000	0.000
Nepal	0.000	0.046	0.929	0.210	0.000
New Zealand	1.000	0.000	0.000	0.000	0.000

Table A5: Posterior Probability of Group Membership for Model 4(cont.)

Country	High (Stable)	High (Unstable)	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Panama	0.000	0.004	0.951	0.247	0.000
Peru	0.000	0.004	0.995	0.206	0.000
Philippines	0.000	0.006	0.655	0.496	0.000
Portugal	0.000	0.526	0.440	0.075	0.000
Paraguay	0.000	0.012	0.870	0.243	0.021
Romania	0.000	0.001	0.306	0.859	0.000
Rwanda	0.000	0.000	0.001	0.000	0.996
Singapore	0.000	0.000	0.036	0.010	0.951
Sierra Leone	0.000	0.004	0.978	0.206	0.010
El Salvador	0.000	0.004	0.995	0.205	0.000
Sweden	0.000	0.986	0.005	0.001	0.000
Syrian Arab Republic	0.000	0.000	0.009	0.003	0.986
Chad	0.000	0.000	0.012	0.011	0.968
Togo	0.000	0.000	0.018	0.014	0.958
Thailand	0.000	0.004	0.995	0.205	0.000
Trinidad and Tobago	0.000	0.973	0.014	0.002	0.000
Tunisia	0.000	0.000	0.002	0.001	0.993
Turkey	0.000	0.004	0.994	0.205	0.000
Taiwan	0.000	0.010	0.942	0.232	0.000
Tanzania	0.000	0.001	0.209	0.564	0.298
Uganda	0.000	0.002	0.406	0.184	0.509

Table A5: Posterior Probability of Group Membership for Model 4(cont.)

Country	High (Stable)	High (Unstable)	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Uruguay	0.000	0.008	0.984	0.205	0.000
United States	1.000	0.000	0.000	0.000	0.000
Venezuela, RB	0.000	0.587	0.356	0.039	0.000
South Africa	0.000	0.004	0.969	0.231	0.000
Congo, Dem. Rep.	0.000	0.000	0.002	0.001	0.993
Zambia	0.000	0.002	0.452	0.716	0.002

Table A6: Posterior Probability of Group Membership for Model 5

Country	High	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Argentina	0.000	1.000	0.002	0.000
Australia	1.000	0.000	0.000	0.000
Austria	1.000	0.000	0.000	0.000
Burundi	0.000	0.000	0.452	0.525
Belgium	1.000	0.000	0.000	0.000
Benin	0.000	0.007	0.956	0.026
Burkina Faso	0.000	0.785	0.206	0.004
Bolivia	0.000	0.999	0.002	0.000
Brazil	0.001	0.904	0.028	0.001
Central African Republic	0.000	0.016	0.947	0.026
Canada	1.000	0.000	0.000	0.000
Switzerland	1.000	0.000	0.000	0.000
Chile	0.000	1.000	0.002	0.000
China	0.000	0.000	0.027	0.952
Cote d'Ivoire	0.000	0.000	0.068	0.914
Cameroon	0.000	0.000	0.034	1.000
Congo, Rep.	0.000	0.050	0.831	0.047
Colombia	0.012	0.695	0.010	0.000
Costa Rica	0.994	0.005	0.000	0.000
Cyprus	0.002	0.984	0.002	0.000
Denmark	1.000	0.000	0.000	0.000

Table A6: Posterior Probability of Group Membership for Model 5(cont.)

Country	High	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Dominican Republic	0.000	0.944	0.002	0.000
Algeria	0.000	0.003	0.091	0.882
Ecuador	0.000	0.991	0.002	0.000
Egypt, Arab Rep.	0.000	0.000	0.083	0.886
Spain	0.003	0.972	0.002	0.000
Finland	0.977	0.021	0.000	0.000
France	1.000	0.000	0.000	0.000
Gabon	0.000	0.002	0.046	0.926
United Kingdom	1.000	0.000	0.000	0.000
Ghana	0.000	1.000	0.002	0.000
Guinea	0.000	0.000	0.085	0.884
Greece	0.000	0.999	0.002	0.000
Guatemala	0.000	0.976	0.021	0.000
Honduras	0.000	0.994	0.002	0.000
Indonesia	0.000	0.054	0.884	0.023
India	0.006	0.857	0.002	0.000
Ireland	1.000	0.000	0.000	0.000
Iran	0.000	0.000	0.105	0.873
Iceland	1.000	0.000	0.000	0.000
Israel	0.975	0.023	0.000	0.000
Italy	1.000	0.000	0.000	0.000

Table A6: Posterior Probability of Group Membership for Model 5(cont.)

