Quasi-Bayesian Inference for Grouped Panels

Jiaming Huang

November 30, 2023

Universitat Pompeu Fabra and Barcelona School of Economics

Grouped Panels

- Linear panel data models

$$y_{it} = x_{it}'\beta_{\gamma_i} + \epsilon_{it} \ , \qquad t = 1, \ldots, T, \quad i = 1, \ldots, N$$

- Group membership $\gamma_i \in \{1, \ldots, G\}$ is latent and needs to be estimated
  - ▷ Examples: y_{it} consumption, x_{it} income; theory predicts different β_{γ_i} marginal propensities to consume for hand-to-mouth households. **But who are the hand-to-mouth?**

- Group-level parameters $\beta_1, \ldots, \beta_G$ conditional on given group membership γ
  - ▷ Pre-defined $\gamma \implies$ potential misspecification bias
  - ▷ Data-driven $\gamma \implies$ potential selection bias

Failure of Existing Approaches

- Conditional on group membership γ, decompose group-level estimation errors into

$$\hat{\beta}_g - \beta_g^0 = \left(\sum_{i,t} \mathbf{1}\{\gamma_i = g\} x_{it} x_{it}' \right)^{-1} \left(\sum_{i,t} \mathbf{1}\{\gamma_i = g\} \left[x_{it} x_{it}' (\beta^0_{\gamma_i^0} - \beta_g^0) + x_{it}\epsilon_{it} \right] \right)$$

- Conventional asymptotic inference relies on γ being *fixed at the true* γ^0
  - ▷ Misspecification bias: $\beta^0_{\gamma_i^0} \neq \beta_g^0$
  - ▷ Selection bias: γ_i a random variable depending on $x_{it}\epsilon_{it}$
- The problem: lack of uncertainty quantification of γ!

Model-Based Clustering

- Model-based clustering provides the desired uncertainty quantification of γ
  - ▷ Example 1: Bayesian clustering with Dirichlet-Categorical prior
  - ▷ Example 2: Finite mixture models
- But it requires full distribution of the data:
  - ▷ Example: the distribution of the error term ϵ_{it}
  - ▷ If misspecified, can we still learn the latent group structure?

This Paper

1. Generic quasi-Bayesian framework for large classes of loss functions and priors
2. First result on consistency and contraction rate for quasi-Bayesian clustering
3. Bootstrap-based learning rate calibration to ensure coverage
4. Revisit heterogeneous wage cyclicalities: the group patterns cast doubt on conventional shock amplification mechanisms

Literature

1. latent group heterogeneity
   - Bonhomme and Manresa, 2015; Cheng et al., 2019; Liu et al., 2020; Su and Ju, 2018; Su et al., 2016; Wang et al., 2018; Y. Zhang et al., 2019
2. Bayesian clustering
   - Duan and Dunson, 2021; Ishwaran and James, 2001; Miller and Harrison, 2018; Ren et al., 2022; Richardson and Green, 1997; Rigon et al., 2023; B. Zhang, 2023
3. Quasi-Bayesian inference
   - Atchadé, 2017; Chernozhukov and Hong, 2003; Kim, 2002; Miller, 2021; Petrova, 2019
4. Heterogeneous wage cyclicalities and income risks
   - Auclert, 2019; Bils, 1985; Busch et al., 2022; Figueiredo, 2022; Guvenen et al., 2014; Low et al., 2010

Quasi-Bayesian Clustering Framework

Setup

- Observe a $p \times 1$ vector w_{it} for unit $i = 1, \ldots, N$ and time $t = 1, \ldots, T$
- Each unit is associated with a parameter $\beta_i \in \mathbb{R}^d$ of interest
- Assume that β_i are grouped into G groups, represented by $(\boldsymbol{\beta}, \boldsymbol{\gamma})$
  - ▷ $dG \times 1$ grouped parameter vector $\boldsymbol{\beta} = \mathrm{vec}(\beta_1, \ldots, \beta_G)$
  - ▷ $N \times 1$ group membership vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)'$ with $\gamma_i \in \{1, \ldots, G\}$

Model

- Frequentist clustering is given by

$$(\hat{\beta}, \hat{\gamma}) = \arg\min \overline{L}(\beta, \gamma)$$

  - $\overline{L}(\beta, \gamma)$ is a sample loss function

- Bayesian clustering is characterized by the posterior

$$\pi_{NT}(\beta, \gamma) \propto \pi(\mathbf{W}|\beta, \gamma)\pi(\beta, \gamma)$$

