Determining the Number of Groups in Latent Panel Structures

Xun Lu[a] and Liangjun Su[b]

[a]Hong Kong University of Science and Technology
[b]Singapore Management University

July 10, 2014

Outline

1. The Panel Structure Model and Hypotheses
2. Estimation and Test Statistic
3. Asymptotic Properties
4. Monte Carlo Simulations
5. Empirical Application
6. Conclusion

1. The Panel Structure Model and Hypotheses

The Panel Structure Model

- Complete slope homogeneity

$$y_{it} = \beta^{0\prime} x_{it} + \mu_i + u_{it}, \ \ i = 1, \dots, N \text{ and } t = 1, \dots, T, \tag{1}$$

  - 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,
  - easy estimation and inference,
  - but frequently questioned and rejected in empirical studies.
- Complete slope heterogeneity (random coefficient):

$$y_{it} = \beta_i^{0\prime} x_{it} + \mu_i + u_{it} \tag{2}$$

  - slow convergence rate $\sqrt{T}$.

The Panel Structure Model

- Latent group structure:

$$y_{it} = \beta_i^{0\prime} x_{it} + \mu_i + u_{it} \tag{3}$$

- Group structure of β_i^0 :

$$\beta_i^0 = \begin{cases} \alpha_1^0 & \text{if } i \in G_1^0 \\ \vdots & \vdots \\ \alpha_K^0 & \text{if } i \in G_K^0 \end{cases} . \tag{4}$$

  - N individuals belong to K groups, where K is unknown.
  - β_i^0's are homogeneous within each groups, but heterogeneous across the K groups.
  - $\alpha_1^0, ..., \alpha_K^0$ are unknown.
  - Classification is unknown.

The Panel Structure Model

- 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.
- 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
  - how to determine the unknown number of groups;
  - how to identify the individual's group membership.

Related Literature

- Known group structure
  - Bester and Hansen (2009): a panel structure model where individuals are grouped according to some external classification, geographic location, or observable explanatory variables.
- Unknown group structure
  - 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 (2012) 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 Panel Structure Model

- Su, Shi, and Phillips (2013, SSP) propose a new **classifier Lasso (C-Lasso)** 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 requires the knowledge on the number of groups.**
- This paper provides a testing procedure to determine the number of groups.

Hypotheses

- Hypotheses

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

where K_0 and $K_{\max}$ are pre-specified by researchers.

- Basic idea:
  - 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 $K = K^*$.
- When $K_0 = 1$, the test becomes a test for homogeneity in the slope coefficients.
- When $K_0 > 1$, we test whether the group structure is correctly specified.

2. Estimation and Test Statistic

Estimation under the null

- $K_0 = 1$, homogenous panel. Standard method.
- $K_0 > 1$, panel structure 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:

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

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

Estimation under the null

- PLS objective function

$$Q_{NT,\lambda}^{(K_0)}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Q_{NT}(\boldsymbol{\beta}) + \frac{\lambda}{N} \sum_{i=1}^{N} \Pi_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \qquad (5)$$

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

- SSP's C-Lasso estimates: $\hat{\boldsymbol{\alpha}} \equiv (\hat{\alpha}_1, ..., \hat{\alpha}_{K_0})$ and $\hat{\boldsymbol{\beta}} \equiv (\hat{\beta}_1, ..., \hat{\beta}_N)$.
- Numerical algorithm: a sequence of convex problems.

Test statistic

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

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

  where $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$, and $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$.
- Under the null, $\hat{\beta}_i$ is a consistent estimator of β_i^0. Hence, $\hat{u}_{it}$ should be close to u_{it} and x_{it} should not have any predictive power for u_{it}.
- Run the regression

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

  and test

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

Test statistic

- Consider the Gaussian quasi-likelihood function for $\hat{u}_{it}$:

$$\ell\left(\boldsymbol{\phi}\right)=\sum_{i=1}^{N}\left(\hat{u}_i-M_0X_i\phi_i\right)'\left(\hat{u}_i-M_0X_i\phi_i\right),$$

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

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

where we make the dependence of $LM_{NT}\left(K_0\right)$ on K_0 explicit.

