Unobserved Clusters of Time-Varying Heterogeneity in Nonlinear Panel Data Models

Martin Mugnier[1]

[1]Center for Research in Economics and Statistics, ENSAE, IP Paris

NASMES 2022, Miami

June 18, 2022

Motivation (1/2)

- Observational **panel data** offer opportunities to control for **unobserved heterogeneity** (UH)
- Workhorse two-way fixed effects (FE) regression model:

$$Y_{it} = X_{it}'\beta + \alpha_i + \xi_t + \varepsilon_{it}, \quad i = 1, ..., N, t = 1, ..., T.$$

  - α_i: time-invariant individual-specific effect; ξ_t: time trend.

1. Parallel trend ($\alpha_i + \xi_t$) may be overly restrictive, especially in large-T panels:
   - Innovation and competition? $\neq$ trajectories of technological change and market structures.
   - Mental health and abortion? $\neq$ trajectories of unobserved risky behaviors.
2. In many economic settings, a linear model is poorly suited
   - Count data (e.g., number of patents), discrete choice (e.g., developing a mental illness)

- Parcimonious and flexible approach to capture diverging time-varying patterns of UH in nonlinear panel models?
  - Focus on "true" panel but many insights extend to directed pseudo-panels with nonlinear interactions (e.g., gravity).

Motivation (2/2)

- Allowing for **time-varying UH** in **nonlinear FE models** is **challenging**.
  - Interactive fixed effects $\alpha_i'\xi_t$, $\alpha_i, \xi_t \in \mathbb{R}^r$, $r \in \mathbb{N}^*$ (Bai, 2009).
  - Large number of fixed effects $\Longrightarrow$ **incidental parameters problem** (small and **large-T**).
  - Semiparametric FE estimators have non-centered asymp. distributions, inference generally requires $N \approx T$, interpretation of FE is difficult (Bonhomme, Lamadon, and Manresa, 2021; Chen, Fernández-Val, and Weidner, 2021; Fernández-Val and Weidner, 2016).
  - Interesting exception is **discrete UH** (Bonhomme and Manresa, 2015; Hahn and Moon, 2010).

Motivation (2/2)

- Allowing for **time-varying UH** in **nonlinear FE models** is **challenging**.
  - Interactive fixed effects $\alpha_i'\xi_t$, $\alpha_i, \xi_t \in \mathbb{R}^r$, $r \in \mathbb{N}^*$ (Bai, 2009).
  - Large number of fixed effects $\implies$ **incidental parameters problem** (small and **large-T**).
  - Semiparametric FE estimators have non-centered asymp. distributions, inference generally requires $N \approx T$, interpretation of FE is difficult (Bonhomme, Lamadon, and Manresa, 2021; Chen, Fernández-Val, and Weidner, 2021; Fernández-Val and Weidner, 2016).
  - Interesting exception is **discrete UH** (Bonhomme and Manresa, 2015; Hahn and Moon, 2010).
- Often plausible that **UH only takes a restricted number of paths across time**.
  - Health types (Deb and Trivedi, 1997; Janys and Siflinger, 2021), innovation clusters
  - **Unobserved clusters = individuals with the same unobserved paths of time-varying UH**.
  - Interest: data-driven clustering + cluster-specific trends + structural parameters.
- For a large class of **discrete outcome models**, popular in empirical research:

1. No clear large-T **nonparametric identification** result.
2. Lack of suitable **estimators & inference** (e.g., allowing T to grow slowly with N).
3. Few empirical evidence on the consequences of neglecting time-varying UH.

This Paper

- **Panel data:** random sample $\{(Y_{it}, X_{it}')'_{1\leq t\leq T} : 1 \leq i \leq N\}$.
- Static **nonlinear grouped fixed effects (NGFE) models** with **single index:**
  - Individual i at time t chooses $Y_{it} \in \mathcal{Y}$ with probability

$$\Pr\left(Y_{it} = y \mid X_i^t, g_i^0, \alpha^0_{g_i^0 t}\right) = h^0\left(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}\right),$$

  - $X_{it} \in \mathcal{X}_i \subset \mathbb{R}^p$: exogeneous/pre-determined explanatory variables, $X_i^t = (X_{i1}', ..., X_{it}')'$;
  - $\beta^0 \in \mathbb{R}^p$: **unknown** common parameter;
  - $g_i^0 \in \{1, ..., G^0\}$: **unobserved** cluster/group membership variable, $\gamma^0 = (g_1^0, ..., g_N^0)'$;
  - $\alpha^0_{gt} \in \mathbb{R}$: **unobserved** cluster-specific time-effect, $\alpha^0 = \{\alpha^0_{gt} : (g, t)\}$;
  - $h^0 \in \left\{h : \mathcal{Y} \times \mathbb{R} \to (0, 1), \sum_{y\in\mathcal{Y}} h(y, \cdot) = 1, \sum_{y\in\mathcal{Y}} |y|\, h(y, \cdot) < +\infty\right\}$: **unknown** link function;
  - FE approach: $X_i^t \mid \gamma^0, \alpha^0$ is **unrestricted**.
- Nest popular models in empirical research (e.g., binary, ordered, count outcome). Examples

This Paper

- **Panel data:** random sample $\{(Y_{it}, X_{it}')'_{1\leq t\leq T} : 1 \leq i \leq N\}$.
- Static **nonlinear grouped fixed effects (NGFE) models** with **single index:**
  - Individual i at time t chooses $Y_{it} \in \mathcal{Y}$ with probability
    $$\Pr\left(Y_{it} = y \mid X_i^t, g_i^0, \alpha^0_{g_i^0 t}\right) = h^0\left(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}\right),$$
  - $X_{it} \in \mathcal{X}_i \subset \mathbb{R}^p$: exogeneous/pre-determined explanatory variables, $X_i^t = (X_{i1}', ..., X_{it}')'$;
  - $\beta^0 \in \mathbb{R}^p$: **unknown** common parameter;
  - $g_i^0 \in \{1, ..., G^0\}$: **unobserved** cluster/group membership variable, $\gamma^0 = (g_1^0, ..., g_N^0)'$;
  - $\alpha^0_{gt} \in \mathbb{R}$: **unobserved** cluster-specific time-effect, $\alpha^0 = \{\alpha^0_{gt} : (g, t)\}$;
  - $h^0 \in \left\{h : \mathcal{Y} \times \mathbb{R} \to (0, 1), \sum_{y\in\mathcal{Y}} h(y, \cdot) = 1, \sum_{y\in\mathcal{Y}} |y|\, h(y, \cdot) < +\infty\right\}$: **unknown** link function;
  - FE approach: $X_i^t \mid \gamma^0, \alpha^0$ is **unrestricted**.
- Nest popular models in empirical research (e.g., binary, ordered, count outcome). Examples
- **Object of interest:** $\theta^0_{NT} := (\beta^0, h^0, G^0, \gamma^0, \alpha^0)$.
- **Research question(s):** identification and (semiparametric) estimation as $N, T \to \infty$? How much allowing for time-varying UH can lead to $\neq$ conclusions in practice? Can we learn meaningful clusters?

Main Results

- **Large-T nonparametric identification:**
  - Provide sufficient conditions for **point identification** of θ^0_{NT} as $N, T \to \infty$.
    $\implies$ All marginal effects are identified.

- **Semiparametric estimation and inference:**
  - Propose **"classification likelihood" estimators** $(\widehat{\beta}, \widehat{\gamma}, \widehat{\alpha})$, assuming (h^0, G^0) known.
  - In some strictly concave models (e.g., Probit, Logit, Poisson), under regularity conditions:
    - $\widehat{\beta}$, $\widehat{\alpha}_{gt}$ are **consistent** and root-(NT) (resp. N) **asymptotically normal (centered at 0)**.
    - $\widehat{\gamma}$ is **uniformly consistent**: $\max_{i \in \{1,\ldots,N\}} \left| \widehat{g}_i - g_i^0 \right| = o_p(1)$.

