Identifying Latent Structures in Panel Data

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

\[ y_{it} = \beta_i \left( \frac{t}{T} \right) + \mu_i + u_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T \]

Data: World Bank annual data from 1960-2012 for 92 countries (\( N = 92, \ T = 53 \)).

Nonparametric sieve estimation with cubic B-spline, \( \# \text{knot} = 3 \).
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
The Model and Literature

The model

\[ y_{it} = \beta_i^0 x_{it} + \mu_i + u_{it} \]  

(1)

where

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

\[ \beta_i^0 = \begin{cases} \alpha_1^0 & \text{if } i \in G_1^0 \\ \vdots & \vdots \\ \alpha_{K_0}^0 & \text{if } i \in G_{K_0}^0 \end{cases} \]  

(2)

Here \( \alpha_j^0 \neq \alpha_k^0 \) for any \( j \neq k \), \( \bigcup_{k=1}^{K_0} G_k^0 = \{1, 2, \ldots, N\} \), and \( G_k^0 \cap G_j^0 = \emptyset \) for any \( j \neq k \). Let \( N_k = \# G_k^0 \).
**Motivation**

- **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.
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.

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 \((N, T) \rightarrow \infty\) 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.
Motivated by the key feature of Lasso to handle parameter sparsity. $\{\beta_i, i = 1, \ldots, N\}$ versus $\{\alpha_k, k = 1, \ldots, K_0\}$.

Contribute to the literature on fused Lasso (e.g., Tibshirani et al. (2005)). No natural ordering across individuals.

Additive-multiplicative penalty terms: ⇒ Classifier-Lasso or C-Lasso.

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.

$K_0$ is unknown: a BIC-type information criterion is proposed.

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.
The Model and Literature

Potential applications

- **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: \( y_{it} = \beta_i x_{it} + \mu_i + u_{it} \).
- Define
  \[
  Q_{0,NT} (\beta, \mu) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \beta_i' x_{it} - \mu_i)^2.
  \]
- Concentrate \( \mu \) out:
  \[
  Q_{1,NT} (\beta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{y}_{it} - \beta_i' \tilde{x}_{it})^2,
  \]
  where \( \tilde{x}_{it} = x_{it} - T^{-1} \sum_{t=1}^{T} x_{it} \) and \( \tilde{y}_{it} = y_{it} - T^{-1} \sum_{t=1}^{T} y_{it} \).
Penalized Least Squares Estimation

PLS Estimation

- PLS objective function

\[
Q^{(K_0)}_{1N_T,\lambda_1}(\beta, \alpha) = Q_{1,NT}(\beta) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \prod_{k=1}^{K_0} \| \beta_i - \alpha_k \|,
\]

(3)

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

- C-Lasso estimates: \( \hat{\alpha} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_K) \) and \( \hat{\beta} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N) \).

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

(ii) \( \hat{Q}_{i,\bar{x}\bar{x}} = \frac{1}{T} \sum_{t=1}^{T} \bar{x}_{it} \bar{x}_{it}' \xrightarrow{P} Q_{i,\bar{x}\bar{x}} > 0 \forall i. \exists c_{\bar{x}\bar{x}} such that \lim_{(N,T) \to \infty} \min_{1 \leq i \leq N} \mu_{\min}(\hat{Q}_{i,\bar{x}\bar{x}}) \geq c_{\bar{x}\bar{x}} > 0. \)

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

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

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

where \( (\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)}) is a suitable permutation of (\hat{\alpha}_{1}, ..., \hat{\alpha}_{K_0}). \)
Note that $Q_{1iNT,\lambda_1}^{(K_0)} (\beta, \alpha) = \frac{1}{N} \sum_{i=1}^{N} Q_{1iNT,\lambda_1}^{(K_0)} (\beta_i, \alpha)$, where

$$Q_{1iNT,\lambda_1}^{(K_0)} (\beta_i, \alpha) = Q_{1NT,i} (\beta_i) + \lambda_1 \prod_{k=1}^{K_0} \| \beta_i - \alpha_k \|,$$

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

Pointwise convergence:

$$Q_{1iNT,\lambda_1}^{(K_0)} (\hat{\beta}_i, \hat{\alpha}) - Q_{1iNT,\lambda_1}^{(K_0)} (\beta_0, \hat{\alpha}) = Q_{1NT,i} (\hat{\beta}_i) - Q_{1NT,i} (\beta_0^0) + \lambda_1 \left\{ \prod_{k=1}^{K_0} \| \hat{\beta}_i - \hat{\alpha}_k \| - \prod_{k=1}^{K_0} \| \beta_0^0 - \hat{\alpha}_k \| \right\} \leq 0$$

Given $\hat{\alpha}$, $\hat{\beta}_i$ must minimize $Q_{1iNT,\lambda_1}^{(K_0)} (\beta_i, \hat{\alpha})$ with respect to $\beta_i$. 
**Mean-square convergence:** relies on the observation

\[ Q^{(K_0)}_{1NT,\lambda_1} (\hat{\beta}, \hat{\alpha}) - Q^{(K_0)}_{1NT,\lambda_1} (\beta^0, \alpha^0) \leq 0. \]  

(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

\[ \frac{1}{N} \sum_{i=1}^{N} \| \hat{\beta}_i - \beta^0_i \|^2 \geq L / T. \]

**Convergence of** \( (\hat{\alpha}_{(1)}, ..., \hat{\alpha}_{(K_0)}) \): relies on the observation

\[ P_{NT} (\hat{\beta}, \hat{\alpha}) - P_{NT} (\beta, \alpha^0) \leq 0 \]