- We can verify that

$$LM_{NT}\left(K_0\right)=\sum_{i=1}^{N}\hat{u}_i'M_0X_i\left(X_i'M_0X_i\right)^{-1}X_i'M_0\hat{u}_i. \tag{6}$$

3. Asymptotic Properties

Asymptotic normality

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 $\mathbb{H}_0(K_0)$,

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

Asymptotic normality

- Consistent estimates of B_{NT} and V_{NT}:

$$\hat{B}_{NT}(K_0) = N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2 h_{i,tt},$$

$$\hat{V}_{NT}(K_0) = 4T^{-2}N^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} \left[\hat{u}_{it} \hat{b}_{it}' \sum_{s=1}^{t-1} \hat{b}_{is} \hat{u}_{is} \right]^2$$

where $\hat{b}_{it} = \hat{\Omega}_i^{-1/2}(X_{it} - T^{-1} \sum_{s=1}^{T} X_{is})$.

- Feasible test statistic:

$$\hat{J}_{NT}(K_0) \equiv \left(N^{-1/2} LM_{NT}(K_0) - \hat{B}_{NT}(K_0) \right) / \sqrt{\hat{V}_{NT}(K_0)}. \quad (7)$$

Theorem

Suppose Assumptions A.1-A.3 hold. Then under $\mathbb{H}_0(K_0)$, $\hat{B}_{NT}(K_0) = B_{NT} + o_P(1)$, $\hat{V}_{NT}(K_0) = V_{NT} + o_P(1)$, *and* $\hat{J}_{NT}(K_0) \xrightarrow{D} N(0,1)$.

Alternatives

- Let $\mathcal{G}_K = \{(G_1, ..., G_K) : \cup_{k=1}^K G_k = \{1, ..., N\}, G_k \cap G_j = \varnothing$ for $j \neq k\}$.
- **Assumption A.4**. (i) $N^{-1} \sum_{i=1}^N \left\|\beta_i^0\right\|^2 = O_P(1)$.
  (ii) $\inf_{(G_1,...,G_{K_0}) \in \mathcal{G}_{K_0}} \min_{(\alpha_1,...,\alpha_{K_0})} N^{-1} \sum_{k=1}^{K_0} \sum_{i \in G_k} \left\|\beta_i^0 - \alpha_k\right\|^2 \xrightarrow{P} \underline{c}_{K_0} > 0$ as $N \to \infty$.
- A.4(i) is trivially satisfied if β_i^0's are uniformly bounded or random with finite second moments.
- A.4(ii) says that one cannot group $\left\{\beta_i^0\right\}_{i=1}^N$ into K_0 groups by leaving out an "insignificant" number of unclassified individuals.

Theorem

Under $\mathbb{H}_1(K_0)$ with possible diverging $K_{\max}$ and random coefficients, $P(\hat{J}_{NT}(K_0) \geq c_{NT}) \to 1$ as $(N, T) \to \infty$ for any non-stochastic sequence $c_{NT} = o\left(N^{1/2}T\right)$.

4. Monte Carlo Simulations

DGPs

- DGP 1: $y_{it} = \beta^0_{1i} X_{1it} + \beta^0_{2i} X_{2it} + \mu_i + u_{it}$,
- DGPs 2-4: $y_{it} = \beta^0_{1i} X_{1it} + \beta^0_{2i} y_{i,t-1} + \mu_i + u_{it}$,
  - where $X_{it} = \xi_{it} + \mu_i$, and μ_i, ξ_{it}, u_{it} are IID $N(0,1)$ variables.
  - In DGPs 1 and 2:
    $$\left(\beta^0_{1i}, \beta^0_{2i}\right) = \begin{cases} (0.5, -0.5) & \text{with probability } 0.3 \\ (-0.5, 0.5) & \text{with probability } 0.3 \\ (0,0) & \text{with probability } 0.4 \end{cases}.$$
  - In DGP 3, we consider a completely heterogeneous panel structure where β^0_{1i} and β^0_{2i} follow $N(0.5, 1)$ and $\mathrm{U}(-0.5, 0.5)$, respectively.
  - In DGP 4:

    $$\left(\beta^0_{1i}, \beta^0_{2i}\right) = \begin{cases} (0.5 + 0.1\nu_{1i}, -0.5 + 0.1\nu_{2i}) & \text{with probability } 0.3 \\ (-0.5 + 0.1\nu_{1i}, 0.5 + 0.1\nu_{2i}) & \text{with probability } 0.3 \\ (0.1\nu_{1i}, 0.1\nu_{2i}) & \text{with probability } 0.4 \end{cases},$$

    where ν_{1i} and ν_{2i} are each IID $N(0,1)$, mutually independent, and independent of μ_i, ξ_{it}, and u_{it}.