- **Monte Carlo simulations:**
  - Good finite sample properties of large-T approximations.

- **Empirical application**:
  - Revisit **inverted-U relationship** between **innovation and competition** (Aghion, Bloom, Blundell, Griffith, and Howitt, 2005, ABBGH hereafter).
  - Find evidence of **time-varying unobserved heterogeneity**, a **mildly inverted-U**, **data-driven clustering** of industries.

Literature and Contributions

- **Nonseparable panel data models with (time-varying) unobserved fixed effects**
  Altonji and Matzkin (2005); Botosaru, Muris, and Pendakur (2021); Chernozhukov, Fernández-Val, Hahn, and Newey (2013); Evdokimov (2010, 2011); Freyberger (2018); Hoderlein and White (2012); Honore and Lewbel (2002); Mugnier and Wang (2021); Zeleneev (2020)

  $\hookrightarrow$ **Contribution:** point identification of all parameters in a large-T setting, with limited time-homogeneity conditions; clustering structure; discrete outcome.

- **Estimation of nonlinear (interactive) fixed effects with time-varying UH**
  Ando and Bai (2022); Bonhomme, Lamadon, and Manresa (2021); Chen, Fernández-Val, and Weidner (2021); Moon and Weidner (2019)

  $\hookrightarrow$ **Contribution:** new semiparametric estimator, retain Bonhomme and Manresa (2015)'s GFE estimator nice asymptotic properties when $T/N \to 0$ (no asymptotic bias).

- **Sparsity/finite mixtures as dimension reduction devices to the incid. param. pb**
  Bester and Hansen (2016); Bonhomme and Manresa (2015); Cheng, Schorfheide, and Shao (2021); Gu and Volgushev (2019); Hahn and Moon (2010); Kock (2016); Moon and Weidner (2019); Saggio (2012); Su, Shi, and Phillips (2016); Su, Wang, and Jin (2019); Vogt and Linton (2017); Wang and Su (2021)

  $\hookrightarrow$ **Contribution:** allow for time-varying UH; nonlinear & nonparametric setting; no tuning-parameter.

Outline

Outline

Large-N, Large-T Nonparametric Identification

Semiparametric Estimation and Inference

Monte Carlo Simulations

Empirical Application: Revisiting the Inverted-U Relationship Between Innovation and Competition

Conclusion

Identification: Some Intuition

- Target is $(G^0, \gamma^0, \beta^0, h^0, \alpha^0) \approx$ high-dimensional nonlinear optimal partitioning problem.
- Split into two sequential problems:

1. Classification: identify the adjency matrix A_N^0 of the unobserved but "easy-to-optimal-cut" graph spanned by group membership variables γ^0:

$$A_N^0 = (\mathbb{1}\{g_i^0 = g_j^0\})_{(i,j)\in\{1,\dots,N\}^2}.$$

  - Conditional time ergodicity/mixing: identify **individual-level conditional probabilities**.
  - **Cluster-separation condition** + overlapping covariates: make $O(N^2)$ **pairwise comparisons** (similar to pairwise differencing in Mugnier, 2022).
    - Can be achieved with special regressor (not necessarily large support) + "completeness".

Identification: Some Intuition

- Target is $(G^0, \gamma^0, \beta^0, h^0, \alpha^0) \approx$ high-dimensional nonlinear optimal partitioning problem.
- Split into two sequential problems:

1. Classification: identify the adjency matrix A_N^0 of the unobserved but "easy-to-optimal-cut" graph spanned by group membership variables γ^0:

$$A_N^0 = (\mathbb{1}\{g_i^0 = g_j^0\})_{(i,j)\in\{1,\ldots,N\}^2}.$$

   - Conditional time ergodicity/mixing: identify **individual-level conditional probabilities**.
   - **Cluster-separation condition** + overlapping covariates: make $O(N^2)$ **pairwise comparisons** (similar to pairwise differencing in Mugnier, 2022).
     - Can be achieved with special regressor (not necessarily large support) + "completeness".

2. Identification of other structural parameters: solve at the identified **cluster**$\times$**time level**.
   - β^0: within cluster i.i.d. sampling + results for single index models (Ichimura, 1993).
   - (h^0, α^0): monotonicity + compensating variations (D'Haultfoeuille, Hoderlein, and Sasaki, 2021; Mugnier and Wang, 2021).

Identification: Assumptions

- NGFE model: $\mathcal{Y}$ is at most countable and

$$\Pr\left(Y_{it} = y \mid X_i^t, g_i^0, \alpha^0_{g_i^0 t}\right) = h^0\left(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}\right), \quad (i,t,y) \in [N] \times [T] \times \mathcal{Y}.$$

- Normalization: $\left\|\beta^0\right\| = 1, \alpha^0_{11} = 0$.

Assumption 1 (Random sampling)

There exist sequences of random vectors of fixed dimensions $\lambda^0 := \{\lambda^0_{gt} : (g,t)\}$, $\mu^0 := \{\mu^0_g : g\}$, $\xi^0 := \{\xi^0_i : i\}$, such that:

(a) $(Y_i', X_i', g_i^0)'$ *is i.i.d. across* i *conditional on* α^0, λ^0, *and* μ^0.

(b) *For all* i: $\{(Y_{it}, X_{it}', \alpha^0_{g_i^0 t}) : t \geq 2\}$ *is a strictly stationary strong mixing process with mixing coefficient* $\tau_i(\cdot)$ *conditional on* $g_i^0, \mu^0_{g_i^0}, \xi^0_i$. *Let* $\tau(\cdot) = \sup_i \tau_i(\cdot)$ *satisfy* $\tau(s) \leq c_\tau m^s$ *for some* $c_\tau > 0$ *and* $m \in (0,1)$.

(c) *For all* t: $Y_{1t} \mid X_1, g_1, \alpha^0, \lambda^0, \mu^0, \xi^0 \overset{d}{=} Y_{1t} \mid X_{1t}, g_1^0, \alpha^0_{g_1^0 t}$.

Identification: Assumptions

Assumption 2 (Latent clustering)

$\mathcal{X} := \bigcap_{i=1}^{\infty} \mathcal{X}_i$ *is not empty and:*

(a) *There exist known* $\mathcal{X}^0 \subset \mathcal{X}$, $y \in \mathcal{Y}$, *and functional* ϕ *such that, for all fixed* $(i,j) \in \mathcal{N}^2$, *letting* $\rho_i(x) : \mathcal{X}^0 \ni x \mapsto \Pr\left(Y_{i2} = y \mid X_{i2} = x, g_i^0, \mu_i^0, \xi_i^0\right)$, $\phi(\rho_i, \rho_j) = \mathbb{1}\{g_i^0 = g_j^0\}$.

(b) *For all* g*: almost surely* $\Pr(g_1^0 = g \mid \alpha^0, \lambda^0, \mu^0, \xi^0) > 0$.

Assumption 3 (Regularity and smoothness)

(a) *Conditional on* $g_i^0, \mu_{g_i^0}^0, \xi_i^0$, X_{i2} *admits a uniformly continuous density function* $f_{X_{i2}|g_i^0, \mu_{g_i^0}^0, \xi_i^0}$ *such that* $\inf_{x \in \mathcal{X}^0} f_{X_{i2}|g_i^0, \mu_{g_i^0}^0, \xi_i^0}(x) \geq \delta > 0$ *and* $\sup_{x \in \mathcal{X}^0} f_{X_{i2}|g_i^0, \mu_{g_i^0}^0, \xi_i^0}(x) < \infty$.

