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

Martin Mugnier[1]

[1]CREST, ENSAE, Institut Polytechnique de Paris

EEA-ESEM, Milan

August 24, 2022

Motivation

- Observational **panel data** offer opportunities to control for **unobserved heterogeneity** (UH)
- Workhorse two-way fixed effects linear 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. Interest: β, α_i, ξ_t.

- In many economic settings:

1. parallel trends $(\alpha_i + \xi_t)$ may be overly restrictive.
   - Innovation (Y) and competition (X): $\neq$ trajectories of unobserved technological change across industries (Aghion, Bloom, Blundell, Griffith, and Howitt, 2005).
   - Mental health (Y) and abortion (X): $\neq$ trajectories of unobserved risky behaviors across young women (Janys and Siflinger, 2021).

2. a linear model is poorly suited.
   - Count data (e.g., number of patents), discrete choice (e.g., developing a mental illness).

Motivation

- Allowing for **time-varying UH** in **nonlinear FE models** is **challenging**.
  - Interactive fixed effects $\alpha_i'\xi_t$, $\alpha_i, \xi_t \in \mathbb{R}^{G^0}$ (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, 2022; 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

- Allowing for **time-varying UH** in **nonlinear FE models** is **challenging**.
  - Interactive fixed effects $\alpha_i'\xi_t$, $\alpha_i, \xi_t \in \mathbb{R}^{G^0}$ (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, 2022; 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**.
  - innovation clusters, health types (Deb and Trivedi, 1997; Janys and Siflinger, 2021),
  - **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-N, 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 $\{(Y_{it}, X_{it}')'_{1\le t\le T} : 1 \le i \le N\}$.
- Static **nonlinear grouped fixed effects (NGFE) models** with **single index:**
  - Individual i at time t chooses $Y_{it} \in \mathcal{Y}$ with probability

$$\Pr\left(Y_{it} = y \mid X_i^t, g_i^0, \alpha^0_{g_i^0 t}\right) = h^0\left(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}\right), \tag{1}$$

  - $X_{it} \in \mathbb{R}^p$: 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, $\alpha^0 = \{\alpha^0_{gt} : (g, t)\}$;
  - $h^0 \in \{h : \mathcal{Y} \times \mathbb{R} \to (0, 1), \sum_{y\in\mathcal{Y}} h(y, \cdot) = 1, \sum_{y\in\mathcal{Y}} |y| h(y, .) < +\infty\}$ **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

$$\Pr\left(Y_{it} = y \mid X_i^t, g_i^0, \alpha^0_{g_i^0 t}\right) = h^0\left(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}\right), \tag{1}$$

  - $X_{it} \in \mathbb{R}^p$: 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, $\alpha^0 = \{\alpha^0_{gt} : (g, t)\}$;
  - $h^0 \in \{h : \mathcal{Y} \times \mathbb{R} \to (0,1), \sum_{y\in\mathcal{Y}} h(y, \cdot) = 1, \sum_{y\in\mathcal{Y}} |y|h(y, .) < +\infty\}$ **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 (parametric rate) 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-N, 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**: $\max_{i \in \{1,\dots,N\}} |\widehat{g}_i - g_i^0| = o_p(1)$.

- **Monte Carlo simulations:**
  - Good finite sample properties of large-T approximations, outperform competing methods in discrete settings.

- **Empirical application**:
  - Revisit the **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**, and provide a **data-driven clustering** of industries. Some well-known results do not hold anymore.
    $\implies$ controlling for unobserved dynamics matters.

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 (2022); 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

Large-N, T Nonparametric Identification

Semiparametric Estimation and Inference

Monte Carlo Simulations

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

Conclusion

Outline

Identification: Assumptions

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

$$\Pr\left(Y_{it} = y \mid X_{i1}, ..., X_{it}, g_i^0, \alpha^0_{g_i^0 t}\right) = h^0\left(y, X_{it}'\beta^0 + \alpha^0_{g_i^0 t}\right), \quad (i,t,y) \in [N] \times [T] \times \mathcal{Y}.$$

- 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: $Y_{1t} \mid X_1, g_1^0, \alpha^0, \lambda^0, \mu^0, \xi^0 \overset{d}{=} Y_{1t} \mid X_{1t}, g_1^0, \alpha^0_{g_1^0 t}$.