Country	High	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Jamaica	0.998	0.002	0.000	0.000
Jordan	0.000	0.010	0.914	0.070
Japan	0.997	0.003	0.000	0.000
Kenya	0.000	0.004	0.562	0.422
Korea, Rep.	0.000	0.918	0.039	0.001
Sri Lanka	0.041	0.727	0.012	0.000
Luxembourg	0.993	0.006	0.000	0.000
Morocco	0.000	0.097	0.581	0.202
Madagascar	0.000	0.740	0.241	0.005
Mexico	0.000	0.650	0.333	0.009
Mali	0.000	0.008	0.956	0.026
Mauritania	0.000	0.000	0.306	0.657
Malawi	0.000	0.000	0.969	0.026
Malaysia	0.000	0.038	0.707	0.029
Niger	0.000	0.012	0.953	0.026
Nigeria	0.000	1.000	0.002	0.000
Nicaragua	0.000	0.898	0.086	0.002
Netherlands	1.000	0.000	0.000	0.000
Norway	1.000	0.000	0.000	0.000
Nepal	0.000	0.962	0.009	0.000
New Zealand	1.000	0.000	0.000	0.000

Table A6: Posterior Probability of Group Membership for Model 5(cont.)

Country	High	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Panama	0.000	0.992	0.010	0.000
Peru	0.000	0.997	0.003	0.000
Philippines	0.000	0.854	0.113	0.002
Portugal	0.000	0.978	0.002	0.000
Paraguay	0.000	0.190	0.742	0.037
Romania	0.000	0.038	0.925	0.025
Rwanda	0.000	0.000	0.029	0.953
Singapore	0.000	0.000	0.054	0.926
Sierra Leone	0.000	0.068	0.881	0.024
El Salvador	0.000	1.000	0.002	0.000
Sweden	0.989	0.009	0.000	0.000
Syrian Arab Republic	0.000	0.000	0.033	0.944
Chad	0.000	0.001	0.221	0.749
Togo	0.000	0.009	0.827	0.146
Thailand	0.000	0.999	0.002	0.000
Trinidad and Tobago	0.979	0.016	0.000	0.000
Tunisia	0.000	0.000	0.046	0.935
Turkey	0.000	0.998	0.002	0.000
Taiwan	0.000	0.865	0.119	0.002
Tanzania	0.000	0.033	0.933	0.026
Uganda	0.000	0.014	0.724	0.224

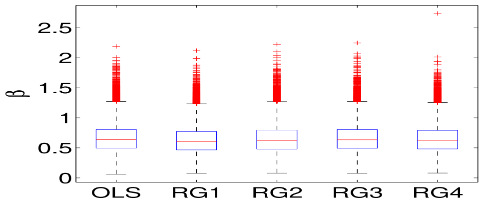
Table A6: Posterior Probability of Group Membership for Model 5(cont.)

Country	High	Transition (Early,Gradual)	Transition (Late,Rapid)	Low
Uruguay	0.000	0.997	0.002	0.000
United States	1.000	0.000	0.000	0.000
Venezuela, RB	0.024	0.789	0.002	0.000
South Africa	0.000	0.056	0.917	0.024
Congo, Dem. Rep.	0.000	0.000	0.027	0.950
Zambia	0.000	0.002	1.000	0.034

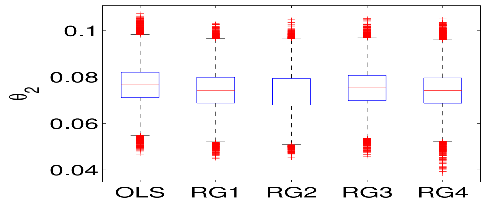
Table A7: Posterior Estimates with Alternative Priors