(b) *Almost surely,* $\mathbb{E}(\|X_{12}\|^2 \mid g_1^0, \alpha^0, \lambda^0, \mu^0)$ *is finite and* $\mathbb{E}(X_{12}X_{12}' \mid g_1^0, \alpha^0, \lambda^0, \mu^0)$ *is nonsingular.*

(c) $\sum_{y \in \mathcal{Y}} y h^0(y, \cdot)$ *is differentiable on* $\mathbb{R}$ *and not constant on the support of* $X_{it}'\beta^0 + \alpha_{g_i^0 t}^0$.

Identification: Assumptions

Assumption 4 (Monotonicity)

There exists $y \in \mathcal{Y}$ such that $h^0(y, v)$ is strictly monotonic in v.

Assumption 5 (Compensating variations)

For all fixed $(g, \widetilde{g}, t)$, all $x_1 \in \mathcal{X}$, there exists $x_2 \in \mathcal{X}$ such that

$$\alpha^0_{\widetilde{g}t} + x_1'\beta^0 = \alpha^0_{gt} + x_2'\beta^0.$$

Similarly, for all $(g, t, \widetilde{t})$, all $x_3 \in \mathcal{X}$, there exists $x_4 \in \mathcal{X}$ such that

$$\alpha^0_{g\widetilde{t}} + x_3'\beta^0 = \alpha^0_{gt} + x_4'\beta^0.$$

Identification: Main Result

Theorem (Identification)

Let Assumptions 1, 2 and 3(a) hold, and let N and T diverge jointly to infinity.

1. $\{A_N^0 : N \in \mathbb{N}^*\}$ *and* G^0 *are identified.*
2. *If Assumptions 3(b)-5 further hold, then*
   - β^0 *is identified.*
   - *For all* $(g, t) \in \left\{1, ..., G^0\right\} \times \mathbb{N}^*$, α_{gt}^0 *is identified up to cluster relabeling.*
   - h^0 *is identified.*

▸ Sketch of Proof

Outline

Estimation: Semiparametric NGFE Estimators

- Nonparametric estimation based on constructive identification?
  - Estimate $\widehat{\rho_i}$ with ML, prove $\left\|\widehat{A}_N - A_n^0\right\|_\infty \xrightarrow{p} 0$...
  - Slow convergence rates might be deterrent for panel data applications.

Estimation: Semiparametric NGFE Estimators

- Nonparametric estimation based on constructive identification?
  - Estimate $\widehat{\rho}_i$ with ML, prove $\left\|\widehat{A}_N - A_n^0\right\|_\infty \xrightarrow{p} 0$...
  - Slow convergence rates might be deterrent for panel data applications.

- Instead, propose a **practically useful** approach: assume **known** (h^0, G^0) (e.g., Probit).

- **Semiparametric Classification Maximum Likelihood Estimator**:

$$(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma}) = \underset{(\beta,\alpha,\gamma)\in\mathcal{B}\times\mathcal{A}^{G^0 T}\times\Gamma_{G^0}}{\operatorname{argmin}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} -\log h^0(Y_{it}, X_{it}'\beta + \alpha_{g_i t}),$$

where Γ_{G^0} = all possible groupings $\gamma = (g_1, ..., g_N)'$ of the N individuals into G^0 groups.

- Extend Bonhomme and Manresa (2015); Bryant and Williamson (1978).
- Non-smooth non-convex discrete optimization problem, but computation for small values of G^0 is feasible (up to local minima). Alternative minimization algorithm
- Alleviating computational burden is work in progress.
- Choice of G^0: AIC/BIC or report results for multiple choices.

Consistency and Large-T Inference: Binary Outcome

- **Strong concavity** of the log-likelihood function with respect to (β, α) is key.
- Semiparametric NGFE model with binary outcome ($|\mathcal{Y}| = 2$):

$$Y_{it} = \mathbb{1}\{X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} \geq 0\},\ i = 1, ..., N,\ t = 1, ..., T. \tag{1}$$

- For any $\mathbf{Z} = (Z_{11}, ..., Z_{1T}, ..., Z_{N1}, ..., Z_{NT})'$, denote $\mathbf{Z}_-^{(t)} = \{Z_{is} : 1 \leq i \leq N, 1 \leq s \leq t\}$, $\mathbf{Z}_+^{(t)} = \{Z_{is} : 1 \leq i \leq N, t \leq s \leq T\}$. $\beta^0 \in \mathcal{B} \subset \mathbb{R}^p$, $\alpha^0_{gt} \in \mathcal{A} \subset \mathbb{R}$.

Assumption (Mod.)

Eq. (1) holds and:

(a) (Weak exogeneity) $(\mathbf{X}_-^{(t)}, \gamma^0, \alpha^0, \varepsilon_-^{(t-1)})$ and $\varepsilon_+^{(t)}$ are independent.

(b). (Parametric noise) The $\{\varepsilon_{it} : (i, t)\}$ are identically distributed with known cumulative distribution function Ψ that is fully supported on $\mathbb{R}$, three times continuously differentiable, strictly increasing, and such that $(\log \Psi)'' < 0$ and Ψ' is symmetric around 0.

Consistency of NGFE Estimators

Assumption (Cons.)

(a) (Compactness) $\mathcal{B}$ and $\mathcal{A}$ are compact convex subsets.

(b) (Bounded covariates) There exists a constant $M > 0$ such that $\left\| X_{it} \right\| \leq M$ almost surely, where $\left\| \cdot \right\|$ denotes the Euclidean norm.

(c) (Noncollinearity) Let $\overline{X}_{g \wedge \widetilde{g}, t}$ denote the mean of X_{it} in the intersection of groups $g_i^0 = g$, and $g_i = \widetilde{g}$. For all groupings $\gamma = \{g_1, ..., g_N\} \in \Gamma_{G^0}$, I define $\widehat{\rho}(\gamma)$ as the minimum eigenvalue of the following matrix:

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \overline{X}_{g_i^0 \wedge g_i, t})(X_{it} - \overline{X}_{g_i^0 \wedge g_i, t})'.$$

Then, $\operatorname{plim}_{N,T \to \infty} \min_{\gamma \in \Gamma_{G^0}} \widehat{\rho}(\gamma) = \rho > 0$.

Consistency of NGFE Estimators

Theorem (Consistency)

Let Assumptions Mod. and Cons. hold. Then, as N and T tend to infinity:

- $\widehat{\beta} \xrightarrow{p} \beta^0$, and
- $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{\alpha}_{\widehat{g_i}t} - \alpha^0_{g_i^0 t}\right)^2 \xrightarrow{p} 0.$

Inference: Asymptotic Normality

- Show asymptotic equivalence to the oracle MLE with known groups. ▸ Details
  - Require well-separation of groups.

- Asymptotic distributions in nonlinear settings $\rightarrow$ typically non-centered.
- If there exists $\nu > 0$ such that $N/T^{\nu} \rightarrow 0$ and $T/N \rightarrow 0$ as $N, T \rightarrow \infty$, then:
  - Static case: under cross-sectional and time independence of Y_{it} given $\mathbf{X}, \gamma^0, \alpha^0$ + regularity conditions, we have

$$\sqrt{NT}(\widehat{\beta} - \beta^0) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$
$$\sqrt{N}(\widehat{\alpha}_{gt} - \alpha^0_{gt}) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{gt}),$$

  using asymptotic expansions from Arellano and Hahn (2007); Hahn and Newey (2004).
- If $T/N \rightarrow c$ as $N, T \rightarrow \infty$:
  - Adapt Chen, Fernández-Val, and Weidner (2021) to derive analytic expressions of asymptotic biases.