Identification: Assumptions

Assumption 2 (Latent clustering & injectivity condition)

$\mathcal{X} := \bigcap_{i=1}^{\infty} \mathcal{X}_i$ *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* $\rho_i(x) : \mathcal{X}^0 \ni x \mapsto \Pr\left(Y_{i2} = y \mid X_{i2} = x, g_i^0, \mu_i^0, \xi_i^0\right)$, $\phi\left(\rho_i, \rho_j\right) = \mathbf{1}\left\{g_i^0 = g_j^0\right\}$.

(b). *For all* $g \in \mathcal{G}^0$, *almost surely* $\Pr\left(g_1^0 = g \mid \alpha^0, \lambda^0, \mu^0, \xi^0\right) > 0$.

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 $0 < \underline{\delta} \leq \inf_{x \in \mathcal{X}^0} f_{X_{i2}|g_i^0,\mu_{g_i^0}^0,\xi_i^0}(x) \leq \sup_{x \in \mathcal{X}^0} f_{X_{i2}|g_i^0,\mu_{g_i^0}^0,\xi_i^0}(x) \leq \overline{\delta} < \infty$.

(b). Almost surely, $\mathbb{E}\left(\left\|X_{12}\right\|^2 \mid g_1^0, \alpha^0, \lambda^0, \mu^0\right)$ is finite and $\mathbb{E}\left(X_{12}X_{12}' \mid g_1^0, \alpha^0, \lambda^0, \mu^0\right)$ 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 $(g, \widetilde{g}, t)$, *there exist* $x_1, x_2 \in \mathcal{X}$ *such that*

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

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

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

Identification: Main Result

Let $W_N^0 := \left(\mathbf{1}\left\{g_i^0 = g_j^0\right\}\right)_{(i,j)\in\{1,\dots,N\}^2}$.

Theorem (Identification)

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

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

Identification: Sketch of Proof

1. Fix $N \in \mathbb{N}^*$ and let $\overline{y} \in \mathcal{Y}$, $\mathcal{X}^0 \subset \mathcal{X}$ verifying 2(a) and $x \in \mathcal{X}^0$.
   - Pooling individual i's choices over time when $(Y_{it}, X_{it}) = (\overline{y}, x)$, Assumption 1(b) and 3(a) ensure that

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

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

Identification: Sketch of Proof

1. Fix $N \in \mathbb{N}^*$ and let $\overline{y} \in \mathcal{Y}$, $\mathcal{X}^0 \subset \mathcal{X}$ verifying 2(a) and $x \in \mathcal{X}^0$.
    - Pooling individual i's choices over time when $(Y_{it}, X_{it}) = (\overline{y}, x)$, Assumption 1(b) and 3(a) ensure that

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

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

Identification: Sketch of Proof

1. Fix $N \in \mathbb{N}^*$ and let $\overline{y} \in \mathcal{Y}$, $\mathcal{X}^0 \subset \mathcal{X}$ verifying 2(a) and $x \in \mathcal{X}^0$.
   - Pooling individual i's choices over time when $(Y_{it}, X_{it}) = (\overline{y}, x)$, Assumption 1(b) and 3(a) ensure that
   $$\mathbb{E}\left[\mathbf{1}\left\{Y_{i2} = \overline{y}\right\} \mid X_{i2} = x, g_i^0, \mu^0_{g_i^0}, \xi_i^0\right] = \Pr\left(Y_{i2} = \overline{y} \mid X_{i2} = x, g_i^0, \mu^0_{g_i^0}, \xi_i^0\right) = \rho_i(x)$$
   is identified.
   - ϕ known $\implies$ $W_N^0 = (\phi(\rho_i, \rho_j))_{(i,j)\in\{1,\dots,N\}^2}$ is identified.
2. Under 1(a) and 2(b): $G^0 = \limsup_{N\to\infty} \mathrm{rank}(W_N^0)$ 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

4 Let $\underline{y}$ such that $h^0(\underline{y}, \cdot)$ is strictly monotonic.

- Pooling units in clusters $(g, \widetilde{g}) \in \{1, ..., G^0\}^2$ such that $(Y_{it}, X_{it}) = (\underline{y}, x_1)$ one identifies:

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

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

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

- Inverting $h(\underline{y}, \cdot)$, $\alpha_{\widetilde{g}t}^0 - \alpha_{gt}^0 = (x_2 - x_1)'\beta^0$ 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 $h^0(\cdot, \cdot)$ as a function of $y \in \mathcal{Y}$ and identified single index $X_{it}'\beta^0 + \alpha_{g_i^0 t}^0$.