(5)

where \( P_{NT} (\beta, \alpha) = \frac{1}{N} \sum_{i=1}^{N} \prod_{k=1}^{K_0} \| \beta_i - \alpha_k \| \), 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

\[
\hat{G}_k = \left\{ i \in \{1, 2, ..., N\} : \hat{\beta}_i = \hat{\alpha}_k \right\} \text{ for } k = 1, ..., K_0,
\]

\[
\hat{E}_{kNT,i} = \left\{ i \notin \hat{G}_k \mid i \in G^0_k \right\}, \quad \hat{F}_{kNT,i} = \left\{ i \notin G^0_k \mid i \in \hat{G}_k \right\},
\]

\[
\hat{E}_{kNT} = \bigcup_{i \in G^0_k} \hat{E}_{kNT,i}, \text{ and } \hat{F}_{kNT} = \bigcup_{i \in \hat{G}_k} \hat{F}_{kNT,i}.
\]

**Definition 1. (Uniform consistency of classification)** We say that a classification method is *individually consistent* if

\[
P(\hat{E}_{kNT,i}) \to 0 \text{ as } (N, T) \to \infty \text{ for each } i \in G^0_k \text{ and } k = 1, ..., K_0, \text{ and } P(\hat{F}_{kNT,i}) \to 0 \text{ as } (N, T) \to \infty \text{ for each } i \in \hat{G}_k \text{ and } k = 1, ..., K_0.
\]

It is *uniformly consistent* if

\[
P\left( \bigcup_{k=1}^{K_0} \hat{E}_{kNT} \right) \to 0 \text{ and } P\left( \bigcup_{k=1}^{K_0} \hat{F}_{kNT} \right) \to 0 \text{ as } (N, T) \to \infty.
\]
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) $P \left( \bigcup_{k=1}^{K_0} \hat{E}_{kNT} \right) \leq \sum_{k=1}^{K_0} P \left( \hat{E}_{kNT} \right) \to 0$ as $(N, T) \to \infty$,
(ii) $P \left( \bigcup_{k=1}^{K_0} \hat{F}_{kNT} \right) \leq \sum_{k=1}^{K_0} P \left( \hat{F}_{kNT} \right) \to 0$ as $(N, T) \to \infty$. 
Assumption A3. (i) \( \Phi_k \equiv \frac{1}{N_k T} \sum_{i \in G_k} \sum_{t=1}^T \tilde{x}_{it} \tilde{x}'_{it} \overset{P}{\to} \Phi_k > 0 \) as \((N, T) \to \infty\).

(ii) \( \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k} \sum_{t=1}^T \tilde{x}_{it} \tilde{u}_{it} - B_{kNT} \overset{D}{\to} N(0, \Psi_k) \) as \((N, T) \to \infty\) where
\[ B_{kNT} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k} \sum_{t=1}^T \mathbb{E}(x_{it} \tilde{u}_{it}) \]
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
\[
\sqrt{N_k T} \left( \hat{\alpha}_k - \alpha_k^0 \right) - \Phi_k^{-1} B_{kNT} \overset{D}{\to} N(0, \Phi_k^{-1} \Psi_k \Phi_k^{-1}) \text{ for } k = 1, \ldots, K_0.
\]
Penalized Least Squares Estimation

Oracle Property

If the individual’s group identity is known, the WG estimator of $\alpha_k^0$ 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 $\sqrt{N_k T} \left( \bar{\alpha}_k - a^0_k \right) - \bar{\Phi}_k^{-1} B_{kNT} \overset{D}{\rightarrow} N \left( 0, \Phi_k^{-1} \Psi_k \Phi_k^{-1} \right)$.

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 $\sqrt{N_k T} \left( \hat{\alpha}_k - a^0_k \right) = \sqrt{N_k T} \left( \hat{\alpha}_{\hat{G}_k} - a^0_k \right) + o_P(1)$, where $\hat{\alpha}_{\hat{G}_k}$ is the post-Lasso estimator:

$$\hat{\alpha}_{\hat{G}_k} = \left( \sum_{i \in \hat{G}_k} \sum_{t=1}^T \tilde{x}_{it} \tilde{x}_{it}' \right)^{-1} \sum_{i \in \hat{G}_k} \sum_{t=1}^T \tilde{x}_{it} \tilde{y}_{it}.$$
Suppose that Assumptions A1-A3 hold. Then
\[
\sqrt{N_k T} \left( \hat{\alpha}_{G_k} - \alpha_k^0 \right) - \Phi_k^{-1} \mathbf{B}_{kNT} \xrightarrow{D} \mathcal{N}(0, \Phi_k^{-1} \Psi_k \Phi_k^{-1}) \text{ for } k = 1, \ldots, K_0.
\]
Consider the following PLS criterion

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

where $1 \leq K \leq K_{\text{max}}$. 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 $\alpha^0_k$ by

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

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

Information criterion:

$$IC_1 (K, \lambda_1) = \ln \left[ \hat{\sigma}^2 \hat{G} (K, \lambda_1) \right] + \rho_1 NT pK, \quad (8)$$
1. Mixed Panel Structure Models:

\[ y_{it} = \beta_{(1)}^{0} x_{it(1)} + \beta_{i(2)}^{0} x_{it(2)} + \mu_i + u_{it}, \]  

(9)

where \( \beta_{i(2)}^{0} = \alpha_{k}^{0} \) if \( i \in G_{k}^{0} \) where \( k = 1, \ldots, K_{0} \) and \( G_{1}^{0}, \ldots, G_{K_{0}}^{0} \) form a partition for \( \{1, 2, \ldots, N\} \). See Pesaran, Shin, and Smith (1999).