  - $\pi(\mathbf{W}|\beta, \gamma)$ is the likelihood and $\pi(\beta, \gamma)$ a prior

Model

- Quasi-Bayesian clustering is characterized by the posterior

$$\pi_{NT}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto \left(\exp\left[-NT\,\overline{L}(\boldsymbol{\beta}, \boldsymbol{\gamma})\right]\right)^{\psi} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma})$$

  - ▷ Why this form?

$$\pi_{NT}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \arg\min_{\tilde{\pi}} \int NT\,\overline{L}(\boldsymbol{\beta}, \boldsymbol{\gamma})\tilde{\pi}(\boldsymbol{\beta}, \boldsymbol{\gamma})\mathrm{d}\boldsymbol{\beta}\mathrm{d}\boldsymbol{\gamma} + \psi^{-1}\mathrm{KL}(\tilde{\pi}||\pi)$$

  - ▷ What does ψ do?
  - ▷ Why need this?

Model

- Quasi-Bayesian clustering is characterized by the posterior

$$\pi_{NT}(\beta, \gamma) \propto \left(\exp\left[-NT\,\overline{L}(\beta, \gamma)\right]\right)^{\psi} \pi(\beta, \gamma)$$

  - ▷ Why this form?
  - ▷ What does ψ do?

    Controls the relative weight on the loss function

  - ▷ Why need this?

Model

- Quasi-Bayesian clustering is characterized by the posterior

$$\pi_{NT}(\beta, \gamma) \propto \left(\exp\left[-NT\,\overline{L}(\beta, \gamma)\right]\right)^{\psi} \pi(\beta, \gamma)$$

  - ▷ Why this form?
  - ▷ What does ψ do?
  - ▷ Why need this?

    Quasi-Bayesian posterior generally leads to incorrect frequentist coverage

Special Cases

- Posterior mode *recovers the K-means clustering* if
  - ▷ flat prior $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto c > 0$
- Posterior mode *recovers pre-defined grouping* $\overline{\gamma}$ if
  - ▷ flat prior on coefficients $\pi(\boldsymbol{\beta}) \propto c > 0$ but hard constraint on grouping $\mathbf{1}\{\gamma = \overline{\gamma}\}$
- Posterior distribution *recovers standard Bayesian clustering* if
  - ▷ $\overline{\ell}_i(\beta_{\gamma_i})$ is the negative log-likelihood and $\psi = 1$

Linear Panel Data Model:
A Worked-out Example

Setup

- Linear model $y_{it} = x'_{it}\beta_{\gamma_i} + \epsilon_{it}$
- Loss $\overline{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{NT}\sum_{i,t}(y_{it} - x'_{it}\beta_{\gamma_i})^2$
- Assume G is known and fixed. Set prior $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})$ where

$$\begin{aligned} \boldsymbol{\eta} = (\eta_1, \ldots, \eta_G) &\sim \text{Dirichlet}(\alpha), \\ \gamma_i|\boldsymbol{\eta} &\sim \text{Categorical}(\eta_1, \ldots, \eta_G), \\ \beta_g &\sim N(\mu, \Sigma) \end{aligned}$$

- Given $\boldsymbol{\gamma}$, denote N_g the number of units in group g, $Y_g = (y_i)_{i\in\mathcal{C}_g}$ and $X_g = (x_i)_{i\in\mathcal{C}_g}$ the vectorized data assigned to group g

Algorithm 1: Blocked Gibbs Sampler

Given $\psi > 0$, iterate between

1. **Sampling Group Membership Given $\boldsymbol{\beta}$.** For $i = 1, \ldots, N$, draw

$$\gamma_i \sim \pi_{NT}(\gamma_i = g|\boldsymbol{\beta}) \propto \exp\left[-\psi \sum_t (y_{it} - x_{it}'\beta_g)^2\right](N_{g,-i} + \alpha)$$

2. **Sampling Grouped Parameters Given $\boldsymbol{\gamma}$.** For $g = 1, \ldots, G$, draw

$$\beta_g \sim N(\tilde{\mu}, \tilde{\Sigma}), \quad \tilde{\mu}' = \tilde{\Sigma}^{-1}\left[2\psi Y_g X_g + \mu'\Sigma^{-1}\right], \quad \tilde{\Sigma}^{-1} = 2\psi X_g' X_g + \Sigma^{-1}$$