	Prior	*Model 1* DP prior $v_{i,t} \sim N(0,\sigma^2)$		*Model 2* DPM prior $v_{i,t} \sim N(0,\sigma^2)$		*Model 3* DP prior $v_{i,t} \sim N(0,\lambda_i^{-1}\sigma^2)$		*Model 4* DP prior $v_{i,t} \sim N(0,\sigma^2_{g_i})$		*Model 5* DP prior $v_{i,t} \sim N(0,\sigma^2_{g_i,t})$	
		Posterior									
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
θ_1	$\mathscr{N}(0,10^2)$	0.473	(0.093)	0.513	(0.092)	0.272	(0.050)	0.509	(0.043)	0.502	(0.041)
θ_2	$\mathscr{N}(0,10^2)$	0.065	(0.009)	0.065	(0.009)	0.061	(0.009)	0.017	(0.006)	0.043	(0.007)
$\frac{\theta_2}{1-\theta_1}$	-	0.127	(0.025)	0.138	(0.025)	0.085	(0.012)	0.035	(0.013)	0.086	(0.014)
		Probability of Group Counts									
$Pr(G \leq 2)$		0.000		0.000		0.000		0.000		0.000	
$Pr(G = 3)$		0.000		0.000		0.000		0.000		0.000	
$Pr(G = 4)$		0.000		0.000		0.000		0.010		0.206	
$Pr(G = 5)$		0.000		0.000		0.000		**0.666**		**0.529**	
$Pr(G = 6)$		0.000		0.000		0.000		0.251		0.232	
$Pr(G = 7)$		0.000		0.000		0.001		0.065		0.030	
$Pr(G = 8)$		0.120		**0.297**		0.071		0.005		0.001	
$Pr(G = 9)$		0.218		0.250		0.239		0.000		0.000	
$Pr(G = 10)$		**0.288**		0.254		**0.336**		0.000		0.000	
$Pr(G = 11)$		0.172		0.151		0.226		0.000		0.000	
$Pr(G = 12)$		0.120		0.035		0.092		0.000		0.000	
$Pr(G = 13)$		0.052		0.010		0.030		0.000		0.000	
$Pr(G \geq 14)$		0.005		0.000		0.000		0.000		0.000	
WAIC		-272.625		-255.202		-345.401		-457.851		-398.817	

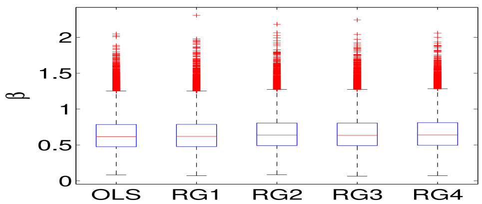
Note: Total Iteration/Burn-in Iteration = 100,000/50,000. We employ the pooled OLS estimation method to obtain the base distribution of the DP prior, $\mathscr{N}(\hat{\alpha}_0, 100 \times \hat{C}ov(\hat{\alpha}_0))$ where $\hat{\alpha}_0$ is the OLS estimate of α_{OLS} and $\hat{C}ov(\hat{\alpha}_0)$ is the OLS estimate of $Cov(\hat{\alpha}_0)$. We employ the pooled OLS estimation method to obtain the base distribution of the DPM prior, $\mathscr{N}(\hat{\delta}_0, 100 \times \hat{C}ov(\hat{\delta}_0))$ where $\hat{\delta}_0$ is the OLS estimate of δ_{OLS} and $\hat{C}ov(\hat{\delta}_0)$ is the OLS estimate of $Cov(\hat{\delta}_0)$. G represents the number of non-empty groups.