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

\[ Q_{1,NT}(\theta, \mu) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi(w_{it}, \theta, \mu_i), \]  

(10)

where \( \theta \) is a common parameter, \( \mu = (\mu_1, \ldots, \mu_N) \), \( \phi = -\ln f \), and \( f(w_{it}, \theta^0, \mu^0_i) \) is the PDF of \( w_{it} \), and \( \mu^0_i = \alpha^0_k \) if \( i \in G_k^0 \) for \( k = 1, \ldots, K_0 \). The PLS objective function here takes the form

\[ Q_{1,NT,\lambda_1}^{(K_0)}(\theta, \mu, \alpha) = Q_{1,NT}(\theta, \mu) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \prod_{k=1}^{K_0} \| \mu_i - \alpha_k \|. \]
3. Group Patterns of Heterogeneity: Bonhomme and Manresa (2014) consider:

$$y_{it} = \theta^0'x_{it} + \mu_{g,t} + u_{it},$$  \hspace{1cm} (11)

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

- Note that $\mu_{g,t} = \lambda'_i f_t$ where $f_t = (\mu_{1t}, \ldots, \mu_{K_0t})'$, $\lambda_i = (0, \ldots, 1, \ldots, 0)'$ with 1 in the $k^{th}$ position if $i \in G^0_k$ for $k = 1, \ldots, K_0$ and zeros elsewhere, we may embed (11) in the more general model

$$y_{it} = \theta^0'x_{it} + \lambda^0'_i f^0_t + u_{it},$$  \hspace{1cm} (12)

where $\lambda^0_i = \alpha^0_k$ if $i \in G^0_k$ for $k = 1, \ldots, K_0$.

- A two-step approach. (1) obtain the Gaussian QMLES $\ddot{\theta}$, $\ddot{\lambda}_i$, and $\ddot{f}_t$ under certain identification restriction, (2) consider $y_{it} = \theta^0'x_{it} + \lambda^0'_i \ddot{f}_t + u_{it}$ by imposing: $\lambda^0_i = \alpha^0_k$ if $i \in G^0_k$ where $k = 1, \ldots, K_0$. 
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
  \[
  \Delta y_{it} = \beta_i^0 \Delta x_{it} + \Delta u_{it}.
  \]  
  \[
  (13)
  \]

- The PGMM criterion function
  \[
  Q_{2NT, \lambda_2}^{(K_0)} (\beta, \alpha) = Q_{2,NT} (\beta) + \frac{\lambda_2}{N} \sum_{i=1}^{N} \prod_{k=1}^{K_0} \| \beta_i - \alpha_k \|,
  \]  
  \[
  (14)
  \]
  where

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

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

If Assumption B1 holds, then

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

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

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

where $\left( \tilde{\alpha}_{(1)}, \ldots, \tilde{\alpha}_{(K_0)} \right)$ is a suitable permutation of $\left( \tilde{\alpha}_1, \ldots, \tilde{\alpha}_{K_0} \right)$.
Penalized GMM Estimation
Classification Consistency

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

**Theorem**

If Assumptions B1-B2 hold, then

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

(ii) \( P \left( \bigcup_{k=1}^{K_0} \tilde{F}_{kNT} \right) \leq \sum_{k=1}^{K_0} P \left( \tilde{F}_{kNT} \right) \to 0 \text{ as } (N, T) \to \infty. \)
The PGMM estimators \( \{ \tilde{\alpha}_k \} \) may fail to possess the oracle property.

If the group identities were known in advance, one could obtain \( \tilde{\alpha}_k \) as the minimizer of

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

Data Generating Processes (DGPs)

- 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)**

\[
y_{it} = \beta_i x_{it} + \mu_i + u_{it},
\]

\[
x_{it1} = 0.2\mu_i + z_{it1},
\]

\[
x_{it2} = 0.2\mu_i + z_{it2},
\]

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

\[
(\alpha_{1}^{0}, \alpha_{2}^{0}, \alpha_{3}^{0}) = \begin{pmatrix} 
0.4 \\
1.6 \\
1 \\
1 \\
0.4 \\
1.6 
\end{pmatrix}.
\]
DGP 2 (Static panel with endogeneity)

\[ y_{it} = \beta_i x_{it} + \mu_i + u_{it}, \]
\[ x_{it1} = 0.2\mu_i + 0.5z_{it1} + 0.5z_{it2} + 0.5e_{it}, \]
\[ x_{it2} \sim N(0, 1) \] 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 \( u_{it} \).

\[
\begin{pmatrix}
  u_{it} \\
  e_{it}
\end{pmatrix}
\sim N
\begin{pmatrix}
  \begin{pmatrix}
    0 \\
    0
  \end{pmatrix}, \\
  \begin{pmatrix}
    0 & 0.3 \\
    0.3 & 0
  \end{pmatrix}
\end{pmatrix}.
\]

\[
(\alpha_1^0, \alpha_2^0, \alpha_3^0) = \begin{pmatrix}
  \begin{pmatrix}
    0.2 \\
    1.8
  \end{pmatrix}, \\
  \begin{pmatrix}
    1 \\
    1
  \end{pmatrix}, \\
  \begin{pmatrix}
    1.8 \\
    0.2
  \end{pmatrix}
\end{pmatrix}.
\]
DGP 3 (PAR(1) with two exogenous regressors)