Implementation

- $\lambda = c \cdot s_Y \cdot T^{-3/4}$, where s_Y is the sample standard deviation of Y_{it} and c is some constant. $c = 0.25$, 0.5 and 1.
- $N = 40, 80$, $T = 10, 20, 40$.
- In practice, recommended nominal level: $1/T$

Table 1: Empirical rejection frequency (DGP 1)

K	N	T	$c = 0.5$			$c = 1$		
			0.01	0.05	0.10	0.01	0.05	0.10
	40	10	0.998	0.999	1	0.998	0.999	1
	40	20	1	1	1	1	1	1
1	40	40	1	1	1	1	1	1
	80	10	1	1	1	1	1	1
	80	20	1	1	1	1	1	1
	80	40	1	1	1	1	1	1
	40	10	0.176	0.400	0.525	0.144	0.358	0.484
	40	20	0.683	0.865	0.935	0.665	0.855	0.932
2	40	40	1	1	1	1	1	1
	80	10	0.385	0.630	0.749	0.323	0.581	0.715
	80	20	0.964	0.996	0.999	0.966	0.993	0.999
	80	40	1	1	1	1	1	1
	40	10	0.016	0.061	0.104	0.012	0.049	0.094
	40	20	0.014	0.039	0.077	0.009	0.037	0.075
3	40	40	0.007	0.034	0.061	0.012	0.036	0.065
	80	10	0.025	0.080	0.152	0.021	0.078	0.138
	80	20	0.020	0.039	0.063	0.019	0.043	0.078
	80	40	0.009	0.042	0.073	0.007	0.041	0.078

Table 1: Empirical rejection frequency (DGP 2)

			$c = 0.5$			$c = 1$		
K	N	T	0.01	0.05	0.10	0.01	0.05	0.10
	40	10	1	1	1	1	1	1
	40	20	1	1	1	1	1	1
1	40	40	1	1	1	1	1	1
	80	10	1	1	1	1	1	1
	80	20	1	1	1	1	1	1
	80	40	1	1	1	1	1	1
	40	10	0.152	0.361	0.482	0.118	0.289	0.421
	40	20	0.766	0.924	0.959	0.739	0.907	0.957
2	40	40	1	1	1	1	1	1
	80	10	0.376	0.623	0.769	0.266	0.503	0.659
	80	20	0.986	0.998	0.999	0.980	0.997	0.999
	80	40	1	1	1	1	1	1
	40	10	0.008	0.036	0.062	0.004	0.026	0.054
	40	20	0.014	0.039	0.075	0.009	0.034	0.073
3	40	40	0.006	0.021	0.052	0.008	0.023	0.050
	80	10	0.014	0.043	0.087	0.008	0.034	0.068
	80	20	0.022	0.044	0.086	0.019	0.045	0.080
	80	40	0.013	0.048	0.088	0.014	0.049	0.086

Table 1: Empirical rejection frequency (DGP 3)

K	N	T	$c = 0.5$			$c = 1$		
			0.01	0.05	0.10	0.01	0.05	0.10
	40	10	1	1	1	1	1	1
	40	20	1	1	1	1	1	1
1	40	40	1	1	1	1	1	1
	80	10	1	1	1	1	1	1
	80	20	1	1	1	1	1	1
	80	40	1	1	1	1	1	1
	40	10	0.999	1	1	0.997	1	1
	40	20	1	1	1	1	1	1
2	40	40	1	1	1	1	1	1
	80	10	1	1	1	1	1	1
	80	20	1	1	1	1	1	1
	80	40	1	1	1	1	1	1
	40	10	0.974	0.995	0.997	0.981	0.996	0.999
	40	20	1	1	1	1	1	1
3	40	40	1	1	1	1	1	1
	80	10	1	1	1	1	1	1
	80	20	1	1	1	1	1	1
	80	40	1	1	1	1	1	1

Table 1: Empirical rejection frequency (DGP 4)