Outline

Monte Carlo Simulations: Static Logit Model

- $N = 90$ and $T = 7$; 500 replications.
- The data generating process is

$$Y_{it} = \mathbb{1}\{X_{it}\beta + \alpha_{g_i t} \geq \varepsilon_{it}\}, \; i = 1, ..., N, \; t = 1, ..., T,$$

where,

- $\beta = 1$, $X_{it} \sim \mathcal{N}(0, 1)$, $g_i \sim \mathcal{U}(\{1, ..., G\})$, for $G \in \{2, 3, 5\}$,
- $\alpha_{gt} = \mu_g + U_t$ with $U_t \sim \mathcal{U}([-\pi/\sqrt{3}, \pi/\sqrt{3}])$, $\mu = (-1, 1)'$ if $G = 2$, $\mu = (-\pi/\sqrt{3}, 0, \pi/\sqrt{3})'$ if $G = 3$, and $\mu = (-2\pi/\sqrt{3}, -\pi/\sqrt{3}, 0, \pi/\sqrt{3}, 2\pi/\sqrt{3})'$ if $G = 5$,
- $\varepsilon_{it} \sim$Logit$(0, \pi^2/3)$.
- U_t, x_{it}, g_i and ε_{it} are independent and i.i.d. across individuals and time periods.
- Competing methods have significant small-T biases in more adversarial settings (correlated effects): OLS, CMLE, Bonhomme and Manresa (2015)'s GFE.
- In progress: compare with Bonhomme, Lamadon, and Manresa (2021)'s 2-step GFE, Chen, Fernández-Val, and Weidner (2021)'s nonlinear factor model.

Monte Carlo Distribution of $\widehat{\beta}$

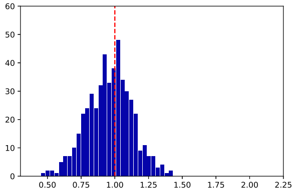


Figure: $G = 2$

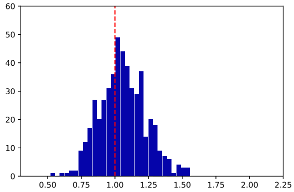


Figure: $G = 3$

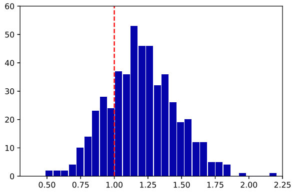


Figure: $G = 5$

- No bias for small G.
- Small-sample bias (and optimization errors) as G increases.

Monte Carlo Simulations: Bias and (Mis)Classification

	β		Clustering				CPU time
G	Bias	RMSE	M	R	P	RI	
2	-0.04	0.17	0.05	0.87	0.83	0.84	14.94
3	0.06	0.18	0.36	0.65	0.55	0.70	21.69
5	0.21	0.33	0.50	0.59	0.47	0.78	34.17

Notes: $N = 90, T = 7$, 500 replications, $\beta = 1$. M=Misclassification rate = minimum of $\sum_{i=1}^{N} |\widehat{g}_i - g_i^0|/N$ over all possible group labelings for $\widehat{g}_i$; R = Recall rate; P = Precision rate; RI = Rand Index. CPU time in seconds on a Microsoft 10 Professional, Intel(R) Core(TM) i7-1165G7MQ CPU@ 2.80GHz, 16GB RAM.

Monte Carlo Simulations: Inference and Coverage

	β		
G	Large-T analytical SE	SD across MC samples	.95
2	0.127	0.167	0.836
3	0.128	0.171	0.850
5	0.173	0.257	0.728

Notes: $N = 90$, $T = 7$, 500 replications, $\beta = 1$. Column (1) reports the median of the estimates of the large-T clustered variance formula, column (2) reports the the actual standard deviation across simulations, and column (3) reports the empirical non-rejection probabilities (nominal size 5%) based on the large-T clustered variance formula.

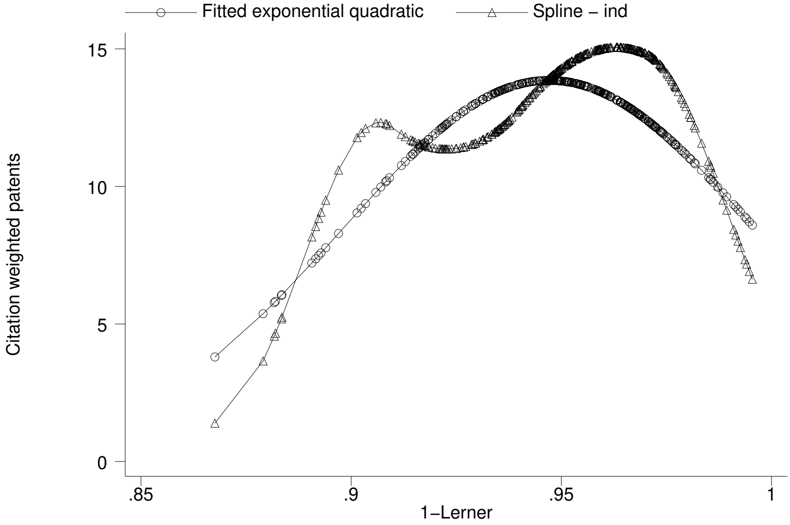
- Small undercoverage, improves as $N, T \to \infty$.

Outline

Innovation and Competition: an Inverted-U Relationship?

- **Does competition lead to more innovation?**
  - Longstanding debate (Griffith and Van Reenen, 2021).
  - Schumpetarian effect (-) v.s. "escape-competition" effect (+).
  - How to measure competition? innovation?
- Influential QJE's paper: Aghion, Bloom, Blundell, Griffith, and Howitt (2005)
  - 17 UK industries i over 22 years t (1973-1994): large-T, moderately large N.
  - Y_{it}=citation-weigthed patents$_{it}$; X_{it}=(1-Lerner)$_{it}$, (1-Lerner)$^2_{it}$.
  - Main specification: $Y_{it} \mid X_{i1}, ..., X_{it}, \alpha_i, \xi_t \sim \text{Poiss}(X_{it}'\beta + \alpha_i + \xi_t)$.
  - **"Inverted-U" relationship**: $\widehat{\beta}_1^{***} > 0$ and $\widehat{\beta}_2^{***} < 0$.
  - Model: *neck-to-neck* and *leader-laggard* firms, incremental incentives.
- **Fragile relationship**, sensitive to:
  - Country (Askenazy, Cahn, and Irac, 2013; Correa and Ornaghi, 2014; Hashmi, 2013), structural breaks (Correa, 2012), controls (Aghion, Van Reenen, and Zingales, 2013).
  - **Unobserved (confounding) dynamics?**

ABBGH's Inverted-U Relationship (Poisson TWFE)



Summary Statistics

	Competition = 1-Lerner index	Innovation = Citation-weighted patents	Technology gap
Mean	0.95	6.66	0.49
SD	0.02	8.43	0.16
p_{10}	0.92	0	0.28
Median	0.95	3.35	0.51
p_{90}	0.98	20.19	0.69

Notes: There are 17 industries, 354 observations and the time period covers 1973-94. Industries

- Point mass at zero & positivity motivate a Poisson-type model...
  - ....though conditional mean $\neq$ conditional variance here.
- Identification theorem: it is in theory valid to relax this parametric assumption.
  - E.g.: the negative-binomial model verifies our monotonicity and smoothness conditions.

Innovation and Competition Revisited

- Challenge the view of "permanent unobserved technological level" + common trend.
  - Telecom/internet revolution might not have affected all industries the same way.

- Substitute ABBGH's **common trend assumption** with that of a finite number of **unobserved clustered trends**.

- We find

  1. Evidence of time-varying UH (low-, increasing-, and high-innovation).
  2. **Mildly inverted-U**, no-evidence of structural break.

- Estimated clusters can be used as input for further analysis:
  - which key variables are missing?

Exercise #1: Evidence of Unobserved Time-Varying Heterogeneity?