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

where \( x_{it2} \) and \( x_{it3} \) are two exogenous regressors and they are independent of all error terms. They follow the standard normal distribution.

\[ y_{i0} = \beta_{i2} x_{i02} + \beta_{i3} x_{i03} + \mu_i + u_{i0} \]

so that the observations in \( i \) is a strictly stationary time series with mean \( \mu_i \).

\[
(a_1^0, a_2^0, a_3^0) = \begin{pmatrix} 0.8 \\ 0.4 \\ 0.4 \end{pmatrix}, \begin{pmatrix} 0.6 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.4 \\ 1.6 \\ 1.6 \end{pmatrix}.
\]

\[ \bar{P}(\hat{E}) = \frac{1}{N} \sum_{i=1}^{N} \hat{P}(\hat{E}_{kNT,i}) \text{ and } \bar{P}(\hat{F}) = \frac{1}{N} \sum_{i=1}^{N} \hat{P}(\hat{F}_{kNT,i}). \]
### Table 1: Classification error for C-Lasso

<table>
<thead>
<tr>
<th>$C_\lambda$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
<th>1.6</th>
<th>3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$T$</td>
<td>$\bar{P}(\hat{E})$</td>
<td>$\bar{P}(\hat{F})$</td>
<td>$\bar{P}(\hat{E})$</td>
<td>$\bar{P}(\hat{F})$</td>
</tr>
<tr>
<td>DGP1</td>
<td>100</td>
<td>10</td>
<td>0.1805</td>
<td>0.0901</td>
<td>0.1899</td>
</tr>
<tr>
<td>PLS</td>
<td>100</td>
<td>20</td>
<td>0.0593</td>
<td>0.0289</td>
<td>0.0585</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>40</td>
<td>0.0103</td>
<td>0.0049</td>
<td>0.0098</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>10</td>
<td>0.1691</td>
<td>0.0848</td>
<td>0.1771</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>20</td>
<td>0.0586</td>
<td>0.0284</td>
<td>0.0556</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>40</td>
<td>0.0092</td>
<td>0.0044</td>
<td>0.0083</td>
</tr>
<tr>
<td>DGP2</td>
<td>100</td>
<td>10</td>
<td>0.2082</td>
<td>0.0993</td>
<td>0.2001</td>
</tr>
<tr>
<td>PGMM</td>
<td>100</td>
<td>20</td>
<td>0.1027</td>
<td>0.0485</td>
<td>0.0958</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>40</td>
<td>0.0321</td>
<td>0.0152</td>
<td>0.0307</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>10</td>
<td>0.2037</td>
<td>0.0980</td>
<td>0.1982</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>20</td>
<td>0.1020</td>
<td>0.0483</td>
<td>0.0942</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>40</td>
<td>0.0332</td>
<td>0.0158</td>
<td>0.0299</td>
</tr>
</tbody>
</table>
Monte Carlo Simulations

Classification error

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

<table>
<thead>
<tr>
<th>$C_\lambda$</th>
<th>$N$</th>
<th>$T$</th>
<th>$\bar{P}(\hat{E})$</th>
<th>$\bar{P}(\hat{F})$</th>
<th>$\bar{P}(\hat{E})$</th>
<th>$\bar{P}(\hat{F})$</th>
<th>$\bar{P}(\hat{E})$</th>
<th>$\bar{P}(\hat{F})$</th>
<th>$\bar{P}(\hat{E})$</th>
<th>$\bar{P}(\hat{F})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP3</td>
<td>100</td>
<td>10</td>
<td>0.2063</td>
<td>0.1038</td>
<td>0.1839</td>
<td>0.0908</td>
<td>0.1913</td>
<td>0.0937</td>
<td>0.2305</td>
<td>0.1092</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>20</td>
<td>0.1000</td>
<td>0.0501</td>
<td>0.0826</td>
<td>0.0404</td>
<td>0.0750</td>
<td>0.0357</td>
<td>0.0800</td>
<td>0.0391</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>40</td>
<td>0.0277</td>
<td>0.0137</td>
<td>0.0222</td>
<td>0.0106</td>
<td>0.0183</td>
<td>0.0085</td>
<td>0.0158</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>10</td>
<td>0.2025</td>
<td>0.1026</td>
<td>0.1714</td>
<td>0.0853</td>
<td>0.1709</td>
<td>0.0844</td>
<td>0.2079</td>
<td>0.0998</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>20</td>
<td>0.0983</td>
<td>0.0490</td>
<td>0.0794</td>
<td>0.0386</td>
<td>0.0703</td>
<td>0.0333</td>
<td>0.0716</td>
<td>0.0347</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>40</td>
<td>0.0255</td>
<td>0.0126</td>
<td>0.0209</td>
<td>0.0100</td>
<td>0.0173</td>
<td>0.0080</td>
<td>0.0151</td>
<td>0.0069</td>
</tr>
<tr>
<td>DGP3</td>
<td>100</td>
<td>10</td>
<td>0.3133</td>
<td>0.1551</td>
<td>0.2969</td>
<td>0.1464</td>
<td>0.2872</td>
<td>0.1406</td>
<td>0.2968</td>
<td>0.1440</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>20</td>
<td>0.1703</td>
<td>0.0838</td>
<td>0.1512</td>
<td>0.0746</td>
<td>0.1361</td>
<td>0.0660</td>
<td>0.1334</td>
<td>0.0628</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>40</td>
<td>0.0694</td>
<td>0.0338</td>
<td>0.0579</td>
<td>0.0283</td>
<td>0.0487</td>
<td>0.0232</td>
<td>0.0432</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>10</td>
<td>0.3081</td>
<td>0.1529</td>
<td>0.2868</td>
<td>0.1425</td>
<td>0.2778</td>
<td>0.1367</td>
<td>0.2826</td>
<td>0.1380</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>20</td>
<td>0.1691</td>
<td>0.0836</td>
<td>0.1527</td>
<td>0.0753</td>
<td>0.1324</td>
<td>0.0645</td>
<td>0.1265</td>
<td>0.0597</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>40</td>
<td>0.0727</td>
<td>0.0357</td>
<td>0.0587</td>
<td>0.0290</td>
<td>0.0490</td>
<td>0.0238</td>
<td>0.0434</td>
<td>0.0202</td>
</tr>
</tbody>
</table>
Monte Carlo Simulations