Comparison with K-means

- Consider the posterior odds ratio

$$\frac{\pi_{NT}(\gamma_i = j|\beta)}{\pi_{NT}(\gamma_i = k|\beta)} = \exp\left[-\psi\sum_t \left((y_{it} - x_{it}'\beta_j)^2 - (y_{it} - x_{it}'\beta_k)^2\right) + \ln\frac{|\mathcal{C}_{j,-i}| + \alpha}{|\mathcal{C}_{k,-i}| + \alpha}\right]$$

- Key difference: quasi-Bayesian clustering updates group membership *probabilistically* while K-means does so *deterministically*

- **Useful when the sample loss is flat or full of local minima**

Learning Rate Calibration

- The loss function is a misspecified likelihood, so posterior inference is not correct
- The learning rate ψ can be calibrated to improve coverage
- Let $\zeta = f(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be the object of interest
  - $f(\cdot, \cdot)$ **must be invariant to permutation of group labels**
  - Example 1: average effects $\zeta = \frac{1}{N}\sum_{i=1}^{N} \beta_{\gamma_i}$
  - Example 2: ordered group parameters $\zeta = (\beta_{\sigma(1)}, \ldots, \beta_{\sigma(G)})$ where $\sigma : [G] \to [G]$ is a permutation such that $\beta_{\sigma(1)} \leq \ldots \leq \beta_{\sigma(G)}$

Algorithm 2: Learning Rate Calibration

Given a target level $\alpha \in (0,1)$, and current learning rate $\psi^{(j)}$

1 Bootstrap Quasi-Posterior.

For $b = 1, \ldots, B$,

1. Sample with replacement from $\{(Y_i, X_i)\}_{i=1}^N$ to obtain $\{(Y_i^{(b)}, X_i^{(b)})\}_{i=1}^N$
2. Sample $\left\{(\beta^{(b,m)}, \gamma^{(b,m)})\right\}_{m=1}^M$ using Algorithm 1, $\psi^{(j)}$, and bootstrapped data
3. Calculate $\zeta^{(b,m)} = f(\beta^{(b,m)}, \gamma^{(b,m)})$; set $\zeta^{(b)}$ the posterior mode and $CS^{(b)} = [q_{\alpha/2}(\{\zeta^{(b,m)}\}_m), q_{1-\alpha/2}(\{\zeta^{(b,m)}\}_m)]$ from posterior quantiles

2 Calculate Empirical Coverage.

3 Check Convergence.

Algorithm 2: Learning Rate Calibration

Given a target level $\alpha \in (0, 1)$, and current learning rate $\psi^{(j)}$

1. **Bootstrap Quasi-Posterior.**

2. **Calculate Empirical Coverage.**

$$\hat{\mathbb{P}}_\psi = \frac{1}{B} \sum_b \mathbf{1}\left\{\bar{\zeta} \in \mathrm{CS}^{(b)}\right\} \text{ where } \bar{\zeta} = \frac{1}{B} \sum_b \zeta^{(b)}$$

3. **Check Convergence.**

Algorithm 2: Learning Rate Calibration

Given a target level $\alpha \in (0, 1)$, and current learning rate $\psi^{(j)}$

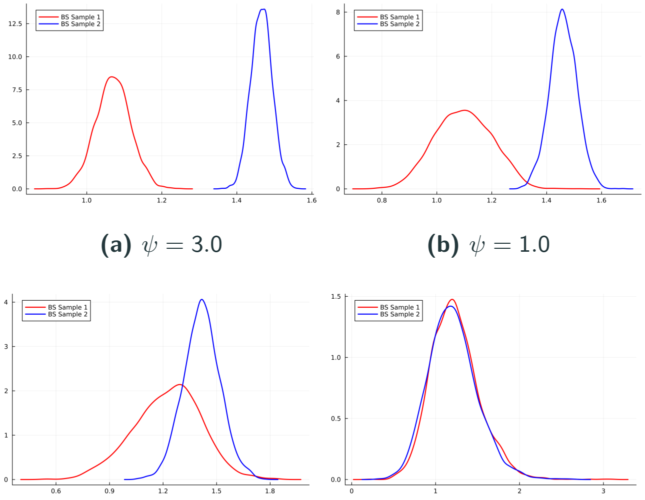
1. **Bootstrap Quasi-Posterior.**

2. **Calculate Empirical Coverage.**

3. **Check Convergence.**

   If $|\hat{\mathbb{P}}_\psi - (1 - \alpha)| < \frac{1}{B}$, stop; otherwise go back to Step 1 with

$$\psi^{(j+1)} = \psi^{(j)} + (j + 1)^{-a}(\hat{\mathbb{P}}_\psi - (1 - \alpha))$$