Outline

Large-N, T Nonparametric Identification

Semiparametric Estimation and Inference

Monte Carlo Simulations

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

Conclusion

Estimation: Semiparametric NGFE Estimators

- Nonparametric estimation based on constructive identification: possible but hard in practice.
  - Estimate $\widehat{\rho}_i$ with ML, prove $\left\|\widehat{W_N} - W_N^0\right\|_\infty \xrightarrow{p} 0$...
  - Slow convergence rates might be deterrent, no inference.

Estimation: Semiparametric NGFE Estimators

- Nonparametric estimation based on constructive identification: possible but hard in practice.
  - Estimate $\widehat{\rho}_i$ with ML, prove $\left\|\widehat{W_N} - W_N^0\right\|_\infty \xrightarrow{p} 0$...
  - Slow convergence rates might be deterrent, no inference.
- Instead, propose a **practically useful** approach, assuming (h^0, G^0) is known (e.g., Probit).
- **Semiparametric Classification Maximum Likelihood Estimator**:

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

where $\gamma = (g_1, \ldots, g_N)'$.

  - 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
  - Choice of G^0: AIC/BIC or report results for multiple choices.
  - Mugnier (2022): computationally trivial estimator + estimation of G^0, but guarantees only in linear/multiplicative models.

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 ($|\mathcal{Y}| = 2$):

$$Y_{it} = \mathbf{1}\left\{X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} \geq 0\right\}, \ i = 1, ..., N, \ t = 1, ..., T. \tag{2}$$

- 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. (2) holds and:

(a). (Weak exogeneity) $(\mathbf{X}_-^{(t)}, \gamma^0, \alpha^0, \boldsymbol{\varepsilon}_-^{(t-1)})$ and $\boldsymbol{\varepsilon}_+^{(t)}$ 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 of $\mathbb{R}^p$ and $\mathbb{R}$, respectively.

(b) (Bounded covariates) There exists a constant $M > 0$ such that $\left\|X_{it}\right\| \leq M$ almost surely.

(c) (Noncollinearity) Let $\overline{X}_{g\wedge\widetilde{g},t}$ denotes the mean of X_{it} in the intersection of clusters $g_i^0 = g$, and $g_i = \widetilde{g}$. For all partitions $\gamma = \{g_1, ..., g_N\} \in \Gamma_{\mathcal{G}^{0N}}$, let $\widehat{\rho}(\gamma)$ denote 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, $\operatorname{plim}_{N,T\to\infty} \min_{\gamma\in\Gamma_{\mathcal{G}^0}} \widehat{\rho}(\gamma) = \rho > 0$.

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

- Under reg. conditions (well-separation of groups, time-dependence, noncollinearity):
  - Uniformly consistent classification of individuals: $\sup_i \left|\widehat{g}_i - g_i^0\right| \xrightarrow{p} 0$,
  - Asymptotic equivalence to the infeasible oracle MLE which knows the clustering $(\widetilde{\beta}, \widetilde{\alpha})$. ▸ Details
- 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

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

  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

Large-N, T Nonparametric Identification

Semiparametric Estimation and Inference

Monte Carlo Simulations

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

Conclusion

Monte Carlo Simulations: Static Logit Model

- $N = 90$ and $T = 7$; 50 replications, 200 iterations (should be both increased in near future).
- The data generating process is

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

where, $\beta = 1$, $\varepsilon_{it} \sim$Logit$(0, \pi^2/3)$, and $g_i \sim$Unif$\left\{1, ..., G^0\right\}$ for $G^0 \in \{2, 3, 5\}$.

  - DGP 1: grouped patterns of time-varying UH (AR(1) processes)
  - DGP 2: grouped patterns of time-invariant UH.
  - DGP 3: continuous time-invariant UH.
  - DGP4 : No UH.

In all DGPs: $(X_{it}, \alpha_{g_i t}) \perp\!\!\!\perp \varepsilon_{it}$, and X_{it} is continuous with limited independent variation.

- Results:
  - Competing methods have significant small-T biases, higher RMSE ,and less coverage in more adversarial settings (correlated time-varying effects): CMLE, Bonhomme, Lamadon, and Manresa (2022)'s 2-step GFE, Bonhomme and Manresa (2015)'s GFE, linear TWFE, nonlinear TWFE, pooled OLS.
  - Also consider a dynamic model in paper: results are similar.