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$C_\lambda$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
<th>1.6</th>
<th>3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>C-Lasso</td>
<td>0.1010</td>
<td>0.0364</td>
<td>0.1116</td>
<td>0.1303</td>
<td>0.0293</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-lasso</td>
<td>0.0907</td>
<td>0.0282</td>
<td>0.1035</td>
<td>0.1274</td>
<td>0.0254</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0583</td>
<td>-0.0033</td>
<td>0.0583</td>
<td>-0.0033</td>
<td>0.0583</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>C-Lasso</td>
<td>0.0590</td>
<td>0.0154</td>
<td>0.0560</td>
<td>0.0183</td>
<td>0.0507</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-lasso</td>
<td>0.0450</td>
<td>0.0066</td>
<td>0.0467</td>
<td>0.0092</td>
<td>0.0470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0399</td>
<td>-0.0021</td>
<td>0.0399</td>
<td>-0.0021</td>
<td>0.0399</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>C-Lasso</td>
<td>0.0347</td>
<td>0.0096</td>
<td>0.0348</td>
<td>0.0047</td>
<td>0.0305</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-lasso</td>
<td>0.0292</td>
<td>0.0012</td>
<td>0.0293</td>
<td>0.0002</td>
<td>0.0291</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0281</td>
<td>-0.0010</td>
<td>0.0281</td>
<td>-0.0010</td>
<td>0.0281</td>
</tr>
</tbody>
</table>
### Table 3: PGMM Estimation of $\beta_1$ in DGP 2

<table>
<thead>
<tr>
<th>$C_\lambda$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
<th>1.6</th>
<th>3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>N T</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>100 10</td>
<td>0.1906</td>
<td>0.1093</td>
<td>0.1907</td>
<td>0.1242</td>
<td>0.1388</td>
</tr>
<tr>
<td>C-Lasso</td>
<td>0.1416</td>
<td>0.0152</td>
<td>0.1368</td>
<td>0.0251</td>
<td>0.1413</td>
</tr>
<tr>
<td>Post-lasso</td>
<td>0.1603</td>
<td>0.0684</td>
<td>0.1586</td>
<td>0.0811</td>
<td>0.0928</td>
</tr>
<tr>
<td>C-Lasso BC</td>
<td>0.1603</td>
<td>-0.0001</td>
<td>0.0993</td>
<td>-0.0001</td>
<td>0.0993</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.0993</td>
<td>0.0993</td>
<td>0.0993</td>
<td>0.0993</td>
<td>0.0993</td>
</tr>
<tr>
<td>100 20</td>
<td>0.1179</td>
<td>0.0560</td>
<td>0.1176</td>
<td>0.0683</td>
<td>0.1182</td>
</tr>
<tr>
<td>C-Lasso</td>
<td>0.0838</td>
<td>0.0138</td>
<td>0.0815</td>
<td>0.0181</td>
<td>0.0810</td>
</tr>
<tr>
<td>Post-lasso</td>
<td>0.0986</td>
<td>0.0374</td>
<td>0.0978</td>
<td>0.0464</td>
<td>0.0986</td>
</tr>
<tr>
<td>C-Lasso BC</td>
<td>0.0986</td>
<td>-0.0004</td>
<td>0.0680</td>
<td>-0.0004</td>
<td>0.0680</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.0680</td>
<td>0.0680</td>
<td>0.0680</td>
<td>0.0680</td>
<td>0.0680</td>
</tr>
<tr>
<td>100 40</td>
<td>0.0712</td>
<td>0.0400</td>
<td>0.0754</td>
<td>0.0422</td>
<td>0.0761</td>
</tr>
<tr>
<td>C-Lasso</td>
<td>0.0519</td>
<td>0.0136</td>
<td>0.0522</td>
<td>0.0129</td>
<td>0.0519</td>
</tr>
<tr>
<td>Post-lasso</td>
<td>0.0614</td>
<td>0.0274</td>
<td>0.0632</td>
<td>0.0282</td>
<td>0.0637</td>
</tr>
<tr>
<td>C-Lasso BC</td>
<td>0.0614</td>
<td>0.0007</td>
<td>0.0492</td>
<td>0.0007</td>
<td>0.0492</td>
</tr>
<tr>
<td>Oracle</td>
<td>0.0492</td>
<td>0.0492</td>
<td>0.0492</td>
<td>0.0492</td>
<td>0.0492</td>
</tr>
<tr>
<td>$N$</td>
<td>$T$</td>
<td>$C_\lambda$</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>C-Lasso</td>
<td>0.1331</td>
<td>-0.1216</td>
<td>0.1264</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-lasso</td>
<td>0.1011</td>
<td>-0.0863</td>
<td>0.1041</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.1220</td>
<td>-0.1088</td>
<td>0.1157</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0922</td>
<td>-0.0745</td>
<td>0.0949</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0928</td>
<td>-0.0855</td>
<td>0.0928</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>C-Lasso</td>
<td>0.0782</td>
<td>-0.0711</td>
<td>0.0740</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-lasso</td>
<td>0.0539</td>
<td>-0.0431</td>
<td>0.0558</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0723</td>
<td>-0.0643</td>
<td>0.0682</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0494</td>
<td>-0.0368</td>
<td>0.0508</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0527</td>
<td>-0.0469</td>
<td>0.0527</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>C-Lasso</td>
<td>0.0428</td>
<td>-0.0372</td>
<td>0.0405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-lasso</td>
<td>0.0289</td>
<td>-0.0224</td>
<td>0.0295</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C-Lasso BC</td>
<td>0.0401</td>
<td>-0.0339</td>
<td>0.0378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-Lasso BC</td>
<td>0.0266</td>
<td>-0.0193</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oracle</td>
<td>0.0285</td>
<td>-0.0236</td>
<td>0.0285</td>
</tr>
</tbody>
</table>
Table 5: PGMM Estimation of $\beta_1$ in DGP 3

