Identifying Latent Structures in Panel Data

Liangjun Su, Zhentao Shi, and Peter C.B. Phillips
Singapore Management University, Yale Univeristy, and Yale University

The 20th International Panel Data Conference, Tokyo

July 9, 2014

A motivating example: heterogenous trending behavior of real GDP per capita

- Model:

$$y_{it} = \beta_i \left(t/T \right) + \mu_i + u_{it}, \;\; i = 1, ..., N, \;\; t = 1, ..., T$$

- Data: World Bank annual data from 1960-2012 for 92 countries ($N = 92, \; T = 53$).
- Nonparametric sieve estimation with cubic B-spline, $\#\text{knot} = 3$.

Heterogenous trending behavior of real GDP per capita

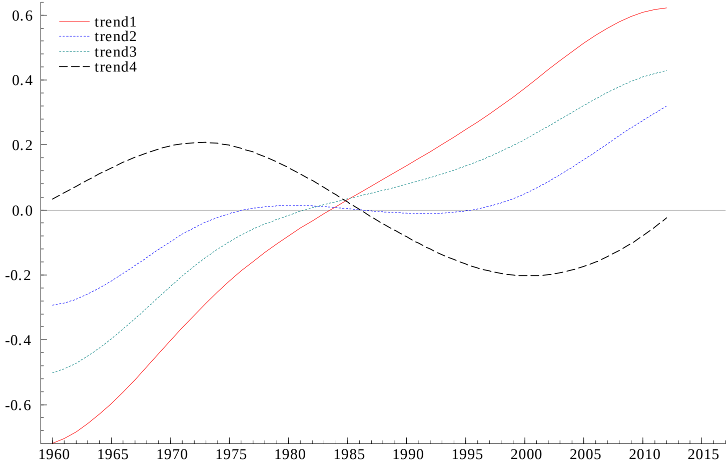

Figure: Four estimated trends of real GDP per capita (logarithm and demeaned) for countries in each of the four estimated groups

Heterogenous trending behavior of real GDP per capita

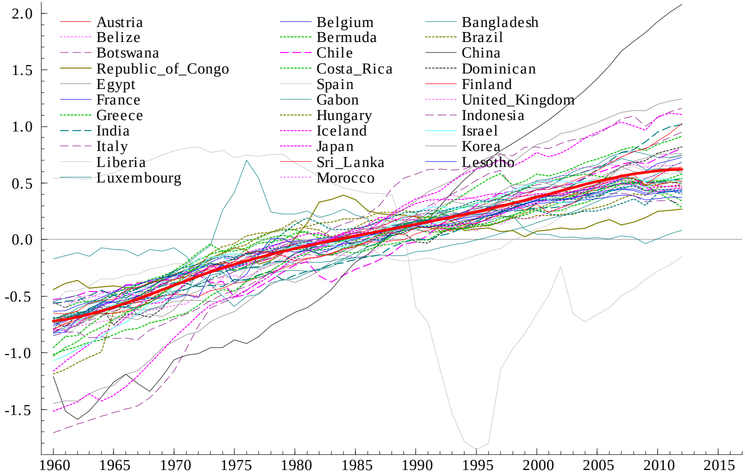

Figure: Trending behavior of the real GDP per capita for countries in Group 1

Heterogenous trending behavior of per capita GDP

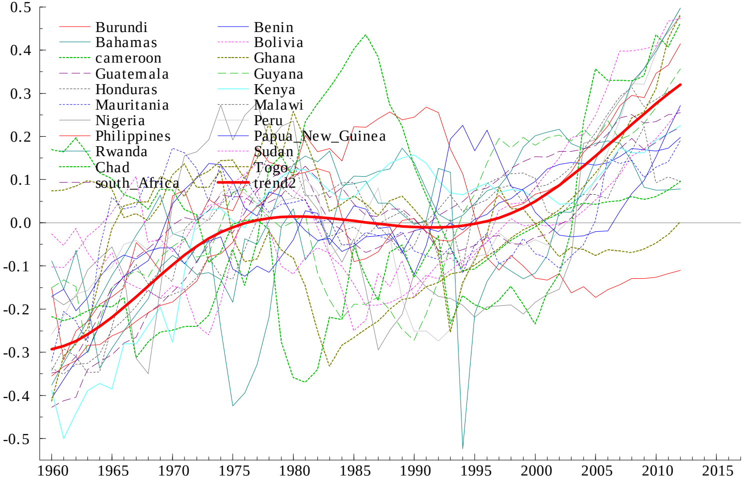

Figure: Trending behavior of the real GDP per capita for countries in Group 2

Heterogenous trending behavior of real GDP per capita

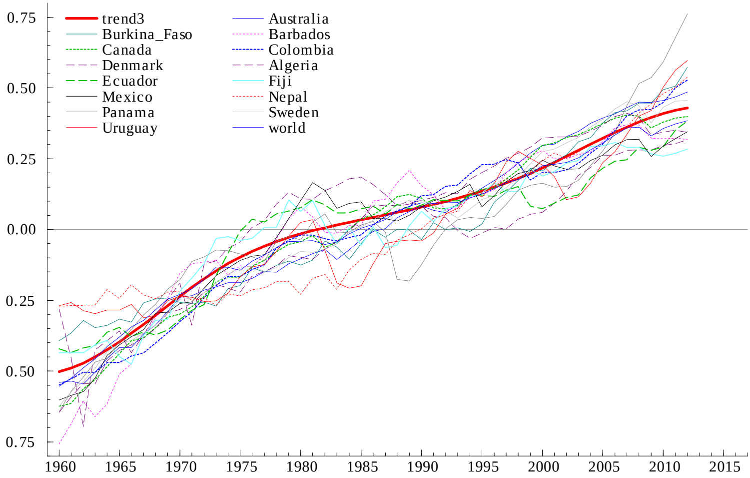

Figure: Trending behavior of the real GDP per capita for countries in Group 3

Heterogenous trending behavior of real GDP per capita

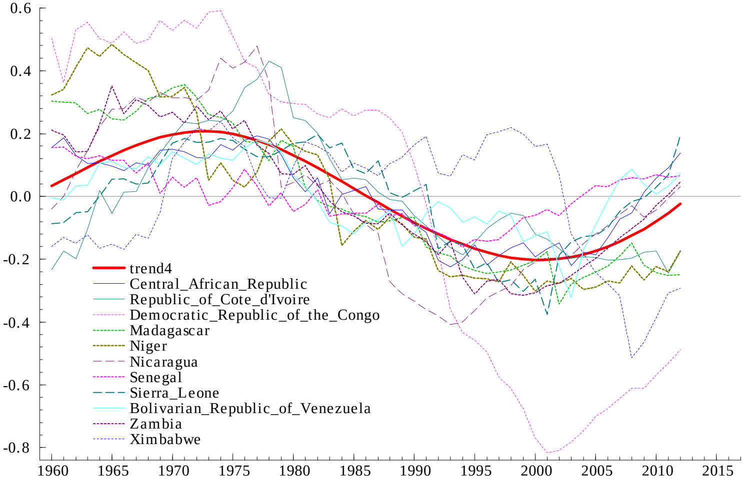

Figure: Trending behavior of the real GDP per capita for countries in Group 4

Outline of the Presentation

- The Model and Literature
- Penalized Least Squares (PLS) Estimation
- Penalized GMM Estimation
- Monte Carlo Simulations
- Empirical Application
- Extension to Models with Cross Section Dependence
- Conclusions

$$y_{it} = \beta_i^{0\prime} x_{it} + \mu_i + u_{it} \tag{1}$$

where
x_{it} is a $p \times 1$ vector of explanatory variables,
μ_i is an individual fixed effect,
u_{it} is the idiosyncratic error term with zero mean, and

$$\beta_i^0 = \begin{cases} \alpha_1^0 & \text{if } i \in G_1^0 \\ \vdots & \vdots \\ \alpha_{K_0}^0 & \text{if } i \in G_{K_0}^0 \end{cases}. \tag{2}$$

Here $\alpha_j^0 \neq \alpha_k^0$ for any $j \neq k$, $\cup_{k=1}^{K_0} G_k^0 = \{1, 2, ..., N\}$, and $G_k^0 \cap G_j^0 = \varnothing$ for any $j \neq k$. Let $N_k = \# G_k^0$.

- **Latent heterogeneity** is an important phenomenon in panel data analysis. Neglecting it can lead to inconsistent estimation and misleading inference; see Hsiao (2003, Chapter 6). But it is challenging to model latent heterogeneity in empirical research: do we allow for heterogeneous slope coefficients in a regression?
- **Complete slope homogeneity**: Easy estimation and inference, but frequently questioned and rejected in empirical studies.
- **Complete slope heterogeneity:**
  1. Random coefficient model: parameters are assumed to be independent draws from a common distribution – see Hsiao and Pesaran (2008).
  2. Use Bayesian methods to shrink the individual slope estimates towards the overall mean – see Maddala, Trost, Li, and Joutz (1997).
  3. Parameterize individual slope coefficients as a function of observed characteristics – see Durlauf, Kourtellos, and Minkin (2001) and Browning, Ejrnæs, and Alvarez (2010).
  4. Estimate the individual slope coefficients using heterogenous time series regressions for each individual.

