Grouped Patterns of Heterogeneity in Panel Data

(Bonhomme and Manresa, 2015)

Sugarkhuu Radnaa

University of Bonn

November 22, 2022

Contents

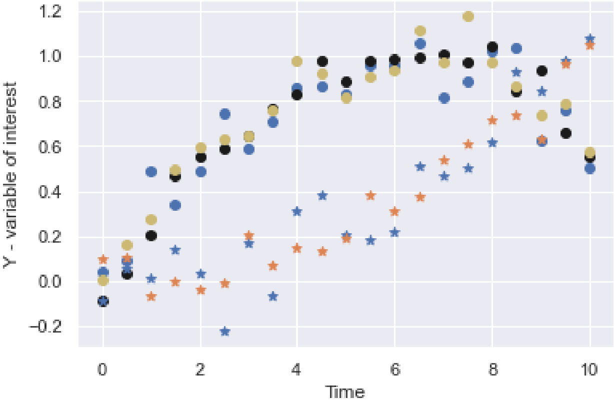
1. Motivation
2. Specification
3. Computation and an estimator
4. Asymptotic properties of the estimator
5. Application: INCOME AND (WAVES OF) DEMOCRACY

Contents

1. Motivation
2. Specification
3. Computation and an estimator
4. Asymptotic properties of the estimator
5. Application: INCOME AND (WAVES OF) DEMOCRACY

Motivation

The panel units may tend to cluster around different trends.



Common problems with usual fixed effects

- 'Incidental parameter' bias may be substantial in short panels.
- Unobserved heterogeneity is constant over time.

Contents

1. Motivation
2. Specification
3. Computation and an estimator
4. Asymptotic properties of the estimator
5. Application: INCOME AND (WAVES OF) DEMOCRACY

Specification: Flexible way to account for unobserved heterogeneity

Group-specific, time-variant effect:

$$y_{it} = x_{it}'\theta + \alpha_{g_i t} + v_{it}, i = 1, \ldots, N, t = 1, \ldots, T$$

- y_{it} - outcome
- x_{it} - covariates
- $\alpha_{g_i t}$ - group-specific time effect, the group membership variables g_i, for all $i \in 1, \ldots, N$
- v_{it} - residual

Assumptions:

- Leaves the correlation of GFE with explanatory variables unrestricted
- No need of parallel trends assumption

Possible extensions

Unit-specific time-invariant heterogeneity:

$$y_{it} = x_{it}'\theta + \alpha_{g_i t} + \eta_i + v_{it}$$

Heterogeneous group-specific effects of covariates:

$$y_{it} = x_{it}'\theta_{g_i} + \alpha_{g_i t} + v_{it}$$

Contents

1. Motivation
2. Specification
3. Computation and an estimator
4. Asymptotic properties of the estimator
5. Application: INCOME AND (WAVES OF) DEMOCRACY

Computation

Objective

$$(\hat{\theta}, \hat{\alpha}, \hat{\gamma}) = \underset{(\theta,\alpha,\gamma)\in\Theta\times\mathscr{A}^{GT}\times\Gamma_G}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\theta - \alpha_{g_i t})^2$$

where $\gamma \in \Gamma_G$ denotes a particular grouping (i.e., partition) of the N units, where Γ_G is the set of all groupings of $\{1, \ldots, N\}$ to at most G groups.

Computation hurdle: Partition of N elements to G groups (explosive) - Stirlings number of the second kind.

Computation

Algorithm 1

1. Let $(\theta^{(s)}, \alpha^{(s)}) \in \Theta \times \mathscr{A}^{GT}$ be some starting value. Set $s = 0$.
2. Compute for all $i \in \{1, \ldots, N\}$:

$$g_i^{(s+1)} = \underset{g \in \{1,\ldots,G\}}{\operatorname{argmin}} \sum_{t=1}^{T} (y_{it} - x_{it}' \theta^{(s)} - \alpha_{g_i t}^{(s)})^2$$

Here, for large problems, one can apply k-means clustering algorithm on residuals after covariates

Computation:

Algorithm 1 - continued

3. Compute:

$$(\theta^{(s+1)}, \alpha^{(s+1)}) = \underset{(\theta,\alpha)\in\Theta\times\mathcal{A}^{GT}}{\operatorname{argmin}} \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - x_{it}'\theta - \alpha_{g_i^{(s+1)}t})^2$$

   Here, one does an OLS regression that controls for the interactions of group indicators and time dummies.

4. Set $s = s + 1$ and go to Step 2 (until numerical convergence)

Contents

1. Motivation
2. Specification
3. Computation and an estimator
4. Asymptotic properties of the estimator
5. Application: INCOME AND (WAVES OF) DEMOCRACY

Asymptotic properties: Consistency

Assumption 1

There exists a constant $M > 0$ such that:

(a) Θ and $\mathscr{A}$ are compact subsets of R^k and R, respectively.