			$c = 0.5$			$c = 1$		
K	N	T	0.01	0.05	0.10	0.01	0.05	0.10
	40	10	1	1	1	1	1	1
	40	20	1	1	1	1	1	1
1	40	40	1	1	1	1	1	1
	80	10	1	1	1	1	1	1
	80	20	1	1	1	1	1	1
	80	40	1	1	1	1	1	1
	40	10	0.243	0.501	0.633	0.181	0.404	0.552
	40	20	0.907	0.970	0.991	0.891	0.958	0.985
2	40	40	1	1	1	1	1	1
	80	10	0.590	0.805	0.893	0.476	0.705	0.839
	80	20	0.999	1	1	0.999	1	1
	80	40	1	1	1	1	1	1
	40	10	0.034	0.089	0.145	0.019	0.079	0.124
	40	20	0.104	0.208	0.290	0.098	0.202	0.282
3	40	40	0.380	0.617	0.733	0.370	0.611	0.725
	80	10	0.066	0.157	0.260	0.052	0.142	0.222
	80	20	0.209	0.385	0.512	0.210	0.379	0.501
	80	40	0.769	0.913	0.954	0.759	0.909	0.947

Table 2: Frequency of number of groups determined: DGP 1

	N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
	40	10	0	0.475	0.421	0.075	0.026	0.003
	40	20	0	0.133	0.828	0.025	0.014	0
$c=0.5$	40	40	0	0	0.981	0.013	0.006	0
	80	10	0	0.251	0.599	0.127	0.023	0
	80	20	0	0.004	0.957	0.031	0.007	0.001
	80	40	0	0	0.976	0.014	0.010	0
	40	10	0	0.516	0.390	0.065	0.026	0.003
	40	20	0	0.144	0.818	0.027	0.010	0.001
$c=1$	40	40	0	0	0.977	0.016	0.006	0.001
	80	10	0	0.286	0.577	0.115	0.020	0.002
	80	20	0	0.007	0.949	0.029	0.015	0
	80	40	0	0	0.977	0.011	0.012	0

Table 2: Frequency of number of groups determined: DGP 2

	N	T	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K>5$
	40	10	0	0.518	0.421	0.048	0.013	0
	40	20	0	0.075	0.886	0.033	0.006	0
$c=0.5$	40	40	0	0	0.987	0.012	0.000	0.001
	80	10	0	0.232	0.682	0.063	0.020	0.003
	80	20	0	0.002	0.954	0.034	0.010	0
	80	40	0	0	0.971	0.024	0.005	0
	40	10	0	0.580	0.367	0.037	0.015	0.001
	40	20	0	0.091	0.875	0.028	0.006	0
$c=1$	40	40	0	0	0.986	0.012	0.001	0.001
	80	10	0	0.342	0.590	0.055	0.012	0.001
	80	20	0	0.003	0.952	0.032	0.013	0
	80	40	0	0	0.974	0.022	0.004	0

Table 2: Frequency of number of groups determined: DGP 3

	N	T	$K < 5$	$K = 5$	$K = 6$	$K = 7$	$K = 8$	$K > 8$
	40	10	0.016	0.042	0.023	0.067	0.016	0.836
	40	20	0	0	0.004	0.013	0.012	0.971
$c = 0.5$	40	40	0	0	0	0	0	1
	80	10	0	0.001	0	0.005	0	0.994
	80	20	0	0	0	0	0	1
	80	40	0	0	0	0	0	1
	40	10	0.008	0.022	0.012	0.027	0.014	0.917
	40	20	0	0	0	0.004	0.004	0.992
$c = 1$	40	40	0	0	0	0	0	1
	80	10	0	0	0	0.002	0	0.998
	80	20	0	0	0	0	0	1
	80	40	0	0	0	0	0	1

Table 2: Frequency of number of groups determined: DGP 4

	N	T	$K < 5$	$K = 5$	$K = 6$	$K = 7$	$K = 8$	$K > 8$
	40	10	0	0.367	0.490	0.094	0.042	0.007
	40	20	0	0.030	0.761	0.140	0.056	0.013
$c = 0.5$	40	40	0	0	0.490	0.262	0.179	0.069
	80	10	0	0.107	0.634	0.152	0.097	0.010
	80	20	0	0	0.615	0.232	0.122	0.031
	80	40	0	0	0.139	0.301	0.314	0.246
	40	10	0	0.450	0.426	0.080	0.038	0.006
	40	20	0	0.041	0.756	0.124	0.071	0.008
$c = 1$	40	40	0	0	0.506	0.231	0.195	0.068
	80	10	0	0.162	0.616	0.136	0.081	0.005
	80	20	0	0	0.618	0.205	0.139	0.038
	80	40	0	0	0.145	0.251	0.356	0.248