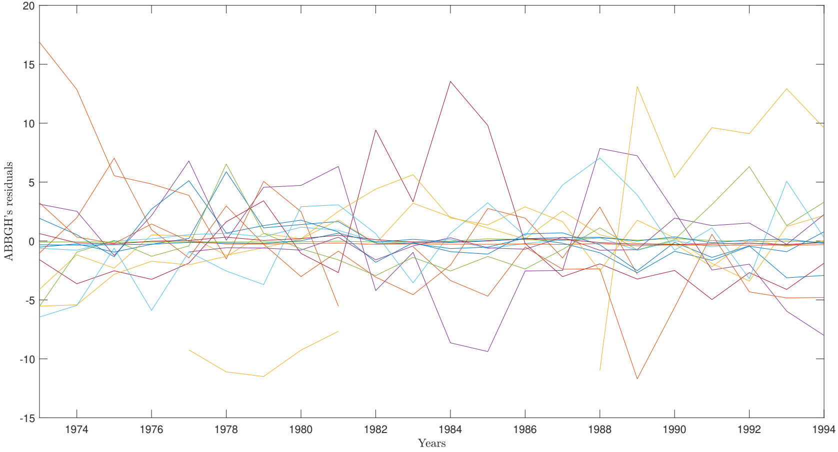
- Suppose we trust ABBGH's estimate $\widehat{\beta}$.
- Do we find evidence of a latent clustering structure in the data? Time-varying UH?
- Apply a smooth clustering method developed in companion paper, the **tetrad pairwise distance** (TPWD) estimator, to the panel of residuals

▸ Details TPWD

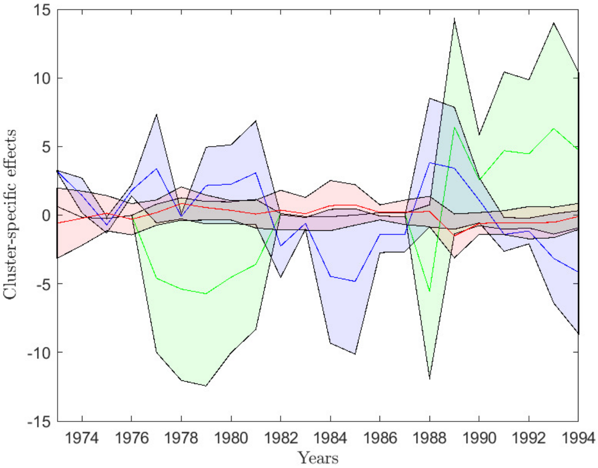
$$Y_{it} - \exp(X_{it}'\widehat{\beta} + \widehat{\alpha}_i + \widehat{\xi}_t).$$

- Unconstrained number of clusters.
- Polynomial time (no optimization).
- Input: regularization parameter $c \in (0, +\infty)$.
- Outputs:
  - regularization path $\{\widehat{G}(c) : c \in (0, +\infty)\}$.
  - cluster-specific time-varying effects.

Exercise #1: ABBGH's Residuals



Exercise #1: Three TPWD Clusters



▸ 2 clusters ▸ 4 clusters

Noisy data, few industries.
→ In progress: use Correa and Ornaghi (2014) and Hashmi (2013)'s U.S. data with more digits to increase the number of industries per clusters.

Exercise #2: Fitting a NGFE Poisson model

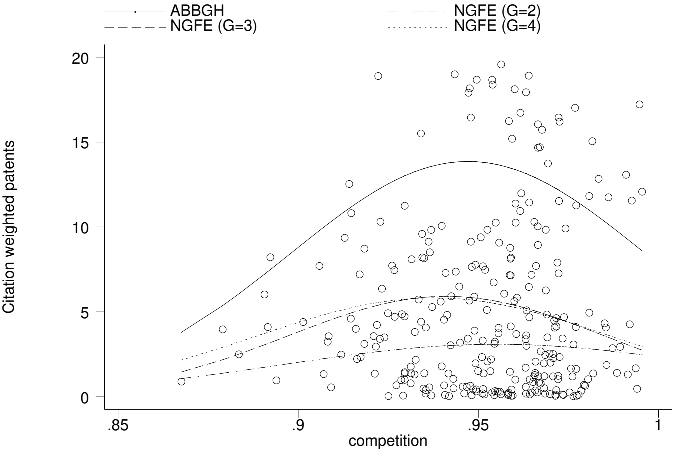
- Suppose we do not trust ABBGH's $\widehat{\beta}$ anymore.
- Allow for unobserved clusters of time-varying heterogeneity, estimate:

$$\Pr\left(patw_{it} = y \middle| \text{comp}_{it}, \text{comp}^2_{it}, g_i, \alpha_{g_i t}\right) = \exp(-\lambda_{it})\lambda_{it}^{y}/y!,$$

where $\lambda_{it} = \exp(\text{comp}_{it}\beta_1 + \text{comp}^2_{it}\beta_2 + \alpha_{g_i t})$ and $G \in \{2, 3, 4\}$.

- We obtain a ***mildly* inverted-U** relationship. ▸ Table

Excercise #2: A mildly Inverted-U



Exercise #2: Unobserved Clustered Dynamics

Figure: Time Effects

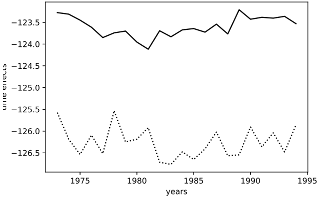


Figure: $G = 2$

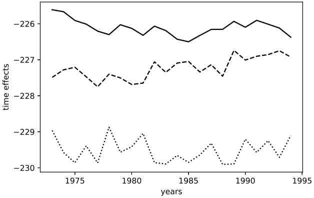


Figure: $G = 3$

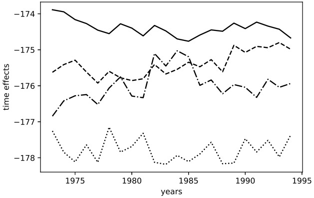
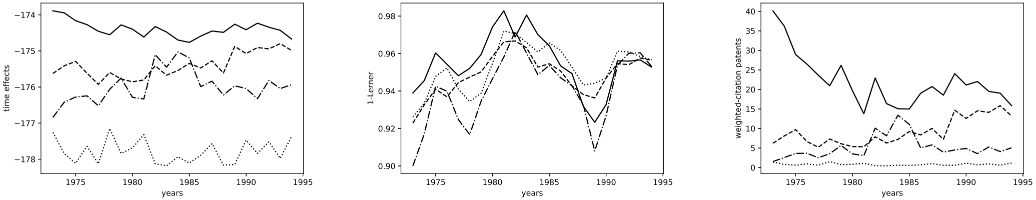


Figure: $G = 4$

Notes: Solid line=High-Innovation, dotted line=Low Innovation, dashed line=Early Catchers, dashdotted line=Late Catchers.

Visiualize Clusters

Exercise #2: Time Effects, Competition, and Innovation ($G = 4$)

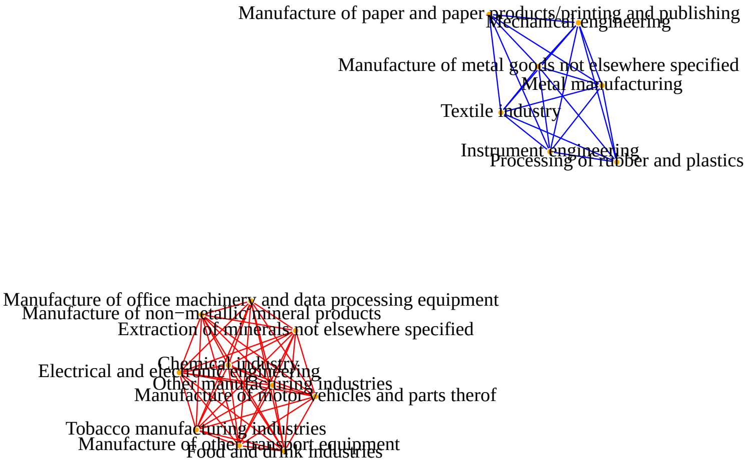


Note: Solid line=High-Innovation, dotted line=Low Innovation, dashed line=Early Catchers, dashdotted line=Late Catchers.