(b) $E(\|x_{it}\|^2) \leq M$, where $\|\cdot\|$ denotes the Euclidean norm.

(c) $E(v_{it}) = 0$, and $E(v_{it}^4) \leq M$.

(d) $|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E(v_{it}v_{is}x_{it}'x_{is})| \leq M$.

(e) $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\sum_{t=1}^{T} E(v_{it}v_{jt})| \leq M$.

(f) $|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(v_{it}v_{jt}, v_{is}v_{js})| \leq M$.

Asymptotic properties: Consistency

Assumption 1 - continued

(g) Let $\bar{x}_{g \wedge \tilde{g},t}$ denote the mean of x_{it} in the intersection of groups $g_i^0 = g$, and $g_i = \tilde{g}$. For all groupings $\gamma = \{g_1, \ldots, g_N\} \in \Gamma_G$, define $\hat{\rho}(\gamma)$ as the minimum eigenvalue of the following matrix [a]:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{g_i^0 \wedge g_i,t})(x_{it} - \bar{x}_{g_i^0 \wedge g_i,t})'$$

Then $plim_{N,T\to\infty} min_{\gamma\in\Gamma_G}\hat{\rho}(\gamma) = \rho > 0$.

[a] $\tilde{x}$ refers to be the infeasible version of the GFE estimator where group membership g_i, instead of being estimated, is fixed to its population counterpart g_i^0, and then estimated with OLS.

Asymptotic properties: Consistency

Theorem 1

Let Assumption 1 hold. Then, as N and T tend to infinity:

$$\hat{\theta} \xrightarrow{p} \theta^0 \quad \text{and}$$

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{\alpha}_{\hat{g}_i t} - \alpha^0_{g_i^0 t})^2 \xrightarrow{p} 0.$$

where the ^{0} superscripts refer to true parameter values.

Asymptotic properties: Distribution

Assumption 2

(a) For all $g \in \{1, \dots, G\}$: $plim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1\{g_i^0 = g\} = \pi_g > 0$.

(b) For all $(g, \tilde{g}) \in \{1, \dots, G\}^2$ such that $g \neq \tilde{g}$: $plim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} (\alpha_{gt}^0 - \alpha_{\tilde{g}t}^0)^2 = c_{g,\tilde{g}} > 0$.

(c) There exist constants $a > 0$ and $d_1 > 0$ and a sequence $\alpha[t] \leq e^{-at^{d_1}}$ such that, for all $i \in \{1, \dots, N\}$ and $(g, \tilde{g}) \in \{1, \dots, G\}^2$ such that $g \neq \tilde{g}, \{v_{it}\}_t, \{\alpha_{gt}^0 - \alpha_{\tilde{g}t}^0\}_t$, and $(\alpha_{gt}^0 - \alpha_{\tilde{g}t}^0)v_{it\,t}$ are strongly mixing processes with mixing coefficients $\alpha[t]$. Moreover, $E((\alpha_{gt}^0 - \alpha_{\tilde{g}t}^0)v_{it}) = 0$.

Asymptotic properties: Distribution

Assumption 2 - continued

(d) There exist constants $b > 0$ and $d_2 > 0$ such that $Pr(|v_{it}| > m) \leq e^{1-(\frac{m}{b})^{d_2}}$ for all i, t, and $m > 0$.

(e) There exists a constant $M^* > 0$ such that, as N, T tend to infinity:

$$\underset{i \in \{1,...,N\}}{sup} Pr\Big(\frac{1}{T}\sum_{t=1}^{T} \|x_{it}\| \geq M^*\Big) = o(T^{-\Sigma})$$

for all $\delta > 0$.

Asymptotic properties: Distribution

Theorem 2

Asymptotic Equivalence: Let Assumptions 1 and 2 hold. Then, for all $\Sigma > 0$ and as N and T tend to infinity:

(16) $Pr(\underset{i \in 1,\ldots,N}{|\hat{g}_i - g_i^0|} > 0) = o(1) + o(NT^{-\Sigma})$

(17) $\hat{\theta} = \tilde{\theta} + o_p(T^{-\Sigma})$

(18) $\hat{\alpha}_{gt} = \tilde{\alpha}_{gt} + o_p(T^{-\Sigma})$ for all g, t.

Asymptotic properties: Distribution

Assumption 3

(a) For all i, j, and t: $E(x_{jt} v_{it}) = 0$

(b) There exist positive definite matrices Σ_θ and Ω_θ such that

$$\Sigma_\theta = \underset{N,T\to\infty}{plim} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_{g_i^0 t})(x_{it} - \bar{x}_{g_i^0 t})',$$

$$\Omega_\theta = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E[v_{it} v_{js} (x_{it} - \bar{x}_{g_i^0 t})(x_{js} - \bar{x}_{g_j^0 s})']$$

Asymptotic properties: Distribution

Assumption 3 - continued

(c) As N and T tend to infinity:
$\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_{g_i^0 t}) v_{it} \xrightarrow{d} N(0, \Omega_\theta)$.