- **Panel structure model:**
  - individuals belong to a number of homogeneous groups or clubs within a broadly heterogeneous population.
  - regression parameters are the same within each group but differ across groups.
  - Two essential questions are:
    how to determine the unknown number of groups;
    how to identify the individual's group membership.

1. Bester and Hansen (2009) consider a panel structure model where individuals are grouped according to some external classification, geographic location, or observable explanatory variables. So the group structure is *completely known* to the researcher.
2. Several approaches have been proposed to determine an *unknown* group structure in modeling unobserved slope heterogeneity in panels.
   - Mixture models/distributions: Sun (2005), Kasahara and Shimotsu (2009), and Browning and Carro (2011), model membership probabilities.
   - K-means algorithm: Lin and Ng (2012) and Sarafidis and Weber (2011) perform conditional clustering to estimate linear panel structure models but provide no asymptotic properties. Bonhomme and Manresa (2014) introduce time-varying grouped patterns of heterogeneity in linear panel data models based on K-means algorithm, and study the asymptotic properties. Both require that N and T pass to infinity jointly.

- The present paper proposes a new method for estimation and inference in panel models when
  - the slope parameters are heterogenous across groups,
  - individual group membership is unknown,
  - classification is to be determined empirically.
- It is an automated data-determined procedure and does not require the specification of any modeling mechanism for the unknown group structure.
- It involves a new variant of Lasso (Tibshirani, 1996).
- Like Lin and Ng (2012), Bonhomme and Manresa (2014) and Phillips and Sul (2007), we assume that $(N, T) \to \infty$ jointly. But in our asymptotic theory T can pass to infinity at a very slow rate, even a slowly varying rate such as $O\left((\ln N)^{1+\epsilon}\right)$ for any $\epsilon > 0$ in the case of uniformly bounded regressors.

1. Motivated by the key feature of Lasso to handle **parameter sparsity**. $\{\beta_i, i = 1, ..., N\}$ versus $\{\alpha_k, k = 1, ..., K_0\}$.
2. Contribute to the literature on **fused Lasso** (e.g.,Tibshirani et al. (2005)). No natural ordering across individuals.
3. Additive-multiplicative penalty terms: $\Rightarrow$ **Classifier-Lasso** or **C-Lasso**.
4. Two classes of estimates: PLS and PGMM. In either case, we show **uniform classification consistency**. Such a uniform result allows us to establish an **oracle property** for the PLS estimator. But our PGMM estimator generally does not have the oracle property.
5. K_0 is unknown: a BIC-type information criterion is proposed.
6. Easy to extend to nonlinear models such as discrete choice models, to SP and NP models, to models where only a subset of parameters are allowed to be group-specific, etc.

- **Economic Growth Convergence:**
  Much of the recent literature on economic growth addresses sources of possible heterogeneity, including the occurrence of multiple steady states and history-dependence in growth trajectories - see Deissenberg, Feichtinger, Semmler, and Wirl (2001) and Durlauf, Johnson and Temple (2005) and Eberhardt and Teal (2011) for overviews of the relevant growth theory and empirics.
- **Subsample Studies of Stability:** Much empirical research is concerned with studying the stability of certain regression coefficients over subsamples of the data.
- **Panel Unit Root Grouping:** Our methodology can be used to classify a subgroup of unit-root processes in the panel from a wider class of stationary and nonstationary processes.

Penalized Least Squares Estimation

Penalized Least Squares Estimation

Within-group Estimation

- Model: $y_{it} = \beta_i^{0\prime} x_{it} + \mu_i + u_{it}$.
- Define

$$Q_{0,NT}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \beta_i' x_{it} - \mu_i \right)^2 .$$

- Concentrate $\boldsymbol{\mu}$ out:

$$Q_{1,NT}(\boldsymbol{\beta}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{y}_{it} - \beta_i' \tilde{x}_{it} \right)^2 ,$$

where $\tilde{x}_{it} = x_{it} - T^{-1} \sum_{t=1}^{T} x_{it}$ and $\tilde{y}_{it} = y_{it} - T^{-1} \sum_{t=1}^{T} y_{it}$.

- PLS objective function

$$Q_{1NT,\lambda_1}^{(K_0)}(\boldsymbol{\beta},\boldsymbol{\alpha}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\|\beta_i - \alpha_k\|, \tag{3}$$

where $\lambda_1 = \lambda_{1NT}$ is a tuning parameter.

- C-Lasso estimates: $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_K)$ and $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)$.
- Numerical algorithm: a sequence of convex problems.

Assumption A1. (i) $\hat{Q}_{i,\tilde{x}\tilde{u}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}u_{it} = O_P\left(T^{-1/2}\right)\ \forall\ i.$

(ii) $\hat{Q}_{i,\tilde{x}\tilde{x}} = \frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}' \xrightarrow{P} Q_{i,\tilde{x}\tilde{x}} > 0\ \forall\ i.\ \exists\ \underline{c}_{\tilde{x}\tilde{x}}$ *such that* $\underline{\lim}_{(N,T)\to\infty}\min_{1\le i\le N}\mu_{\min}(\hat{Q}_{i,\tilde{x}\tilde{x}}) \ge \underline{c}_{\tilde{x}\tilde{x}} > 0.$

(iii) $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{Q}_{i,\tilde{x}\tilde{u}}\right\|^2 = O_P\left(T^{-1}\right).$

(iv) $N_k/N \to \tau_k \in (0,1)$ *for each* $k = 1, ..., K_0$ *as* $N \to \infty.$

(v) $\lambda_1 \to 0$ *as* $(N,T) \to \infty.$

Theorem

Suppose that Assumption A1 holds. Then

(i) $\hat{\beta}_i - \beta_i^0 = O_P\left(T^{-1/2} + \lambda_1\right)$ *for* $i = 1, 2, ..., N,$

(ii) $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\beta}_i - \beta_i^0\right\|^2 = O_P\left(T^{-1}\right),$

(iii) $\left(\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)}\right) - (\alpha_1^0, ..., \alpha_{K_0}^0) = O_P\left(T^{-1/2}\right),$

where $(\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)})$ *is a suitable permutation of* $(\hat{\alpha}_1, ..., \hat{\alpha}_{K_0}).$

- Note that $Q_{1NT,\lambda_1}^{(K_0)}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^{N} Q_{1iNT,\lambda_1}^{(K_0)}(\beta_i,\boldsymbol{\alpha})$, where

$$\begin{aligned} Q_{1iNT,\lambda_1}^{(K_0)}(\beta_i,\boldsymbol{\alpha}) &= Q_{1NT,i}(\beta_i) + \lambda_1 \Pi_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \\ Q_{1NT,i}(\beta_i) &= \frac{1}{T}\sum_{t=1}^{T}\left(\tilde{y}_{it} - \beta_i'\tilde{x}_{it}\right)^2. \end{aligned}$$

- **Pointwise convergence**:

$$\begin{aligned} & Q_{1iNT,\lambda_1}^{(K_0)}(\hat{\beta}_i,\hat{\boldsymbol{\alpha}}) - Q_{1iNT,\lambda_1}^{(K_0)}(\beta_i^0,\hat{\boldsymbol{\alpha}}) \\ = & \; Q_{1NT,i}(\hat{\beta}_i) - Q_{1NT,i}(\beta_i^0) \\ & + \lambda_1 \left\{ \Pi_{k=1}^{K_0} \|\hat{\beta}_i - \hat{\alpha}_k\| - \Pi_{k=1}^{K_0} \|\beta_i^0 - \hat{\alpha}_k\| \right\} \\ \leq & \; 0 \end{aligned}$$

Given $\hat{\boldsymbol{\alpha}}$, $\hat{\beta}_i$ must minimize $Q_{1iNT,\lambda_1}^{(K_0)}(\beta_i,\hat{\boldsymbol{\alpha}})$ with respect to β_i.

- **Mean-square convergence:** relies on the observation

$$Q_{1NT,\lambda_1}^{(K_0)}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}\right) - Q_{1NT,\lambda_1}^{(K_0)}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) \leq 0. \quad (4)$$

We prove it by showing that $\forall\ \epsilon^* > 0$, $\exists\ L = L(\epsilon^*)$ s.t. the above inequality cannot hold with probability $1 - \epsilon^*$ if $\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\beta}_i - \beta_i^0\right\|^2 \geq L/T$.

- **Convergence of** $\left(\hat{\alpha}_{(1)}, \ldots, \hat{\alpha}_{(K_0)}\right)$: relies on the observation

$$P_{NT}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}\right) - P_{NT}\left(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}^0\right) \leq 0 \quad (5)$$

where $P_{NT}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\|\beta_i - \alpha_k\|$, and the fact that the convergence rate of $\hat{\alpha}_k$ (up to permutation) fully depends on the mean-square convergence rate of $\hat{\beta}_i$.