4. Empirical Application

Relationship between income and democracy

- Empirical research: Lipset (1959), Barro (1999), Acemoglu, Johnson, Robinson, and Yared (2008, AJRY hereafter), and Bonhomme, and Manresa (2012, BM hereafter).
- None of the existing studies allows for heterogeneity in the slope coefficients. As discussed in AJRY, "*societies may embark on divergent political-economic development paths*". Thus, ignoring the heterogeneity in the slope coefficients may result in model misspecification and invalidates subsequent inferences.
- Data: BM
- Model:

$$y_{it} = \beta_{1i} X_{i,t-1} + \beta_{2i} y_{i,t-1} + \mu_i + u_{it},$$

  where

  y_{it} : a measure of democracy for country i in period t

  X_{it} be the logarithm of its real GDP per capita.
- 82 countries over the period 1961-2000, five year average, e.g.,$t = 1$ refers to years 1961-1965.

Testing result

Table 3: Test statistics

	$c = 0.5$			$c = 1$		
Null hypothesis: $K =$	1	2	3	1	2	3
Statistics	3.706	1.975	0.944	3.706	2.323	1.699
p-values	0.0001	0.024	0.173	0.0001	0.010	0.045

Table 4: Classification of countries

Group 1 ("negative effect" group) ($N_1 = 19$)				
Burkina Faso	Central Africa	Colombia	Guatemala	Iran
Kenya	Sri Lanka	Madagascar	Mauritania	Malaysia
Niger	Nicaragua	Sierra Leone	El Salvador	Syrian Arab
Chad	Togo	Turkey	South Africa	
Group 2 ("small effect" group) ($N_2 = 30$)				
Argentina	Austria	Burundi	China	Cote d'Ivoire
Cameroon	Congo Rep.	Costa Rica	Dominican	Egypt Arab
France	Gabon	U.K.	Ghana	Indonesia
Ireland	Italy	Japan	Luxembourg	Mexico
Nigeria	Rwanda	Singapore	Sweden	Thailand
Tunisia	Uganda	U.S.	Congo Dem.	Zambia
Group 3 ("large effect" group) ($N_3 = 33$)				
Benin	Bolivia	Brazil	Chile	Cyprus
Algeria	Ecuador	Spain	Finland	Guinea
Greece	Guyana	Honduras	India	Israel
Jamaica	Jordan	Korea Rep.	Morocco	Mali
Malawi	Nepal	Panama	Peru	Philippines
Portugal	Paraguay	Romania	Trinidad&Tobago	Taiwan
Tanzania	Uruguay	Venezuela		

Estimation result

Table 5: Post-Lasso estimation

		β_{1i}			β_{2i}			CIE
		estimates	s.e.	t-stat	estimates	s.e.	t-stat	estimates
common estimation		0.130	0.031	4.160	0.290	0.043	6.770	0.183
$c = 0.5$	Group 1	-0.427	0.058	-7.316	0.115	0.084	1.307	-0.482
	Group 2	0.078	0.035	2.138	-0.107	0.074	-1.351	0.070
	Group 3	0.341	0.051	5.976	0.380	0.063	5.983	0.550
$c = 1$	Group 1	-0.363	0.057	-4.740	0.139	0.089	1.405	-0.422
	Group 2	0.021	0.022	0.828	-0.240	0.074	-2.352	0.017
	Group 3	0.286	0.041	6.677	0.350	0.057	6.157	0.440

Note: CIE stands for cumulative income effect, which is defined as $(\beta_{1i}/(1-\beta_{2i}))$.

Explaining the group pattern

- Based on the estimates of β_{1i}, we refer to groups 1, 2 and 3 as the "negative effect", "small effect" and "large effect" groups.
- How do we explain the group membership? E.g., why are China and USA classified into the same "small effect" group?
- Cross-section multinomial logit model using the following covariates
  (i) initial education level in 1965,
  (ii) initial income level in 1965,
  (iii) initial democracy level in 1965,
  (iv) a measure of constraints on the executive at independence,
  (v) independence year/100,
  (vi) 500-year change in income per capita over 1500-2000,
  (vii) 500-year change in democracy over 1500-2000.
- (i), (ii) and (iii) are the initial key economic variables. Acemoglu, Johnson and Robinson (2005) suggest that (iv) is an important determinant of democracy. (v) measures how recent a country became independent. (vi) and (vii) present long-run changes in income and democracy levels.