(a) $\psi = 3.0$

(b) $\psi = 1.0$

(c) $\psi = 0.25$

(d) $\psi = 0.05$

Figure 1: Learning Rate Calibration

Table 1: Clustering Accuracy, RMSE, Band Ratio, and Coverage

Sample Size	Metrics	$\sigma^2 = 1.0$		$\sigma^2 = 2.0$	
		QB	KM	QB	KM
$N = 100, \ T = 10$	AC	79.96	82.89	54.20	58.65
	BR	36.75	20.68	51.18	17.71
	RMSE	30.19	21.90	80.08	90.25
	Coverage	96.56	74.39	97.44	34.33
	RMSE (AE)	3.37	3.46	7.15	7.43
	Coverage (AE)	98.00	-	97.50	-
$N = 100, \ T = 20$	AC	94.53	94.64	64.56	71.61
	BR	24.47	19.07	42.26	16.68
	RMSE	16.26	10.80	53.82	48.94
	Coverage	96.06	89.00	98.06	52.50
	RMSE (AE)	2.38	2.39	4.91	4.98
	Coverage (AE)	96.33	-	97.50	-

DGP: $y_{it} = x_{it}'\beta_{\gamma_i} + \epsilon_{it}, \ \epsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$. Nominal level for coverage: 95%.

Selection of G

- In practice G is unknown and needs to be estimated from the data
- The quasi-Bayesian framework can easily accommodate this, by augmenting the prior:

$$\pi(\boldsymbol{\beta}, \gamma, G) = \pi(\gamma | G)\pi(\boldsymbol{\beta} | G)\pi(G)$$

  - ▷ Previous discussion is the same as $\pi(G^0) = 1$
  - ▷ Example (Uniform): $\pi(G) = \frac{1}{G_{\text{max}}}$ for $G = 1, \ldots, G_{\text{max}}$
  - ▷ Sample $(\boldsymbol{\beta}, \gamma, G)$ from the augmented quasi-posterior ▸ Algorithm (Unknown G)

- Select G as the posterior mode

Asymptotic Properties

Assumption

- Denote $\theta = (\beta, \gamma) \in \Theta$ and $d(\cdot, \cdot)$ some (pseudo) metric on $\Theta \times \Theta$
- $\Pi(\cdot)$ prior, $\Pi_{NT}(\cdot)$ posterior
- $L(\theta) = \overline{L}(\theta)$ the population loss

1. **Assumption 1:** (Identification)

$$\inf_{\theta:\ d(\theta,\theta^0)>\epsilon} L(\theta) - L(\theta^0) > \tilde{\chi}(\epsilon)$$

2. **Assumption 2:** (Uniform Convergence)

$$\sup_{\theta\in\Theta} \left|\overline{L}(\theta) - L(\theta)\right| = o_p(1)$$

3. **Assumption 3:** (Prior Mass) For some non-stochastic sequence $\tilde{c}_{NT}$

$$\Pi\left(\left\{\theta\colon L(\theta) - L(\theta^0) \leq \epsilon\right\}\right) \geq \tilde{c}_{NT}(\epsilon)$$

Consistency

If for some $\delta > 0$

$$\tilde{\chi}(\epsilon) - o(1) - \delta + \frac{\ln \tilde{c}_{NT}(\delta)}{NT\psi} > 0 \ .$$

Then $\mathbb{E}_0 \Pi_{NT} \left(\{\theta \colon d(\theta, \theta^0) > \epsilon\} \right) \to 0$.