Define

$$\begin{aligned}
\hat{G}_k &= \left\{ i \in \{1, 2, ..., N\} : \hat{\beta}_i = \hat{\alpha}_k \right\} \text{ for } k = 1, ..., K_0, \\
\hat{E}_{kNT,i} &= \left\{ i \notin \hat{G}_k \mid i \in G_k^0 \right\}, \ \hat{F}_{kNT,i} = \left\{ i \notin G_k^0 \mid i \in \hat{G}_k \right\}, \\
\hat{E}_{kNT} &= \cup_{i \in G_k^0} \hat{E}_{kNT,i}, \text{ and } \hat{F}_{kNT} = \cup_{i \in \hat{G}_k} \hat{F}_{kNT,i}.
\end{aligned}$$

Definition 1. (Uniform consistency of classification) We say that a classification method is *individually consistent* if $P\left(\hat{E}_{kNT,i}\right) \to 0$ as $(N, T) \to \infty$ for each $i \in G_k^0$ and $k = 1, ..., K_0$, and $P\left(\hat{F}_{kNT,i}\right) \to 0$ as $(N, T) \to \infty$ for each $i \in \hat{G}_k$ and $k = 1, ..., K_0$. It is *uniformly consistent* if $P\left(\cup_{k=1}^{K_0} \hat{E}_{kNT}\right) \to 0$ and $P\left(\cup_{k=1}^{K_0} \hat{F}_{kNT}\right) \to 0$ as $(N, T) \to \infty$.

Assumption A2. (i) $T\lambda_1 \to \infty$ *and* $T\lambda_1^4 \to c_0 \in [0, \infty)$ *as* $(N, T) \to \infty$.
(ii) *For any* $c > 0$, $N \max_{1 \leq i \leq N} P\left(\left\| T^{-1} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{u}_{it} \right\| \geq c\sqrt{\lambda_1}\right) \to 0$ *as* $(N, T) \to \infty$.

Theorem

Suppose that Assumptions A1-A2 hold. Then
(i) $P\left(\cup_{k=1}^{K_0} \hat{E}_{kNT}\right) \leq \sum_{k=1}^{K_0} P\left(\hat{E}_{kNT}\right) \to 0$ *as* $(N, T) \to \infty$,
(ii) $P\left(\cup_{k=1}^{K_0} \hat{F}_{kNT}\right) \leq \sum_{k=1}^{K_0} P\left(\hat{F}_{kNT}\right) \to 0$ *as* $(N, T) \to \infty$.

Assumption A3. (i) $\bar{\Phi}_k \equiv \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \xrightarrow{P} \Phi_k > 0$ *as* $(N, T) \to \infty$.

(ii) $\frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{u}_{it} - \mathbb{B}_{kNT} \xrightarrow{D} N(0, \Psi_k)$ *as* $(N, T) \to \infty$ *where* $\mathbb{B}_{kNT} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \mathbb{E}(x_{it} \tilde{u}_{it})$ *is either* 0 *or* $O(\sqrt{N_k / T})$ *depending on whether* x_{it} *is strictly exogenous.*

Theorem

Suppose that Assumptions A1-A3 hold. Then $\sqrt{N_k T} (\hat{\alpha}_k - \alpha_k^0) - \bar{\Phi}_k^{-1} \mathbb{B}_{kNT} \xrightarrow{D} N(0, \Phi_k^{-1} \Psi_k \Phi_k^{-1})$ *for* $k = 1, ..., K_0$.

- If the individual's group identity is known, the WG estimator of α_k^0 is

$$\bar{\alpha}_k = \left(\sum_{i \in G_k^0} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \sum_{i \in G_k^0} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it}$$

and then $\sqrt{N_k T} \left(\bar{\alpha}_k - \alpha_k^0 \right) - \bar{\Phi}_k^{-1} \mathbb{B}_{kNT} \xrightarrow{D} N \left(0, \Phi_k^{-1} \Psi_k \Phi_k^{-1} \right)$.

- The proof is done by the inspection of the Karush-Kuhn-Tucker (KKT) optimality conditions based on subdifferential calculus (e.g., Bertsekas, 1995).
- Then we show that $\sqrt{N_k T} \left(\hat{\alpha}_k - \alpha_k^0 \right) = \sqrt{N_k T} (\hat{\alpha}_{\hat{G}_k} - \alpha_k^0) + o_P(1)$, where $\hat{\alpha}_{\hat{G}_k}$ is the post-Lasso estimator:

$$\hat{\alpha}_{\hat{G}_k} = \left(\sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{y}_{it}.$$

Theorem

Suppose that Assumptions A1-A3 hold. Then
$\sqrt{N_k T}\left(\hat{\alpha}_{\hat{G}_k} - \alpha_k^0\right) - \bar{\Phi}_k^{-1}\mathbb{B}_{kNT} \xrightarrow{D} N(0, \Phi_k^{-1}\Psi_k\Phi_k^{-1})$ *for* $k = 1, ..., K_0$.

- Consider the following PLS criterion

$$Q_{1NT,\lambda_1}^{(K)}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K} \|\beta_i - \alpha_k\|, \qquad (6)$$

where $1 \leq K \leq K_{\max}$. C-Lasso estimates: $\{\hat{\beta}_i(K, \lambda_1), \hat{\alpha}_k(K, \lambda_1)\}$ of $\{\beta_i, \alpha_k\}$. As above, we can classify individual i into group $\hat{G}_k(K, \lambda_1)$ if and only if $\hat{\beta}_i(K, \lambda_1) = \hat{\alpha}_k(K, \lambda_1)$.
- Define the post-Lasso estimate of α_k^0 by

$$\hat{\alpha}_{\hat{G}_k(K,\lambda_1)} = \left(\sum_{i \in \hat{G}_k(K,\lambda_1)} \sum_{t=1}^{T} \tilde{x}_{it}\tilde{x}_{it}' \right)^{+} \sum_{i \in \hat{G}_k(K,\lambda_1)} \sum_{t=1}^{T} \tilde{x}_{it}\tilde{y}_{it}. \qquad (7)$$

Let $\hat{\sigma}^2_{\hat{G}(K,\lambda_1)} = \frac{1}{NT} \sum_{k=1}^{K} \sum_{i \in \hat{G}_k(K,\lambda_1)} \sum_{t=1}^{T} [\tilde{y}_{it} - \tilde{\alpha}'_{\hat{G}_k(K,\lambda_1)} \tilde{x}_{it}]^2$.
- Information criterion:

$$IC_1(K, \lambda_1) = \ln\left[\hat{\sigma}^2_{\hat{G}(K,\lambda_1)}\right] + \rho_{1NT}\, pK, \qquad (8)$$

1. Mixed Panel Structure Models:

$$y_{it} = \beta_{(1)}^{0\prime} x_{it(1)} + \beta_{i(2)}^{0\prime} x_{it(2)} + \mu_i + u_{it}, \tag{9}$$

where $\beta_{i(2)}^{0} = \alpha_k^0$ if $i \in G_k^0$ where $k = 1, ..., K_0$ and $G_1^0, ..., G_{K_0}^0$ form a partition for $\{1, 2, ..., N\}$. See Pesaran, Shin, and Smith (1999).

2. Nonlinear Panel Data Models: Following Bester and Hansen (2009), we can consider

$$Q_{1,NT}(\theta, \boldsymbol{\mu}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varphi(w_{it}, \theta, \mu_i), \tag{10}$$

where θ is a common parameter, $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)$, $\varphi = -\ln f$, and $f\left(w_{it}, \theta^0, \mu_i^0\right)$ is the PDF of w_{it}, and $\mu_i^0 = \alpha_k^0$ if $i \in G_k^0$ for $k = 1, ..., K_0$. The PLS objective function here takes the form

$$Q_{1NT,\lambda_1}^{(K_0)}(\theta, \boldsymbol{\mu}, \boldsymbol{\alpha}) = Q_{1,NT}(\theta, \boldsymbol{\mu}) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\mu_i - \alpha_k\|.$$

3. Group Patterns of Heterogeneity: Bonhomme and Manresa (2014) consider:

$$y_{it} = \theta^{0\prime} x_{it} + \mu_{g_i t} + u_{it}, \tag{11}$$

where $g_i \in \{1, ..., K_0\}$ map individual units into groups.