Explaining the group pattern

Table 8: Summary statistics by groups

		Group 1 "negative effect"		Group2 "small effect"		Group 3 "large effect"	
variable	variable description	mean	s.d.	mean	s.d.	mean	s.d.
edu65	education level in 1965	1.678	1.160	3.967	2.713	3.232	1.634
inc65	logarithm of real GDP per capita in 1965	7.568	0.582	7.903	1.073	7.852	0.783
dem65	measure of democracy in 1965	0.542	0.233	0.625	0.290	0.585	0.278
constr	constraints on the executive at independence	0.353	0.343	0.295	0.338	0.335	0.367
indcent	year of independence/100	19.094	0.690	18.889	0.735	18.951	0.685
democ	500 year democracy change	0.616	0.274	0.661	0.303	0.826	0.211
growth	500 year income per capita change	1.288	0.931	2.157	1.237	2.091	1.014

Explaining the group pattern

Table 9: Determinants of the group pattern

	Group 1 ("negative effect" group)						
edu65	-0.566***	-0.847**	-1.013***	-1.294***	-1.491***	-1.249***	-0.990**
	(0.199)	(0.339)	(0.347)	(0.388)	(0.418)	(0.415)	(0.423)
inc65	-	0.861	0.869	1.363 *	1.249	1.144	1.817*
	-	(0.718)	(0.727)	(0.778)	(0.845)	(0.873)	(0.962)
dem65	-	-	1.740	0.781	0.549	0.068	-0.356
	-	-	(1.606)	(1.718)	(1.717)	(1.720)	(1.904)
constr	-	-	-	2.116	3.303**	4.282**	4.838***
	-	-	-	(1.313)	(1.487)	(1.704)	(1.716)
indcent	-	-	-	-	-0.906	-1.723*	-1.963**
	-	-	-	-	(0.694)	(0.978)	(0.971)
demco	-	-	-	-	-	-5.099*	-5.627**
	-	-	-	-	-	(2.766)	(2.778)
growth	-	-	-	-	-	-	-1.074
	-	-	-	-	-	-	(0.684)

Explaining the group pattern

Table 9: Determinants of the group pattern

	Group 2 ("small effect" group)						
edu65	0.167	0.260	0.249	0.299	0.301	0.795**	0.780**
	(0.141)	(0.225)	(0.260)	(0.259)	(0.262)	(0.334)	(0.364)
inc65	-	-0.295	-0.298	-0.427	-0.380	-0.714	-0.906
	-	(0.606)	(0.601)	(0.618)	(0.674)	(0.730)	(0.788)
dem65	-	-	0.142	0.535	0.495	-0.157	0.145
	-	-	(1.401)	(1.580)	(1.583)	(1.760)	(1.784)
constr	-	-	-	-0.862	-0.939	-0.340	-0.245
	-	-	-	(0.839)	(1.099)	(1.211)	(1.216)
indcent	-	-	-	-	0.084	-0.983	-1.123
	-	-	-	-	(0.653)	(0.753)	(0.788)
democ	-	-	-	-	-	-7.445***	-7.529***
	-	-	-	-	-	(2.572)	(2.532)
growth	-	-	-	-	-	-	0.105
	-	-	-	-	-	-	(0.540)

Conclusions

- Propose a testing procedure to determine the number of groups in panel structure models
- Derive the asymptotic distribution of the test statistic and prove its consistency
- Apply the method to study the relationship between income and democracy

Thanks!

Supplement 1: Penalized Least Squares Estimation

Numerical Algorithm

1. Start with $\hat{\alpha}^{(0)} = (\hat{\alpha}_1^{(0)}, ..., \hat{\alpha}_{K_0}^{(0)})$ and $\hat{\beta}^{(0)} = (\hat{\beta}_1^{(0)}, ..., \hat{\beta}_N^{(0)})$ such that $\sum_{i=1}^N ||\hat{\beta}_i^{(0)} - \hat{\alpha}_k^{(0)}|| \neq 0$ for each $k = 2, ..., K_0$.
2. Given $\hat{\alpha}^{(r-1)} \equiv (\hat{\alpha}_1^{(r-1)}, ..., \hat{\alpha}_{K_0}^{(r-1)})$ and $\hat{\beta}^{(r-1)} \equiv (\hat{\beta}_1^{(r-1)}, ..., \hat{\beta}_N^{(r-1)})$,
   - 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 $(\boldsymbol{\beta}, \alpha_2)$.

Supplement 1: Penalized Least Squares Estimation

Numerical Algorithm

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

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

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

3 Repeat step 2 until a convergence criterion is met.