(d) For all (g, t):
$lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} E(1\{g_i^0 = g\} 1\{g_j^0 = g\} v_{it} v_{jt}) = \omega_{gt} > 0$

(e) For all (g, t), and as N and T tend to infinity:
$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} 1\{g_i^0 = g\} v_{it} \xrightarrow{d} N(0, \omega_{gt})$

Asymptotic properties: Distribution

Corollary 1 (Theorem 2): Asymptotic Distribution

Let Assumptions 1, 2, and 3 hold, and let N and T tend to infinity such that, for some $v > 0, N/T^v \longrightarrow 0$. Then, we have

(19) $\sqrt{NT}(\hat{\theta} - \theta^0) \xrightarrow{d} N(0, \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1})$, and, for all (g, t),

(20) $\sqrt{N}(\hat{\alpha}_{gt} - \alpha_{gt}^0) \xrightarrow{d} N(0, \frac{\omega_{gt}}{\pi_g^2})$,

where π_g is defined in Assumption 2, and where $\Sigma_\theta, \Omega_\theta$, and ω_{gt} are defined in Assumption 3.

Contents

1. Motivation
2. Specification
3. Computation and an estimator
4. Asymptotic properties of the estimator
5. Application: INCOME AND (WAVES OF) DEMOCRACY

Does income affect democracy?

Acemoglu et al. (2008) found that the positive association between income and democracy disappears when controlling for additive country and time-effects.

$$democracy_{it} = \theta_1 democracy_{it-1} + \theta_2 logGDPpc_{it-1} + \alpha_i + v_{it}$$

The authors interpreted the country fixed-effects as reflecting long-run, historical factors that have shaped the political and economic development of countries.

Does income affect democracy?

Estimate an alternative grouped fixed effects for countries:

$$democracy_{it} = \theta_1 democracy_{it-1} + \theta_2 logGDPpc_{it-1} + \alpha_{g_i t} + v_{it}$$

The cumulative income effect on democracy and the number of groups

After controlling for group-specific, time-variant effects, the impact of the cumulative income becomes trivial.

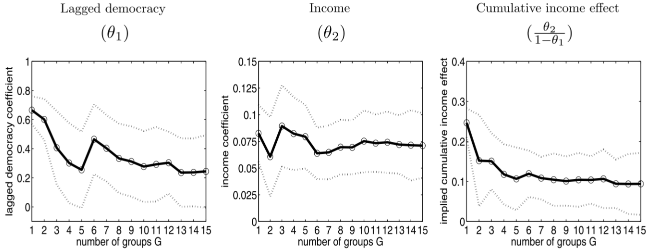


FIGURE 1.—Coefficients of income and lagged democracy. *Note*: Balanced panel from Acemoglu et al. (2008). The x-axis shows the number of groups G used in estimation, the y-axis reports parameter values. 95%-confidence intervals clustered at the country level are shown in dashed lines. Confidence intervals are based on bootstrapped standard errors (100 replications).

Group-specific time-effects explain the majority of differences in democratic changes

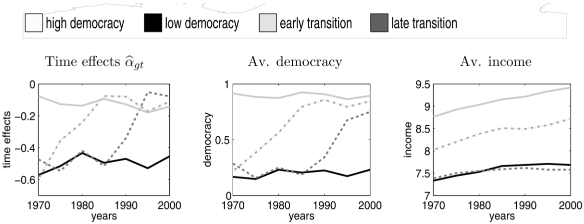


FIGURE 2.—Patterns of heterogeneity, $G = 4$. *Note*: See the notes to Figure 1. On the bottom panel, the left graph shows the group-specific time effects $\widehat{\alpha}_{gt}$. The other two graphs show the group-specific averages of democracy and lagged log-GDP per capita, respectively. Calendar years (1970–2000) are shown on the x-axis. Light solid lines correspond to Group 1 ("high-democracy"), dark solid lines to Group 2 ("low-democracy"), light dashed lines to Group 3 ("early transition"), and dark dashed lines to Group 4 ("late transition"). The top panel shows group membership. The list of countries by group is given in the Supplemental Material.

Simulation

Conclusion and discussion

- Offers a flexible yet parsimonious approach to model unobserved heterogeneity.
- Useful in applications where time-varying grouped effects may be present in the data.
- GFE should be well-suited in difference-in-difference designs, as a way to relax parallel trends assumptions.
- The extension to nonlinear models is a natural next step, however, statistical and computational challenges increase.

Thank you!

Bibliography

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American economic review*, 98(3):808–42.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.