- Note that $\mu_{g_i t} = \lambda_i' f_t$ where $f_t = (\mu_{1t}, ..., \mu_{K_0 t})'$, $\lambda_i = (0, ...1, ...0)'$ with 1 in the k^{th} position if $i \in G_k^0$ for $k = 1, ..., K_0$ and zeros elsewhere, we may embed (11) in the more general model

$$y_{it} = \theta^{0\prime} x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it}, \tag{12}$$

  where $\lambda_i^0 = \alpha_k^0$ if $i \in G_k^0$ for $k = 1, ..., K_0$.
- A two-step approach. (1) obtain the Gaussian QMLEs $\breve{\theta}$, $\breve{\lambda}_i$, and $\breve{f}_t$ under certain identification restriction, (2) consider $y_{it} = \theta^{0\prime} x_{it} + \lambda_i^{0\prime} \breve{f}_t + u_{it}$ by imposing: $\lambda_i^0 = \alpha_k^0$ if $i \in G_k^0$ where $k = 1, ..., K_0$.

4. Granger-causality, Unit Root, and Cointegration in Heterogenous Panels:

The C-Lasso approach is also well suited

- to testing for structural change in heterogeneous panel data models,
- to nonparametric and semiparametric panel data models, and
- to models with heterogeneous parametric or nonparametric time trends (e.g., Kneip, Sickles, and Song (2012), Zhang, Su, and Phillips (2012)).

Penalized GMM Estimation

Penalized GMM Estimation

- Consider the first differenced system

$$\Delta y_{it} = \beta_i^{0\prime} \Delta x_{it} + \Delta u_{it}. \quad (13)$$

- The PGMM criterion function

$$Q_{2NT,\lambda_2}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Q_{2,NT}(\boldsymbol{\beta}) + \frac{\lambda_2}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \quad (14)$$

where

$$\begin{aligned} & Q_{2,NT}(\boldsymbol{\beta}) \\ = \; & \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \sum_{t=1}^{T} z_{it} \left(\Delta y_{it}\text{-}\beta_i' \Delta x_{it} \right) \right]' W_{iNT} \left[\frac{1}{T} \sum_{t=1}^{T} z_{it} \left(\Delta y_{it}\text{-}\beta_i' \Delta x_{it} \right) \right] \\ \neq \; & \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \left(\Delta y_{it}\text{-}\beta_i' \Delta x_{it} \right)' \right] W_{NT} \left[\frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \left(\Delta y_{it}\text{-}\beta_i' \Delta x_{it} \right) \right] \end{aligned}$$

- The PGMM estimates: $\tilde{\boldsymbol{\alpha}} \equiv (\tilde{\alpha}_1, ..., \tilde{\alpha}_{K_0})$ and $\tilde{\boldsymbol{\beta}} \equiv (\tilde{\beta}_1, ..., \tilde{\beta}_N)$.

Theorem

If Assumption B1 holds, then

(i) $\tilde{\beta}_i - \beta_i^0 = O_P\left(T^{-1/2} + \lambda_2\right)$ *for* $i = 1, ..., N$,

(ii) $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\beta}_i - \beta_i^0\right\|^2 = O_P\left(T^{-1}\right)$,

(iii) $\left(\tilde{\alpha}_{(1)}, ..., \tilde{\alpha}_{(K_0)}\right) - (\alpha_1^0, ..., \alpha_{K_0}^0) = O_P\left(T^{-1/2}\right)$,

where $\left(\tilde{\alpha}_{(1)}, ..., \tilde{\alpha}_{(K_0)}\right)$ *is a suitable permutation of* $\left(\tilde{\alpha}_1, ..., \tilde{\alpha}_{K_0}\right)$.

$$\tilde{G}_k = \left\{ i \in \{1, 2, ..., N\} : \tilde{\beta}_i = \tilde{\alpha}_k \right\} \text{ for } k = 1, ..., K_0,$$
$$E_{kNT,i} = \left\{ i \notin \tilde{G}_k \mid i \in G_k^0 \right\}, \ \tilde{F}_{kNT,i} = \left\{ i \notin G_k^0 \mid i \in \tilde{G}_k \right\},$$
$$\tilde{E}_{kNT} = \cup_{i \in G_k^0} \tilde{E}_{kNT,i}, \text{ and } \tilde{F}_{kNT} = \cup_{i \in \tilde{G}_k} \tilde{F}_{kNT,i}.$$

Theorem

If Assumptions B1-B2 hold, then

(i) $P\left(\cup_{k=1}^{K_0} \tilde{E}_{kNT}\right) \leq \sum_{k=1}^{K_0} P\left(\tilde{E}_{kNT}\right) \to 0$ *as* $(N, T) \to \infty$,

(ii) $P\left(\cup_{k=1}^{K_0} \tilde{F}_{kNT}\right) \leq \sum_{k=1}^{K_0} P\left(\tilde{F}_{kNT}\right) \to 0$ *as* $(N, T) \to \infty$.

Theorem

Suppose that Assumptions B1-B3 hold. Then $\sqrt{N_k T}\left(\tilde{\alpha}_k - \alpha_k^0\right) - \bar{A}_k^{-1} B_{kNT} \xrightarrow{D} N(0, A_k^{-1} C_k A_k^{-1})$ *for* $k = 1, \ldots, K_0$.

- The PGMM estimators $\{\tilde{\alpha}_k\}$ may fail to possess the oracle property.
- If the group identities were known in advance, one could obtain $\breve{\alpha}_k$ as the minimizer of

$$\tilde{Q}_{NT}\left(\alpha_k\right) = \left[\frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} z_{it}\left(\Delta y_{it} - \alpha_k' \Delta x_{it}\right)\right]' W_{NT}^{(k)} \times \left[\frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} z_{it}\left(\Delta y_{it} - \alpha_k' \Delta x_{it}\right)\right].$$

- The Post-Lasso estimator $\tilde{\alpha}_{\tilde{G}_k}$ is asymptotically equivalent to $\breve{\alpha}_k$.

Monte Carlo Simulations

- Three DGPs, each with three groups.
- $N_1 : N_2 : N_3 = 0.3 : 0.3 : 0.4.$
- $N = 100,\ 200$ and $T = 10,\ 20,\ 40.$

DGP 1 (Static panel with two exogenous regressors)

$$\begin{aligned} y_{it} &= \beta_i^{0\prime} x_{it} + \mu_i + u_{it}, \\ x_{it1} &= 0.2\mu_i + z_{it1}, \\ x_{it2} &= 0.2\mu_i + z_{it2}, \end{aligned}$$

with $(z_{it1}, z_{it2}) \sim \text{IID } N(0,1)$.

$$(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \left(\begin{pmatrix} 0.4 \\ 1.6 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.6 \\ 0.4 \end{pmatrix} \right).$$

DGP 2 (Static panel with endogeneity)

$$\begin{aligned} y_{it} &= \beta_i^{0\prime} x_{it} + \mu_i + u_{it}, \\ x_{it1} &= 0.2\mu_i + 0.5 z_{it1} + 0.5 z_{it2} + 0.5 e_{it}, \end{aligned}$$

$x_{it2} \sim N(0,1)$ is independent of the idiosyncratic shock u_{it}, where $(z_{it1}, z_{it2}) \sim$ IID $N(0,1)$ are two excluded instrumental variables independent of u_{it}.

$$\begin{pmatrix} u_{it} \\ e_{it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0.3 \\ 0.3 & 0 \end{pmatrix} \right).$$

$$(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \left(\begin{pmatrix} 0.2 \\ 1.8 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1.8 \\ 0.2 \end{pmatrix} \right).$$

DGP 3 (PAR(1) with two exogenous regressors)

$$y_{it} = \beta_{i1}^0 y_{i,t-1} + \beta_{i2}^0 x_{it2} + \beta_{i3}^0 x_{it3} + \mu_i (1 - \beta_{i1}^0) + u_{it}$$

where x_{it2} and x_{it3} are two exogenous regressors and they are independent of all error terms. They follow the standard normal distribution. $y_{i0} = \beta_{i2}^0 x_{i02} + \beta_{i3}^0 x_{i03} + \mu_i + u_{i0}$ so that the observations in i is a strictly stationary time series with mean μ_i.

$$(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \left(\begin{pmatrix} 0.8 \\ 0.4 \\ 0.4 \end{pmatrix}, \begin{pmatrix} 0.6 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.4 \\ 1.6 \\ 1.6 \end{pmatrix} \right).$$

- $\bar{P}(\hat{E}) = \frac{1}{N} \sum_{i=1}^N \hat{P}(\hat{E}_{kNT,i})$ and $\bar{P}(\hat{F}) = \frac{1}{N} \sum_{i=1}^N \hat{P}(\hat{F}_{kNT,i})$.