Exercise #2: Data-driven Industries Clustering ($G = 2$)

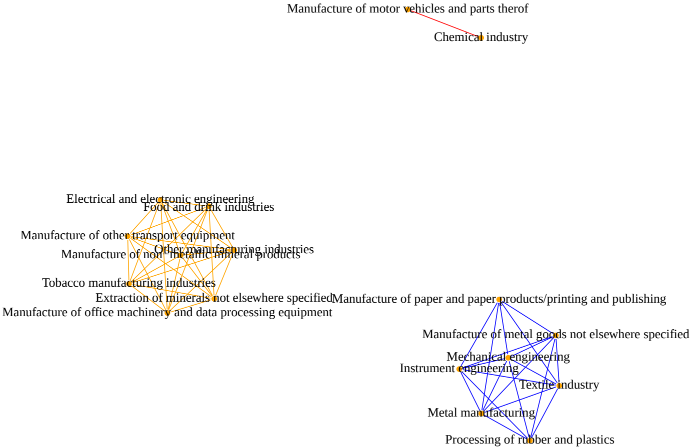
Go back

Low-Innovation/High-innovation



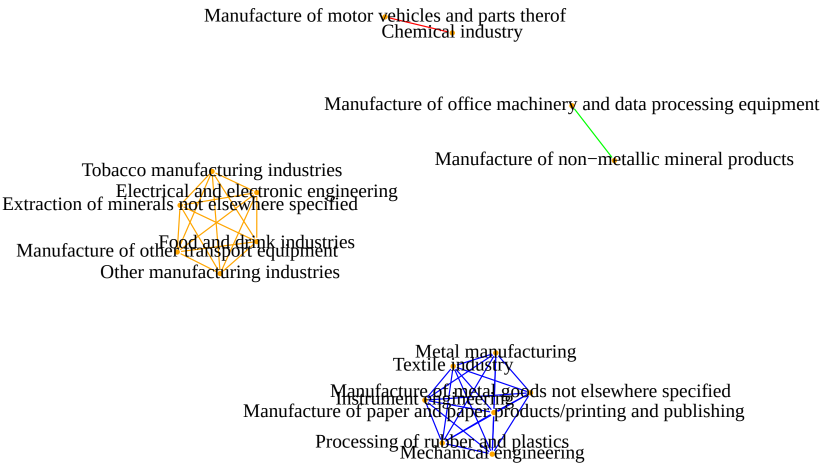
Exercise #2: Data-driven Industries Clustering ($G = 3$)

Low-Innovation/Catching/High-innovation



Exercise #2: Data-driven Industries Clustering ($G = 4$)

Low-Innovation/Late-catching/Early-Catching/High-innovation



Exercise #3: Robustness to Endogeneity and Structural Break

- **Simultaneity** is a key issue here. ABBGH use a **control function** approach based on a set of policy instruments (deregulation policies).
- **Structural break?** Establishment of CAFC courts in 1981-1982: granting patents becomes easier (Correa, 2012).
- We replicate each analysis, letting $G = 4$.

Table: Dependent Variable: Citation-Weighted Patents (Control Function)

	Poisson FE			Poisson GFE		
	Annual	Before 1983	After 1983	Annual	Before 1983	After 1983
Competition$_{it}$	386.59***	229.18*	113.42	527.52***	205.31	8.00
	(67.61)	(122.68)	(100.73)	(95.01)	(134.15)	(104.83)
Competition squared$_{it}$	-205.32***	-114.89*	-60.85	-278.75***	-112.13	-6.89
	(36.11)	(66.49)	(53.37)	(50.69)	(71.01)	(56.50)
Curve shape	sharp inv-U	increasing	·	mildly inv-U	·	·

- Mildly inverted-U robust to control function approach, no structural break.

Outline

Conclusion

- We derive sufficient conditions for **nonparametric identification** of all structural parameters of a class of **nonlinear GFE single-index panel data models with large-N, large-T**.
    - Cover binary choice, ordered choice, count data...
    - Nonparametric estimation of the link function is theoretically justified.
- We propose and study **semiparametric NGFE estimators**.
    - Consistent & asymptotically normal under regularity conditions.
    - Reasonable large-T approximations in small sample.
- We revisit ABBGH's "inverted-U" relationship between **innovation and competition**.
    - Evidence of **significant unobserved time-varying heterogeneity**, mildly inverted-U.
- Future research:
    - Estimation of h^0, G^0.
    - Reducing computational burden.
    - Increase sample size in empirical application (e.g., using U.S. data).

Thank you!

martin.mugnier@ensae.fr

Examples

Go Back

- **Binary outcome**: $Y_{it} = \mathbb{1}\{X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} \geq 0\}$, where the $(\varepsilon_{it})_{i,t}$ are independent from $(X_i, \gamma^0, \alpha^0)$ and i.i.d. with (unknown) cumulative distribution function Ψ^0. Then,

$$h^0(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}) = \mathbb{1}\{y = 1\}\Psi^0(X_{it}'\beta^0 + \alpha^0_{g_i^0 t}) + \mathbb{1}\{y = 0\}(1 - \Psi^0(X_{it}'\beta^0 + \alpha^0_{g_i^0 t})).$$

- **Ordered outcome**:

$$Y_{it} = \begin{cases} 0 & \text{if } X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} < d_1^0. \\ 1 & \text{if } d_1^0 \leq X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} < d_2^0. \\ 2 & \text{if } X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} \geq d_2^0, \end{cases} \tag{2}$$

where $d_2^0 > d_1^0$, and the $(\varepsilon_{it})_{i,t}$ are independent from $(X_i, \gamma^0, \alpha^0)$ and i.i.d. with (unknown) cumulative distribution function Ψ^0. Then,

$$h^0(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}) = \begin{cases} 1 - \Psi^0(X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - d_1^0) & \text{if } y = 0. \\ \Psi^0(X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - d_1^0) - \Psi^0(X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - d_2^0) & \text{if } y = 1. \\ \Psi^0(X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - d_2^0) & \text{if } y = 2. \end{cases}$$

- **Count outcome**: $\mathcal{Y} = \{0, 1, 2, ...\}$. A Poisson parametrization assumes

$$h^0(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}) = \frac{(\lambda_{it}^0)^y \exp(-\lambda_{it}^0)}{y!},$$

where $\lambda_{it}^0 = \exp(X_{it}'\beta^0 + \alpha^0_{g_i^0 t})$.

Identification: Sketch of Proof (1/2)

1. Fix $N \in \mathbb{N}^*$, let $T \to \infty$ to identify A_N^0:
   - Let y and $\mathcal{X}^0$ verifying 2(a) and $x \in \mathcal{X}^0$. Pooling individual i's choices over time when $(Y_{it}, X_{it}) = (y, x)$, Assumption 1(b) and 3(a) ensure that

$$\mathbb{E}\left[\mathbb{1}\{Y_{i2} = y\} \mid X_{i2} = x, g_i^0, \mu^0_{g_i^0}, \xi_i^0\right] = \Pr\left(Y_{i2} = y \mid X_{i2} = x, g_i^0, \mu^0_{g_i^0}, \xi_i^0\right) = \rho_i(x)$$

     is identified.
   - ϕ known $\implies A_N^0 = (\phi(\rho_i, \rho_j))_{(i,j)\in\{1,\ldots,N\}^2}$ is identified.