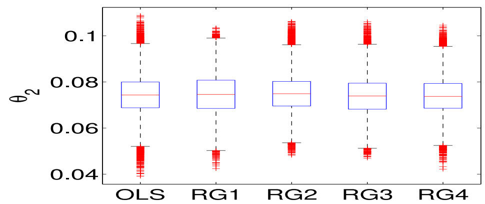
(a) Model 1: β



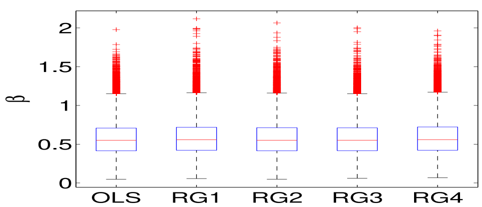
(b) Model 1: θ_2



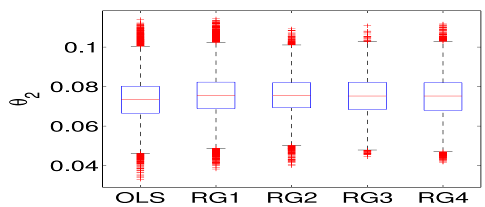
(c) Model 2: β



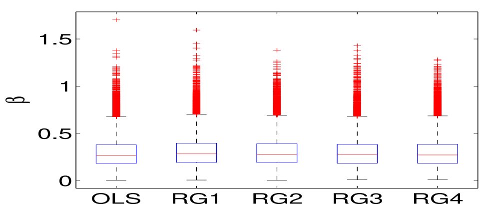
(d) Model 2: θ_2



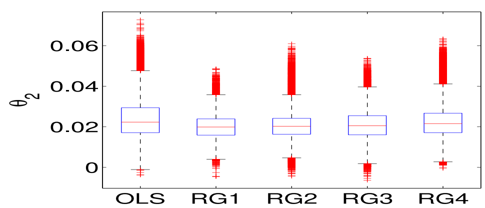
(e) Model 3: β



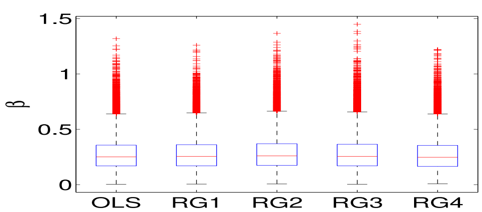
(f) Model 3: θ_2



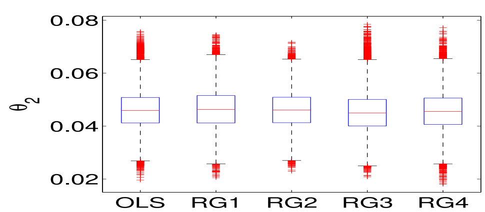
(g) Model 4: β



(h) Model 4: θ_2

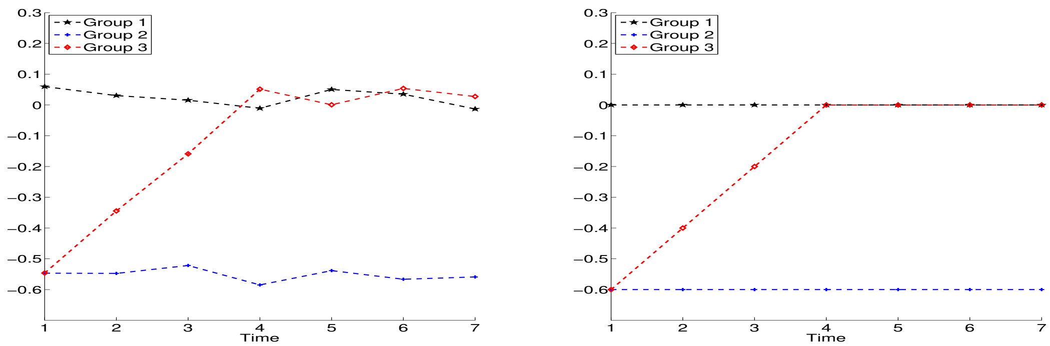


(i) Model 5: β

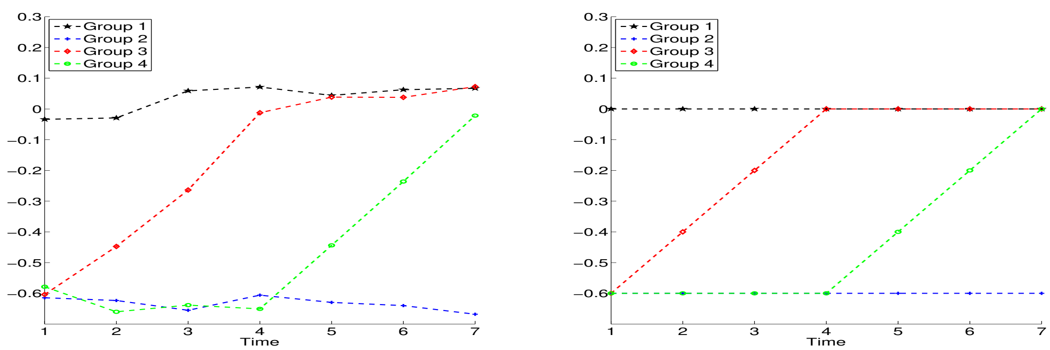


(j) Model 5: θ_2

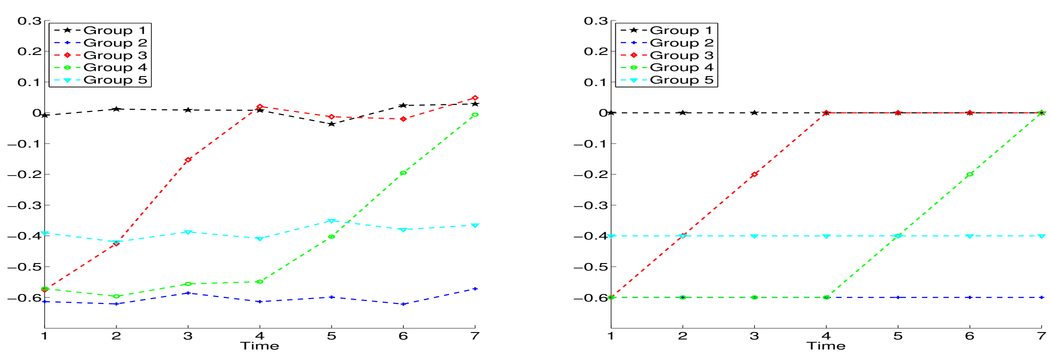
Figure A1: Convergence Check: The first box plot labeled 'OLS' represents the posterior distribution of a selected model parameter obtained using the OLS estimates and $G = 1$ as initial values in the MCMC algorithm. The remaining box plots labeled 'RG' represent the posterior distribution of a selected model