Monte Carlo Simulations

Classification error

Table 1: Classification error for C-Lasso

	C_λ		0.2		0.4		0.8		1.6		3.2	
	N	T	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$
DGP1	100	10	0.1805	0.0901	0.1899	0.0954	0.2236	0.1115	0.2777	0.1305	0.4216	0.1897
PLS	100	20	0.0593	0.0289	0.0585	0.0292	0.0576	0.0290	0.0805	0.0396	0.1304	0.0598
	100	40	0.0103	0.0049	0.0098	0.0046	0.0093	0.0045	0.0094	0.0048	0.0149	0.0070
	200	10	0.1691	0.0848	0.1771	0.0894	0.2097	0.1054	0.2766	0.1322	0.3976	0.1746
	200	20	0.0586	0.0284	0.0556	0.0275	0.0552	0.0277	0.0719	0.0362	0.1338	0.0613
	200	40	0.0092	0.0044	0.0083	0.0040	0.0081	0.0039	0.0078	0.0040	0.0141	0.0066
DGP2	100	10	0.2082	0.0993	0.2001	0.0974	0.2024	0.1004	0.2145	0.1076	0.2527	0.1274
PGMM	100	20	0.1027	0.0485	0.0958	0.0462	0.0888	0.0437	0.0878	0.0440	0.0996	0.0504
	100	40	0.0321	0.0152	0.0307	0.0147	0.0266	0.0130	0.0230	0.0115	0.0227	0.0116
	200	10	0.2037	0.0980	0.1982	0.0971	0.1968	0.0984	0.2113	0.1071	0.2482	0.1257
	200	20	0.1020	0.0483	0.0942	0.0456	0.0872	0.0432	0.0841	0.0424	0.0942	0.0480
	200	40	0.0332	0.0158	0.0299	0.0144	0.0266	0.0130	0.0222	0.0111	0.0212	0.0109

Table 1: Classification error for C-Lasso (cont.)

	C_λ		0.2		0.4		0.8		1.6		3.2	
	N	T	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$	$\bar{P}(\hat{E})$	$\bar{P}(\hat{F})$
DGP3	100	10	0.2063	0.1038	0.1839	0.0908	0.1913	0.0937	0.2305	0.1092	0.4058	0.1715
PLS	100	20	0.1000	0.0501	0.0826	0.0404	0.0750	0.0357	0.0800	0.0391	0.1968	0.0886
	100	40	0.0277	0.0137	0.0222	0.0106	0.0183	0.0085	0.0158	0.0072	0.0373	0.0177
	200	10	0.2025	0.1026	0.1714	0.0853	0.1709	0.0844	0.2079	0.0998	0.3539	0.1498
	200	20	0.0983	0.0490	0.0794	0.0386	0.0703	0.0333	0.0716	0.0347	0.1451	0.0657
	200	40	0.0255	0.0126	0.0209	0.0100	0.0173	0.0080	0.0151	0.0069	0.0220	0.0103
DGP3	100	10	0.3133	0.1551	0.2969	0.1464	0.2872	0.1406	0.2968	0.1440	0.3317	0.1617
PGMM	100	20	0.1703	0.0838	0.1512	0.0746	0.1361	0.0660	0.1334	0.0628	0.1428	0.0664
	100	40	0.0694	0.0338	0.0579	0.0283	0.0487	0.0232	0.0432	0.0198	0.0416	0.0185
	200	10	0.3081	0.1529	0.2868	0.1425	0.2778	0.1367	0.2826	0.1380	0.3173	0.1550
	200	20	0.1691	0.0836	0.1527	0.0753	0.1324	0.0645	0.1265	0.0597	0.1333	0.0620
	200	40	0.0727	0.0357	0.0587	0.0290	0.0490	0.0238	0.0434	0.0202	0.0412	0.0186

Monte Carlo Simulations

Table 2: PLS Estimation of β_1 in DGP 1

N	T	C_λ	0.2		0.4		0.8		1.6		3.2	
			RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
100	10	C-Lasso	0.1010	0.0364	0.1116	0.0364	0.1303	0.0293	0.1780	-0.0150	0.3206	-0.0968
	10	Post-lasso	0.0907	0.0282	0.1035	0.0293	0.1274	0.0254	0.1788	-0.0162	0.3216	-0.0984
	10	Oracle	0.0583	-0.0033	0.0583	-0.0033	0.0583	-0.0033	0.0583	-0.0033	0.0583	-0.0033
100	20	C-Lasso	0.0590	0.0154	0.0560	0.0183	0.0507	0.0154	0.0690	0.0054	0.0856	0.0012
	20	Post-lasso	0.0450	0.0066	0.0467	0.0092	0.0470	0.0090	0.0687	0.0038	0.0846	0.0012
	20	Oracle	0.0399	-0.0021	0.0399	-0.0021	0.0399	-0.0021	0.0399	-0.0021	0.0399	-0.0021
100	40	C-Lasso	0.0347	0.0096	0.0348	0.0047	0.0305	0.0053	0.0301	0.0023	0.0347	0.0011
	40	Post-lasso	0.0292	0.0012	0.0293	0.0002	0.0291	0.0010	0.0290	0.0008	0.0337	0.0010
	40	Oracle	0.0281	-0.0010	0.0281	-0.0010	0.0281	-0.0010	0.0281	-0.0010	0.0281	-0.0010

Monte Carlo Simulations

Table 3: PGMM Estimation of β_1 in DGP 2

		c_λ	0.2		0.4		0.8		1.6		3.2	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
100	10	C-Lasso	0.1906	0.1093	0.1907	0.1242	0.2018	0.1388	0.2096	0.1490	0.2220	0.1581
		Post-lasso	0.1416	0.0152	0.1368	0.0251	0.1413	0.0325	0.1421	0.0381	0.1533	0.0443
		C-Lasso BC	0.1603	0.0684	0.1586	0.0811	0.1679	0.0928	0.1737	0.1009	0.1858	0.1085
		Oracle	0.0993	-0.0001	0.0993	-0.0001	0.0993	-0.0001	0.0993	-0.0001	0.0993	-0.0001
100	20	C-Lasso	0.1179	0.0560	0.1176	0.0683	0.1182	0.0799	0.1239	0.0898	0.1321	0.0985
		Post-lasso	0.0838	0.0138	0.0815	0.0181	0.0810	0.0200	0.0826	0.0212	0.0871	0.0216
		C-Lasso BC	0.0986	0.0374	0.0978	0.0464	0.0986	0.0539	0.1021	0.0600	0.1083	0.0652
		Oracle	0.0680	-0.0004	0.0680	-0.0004	0.0680	-0.0004	0.0680	-0.0004	0.0680	-0.0004
100	40	C-Lasso	0.0712	0.0400	0.0754	0.0422	0.0761	0.0464	0.0753	0.0504	0.0772	0.0557
		Post-lasso	0.0519	0.0136	0.0522	0.0129	0.0519	0.0122	0.0516	0.0112	0.0522	0.0108
		C-Lasso BC	0.0614	0.0274	0.0632	0.0282	0.0637	0.0301	0.0634	0.0317	0.0645	0.0343
		Oracle	0.0492	0.0007	0.0492	0.0007	0.0492	0.0007	0.0492	0.0007	0.0492	0.0007

Table 4: PLS Estimation of β_1 in DGP 3

N	T	C_λ	0.2		0.4		0.8		1.6		3.2	
			RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
100	10	C-Lasso	0.1331	-0.1216	0.1264	-0.1143	0.1189	-0.1028	0.1120	-0.0858	0.1557	-0.0561
		Post-lasso	0.1011	-0.0863	0.1041	-0.0897	0.1059	-0.0866	0.1077	-0.0784	0.1573	-0.0560
		C-Lasso BC	0.1220	-0.1088	0.1157	-0.1022	0.1088	-0.0909	0.1033	-0.0740	0.1532	-0.0443
		Post-Lasso BC	0.0922	-0.0745	0.0949	-0.0782	0.0971	-0.0751	0.0998	-0.0667	0.1548	-0.0441
		Oracle	0.0928	-0.0855	0.0928	-0.0855	0.0928	-0.0855	0.0928	-0.0855	0.0928	-0.0855
100	20	C-Lasso	0.0782	-0.0711	0.0740	-0.0670	0.0671	-0.0603	0.0580	-0.0505	0.0711	-0.0254
		Post-lasso	0.0539	-0.0431	0.0558	-0.0471	0.0558	-0.0482	0.0529	-0.0444	0.0713	-0.0233
		C-Lasso BC	0.0723	-0.0643	0.0682	-0.0605	0.0614	-0.0540	0.0527	-0.0443	0.0691	-0.0191
		Post-Lasso BC	0.0494	-0.0368	0.0508	-0.0410	0.0507	-0.0421	0.0479	-0.0382	0.0694	-0.0170
		Oracle	0.0527	-0.0469	0.0527	-0.0469	0.0527	-0.0469	0.0527	-0.0469	0.0527	-0.0469
100	40	C-Lasso	0.0428	-0.0372	0.0405	-0.0351	0.0363	-0.0310	0.0321	-0.0270	0.0315	-0.0213
		Post-lasso	0.0289	-0.0224	0.0295	-0.0236	0.0297	-0.0241	0.0293	-0.0238	0.0313	-0.0204
		C-Lasso BC	0.0401	-0.0339	0.0378	-0.0319	0.0336	-0.0279	0.0295	-0.0239	0.0294	-0.0182
		Post-Lasso BC	0.0266	-0.0193	0.0272	-0.0206	0.0273	-0.0210	0.0269	-0.0207	0.0294	-0.0173
		Oracle	0.0285	-0.0236	0.0285	-0.0236	0.0285	-0.0236	0.0285	-0.0236	0.0285	-0.0236