Identification: Sketch of Proof (1/2)

1. Fix $N \in \mathbb{N}^*$, let $T \to \infty$ to identify A_N^0:
   - Let y and $\mathcal{X}^0$ verifying 2(a) and $x \in \mathcal{X}^0$. Pooling individual i's choices over time when $(Y_{it}, X_{it}) = (y, x)$, Assumption 1(b) and 3(a) ensure that

     $$\mathbb{E}\left[\mathbb{1}\{Y_{i2} = y\} \mid X_{i2} = x, g_i^0, \mu_{g_i^0}^0, \xi_i^0\right] = \Pr\left(Y_{i2} = y \mid X_{i2} = x, g_i^0, \mu_{g_i^0}^0, \xi_i^0\right) = \rho_i(x)$$

     is identified.
   - ϕ known $\implies A_N^0 = (\phi(\rho_i, \rho_j))_{(i,j) \in \{1,\dots,N\}^2}$ is identified.
2. Under 1(a) and 2(b): $G^0 = \limsup_{N \to \infty} \mathrm{rank}(A_N^0)$ is identified.

Identification: Sketch of Proof (1/2)

1. Fix $N \in \mathbb{N}^*$, let $T \to \infty$ to identify A_N^0:
   - Let y and $\mathcal{X}^0$ verifying 2(a) and $x \in \mathcal{X}^0$. Pooling individual i's choices over time when $(Y_{it}, X_{it}) = (y, x)$, Assumption 1(b) and 3(a) ensure that

     $$\mathbb{E}\left[\mathbb{1}\{Y_{i2} = y\} \mid X_{i2} = x, g_i^0, \mu^0_{g_i^0}, \xi_i^0\right] = \Pr\left(Y_{i2} = y \mid X_{i2} = x, g_i^0, \mu^0_{g_i^0}, \xi_i^0\right) = \rho_i(x)$$

     is identified.
   - ϕ known $\implies A_N^0 = (\phi(\rho_i, \rho_j))_{(i,j)\in\{1,\ldots,N\}^2}$ is identified.
2. Under 1(a) and 2(b): $G^0 = \limsup_{N\to\infty} \mathrm{rank}(A_N^0)$ is identified.
3. Let $(i, t) \in \mathbb{N}^{*2}$. Under 1(a) and 2(b), conditional on $(\gamma^0, \alpha^0, \lambda^0, \mu^0)$, $\{Y_{jt}, X_{jt} : g_j^0 = g_i^0\}$ is an identified infinite sequence of i.i.d. random variables. Theorem 4.1 in Ichimura (1993) with $\varphi(\cdot) = \sum_{y\in\mathcal{Y}} y h^0(y, \cdot + \alpha^0_{g_i^0 t}) \implies \beta^0/\|\beta^0\| = \beta^0$ is identified.

Identification: Sketch of Proof (2/2) ▸ Go Back

4. Let $\bar{y}$ such that $h^0(\bar{y}, \cdot)$ is strictly monotonic.
   - Pooling units in clusters $(g, \widetilde{g}) \in \{1, ..., G^0\}^2$ such that $(Y_{it}, X_{it}) = (\bar{y}, x_1)$ one identifies:

$$\Pr(Y_{1t} = \bar{y} | X_{1t} = x_1, g_1^0 = g, \alpha_{gt}^0) = h^0(\bar{y}, x_1'\beta^0 + \alpha_{gt}^0),$$
$$\Pr(Y_{1t} = \bar{y} | X_{1t} = x_1, g_1^0 = \widetilde{g}, \alpha_{\widetilde{g}t}^0) = h^0(\bar{y}, x_1'\beta^0 + \alpha_{\widetilde{g}t}^0).$$

   - Compensating variations: $x_2 \in \mathcal{X}$ is identified from

$$\Pr(Y_{1t} = \bar{y} | X_{1t} = x_2, g_1^0 = g, \alpha_{gt}^0) = \Pr(Y_{1t} = \bar{y} | X_{1t} = x_1, g_i^0 = \widetilde{g}, \alpha_{\widetilde{g}t}^0)$$
$$\iff h^0(\bar{y}, x_1'\beta^0 + \alpha_{\widetilde{g}t}^0) = h^0(\bar{y}, x_2'\beta^0 + \alpha_{gt}^0).$$

   - Inverting $h(\bar{y}, \cdot)$: $\alpha_{\widetilde{g}t}^0 - \alpha_{gt}^0 = (x_2 - x_1)'\beta^0$ is identified for all $g, \widetilde{g}, t$. Same reasoning fixing g yields identification of $\alpha_{gt}^0 - \alpha_{g\widetilde{t}}^0$ for all $(g, t, \widetilde{t})$.
   - Result follows since $\alpha_{11}^0 = 0$ implies that (α_{1t}^0) is identified for all t and (α_{g1}^0) is identified for all g so that, for all $g \neq 1, t \neq 1$,

$$\alpha_{gt}^0 = \underbrace{\alpha_{gt}^0 - \alpha_{1t}^0}_{:=a} + \underbrace{\alpha_{1t}^0}_{:=b}, \quad \text{where } a, b \text{ are identified.}$$

5. Identify h^0 as a function of $y \in \mathcal{Y}$ and identified single index $X_{it}'\beta^0 + \alpha_{g_i^0 t}^0$.

Estimation: Computation

Go Back

An Iterative (Heuristic) Algorithm

1. Let $(\beta^{(0)}, \alpha^{(0)}) \in \mathcal{B} \times \mathcal{A}^{G^0 T}$ be some starting value. Set $s = 0$.
2. Compute for all $i \in \{1, \ldots, N\}$:

$$g_i^{(s+1)} = \underset{g \in \{1,\ldots,G^0\}}{\operatorname{argmin}} \sum_{t=1}^{T} -\ln h^0 \left(Y_{it}, X_{it}' \beta^{(s)} + \alpha_{gt}^{(s)} \right).$$

3. Compute:

$$\left(\beta^{(s+1)}, \alpha^{(s+1)} \right) = \underset{(\beta,\alpha) \in \mathcal{B} \times \mathcal{A}^{G^0 T}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} -\log h^0 \left(Y_{it}, X_{it}' \beta + \alpha_{g_i^{(s+1)} t} \right).$$

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

(straightforward adaptation of Bonhomme and Manresa, 2015)

Inference: Asymptotic Equivalence to the Oracle MLE

- The infeasible oracle MLE $(\widetilde{\beta}, \widetilde{\alpha})$ verifies:

$$(\widetilde{\beta}, \widetilde{\alpha}) = \underset{(\beta,\alpha)\in\mathcal{B}\times\mathcal{A}^{GT}}{\operatorname{argmin}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} -\ln\Psi\left((2Y_{it}-1)(X_{it}'\beta+\alpha_{g_i^0 t})\right).$$

$\Longrightarrow$ MLE with known group dummies.

Assumption (A.N. 1)

(a) (Non-neglible clusters) For all $g \in \{1, ..., G^0\}$: $\operatorname{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\} = \pi_g > 0$.

(b) (Well-separated clusters) For all $(g, \widetilde{g}) \in \{1, ..., G^0\}^2$ such that $g \neq \widetilde{g}$: $\operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)^2 = c_{g,\widetilde{g}} > 0$.

(c) (Mixing) There exist constants $a > 0$ and $d > 0$ and a sequence $\alpha[t] \leq \exp(-at^d)$ such that, for all $(g, \widetilde{g}) \in \{1, ..., G^0\}^2$ such that $g \neq \widetilde{g}$, $\{\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0 : t\}$ is a strongly mixing process with mixing coefficients $\alpha[t]$.

- Same as Bonhomme and Manresa (2015).