<table>
<thead>
<tr>
<th>$C_\lambda$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.8</th>
<th>1.6</th>
<th>3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N \ T$</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>100 10 C-Lasso</td>
<td>0.1823 -0.1065 0.192 -0.1241 0.1980 -0.1417 0.2090 -0.1627 0.2271 -0.1817</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-lasso</td>
<td>0.1304 -0.0352 0.1231 -0.0331 0.1161 -0.0311 0.1137 -0.0352 0.1202 -0.0427</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-Lasso BC</td>
<td>0.1494 -0.0698 0.1509 -0.0800 0.1516 -0.0897 0.1572 -0.1047 0.1729 -0.1206</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>0.0664 -0.0013 0.0664 -0.0013 0.0664 -0.0013 0.0664 -0.0013 0.0664 -0.0013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 20 C-Lasso</td>
<td>0.0808 -0.0319 0.0858 -0.0478 0.0974 -0.0687 0.1114 -0.0888 0.1247 -0.1035</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-lasso</td>
<td>0.0584 -0.0010 0.0565 -0.0031 0.0546 -0.0068 0.0538 -0.0109 0.0554 -0.0138</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-Lasso BC</td>
<td>0.0678 -0.0175 0.0690 -0.0275 0.0739 -0.0411 0.0814 -0.0548 0.0904 -0.0648</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>0.0399 -0.0027 0.0399 -0.0027 0.0399 -0.0027 0.0399 -0.0027 0.0399 -0.0027</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 40 C-Lasso</td>
<td>0.0442 -0.0126 0.0447 -0.0198 0.0519 -0.0329 0.0646 -0.0491 0.0742 -0.0606</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-lasso</td>
<td>0.0356 0.0025 0.0334 0.0006 0.0327 -0.0018 0.0325 -0.0037 0.0320 -0.0046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-Lasso BC</td>
<td>0.0395 -0.0047 0.0384 -0.0094 0.0406 -0.0173 0.0459 -0.0268 0.0507 -0.0333</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle</td>
<td>0.0274 -0.0011 0.0274 -0.0011 0.0274 -0.0011 0.0274 -0.0011 0.0274 -0.0011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

\[ S_{it} = \beta_1 i S_{i,t-1} + \beta_2 i I_{it} + \beta_3 i R_{it} + \beta_4 i G_{it} + \mu_i + u_{it}, \]  

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)
Empirical Application

World Development Indicators: 1995–2000, 56 countries.

Table 6: Summary statistics for the savings data set

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th>s.e.</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Savings rate</td>
<td>22.099</td>
<td>20.790</td>
<td>8.833</td>
<td>-3.207</td>
<td>53.434</td>
</tr>
<tr>
<td>Inflation rate</td>
<td>7.724</td>
<td>4.853</td>
<td>15.342</td>
<td>-3.846</td>
<td>293.679</td>
</tr>
<tr>
<td>Real interest rate</td>
<td>7.422</td>
<td>5.927</td>
<td>10.062</td>
<td>-63.761</td>
<td>93.915</td>
</tr>
<tr>
<td>Per capita GDP growth rate</td>
<td>2.855</td>
<td>2.971</td>
<td>3.865</td>
<td>-17.545</td>
<td>14.060</td>
</tr>
</tbody>
</table>
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:

\[ H_0(K_0) : K = K_0 \text{ versus } H_1(K_0) : K_0 < K \leq K_{\text{max}}. \]  

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

Table 7: Test statistics

<table>
<thead>
<tr>
<th>( H_0(K_0) )</th>
<th>( c = 1 )</th>
<th>( c = 1.5 )</th>
<th>( c = 2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics</td>
<td>3.040</td>
<td>1.397</td>
<td>0.715</td>
</tr>
<tr>
<td>p-values</td>
<td>0.001</td>
<td>0.081</td>
<td>0.237</td>
</tr>
<tr>
<td>Holm adjusted p-value</td>
<td>0.002</td>
<td>0.081</td>
<td>NA</td>
</tr>
</tbody>
</table>
Empirical Application