Monte Carlo Simulations

Table 5: PGMM Estimation of β_1 in DGP 3

		C_λ	0.2		0.4		0.8		1.6		3.2	
N	T		RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
100	10	C-Lasso	0.1823	-0.1065	0.1892	-0.1241	0.1980	-0.1417	0.2090	-0.1627	0.2271	-0.1817
		Post-lasso	0.1304	-0.0352	0.1231	-0.0331	0.1161	-0.0311	0.1137	-0.0352	0.1202	-0.0427
		C-Lasso BC	0.1494	-0.0698	0.1509	-0.0800	0.1516	-0.0897	0.1572	-0.1047	0.1729	-0.1206
		Oracle	0.0664	-0.0013	0.0664	-0.0013	0.0664	-0.0013	0.0664	-0.0013	0.0664	-0.0013
100	20	C-Lasso	0.0808	-0.0319	0.0858	-0.0478	0.0974	-0.0687	0.1114	-0.0888	0.1247	-0.1035
		Post-lasso	0.0584	-0.0010	0.0565	-0.0031	0.0546	-0.0068	0.0538	-0.0109	0.0554	-0.0138
		C-Lasso BC	0.0678	-0.0175	0.0690	-0.0275	0.0739	-0.0411	0.0814	-0.0548	0.0904	-0.0648
		Oracle	0.0399	-0.0027	0.0399	-0.0027	0.0399	-0.0027	0.0399	-0.0027	0.0399	-0.0027
100	40	C-Lasso	0.0442	-0.0126	0.0447	-0.0198	0.0519	-0.0329	0.0646	-0.0491	0.0742	-0.0606
		Post-lasso	0.0356	0.0025	0.0334	0.0006	0.0327	-0.0018	0.0325	-0.0037	0.0320	-0.0046
		C-Lasso BC	0.0395	-0.0047	0.0384	-0.0094	0.0406	-0.0173	0.0459	-0.0268	0.0507	-0.0333
		Oracle	0.0274	-0.0011	0.0274	-0.0011	0.0274	-0.0011	0.0274	-0.0011	0.0274	-0.0011

Empirical Application

Motivation

- Across countries savings rates vary widely: on average East Asia saves more than 30 percent of gross national disposable income while Sub-Saharan Africa saves less than 15 percent.
- Understanding the disparate saving behavior across countries is of long-lasting research interest in development economics. Theoretical advancement and empirical studies have been accumulating over the years; see Feldstein (1980), Deaton (1990), Edwards (1996) Bosworth, Collins, and Reinhart (1999), Rodrik (2000), and Li, Zhang, and Zhang (2007), among others.
- Empirical research either employs standard panel data methods to handle the heterogeneity, or relies on prior information to categorize countries into groups. Classification criteria vary from geographic locations to the notion of developed countries versus developing countries (Loayza, Schmidt-Hebbel and Servén, 2000).
- Here we apply the new methodology developed in this paper to revisit this empirical problem.

Following Edwards (1996), we consider the following simple regression model

$$S_{it} = \beta_{1i} S_{i,t-1} + \beta_{2i} I_{it} + \beta_{3i} R_{it} + \beta_{4i} G_{it} + \mu_i + u_{it}, \tag{15}$$

where
S_{it} is the ratio of savings to GDP,
$S_{i,t-1}$: capture the persistence of the savings rate.
I_{it} is the CPI-based inflation rate (measure the degree of the macroeconomic stability)
R_{it} is the real interest rate (reflects the price of money)
G_{it} is the per capita GDP growth rate (*conventional wisdom:* across countries higher saving rates tend to go hand in hand with higher income growth, e.g., Loayza, Schmidt-Hebbel, and Servén, 2000)

World Development Indicators: 1995–2000, 56 countries.

Table 6: Summary statistics for the savings data set

	mean	median	s.e.	min	max
Savings rate	22.099	20.790	8.833	-3.207	53.434
Inflation rate	7.724	4.853	15.342	-3.846	293.679
Real interest rate	7.422	5.927	10.062	-63.761	93.915
Per capita GDP growth rate	2.855	2.971	3.865	-17.545	14.060

Empirical Application

Data

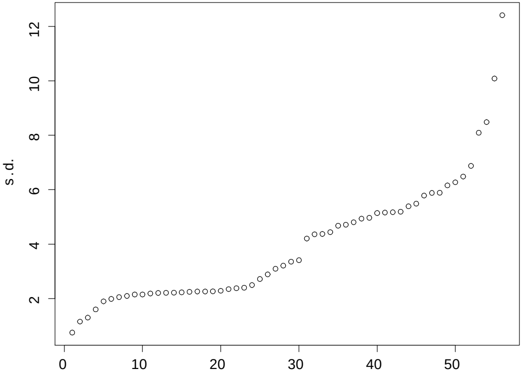

Figure: The time series standard deviations of the saving rates for the 56 countries

Empirical Application

Determination of the number of groups

- Lu and Su's (2014) LM test. Basic idea:

$$\mathbb{H}_0\left(K_0\right): K = K_0 \text{ versus } \mathbb{H}_1\left(K_0\right): K_0 < K \leq K_{\max}. \tag{16}$$

  - Suppose $K_{\min} \leq K \leq K_{\max}$, where $K_{\min}$ is typically 1.
  - First test: $H_0(K_{\min})$ against $H_1(K_{\min})$. If we fail to reject the null, then we conclude that $K = K_{\min}$.
  - Otherwise, we continue to test $H_0(K_{\min}+1)$ against $H_1(K_{\min}+1)$.
  - Repeat this procedure until we fail to reject the null $H_0(K^*)$ and conclude that $K = K^*$.

Table 7: Test statistics

	$c=1$			$c=1.5$			$c=2$		
$\mathbb{H}_0\left(K_0\right)$	1	2	3	1	2	3	1	2	3
Statistics	3.040	1.397	0.715	3.040	1.265	1.069	3.040	2.396	1.411
p-values	0.001	0.081	0.237	0.001	0.102	0.142	0.001	0.008	0.079
Holm adjusted p-value	0.002	0.081	NA	0.0024	0.102	NA	0.002	0.008	NA

Two estimated groups:

- Group 1 (36 countries): Armenia, Australia, Bangladesh, Bolivia, Botswana, Cape Verde, **China**, Costa Rica, Czech, Guatemala, Honduras, Hungary, **Indonesia**, Israel, Italy, **Japan**, Jordan, Latvia, Malawi, Malaysia, Mauritius, Mexico, Mongolia, Panama, Paraguay, **Philippines**, Romania, Russian, South Africa, **Sri Lanka**, Switzerland, Syrian, **Thailand**, Uganda, Ukraine, United Kingdom;
- Group 2 (20 countries): Bahamas, Belarus, Canada, Dominican, Egypt, Guyana, Iceland, India, Kenya, **South Korea**, Lithuania, Malta, Netherlands, Papua New Guinea, Peru, **Singapore**, Swaziland, Tanzania, United States, Uruguay.

Empirical Application

Table 8: Estimation results

Slope coefficients	Common	Group 1		Group 2	
	FE	C-Lasso	post-Lasso	C-Lasso	post-Lasso
β_1	0.6203^{***}	0.5510^{***}	0.5548^{***}	0.6090^{***}	0.6156^{***}
	(0.1330)	(0.1090)	(0.1057)	(0.1060)	(0.1057)
β_2	0.0303	-0.1154^{**}	-0.1068^{**}	0.2712^{***}	0.2661^{***}
	(0.0484)	(0.0464)	(0.0458)	(0.0515)	(0.0514)
β_3	0.0068	-0.0419	-0.0273	0.0525	0.0533
	(0.0432)	(0.0490)	(0.0476)	(0.0406)	(0.0401)
β_4	0.1880^{***}	0.2771^{***}	0.3055^{***}	0.0625	0.0291
	(0.0450)	(0.0470)	(0.0452)	(0.0459)	(0.0442)

Note: *** 1% significant; ** 5% significant; * 10% significant.

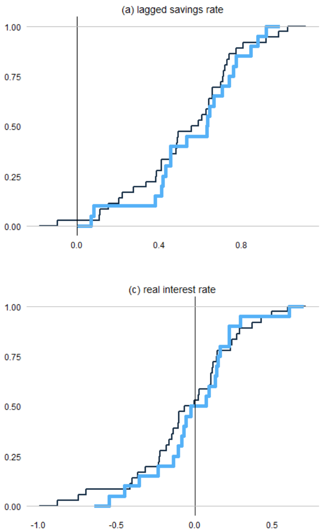

Figure: Empirical distribution functions of the time series estimates of regression coefficients for the two estimated groups (thin line: Group 1; thick line: Group 2)

Panel Structure Models with Interactive Fixed Effects (IFEs)

Panel Structure Models with IFEs

- **Model:**

$$y_{it} = \beta_i^{0\prime} x_{it} + \lambda_i^{0\prime} f_t^0 + \varepsilon_{it},$$

  where λ_i^0 and f_t^0 denote an $R_0 \times 1$ vector of factor loadings and common factors, respectively.