- $\tilde{\chi}(\epsilon)$ measures the signal strength ▸ M-estimation (ident)
- $o(1)$ arises from uniform convergence ▸ M-estimation (conv)
- $\tilde{c}_{NT}(\delta)$ presents a tradeoff
  - ▷ We would like δ sufficiently small; but doing this inflates $\tilde{c}_{NT}(\delta)$
  - ▷ In practice δ determined by the loss function ▸ Mixture-Normal

Contraction Rate

1. **Assumption 1:** Identification $\{\theta\colon d(\theta,\theta^0) \geq \epsilon_{NT}\} \subseteq \{\theta\colon L_N(\theta) - L_N(\theta^0) \geq a(\epsilon_{NT})\}$
2. **Assumption 2:** Uniform Convergence $\mathbb{P}_0\left(\sup_{\theta\in\Theta}\left|L_{NT}(\theta) - L_N(\theta)\right| \geq \frac{a(\epsilon_{NT})}{5}\right) = b_{NT}$
3. **Assumption 3:** Smoothness $0 < \tilde{c}_M < \infty$ such that $\left|L_N(\theta) - L_N(\tilde{\theta})\right| \leq \tilde{c}_M d(\theta,\tilde{\theta})$ for any $\theta, \tilde{\theta} \in \Theta$
4. **Assumption 3:** Prior Mass $\Pi\left(\{\theta\colon d(\theta,\theta^0) \leq \epsilon_{NT}\}\right) \geq c(\epsilon_{NT})$ for some non-stochastic sequence c_{NT} possibly depending on the sample size

If $b_{NT} = o_p(1)$ and $a(\epsilon_{NT}) + \frac{5}{2}\frac{\ln \tilde{c}_{NT}(a(\epsilon_{NT})/5\tilde{c}_M)}{NT\psi} > 0$, then

$$\mathbb{E}_0\Pi_{NT}\left(\{\theta\colon d(\theta,\theta^0) > \epsilon_{NT}\}\right) \to 0$$

▸ M-estimation

Empirical Studies

Heterogeneous Wage Cyclicality

- Extensive empirical literature on heterogeneous wage cyclicality
- Dominant approach: group workers by observables
  - ▷ Example: Poor households have high MPC & high risks (Patterson, 2023)
- Revisit using six waves of PSID data (1999-2009), estimate wage cyclicality

$$\ln y_{it}^{resid} = Y_t \beta_{\gamma_i} + \epsilon_{it}$$

  - ▷ Residualized on a large set of dummies: cohort, education, race, loc, fam size

Table 2: Household Characteristics (Quasi-Bayes)

	Group 1	Group 2	Group 3
Age	45	45	45
Education	14	14	14
Family Size	4	3	4
Total Family Income	145120	117048	129835
Transfer Income	884	2565	3607
Labor Income (Head)	45069	73857	80745
Hours Worked (Head)	2253	2247	2170
House Value	244989	245389	309352
Stocks Value	9564	46405	99193
Pensions & Annuities	46000	46016	87341
Cash	27291	27924	49622
Bonds	12545	15270	22379
Financial Assets	228621	161973	280191
Total Assets	575585	318244	488282
Monthly Rent	8140	9659	14400
Total Consumption	40394	38204	40268
$CS(\beta_g)$ (%)	[17.19,43.72]	[-6.53,-1.76]	[-49.73,-29.27]

▸ K-means

Summary

1. Existing methods that recover latent group structure suffer from bias and poor coverage
2. This paper proposes a generic quasi-Bayesian clustering and provides statistical rationale: consistency and contraction rate of quasi-posterior
   - Easy to implement: blocked Gibbs
   - Large class of loss functions: M-estimation, GMM, etc
   - Large class of priors: mixture-normal, graphical sparsity etc
3. Bootstrap-based learning rate calibration significantly improves coverage
4. Application casts doubt on the hand-to-mouth story of shock amplification

Thank you!

Algorithm 3: Posterior Sampling with Unknown G

Given $\psi > 0$, do Blocked Gibbs

1. **Sample Group Membership** For $i = 1, \ldots, N$

   Let $G = |\mathcal{C}_{-i}|$ be the number of groups when unit i is excluded. Sample γ_i from

$$\pi_{NT}\left(\gamma_i | \gamma_{-i}, \boldsymbol{\beta}\right) \propto \begin{cases} \exp\left[-T\psi l_{iT}(\beta_g)\right]\left(|\mathcal{C}_{g,-i}| + \alpha\right) & g \in \{1, \ldots, G\} \\ \exp\left[-T\psi l_{iT}(\beta_g)\right] \frac{V_{N,G+1}}{V_{N,G}} \frac{\alpha}{H} & g \in \{G+1, \ldots, G+H\} \end{cases} .$$

   Whenever $g > G$, a new β_g is sampled from its prior.