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.
### Table 8: Estimation results

<table>
<thead>
<tr>
<th>Slope coefficients</th>
<th>Common</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FE</td>
<td>C-Lasso</td>
<td>post-Lasso</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>0.6203***</td>
<td>0.5510***</td>
<td>0.5548***</td>
</tr>
<tr>
<td></td>
<td>(0.1330)</td>
<td>(0.1090)</td>
<td>(0.1057)</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>0.0303</td>
<td>-0.1154**</td>
<td>-0.1068**</td>
</tr>
<tr>
<td></td>
<td>(0.0484)</td>
<td>(0.0464)</td>
<td>(0.0458)</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.0068</td>
<td>-0.0419</td>
<td>-0.0273</td>
</tr>
<tr>
<td></td>
<td>(0.0432)</td>
<td>(0.0490)</td>
<td>(0.0476)</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>0.1880***</td>
<td>0.2771***</td>
<td>0.3055***</td>
</tr>
<tr>
<td></td>
<td>(0.0450)</td>
<td>(0.0470)</td>
<td>(0.0452)</td>
</tr>
</tbody>
</table>

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:

\[ y_{it} = \beta_{i0}'x_{it} + \lambda_{i0}'f^0_t + \epsilon_{it}, \]

where \( \lambda^0_i \) and \( f^0_t \) denote an \( R_0 \times 1 \) vector of factor loadings and common factors, respectively.

- **Homogenous**: \( \beta^0_i = \beta^0 \). 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.


Inefficient and slow convergence rate if the models have homogeneous slopes.
Panel Structure Models with IFEs

- Penalized principal component (PPC) estimation

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

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

- Concentrate \( \Lambda \) out:

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

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

- Further concentrate \( F \) out:

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

where \( Q_{NT} (\beta) = \frac{1}{T} \sum_{r=R_0+1}^{T} \mu_r \left[ \frac{1}{N} \sum_{i=1}^{N} (Y_i - X_i \beta_i) (Y_i - X_i \beta_i)' \right] \).
C-Lasso estimates: $\hat{\alpha} \equiv (\hat{\alpha}_1, \ldots, \hat{\alpha}_K)$ and $\hat{\beta} \equiv (\hat{\beta}_1, \ldots, \hat{\beta}_N)$.

Estimate $(\Lambda, F)$ via PC analysis (e.g., Bai and Ng’s 2002) under the identification restrictions: $F'F/T = I_{R_0}$ and $\Lambda'\Lambda = \text{diag}$:

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

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 $(\checkmark)$
- Oracle property $(\checkmark)$
Conclusions and Future Work
Conclusions and Future Work

Conclusions

- 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
Conclusions and Future Work

Future Work

- 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!
Supplement 1: Penalized Least Squares Estimation

Numerical Algorithm

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
   \[\sum_{i=1}^N \| \hat{\beta}_i^{(0)} - \hat{\alpha}_k^{(0)} \| \neq 0 \text{ for each } k = 2, \ldots, K_0.\]

2. Given $\hat{\alpha}^{(r-1)} = (\hat{\alpha}_1^{(r-1)}, \ldots, \hat{\alpha}_{K_0}^{(r-1)})$ and $\hat{\beta}^{(r-1)} = (\hat{\beta}_1^{(r-1)}, \ldots, \hat{\beta}_N^{(r-1)})$,
   - In Step $r \geq 1$, we first choose $(\beta, \alpha_1)$ to minimize
     \[Q_{K_0 N T}^{(r,1)} (\beta, \alpha_1) = Q_{1, N T} (\beta) + \frac{\lambda_1}{N} \sum_{i=1}^N \| \beta_i - \alpha_1 \| \prod_{k \neq 1}^{K_0} \| \hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)} \|\]
     and obtain the updated estimate $(\hat{\beta}^{(r,1)}, \hat{\alpha}_1^{(r)})$ of $(\beta, \alpha_1)$.
   - Next choose $(\beta, \alpha_2)$ to minimize
     \[Q_{K_0 N T}^{(r,2)} (\beta, \alpha_2) = Q_{1, N T} (\beta) + \frac{\lambda_1}{N} \sum_{i=1}^N \| \beta_i - \alpha_2 \| \| \hat{\beta}_i^{(r,1)} - \hat{\alpha}_1^{(r)} \|\]
     \[\times \prod_{k \neq 1, 2}^{K_0} \| \hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)} \|\]
     to obtain the updated estimate $(\hat{\beta}^{(r,2)}, \hat{\alpha}_2^{(r)})$ of $(\beta, \alpha_2)$.
Supplement 1: Penalized Least Squares Estimation
Numerical Algorithm

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

\[
Q^{(r,K_0)}_{K_0NT}(\beta, \alpha_{K_0}) = Q_{1,NT}(\beta) + \frac{\lambda_1}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_K\| \prod_{k=1}^{K_0-1} \left\|\hat{\beta}^{(r,K_0-1)}_i - \hat{\alpha}_k^{(r)}\right\|
\]

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

3 Repeat step 2 until a convergence criterion is met.
Supplement 2: Penalized GMM Estimation

Assumptions

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 $\breve{\xi}_{it} = (\Delta y_{it}, (\Delta x_{it})', z_{it}')'$. Define $\rho (\breve{\xi}_{it}, \beta) = z_{it} (\Delta y_{it} - \beta' \Delta x_{it})$ and