- **Homogenous:** $\beta_i^0 = \beta^0$. Bai (2009), Moon and Weidner (2010), Greenaway-McGrevy et al. (2012), Lu and Su (2013), Su et al. (2013)... Inference is misleading if the slopes are heterogenous.
- **Heterogenous:** Pesaran (2006), Kapetanios and Pesaran (2007), Chudik et al. (2011), Kapetanios et al. (2011), Pesaran and Tosetti (2011), Su and Jin (2012), Ando (2013), Chudik and Pesaran (2013), Song (2013)... Inefficient and slow convergence rate if the models have homogeneous slopes.

- **Penalized principal component (PPC) estimation**

$$Q^{(K_0)}_{0NT,\kappa}(\boldsymbol{\beta},\boldsymbol{\alpha},\Lambda,F) = Q_{0NT}(\boldsymbol{\beta},\Lambda,F) + \frac{\kappa}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\left\|\beta_i - \alpha_k\right\|, \quad (17)$$

  where $Q_{0NT}(\boldsymbol{\beta},\Lambda,F) = \frac{1}{NT}\sum_{i=1}^{N}\left\|Y_i - X_i\beta_i - F\lambda_i\right\|^2$, and κ is a tuning parameter.

- Concentrate Λ out:

$$Q^{(K_0)}_{1NT,\kappa}(\boldsymbol{\beta},\boldsymbol{\alpha},F) = Q_{1NT}(\boldsymbol{\beta},F) + \frac{\kappa}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\left\|\beta_i - \alpha_k\right\|, \quad (18)$$

  where $Q_{1NT}(\boldsymbol{\beta},F) = \frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\beta_i)' M_F (Y_i - X_i\beta_i)$.

- Further concentrate F out:

$$Q^{(K_0)}_{NT,\kappa}(\boldsymbol{\beta},\boldsymbol{\alpha}) = Q_{NT}(\boldsymbol{\beta}) + \frac{\kappa}{N}\sum_{i=1}^{N}\Pi_{k=1}^{K_0}\left\|\beta_i - \alpha_k\right\|, \quad (19)$$

  where $Q_{NT}(\boldsymbol{\beta}) = \frac{1}{T}\sum_{r=R_0+1}^{T}\mu_r\left[\frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i\beta_i)(Y_i - X_i\beta_i)'\right]$.

Panel Structure Models with IFEs

- C-Lasso estimates: $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_K)$ and $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)$.
- Estimate (Λ, F) via PC analysis (e.g., Bai and Ng's (2002)) under the identification restrictions: $F'F/T = I_{R_0}$ and $\Lambda'\Lambda =$ diag:

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\hat{\beta}_i\right)\left(Y_i - X_i\hat{\beta}_i\right)'\right]\hat{F} = \hat{F}V_{NT}, \ \hat{\Lambda} = (\hat{\lambda}_1, \hat{\lambda}_2, ..., \hat{\lambda}_N) \tag{20}$$

  where V_{NT} is a diagonal matrix consisting of the R_0 largest eigenvalues of the above matrix in the square bracket, arranged in descending order, and $\hat{\lambda}_i = T^{-1}\hat{F}'(Y_i - X_i\hat{\beta}_i)$.
- **Numerical difficulty:** Non-convex/nonsmooth
- **Asymptotic properties**:
  - Uniform classification consistency ($\surd$)
  - Oracle property ($\surd$)

Conclusions and Future Work

- Propose a novel approach to study panel structure model motivated by the Lasso principle
- Penalized least squares estimation: work for static or dynamic panel data models without endogeneity
    - uniform selection consistency
    - oracle property
    - IC for determining the number of groups
- Penalized GMM estimation: work for panel data models with endogeneity or dynamic panel without endogeneity
    - uniform selection consistency
    - oracle property in special case
    - IC for determining the number of groups
- Extension to panel data models with cross sectional dependence
- Determining the number of groups

- Parametric framework
  - Panel data models with interactive fixed effects
  - Quantile regression models
  - Non-linear panel data models
  - Panel unit root and cointegration analysis
  - Panel trend/cotrend modeling
- NP and SP framework: easy for sieve estimation

Thanks

Thanks!

1 Start with $\hat{\alpha}^{(0)} = (\hat{\alpha}_1^{(0)}, \ldots, \hat{\alpha}_{K_0}^{(0)})$ and $\hat{\beta}^{(0)} = (\hat{\beta}_1^{(0)}, \ldots, \hat{\beta}_N^{(0)})$ such that $\sum_{i=1}^N ||\hat{\beta}_i^{(0)} - \hat{\alpha}_k^{(0)}|| \neq 0$ for each $k = 2, \ldots, K_0$.

2 Given $\hat{\alpha}^{(r-1)} \equiv (\hat{\alpha}_1^{(r-1)}, \ldots, \hat{\alpha}_{K_0}^{(r-1)})$ and $\hat{\beta}^{(r-1)} \equiv (\hat{\beta}_1^{(r-1)}, \ldots, \hat{\beta}_N^{(r-1)})$,

- In Step $r \geq 1$, we first choose $(\boldsymbol{\beta}, \alpha_1)$ to minimize

$$Q_{K_0 NT}^{(r,1)}(\boldsymbol{\beta}, \alpha_1) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^N \|\beta_i - \alpha_1\| \Pi_{k \neq 1}^{K_0} \left\| \hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)} \right\|$$

  and obtain the updated estimate $(\hat{\beta}^{(r,1)}, \hat{\alpha}_1^{(r)})$ of $(\boldsymbol{\beta}, \alpha_1)$.

- Next choose $(\boldsymbol{\beta}, \alpha_2)$ to minimize

$$\begin{aligned} Q_{K_0 NT}^{(r,2)}(\boldsymbol{\beta}, \alpha_2) &= Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^N \|\beta_i - \alpha_2\| \left\| \hat{\beta}_i^{(r,1)} - \hat{\alpha}_1^{(r)} \right\| \\ &\quad \times \Pi_{k \neq 1,2}^{K_0} \left\| \hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)} \right\| \end{aligned}$$

  to obtain the updated estimate $(\hat{\beta}^{(r,2)}, \hat{\alpha}_2^{(r)})$ of $(\boldsymbol{\beta}, \alpha_2)$.

- Repeat this procedure $(\boldsymbol{\beta}, \alpha_{K_0})$ is chosen to minimize

$$Q_{K_0 NT}^{(r,K_0)}(\boldsymbol{\beta}, \alpha_{K_0}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_K\| \Pi_{k=1}^{K_0-1} \left\| \hat{\beta}_i^{(r,K_0-1)} - \hat{\alpha}_k^{(r)} \right\|$$

  to obtain the updated estimate $(\hat{\beta}^{(r,K_0)}, \hat{\alpha}_{K_0}^{(r)})$ of $(\boldsymbol{\beta}, \alpha_{K_0})$. Let $\hat{\beta}^{(r)} = \hat{\beta}^{(r,K_0)}$ and $\hat{\alpha}^{(r)} = (\hat{\alpha}_1^{(r)}, \ldots, \hat{\alpha}_{K_0}^{(r)})$.

3 Repeat step 2 until a convergence criterion is met.

Let $\tilde{Q}_{i,z\Delta x} = \frac{1}{T}\sum_{t=1}^{T} z_{it}(\Delta x_{it})'$, $\tilde{Q}_{i,z\Delta y} = \frac{1}{T}\sum_{t=1}^{T} z_{it}\Delta y_{it}$, $\bar{Q}_{i,z\Delta x} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[z_{it}(\Delta x_{it})']$, and $\bar{Q}_{i,z\Delta y} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[z_{it}\Delta y_{it}]$. Let $\xi_{it} = (\Delta y_{it}, (\Delta x_{it})', z_{it}')'$. Define $\rho(\xi_{it}, \beta) = z_{it}\left(\Delta y_{it} - \beta'\Delta x_{it}\right)$ and

$$\bar{\rho}_{i,T}(\beta) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{\rho(\xi_{it}, \beta) - \mathbb{E}[\rho(\xi_{it}, \beta)]\}.$$

ASSUMPTION B1. (i) $\mathbb{E}\left[\rho\left(\xi_{it}, \beta_i^0\right)\right] = 0$.
(ii) $\sup_{\beta\in\mathcal{B}_i} \bar{\rho}_{i,T}(\beta) = O_P(1)$ *and* $\frac{1}{N}\sum_{i=1}^{N}\left\|\bar{\rho}_{i,T}(\beta_i)\right\|^2 = O_P(1)$ *for any* $\beta_i \in \mathcal{B}_i$ *and* $i = 1, ..., N$.
(iii) $\tilde{Q}_{i,z\Delta x} = \bar{Q}_{i,z\Delta x} + o_P(1)$ *for each* $i = 1, ..., N$ *and* $\liminf_{(N,T)\to\infty} \min_{1\le i\le N} \mu_{\min}\left(\bar{Q}_{i,z\Delta x}'\bar{Q}_{i,z\Delta x}\right) = \underline{c}_{\bar{Q}} > 0$.
(iv) *There exist* W_i *such that* $\max_{1\le i\le N}\|W_{iNT} - W_i\| = o_P(1)$ *and* $\liminf_{N\to\in} \min_{1\le i\le N} \mu_{\min}(W_i) = \underline{c}_W > 0$.
(v) $N_k/N \to \tau_k \in (0,1)$ *for each* $k = 1, ..., K_0$ *as* $N \to \infty$.