parameter obtained using randomly generated parameter values and G as initial values in the MCMC algorithm. The red line indicates the median. The bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.



(a) DGP 1: True $G = 3$

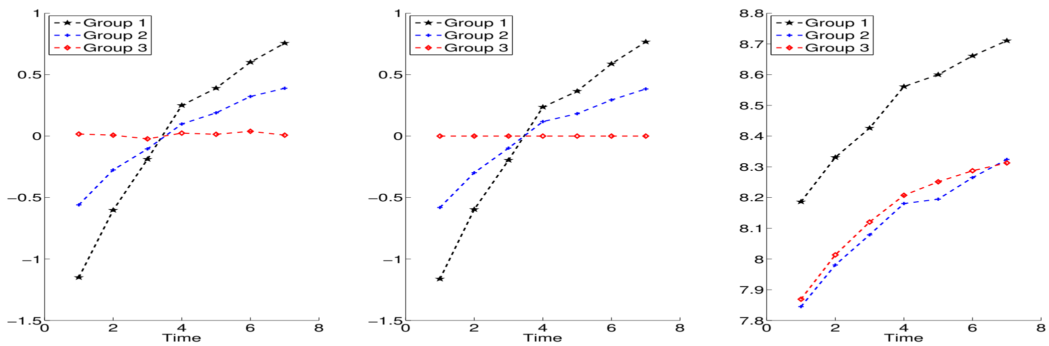


(b) DGP 1: True $G = 4$

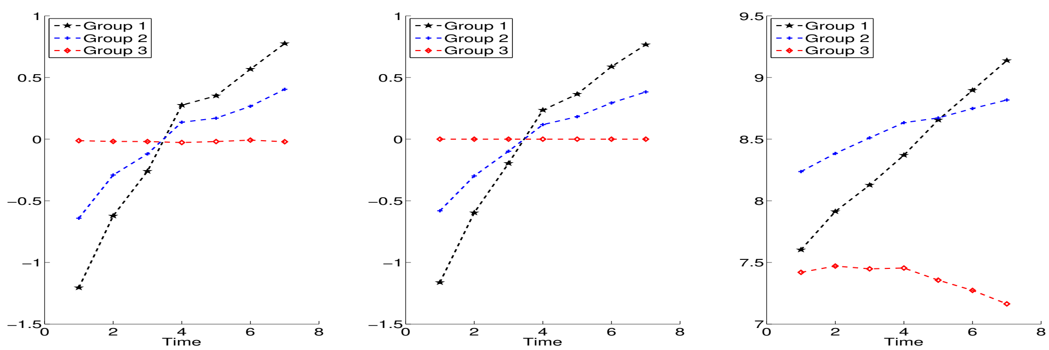


(c) DGP 1: True $G = 5$

Figure A2: Grouped Pattern of Heterogeneity for DGP 1: The panels in the first column show the posterior estimates of α_c. The panels in the second column represent the true values of α_c.



(a) DGP 2: True $G = 3$



(b) DGP 3: True $G = 3$

Figure A3: Grouped Pattern of Heterogeneity for DGP 2 and DGP3: The panels in the first column show the posterior estimates of α_c. The panels in the second column represent the true values of α_c. The panels in the third column represent the across-sectional averages of $x_{i,t}$ in each group.