2. **Sample Grouped Parameters** as in Algorithm 1

back

Algorithm 4: Learning Rate Calibration with Unknown G

Given a target level $\alpha \in (0,1)$, and current learning rate $\psi^{(j)}$

1 Bootstrap Quasi-Posterior.

For $b = 1, \ldots, B$,

1. Draw bootstrapped data $\{(Y_i^{(b)}, X_i^{(b)})\}_{i=1}^N$ as in Algorithm 2
2. Sample $\left\{(\beta^{(b,m)}, \gamma^{(b,m)})\right\}_{m=1}^M$ using **Algorithm 3**, $\psi^{(j)}$, and bootstrapped data
3. Let $\mathbf{G}$ be the set of group number in posterior samples, for each $G \in \mathbf{G}$: Calculate $\zeta^{(b,m,G)}$, $CS^{(b,G)}$ as in Algorithm 2

2 Calculate Empirical Coverage for each G as in Algorithm 2, gives $\hat{\mathbb{P}}_{\psi,G}$

3 Check Convergence for $\min\{\hat{\mathbb{P}}_{\psi,G}\}$ as in Algorithm 2

back

Linear Model: Identification

Loss function

$$\overline{L}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - x_{it}'\beta\right)^2$$

One identification condition is for any $\epsilon > 0$

$$\chi(\epsilon) \geq \epsilon \min_i \underline{\lambda}_i > 0$$

where $\underline{\lambda}_i$ is the smallest eigenvalue of $\mathbb{E}[x_{it}x_{it}']$

M-estimation: Identification

More generally, for loss function

$$\overline{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} h(w_{it}; \beta_{\gamma_i})$$

Identification requires that for any $\epsilon > 0$

$$\min_i \left[\inf_{\|\beta - \beta^0_{\gamma^0_i}\|^2 > \epsilon} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[h(w_{it}; \beta) - h(w_{it}; \beta^0_{\gamma^0_i}) \right] \right] = \chi(\epsilon) > 0$$

▸ back

Linear Model: Uniform Convergence

One sufficient condition is to consider the convergence of individual loss: For any $\epsilon > 0$, we have as $N, T \to \infty$,

$$\mathbb{P}_0 \left(\max_i \sup_{\beta \in \mathcal{B}} \left| \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - x_{it}' \beta \right)^2 - \mathbb{E} \left(y_{it} - x_{it}' \beta \right)^2 \right| \geq \epsilon \right) = o(1) \ .$$

which in turn is determined by the rate at which

$$\frac{1}{T} \sum_{t=1}^{T} x_{it} x_{it}' \to \mathbb{E} x_{it} x_{it}', \quad \frac{1}{T} \sum_{t=1}^{T} \epsilon_{it}^2 \to \mathbb{E} \epsilon_{it}^2, \quad \frac{1}{T} \sum_{t=1}^{T} x_{it} \epsilon_{it} \to 0$$

$\Longrightarrow$ We need primitive assumptions on the dependency and moments

M-estimation: Uniform Convergence

1. $\beta_g^0 \in \mathcal{B}$ for all $g = 1, \ldots, G^0$ where $\mathcal{B}$ is a convex compact subset of $\mathbb{R}^d$.
2. $\{w_{it}\}_{t=1,\ldots,T}$ are independent across i. For each i, it is stationary strong mixing with mixing coefficient α_i, and $\alpha \equiv \max_i \alpha_i$ satisfies $\alpha(\tau) \leq c_\alpha \rho^\tau$ for some $c_\alpha > 0$ and $\rho \in (0,1)$.
3. There exists a non-negative function $M(\cdot)$ such that $\sup_{\beta\in\mathcal{B}} |h(w;\beta)| \leq M(w)$, and $|h(w;\beta) - h(w;\tilde{\beta})| \leq M(w)\|\beta - \tilde{\beta}\|$ for all $\beta, \tilde{\beta} \in \mathcal{B}$. Moreover, $\sup_i \mathbb{E}\left|M(w_{it})\right|^q < c_M$ for some $c_M < \infty$ and $q \geq 6$.
4. Assume that $N^2 = O(T^{q/2-1})$ where $q \geq 6$ is the same constant in 3.