Inference: Asymptotic Equivalence to the Oracle MLE

Go Back

Lemma (Sup-Norm Consistency and Asymptotic Equivalence)

Let Assumptions Mod., Cons., and A.N.1 hold. Then, for all $\delta > 0$ and as N and T tend to infinity

$$\Pr\left(\sup_{i \in \{1, \dots, N\}} \left| \widehat{g}_i - g_i^0 \right| > 0 \right) = o(1) + o(NT^{-\delta}),$$

and

$$\widehat{\beta} = \widetilde{\beta} + o_p(T^{-\delta}),$$

and

$$\widehat{\alpha}_{gt} = \widetilde{\alpha}_{gt} + o_p(T^{-\delta}) \text{ for all } g, t.$$

- If $\sqrt{N} T^{-\delta} \to 0$ for some $\delta > 0$: sufficient to derive limiting distribution of infeasible MLE!

Monte Carlo Simulations: Time Effects

- One realization

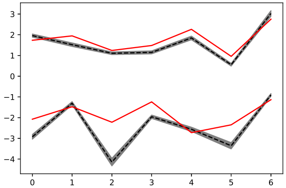


Figure: $G = 2$

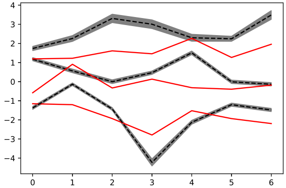


Figure: $G = 3$

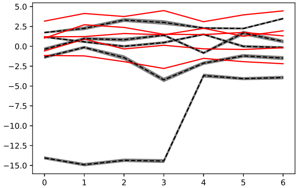


Figure: $G = 5$

- Small-sample bias as G increases, tight 95% CIs.
- Impact of finite-sample misclassification.

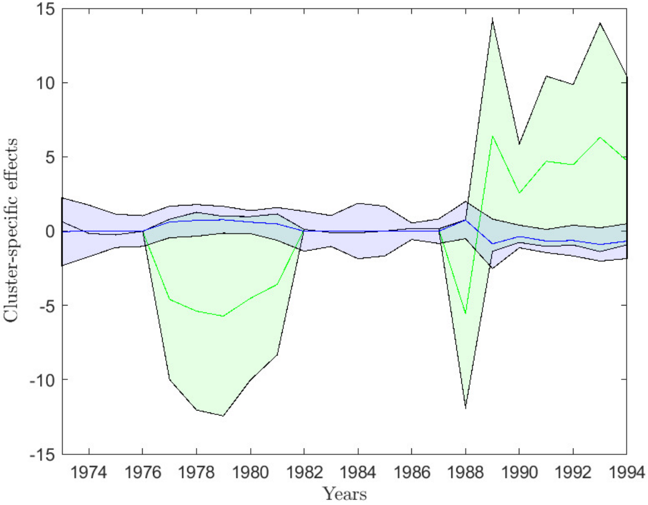
Industries at the 2 Digits

Go Back

SIC 2	Name
22	Metal manufacturing
23	Extraction of minerals not elsewhere specified
24	Manufacture of non-metallic mineral products
25	Chemical industry
31	Manufacture of metal goods not elsewhere specified
32	Mechanical engineering
33	Manufacture of office machinery and data processing equipment
34	Electrical and electronic engineering
35	Manufacture of motor vehicles and parts therof
36	Manufacture of other transport equipment
37	Instrument engineering
41/42	Food, drink and tobacco manufacturing industries
43	Textile industry
47	Manufacture of paper and paper products; printing and publishing
48	Processing of rubber and plastics
49	Other manufacturing industries

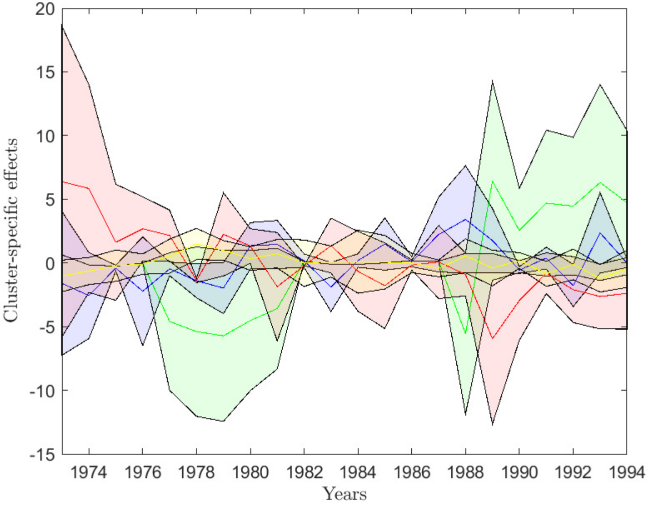
Two Clusters

Go Back



Four Clusters

Go Back



Mildly Inverted-U

Go Back

	Poisson FE		Poisson GFE		
Dependent Variable: Citation-Weighted Patents	(1)	(2)	(3)	(4)	(5)
Competition$_{it}$	152.80	387.46	263.93	487.52	378.96
	(55.74)	(67.74)	(68.36)	(76.44)	(79.30)
Competition squared$_{it}$	-80.99	-204.55	-138.15	-259.02	-202.09
	(29.61)	(36.17)	(36.45)	(40.84)	(42.33)
Jointly significant	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes			
Industry effects		Yes			
Time-varying clustered effects			Yes	Yes	Yes
Number of clusters (G)			2	3	4
Observations	354	354	354	354	354

Note: numbers in brackets are analytical standard errors.

Tetrad Pairwise Differencing Estimator (1/2)

Go Back

- For any tetrad (i, j, k, l), let

$$S_{NT}(i,j,k,l) := \frac{1}{T}\sum_{t=1}^{T}(y_{it} - y_{jt})(y_{kt} - y_{lt}).$$

- The tetrad pairwise differencing (TPWD) estimator is obtained from the following two steps:

1. Let $c_T \in (0, +\infty)$ and compute $\widehat{\mathbf{W}}^{TPWD} \in \{0,1\}^{N^2}$ with entries:

$$\widehat{W}_{ij}^{TPWD} = \mathbb{1}\left\{\max_{(k,l)\in(\{1,\dots,N\}\setminus\{i,j\})^2} |S_{NT}(i,j,k,l)| \leq c_T\right\}, \quad i = 1, \dots, N,\ j = 1, \dots, N.$$

Set $\widehat{G}^{TPWD} = |\{\widehat{\mathbf{W}}_{1,\cdot}^{TPWD}, \dots, \widehat{\mathbf{W}}_{N,\cdot}^{TPWD}\}|$ and pick $(\widehat{g}_1^{TPWD}, \dots, \widehat{g}_N^{TPWD}) \in \{1, \dots, \widehat{G}^{TPWD}\}^N$ satisfying constraints:

$$\left[\widehat{g}_i^{TPWD} = \widehat{g}_j^{TPWD} \iff \widehat{\mathbf{W}}_{i,\cdot}^{TPWD} = \widehat{\mathbf{W}}_{j,\cdot}^{TPWD}\right], \quad i = 1, \dots, N,\ j = 1, \dots, N.$$

Tetrad Pairwise Differencing Estimator (2/2)

2. Compute $\widehat{\alpha} := (\widehat{\alpha}_{11}^{TPWD}, ..., \widehat{\alpha}_{1T}^{TPWD}, ..., \widehat{\alpha}_{\widehat{G}^{TPWD}1}^{TPWD}, ..., \widehat{\alpha}_{\widehat{G}^{TPWD}T}^{TPWD})$ from:

$$\widehat{\alpha} = \underset{\alpha \in \mathcal{A}^{\widehat{G}^{TPWD}T}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \alpha_{\widehat{g_i}^{TPWD}t} \right)^2 .$$

- Asymptotic guarantees under correct (linear) GFE specification (see Mugnier, 2022).