ASSUMPTION B2. (i) $T\lambda_2 \to \infty$ *and* $T\lambda_2^4 \to c_0 \in [0, \infty)$ *as* $(N, T) \to \infty$.

(ii) *For any* $c > 0$, $N \max_{1 \leq i \leq N} P\left(\left\| T^{-1} \sum_{t=1}^{T} z_{it} \Delta u_{it} \right\| \geq c\sqrt{\lambda_2}\right) \to 0$ *as* $(N, T) \to \infty$.

ASSUMPTION B3. (i) $\frac{1}{N_k} \sum_{i \in G_k^0} \left\| \tilde{Q}_{i,z\Delta x} - \bar{Q}_{i,z\Delta x} \right\|^2 = o_P(1)$.

(ii) $\bar{A}_k \equiv \frac{1}{N_k} \sum_{i \in G_k^0} \bar{Q}_{i,z\Delta x}' W_i \bar{Q}_{i,z\Delta x} \to A_k > 0$ *as* $(N, T) \to \infty$.

ASSUMPTION B4. (i) $W_{NT}^{(k)} \xrightarrow{P} W^{(k)} > 0$ as $(N, T) \to \infty$.

(ii) $Q_{z\Delta x, NT}^{(k)} \xrightarrow{P} Q_{z\Delta x}^{(k)}$ *where* $Q_{z\Delta x}^{(k)}$ *has rank* p.

(iii) $\frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} z_{it} \Delta u_{it} \xrightarrow{D} N(0, V_k)$.

Theorem

Suppose that Assumptions B1-B3 hold. Then $\sqrt{N_k T}\left(\tilde{\alpha}_k - \alpha_k^0\right) - \bar{A}_k^{-1} B_{kNT} \xrightarrow{D} N(0,\ A_k^{-1} C_k A_k^{-1})$ *for* $k = 1, ..., K_0$.

Theorem

Suppose that Assumptions B1-B4 hold. Then $\sqrt{N_k T}\left(\tilde{\alpha}_{\tilde{G}_k} - \alpha_k^0\right) \xrightarrow{D} N\left(0, \Omega_k\right)$ *where*

$$\Omega_k = \left[Q_{z\Delta x}^{(k)\prime} W^{(k)} Q_{z\Delta x}^{(k)}\right]^{-1} Q_{z\Delta x}^{(k)\prime} W^{(k)} V_k W^{(k)} Q_{z\Delta x}^{(k)} \left[Q_{z\Delta x}^{(k)\prime} W^{(k)} Q_{z\Delta x}^{(k)}\right]^{-1}$$

and $k = 1, ..., K_0$.

- Note that $\sqrt{N_k T}\left(\tilde{\alpha}_{\tilde{G}_k} - \alpha_k^0\right) = \sqrt{N_k T}\left(\breve{\alpha}_k - \alpha_k^0\right) + o_P\left(1\right)$. That is, the post-Lasso GMM estimator $\tilde{\alpha}_{\tilde{G}_k}$ is asymptotically equivalent to the infeasible estimate $\breve{\alpha}_k$.

- Basic idea:

$$\mathbb{H}_0\left(K_0\right): K = K_0 \text{ versus } \mathbb{H}_1\left(K_0\right): K_0 < K \leq K_{\max}. \tag{21}$$

  - Suppose $K_{\min} \leq K \leq K_{\max}$, where $K_{\min}$ is typically 1.
  - First test: $\mathbb{H}_0(K_{\min})$ against $\mathbb{H}_1(K_{\min})$. If we fail to reject the null, then we conclude that $K = K_{\min}$.
  - Otherwise, we continue to test $\mathbb{H}_0(K_{\min}+1)$ against $\mathbb{H}_1(K_{\min}+1)$.
  - Repeat this procedure until we fail to reject the null $\mathbb{H}_0(K^*)$ and conclude that $K = K^*$.

- Estimation:

$$\hat{\mu}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\beta}_i' X_{it}), \;\; \hat{u}_{it} \equiv y_{it} - \hat{\beta}_i' X_{it} - \hat{\mu}_i.$$

  - $K_0 = 1$: Set $\hat{\beta}_i = \hat{\beta}$, the within-group estimator of the homogeneous slope coefficient. Note that we also suppress the dependence of $\hat{\mu}_i$ on K_0.

Supplement 3: Determining the number of groups

Motivation for the test:

$$\begin{aligned} \hat{u}_{it} &= (y_{it} - \bar{y}_i) - (X_{it} - \bar{X}_i)' \hat{\beta}_i \\ &= u_{it} - \bar{u}_i + (X_{it} - \bar{X}_i)' (\beta_i^0 - \hat{\beta}_i), \end{aligned} \tag{22}$$

where, e.g., $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$. Under the null hypothesis, $\hat{\beta}_i$ is a consistent estimator of β_i^0 and $\hat{u}_{it}$ should be close to u_{it}. By the assumption, x_{it} should not have any predictive power for u_{it}. This motivates us to run the following auxiliary regression model

$$\hat{u}_{it} = v_i + \phi_i' X_{it} + \eta_{it}, \ \ i = 1, ..., N, \ \ t = 1, ..., T, \tag{23}$$

and test the null hypothesis

$$\mathbb{H}_0^* : \phi_i = 0 \text{ for all } i = 1, ..., N.$$

Supplement 3: Determining the number of groups

- We construct an LM-type test statistic by concentrating the intercept υ_i out in (23). Consider the Gaussian quasi-likelihood function for $\hat{u}_{it}$:

$$\ell(\boldsymbol{\phi}) = \sum_{i=1}^{N} \left(\hat{u}_i - M_0 X_i \phi_i\right)' \left(\hat{u}_i - M_0 X_i \phi_i\right),$$

where $\boldsymbol{\phi} \equiv (\phi_1, ..., \phi_N)'$, $\hat{u}_i \equiv (\hat{u}_{i1}, ..., \hat{u}_{iT})'$, and $X_i \equiv (X_{i1}, ..., X_{iT})'$. Define the LM statistic:

$$LM_{NT}(K_0) = \left(T^{-1/2}\frac{\partial \ell(0)}{\partial \boldsymbol{\phi}}\right)' \left(-T^{-1}\frac{\partial^2 \ell(0)}{\partial \boldsymbol{\phi}\, \partial \boldsymbol{\phi}'}\right) \left(T^{-1/2}\frac{\partial \ell(0)}{\partial \boldsymbol{\phi}}\right). \tag{24}$$

We can verify that

$$LM_{NT}(K_0) = \sum_{i=1}^{N} \hat{u}_i' M_0 X_i \left(X_i' M_0 X_i\right)^{-1} X_i' M_0 \hat{u}_i. \tag{25}$$

Supplement 3: Determining the number of groups

Let $h_{i,ts}$ denote the (t,s)'th element of $H_i \equiv M_0 X_i (X_i' M_0 X_i)^{-1} X_i' M_0$. Let $\Omega_i \equiv E(T^{-1} X_i' M_0 X_i)$, $X_{it}^{\dagger} \equiv X_{it} - T^{-1} \sum_{s=1}^{T} E(X_{is})$, and $\bar{b}_{it} \equiv \Omega_i^{-1/2} X_{it}^{\dagger}$. Define

$$
\begin{aligned}
B_{NT} &\equiv N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it}^2 h_{i,tt} \text{ and} \\
V_{NT} &\equiv 4T^{-2} N^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} E\left[u_{it} \bar{b}_{it}' \sum_{s=1}^{t-1} \bar{b}_{is} u_{is} \right]^2 .
\end{aligned}
$$

Theorem

Suppose Assumptions A.1-A.3 hold. Then under $\mathbb{H}_0 (K_0)$,

$$
J_{NT}(K_0) \equiv \left(N^{-1/2} LM_{NT}(K_0) - B_{NT} \right) / \sqrt{V_{NT}} \xrightarrow{D} N(0,1).
$$

Feasible version:
$\hat{J}_{NT}(K_0) \equiv (N^{-1/2} LM_{NT}(K_0) - \hat{B}_{NT}(K_0)) / \sqrt{\hat{V}_{NT}(K_0)}$.