Then for any $\epsilon > 0$, we have as $N, T \to \infty$,

$$\mathbb{P}_0\left(\max_i \sup_{\beta\in\mathcal{B}} \left|\frac{1}{T}\sum_{t=1}^{T} h(w_{it};\beta) - \mathbb{E}h(w_{it};\beta)\right| \geq \epsilon\right) = o(N^{-1}) \; .$$

back

Assumption Mixture-Normal Prior

Assume finite mixture prior on the γ, and normal prior on β, we have

$$\begin{aligned}\Pi\left(\left\{\theta\colon d_{MS}(\theta,\theta^0)\le\epsilon\right\}\right) &\ge \exp\left[-C(N\ln G^0+|\ln\epsilon|)\right]\\ \Pi\left(\left\{\theta\colon d_{H}(\theta,\theta^0)\le\epsilon\right\}\right) &\ge \exp\left[-C(N\ln G^0+|\ln\epsilon|)\right]\end{aligned}$$

where

$$d_{MS}(\theta,\tilde\theta) = \frac{1}{N}\sum_{i=1}^{N}\|\beta_{\gamma_i}-\tilde\beta_{\tilde\gamma_i}\|^2$$

$$d_H(\theta,\tilde\theta) = \max\left\{\max_{g\in\{1,\dots,G_2\}}\min_{\tilde g\in\{1,\dots,G_1\}}\|\tilde\beta_{\tilde g}-\beta_g\|,\ \max_{\tilde g\in\{1,\dots,G_1\}}\min_{g\in\{1,\dots,G_2\}}\|\tilde\beta_{\tilde g}-\beta_g\|\right\}$$

It does not restrict the excess loss $L(\theta)-L(\theta^0)$! This is why smoothness condition is required.

back

Assumption Group Structure

1. G^0 is fixed and $\min_{g \neq l} \|\beta_g^0 - \beta_l^0\| > 0$ for all $g, l \in \{1, \ldots, G^0\}$.
2. For all $g \in \{1, \ldots, G^0\}$, $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{\gamma_i^0 = g\} > \underline{\eta} > 0$.
3. $\Pi(G = k) > 0$.
4. $\Pi(\{\mathcal{C} : |\mathcal{C}_g|/N > 0\}) = 1$.
5. $\Pi\left(\beta_g = \beta_l | G = k\right) = 0$ for $1 \leq g < l \leq k$.
6. $\inf_{\|\beta - \beta^0_{\gamma_i^0}\| \geq \epsilon} \mathbb{E}[l_{iT}(\beta) - l_{iT}(\beta^0_{\gamma_i^0})] > \check{\chi}(\epsilon) > 0$.

back

Result M-estimation

Under Assumptions on identification, uniform convergence, prior and group structure

1. for any $\epsilon > 0$, as N, T go to infinity

$$\Pi_{NT}\left(\{\zeta : d(\theta, \theta^0) > \epsilon\}\right) \xrightarrow{P_0} 0 \tag{1}$$

2. for $\epsilon_{NT} = O(T^{-1})$, as N, T go to infinity

$$\Pi_{NT}\left(\{\theta : d(\theta, \theta^0) > \epsilon_{NT}\}\right) \xrightarrow{P_0} 0 \tag{2}$$

for $d(\cdot, \cdot)$ being d_{MS} and d_H.

back

Empirical Results: K-means

Table 3: Household Characteristics (Quasi-Bayes)

	Group 1	Group 2	Group 3	Group 4
Age	44	44	46	44
Education	14	14	14	14
Family Size	3	4	3	3
Total Family Income	123686	123952	110708	125468
Transfer Income	2459	2034	2154	5171
Labor Income	80504	70918	67832	80701
Hours Worked	2263	2191	2245	2246
House Value	254970	233456	242061	271904
Stocks Value	55928	47058	35643	75864
Pensions & Annuities	40637	44122	49785	63978
Cash	28568	32506	26542	34103
Bonds	19086	12487	13222	18781
Financial Assets	190936	182624	138470	208521
Total Assets	351140	360056	298291	368299
Monthly Rent	11014	8331	9040	12137
Total Consumption	38335	38148	37878	40352
$CS(\beta_g)$ (%)	[-15.67,-14.23]	[23.86,27.78]	[0.77,1.93]	[-41.44,-37.31]

back

Atchadé, Y. A. (2017). **On the contraction properties of some high-dimensional quasi-posterior distributions.** *The Annals of Statistics*, *45*(5).