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

ASSUMPTION B1. (i) $\mathbb{E} [\rho (\breve{\xi}_{it}, \beta_{i}^0)] = 0$.
(ii) $\sup_{\beta \in B_i} \bar{\rho}_{i, T} (\beta) = O_P (1)$ and $\frac{1}{N} \sum_{i=1}^{N} \| \bar{\rho}_{i, T} (\beta; i) \|^2 = O_P (1)$ for any $\beta; i \in B_i$ and $i = 1, \ldots, N$.
(iii) $\tilde{Q}_{i, z\Delta x} = \bar{Q}_{i, z\Delta x} + o_P (1)$ for each $i = 1, \ldots, N$ and

$$
\liminf_{(N, T) \to \infty} \min_{1 \leq i \leq N} \mu_{\min} (\tilde{Q}', z\Delta x \bar{Q}, z\Delta x) = c_{\tilde{Q}} > 0.
$$

(iv) There exist $W_i$ such that $\max_{1 \leq i \leq N} \| W_{iNT} - W_i \| = O_P (1)$ and

$$
\liminf_{N \to \infty} \min_{1 \leq i \leq N} \mu_{\min} (W_i) = c_W > 0.
$$

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

Su, Shi, and Phillips (SMU and Yale) Identifying Latent Structures in Panel Data July 9, 2014 64 / 70
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) \( \tilde{A}_k \equiv \frac{1}{N_k} \sum_{i \in G_k^0} \bar{Q}'_{i,z\Delta x} W_i \bar{Q}_{i,z\Delta x} \to A_k > 0 \) 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) \( \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} z_{it} \Delta u_{it} \xrightarrow{D} N(0, V_k) \).
Supplement 2: Penalized GMM Estimation
Improved Convergence and Asymptotic Properties of Post-Lasso

Theorem

Suppose that Assumptions B1-B3 hold. Then
\[
\sqrt{N_k T} \left( \tilde{\alpha}_k - \alpha_0^k \right) - \tilde{A}_k^{-1} B_{kNT} \xrightarrow{D} N(0, A_k^{-1} C_k A_k^{-1}) \text{ for } k = 1, \ldots, K_0.
\]

Theorem

Suppose that Assumptions B1-B4 hold. Then
\[
\sqrt{N_k T} \left( \tilde{\alpha}_{\tilde{G}_k} - \alpha_0^k \right) \xrightarrow{D} N(0, \Omega_k) \quad \text{where}
\]
\[
\Omega_k = \left[ Q_{z\Delta x}^{(k)} \ W(k) \ Q_{z\Delta x}^{(k)} \right]^{-1} Q_{z\Delta x}^{(k)} \ W(k) \ V_k \ W(k) \ Q_{z\Delta x}^{(k)} \left[ Q_{z\Delta x}^{(k)} \ W(k) \ Q_{z\Delta x}^{(k)} \right]^{-1}
\]
and \( k = 1, \ldots, K_0. \)

Note that \( \sqrt{N_k T} \left( \tilde{\alpha}_{\tilde{G}_k} - \alpha_0^k \right) = \sqrt{N_k T} \left( \tilde{\alpha}_k - \alpha_0^k \right) + o_P(1). \) That is, the post-Lasso GMM estimator \( \tilde{\alpha}_{\tilde{G}_k} \) is asymptotically equivalent to the infeasible estimate \( \tilde{\alpha}_k. \)
Supplement 3: Determining the number of groups

- Basic idea:

\[ H_0(K_0) : K = K_0 \text{ versus } H_1(K_0) : K_0 < K \leq K_{\text{max}}. \]  \hfill (21)

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

- Estimation:

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

- \( K_0 = 1 \) : 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:

\[
\hat{u}_{it} = (y_{it} - \bar{y}_i) - (X_{it} - \bar{X}_i)' \hat{\beta}_i
\]

\[
= u_{it} - \bar{u}_i + (X_{it} - \bar{X}_i)' (\beta_0^i - \hat{\beta}_i),
\]

(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 \( \beta_0^i \) 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

\[
\hat{u}_{it} = \upsilon_i + \phi_i' X_{it} + \eta_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T,
\]

(23)

and test the null hypothesis

\[
\mathbb{H}_0^*: \phi_i = 0 \text{ for all } i = 1, ..., N.
\]
We construct an LM-type test statistic by concentrating the intercept \( v_i \) out in (23). Consider the Gaussian quasi-likelihood function for \( \hat{u}_{it} \):

\[
\ell(\phi) = \sum_{i=1}^{N} (\hat{u}_i - M_0X_i\phi_i)'(\hat{u}_i - M_0X_i\phi_i),
\]

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

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

We can verify that

\[
LM_{NT}(K_0) = \sum_{i=1}^{N} \hat{u}_i'M_0X_i \left( X_i'M_0X_i \right)^{-1} X_i'M_0\hat{u}_i.
\] (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

$$B_{NT} \equiv N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it}^2 h_{i(tt)}$$

and

$$V_{NT} \equiv 4 T^{-2} N^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} E\left[u_{it} \bar{b}_{it}' \sum_{s=1}^{t-1} \bar{b}_{is} u_{is}\right]^2.$$

**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:

$$\hat{J}_{NT} (K_0) \equiv \left(N^{-1/2} \hat{L} M_{NT} (K_0) - \hat{B}_{NT} (K_0)\right) / \sqrt{\hat{V}_{NT} (K_0)}.$$