Auclert, A. (2019). **Monetary Policy and the Redistribution Channel.** *American Economic Review*, *109*(6), 2333–2367.

Bils, M. J. (1985). **Real Wages over the Business Cycle: Evidence from Panel Data.** *Journal of Political Economy*, *93*(4), 666–689.

Bonhomme, S., & Manresa, E. (2015). **Grouped Patterns of Heterogeneity in Panel Data.** *Econometrica*, *83*(3), 1147–1184.

Busch, C., Domeij, D., Guvenen, F., & Madera, R. (2022). **Skewed Idiosyncratic Income Risk over the Business Cycle: Sources and Insurance.** *American Economic Journal: Macroeconomics*, *14*(2), 207–242.

Cheng, X., Schorfheide, F., & Shao, P. (2019). **Clustering for Multi-Dimensional Heterogeneity.**

- Chernozhukov, V., & Hong, H. (2003). **An MCMC approach to classical estimation.** *Journal of Econometrics*, *115*(2), 293–346.
- Duan, L. L., & Dunson, D. B. (2021). **Bayesian Distance Clustering.** *Journal of Machine Learning Research*, *22*(1), 10228–10254.
- Figueiredo, A. (2022). **Wage Cyclicality and Labor Market Sorting.** *American Economic Review: Insights*, *4*(4), 425–442.
- Guvenen, F., Ozkan, S., & Song, J. (2014). **The Nature of Countercyclical Income Risk.** *Journal of Political Economy*, *122*(3), 621–660.
- Ishwaran, H., & James, L. F. (2001). **Gibbs Sampling Methods for Stick-Breaking Priors.** *Journal of the American Statistical Association*, *96*(453), 161–173.
- Kim, J.-Y. (2002). **Limited information likelihood and Bayesian analysis.** *Journal of Econometrics*, *107*(1), 175–193.

References iii

- Liu, R., Shang, Z., Zhang, Y., & Zhou, Q. (2020). **Identification and Estimation in Panel Models with Overspecified Number of Groups.** *Journal of Econometrics*, *215*(2), 574–590.
- Low, H., Meghir, C., & Pistaferri, L. (2010). **Wage Risk and Employment Risk over the Life Cycle.** *American Economic Review*, *100*(4), 1432–1467.
- Miller, J. W. (2021). **Asymptotic normality, concentration, and coverage of generalized posteriors.** *The Journal of Machine Learning Research*, *22*(1), 168:7598–168:7650.
- Miller, J. W., & Harrison, M. T. (2018). **Mixture Models With a Prior on the Number of Components.** *Journal of the American Statistical Association*, *113*(521), 340–356.
- Patterson, C. (2023). **The Matching Multiplier and the Amplification of Recessions.** *American Economic Review*, *113*(4), 982–1012.

- Petrova, K. (2019). **A quasi-Bayesian local likelihood approach to time varying parameter VAR models.** *Journal of Econometrics*, *212*(1), 286–306.
- Ren, Y., Zhu, X., Lu, X., & Hu, G. (2022). **Graphical Assistant Grouped Network Autoregression Model: A Bayesian Nonparametric Recourse.** *Journal of Business & Economic Statistics*, 1–15.
- Richardson, S., & Green, P. J. (1997). **On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion).** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(4), 731–792.
- Rigon, T., Herring, A. H., & Dunson, D. B. (2023). **A generalized Bayes framework for probabilistic clustering.** *Biometrika*, asad004.
- Su, L., & Ju, G. (2018). **Identifying latent grouped patterns in panel data models with interactive fixed effects.** *Journal of Econometrics*, *206*(2), 554–573.

References v

- Su, L., Shi, Z., & Phillips, P. C. B. (2016). **Identifying Latent Structures in Panel Data.** *Econometrica*, *84*(6), 2215–2264.
- Wang, W., Phillips, P. C. B., & Su, L. (2018). **Homogeneity Pursuit in Panel Data Models: Theory and Application.** *Journal of Applied Econometrics*, *33*(6), 797–815.
- Zhang, B. (2023, March). **Incorporating Prior Knowledge of Latent Group Structure in Panel Data Models.**
- Zhang, Y., Wang, H. J., & Zhu, Z. (2019). **Quantile-regression-based clustering for panel data.** *Journal of Econometrics*, *213*(1), 54–67.