Monte Carlo Simulations: Bias and RMSE of $\widehat{\beta}$

		NGFE		CMLE		NLTWFE		2STEPGFE		Pooled OLS		LTWFE		GFE	
DGP	G^0	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
1	2	-0.072	0.268	-0.104	0.551	0.217	0.950	-0.252	1.516	-0.407	0.411	-0.790	0.812	-0.798	0.814
	3	-0.089	0.294	0.294	0.637	0.669	1.000	0.355	0.893	-0.363	0.366	-0.724	0.734	-0.853	0.874
	5	-0.022	0.264	0.167	0.538	0.359	0.824	0.104	0.779	-0.369	0.373	-0.766	0.776	-0.784	0.839
2	2	0.106	0.171	0.010	0.161	0.223	0.302	-0.278	0.309	-0.779	0.780	-0.831	0.831	-0.816	0.818
	3	0.236	0.289	0.014	0.160	0.238	0.309	-0.300	0.345	-0.768	0.769	-0.867	0.867	-0.837	0.841
	5	0.601	0.637	-0.004	0.169	0.250	0.332	-0.324	0.358	-0.747	0.747	-0.916	0.916	-0.853	0.860
3	2	0.352	0.385	-0.001	0.169	0.221	0.313	-0.110	0.211	-0.776	0.777	-0.857	0.857	-0.826	0.827
	3	0.432	0.486	-0.002	0.170	0.219	0.308	-0.066	0.192	-0.788	0.789	-0.859	0.859	-0.845	0.846
	5	0.471	0.499	0.011	0.156	0.235	0.309	-0.057	0.186	-0.787	0.788	-0.858	0.858	-0.833	0.836
4	2	0.040	0.151	-0.002	0.152	0.195	0.269	0.085	0.221	-0.789	0.789	-0.783	0.784	-0.788	0.789
	3	0.095	0.159	0.016	0.124	0.223	0.269	0.109	0.213	-0.776	0.776	-0.778	0.779	-0.790	0.792
	5	0.114	0.178	0.018	0.118	0.222	0.266	0.094	0.204	-0.775	0.775	-0.778	0.779	-0.803	0.809

- Small bias in DGPs 1 and 4 ($\approx$ CMLE/2 and 2CMLE resp.)
- Small RMSE in DGPs 1 and 4 ($\approx$ CMLE/2 and CMLE resp.).
- If time-invariant heterogeneity (DGPs 2 and 3), CMLE is better.
- 2STEPGFE is dominated by the union of both methods.

Monte Carlo Simulation: Classification Accuracy and CPU time

		NGFE					2STEPGFE						GFE				
DGP	G^0	P	R	RI	M	CPU	P	R	RI	M	CPU	$\widehat{G}$	P	R	RI	M	CPU
1	2	0.51	0.87	0.51	0.44	10.62	0.54	0.24	0.51	0.77	10.19	5.38	0.54	0.55	0.54	0.38	29.27
	3	0.35	0.81	0.42	0.57	11.42	0.37	0.24	0.60	0.75	11.34	5.48	0.36	0.38	0.57	0.55	29.63
	5	0.21	0.80	0.35	0.70	14.75	0.24	0.25	0.69	0.71	11.73	5.88	0.24	0.25	0.69	0.63	83.18
2	2	0.56	0.86	0.57	0.36	8.02	0.64	0.45	0.60	0.53	3.57	3.06	0.61	0.61	0.61	0.29	21.95
	3	0.40	0.85	0.49	0.51	8.52	0.57	0.49	0.70	0.44	4.70	3.64	0.46	0.49	0.64	0.42	22.00
	5	0.22	0.87	0.34	0.69	10.15	0.44	0.53	0.77	0.44	5.78	4.44	0.35	0.40	0.74	0.54	20.93

Notes: Static logit model with $\beta = 1$, $N = 90$, and $T = 7$. G^0 = true number of groups, P = Precision rate, R = Recall rate, RI = Rand Index, M = Misclassification Rate = minimum of $\sum_{i=1}^{N} \mathbf{1}\left\{\widehat{g}_i \neq g_i^0\right\}/N$ over all possible cluster relabelings, CPU = CPU time in seconds computed with Python's `time` command `time.perf_counter()`, $\widehat{G}$ = number of groups estimated by 2STEPGFE.

- Classification performance is uniformly poor.
- Accuracy of NGFE improves with number of initialization points (only 200 here), with $T \to \infty$, or less stringent UH.

Monte Carlo Simulations: Inference and Coverage

		NGFE			CMLE		
DGP	G^0	SE	SD	.95	SE	SD	.95
1	2	0.16	0.26	0.86	0.15	0.54	0.38
	3	0.17	0.28	0.80	0.16	0.56	0.40
	5	0.17	0.26	0.84	0.15	0.51	0.42
2	2	0.12	0.13	0.82	0.06	0.16	0.52
	3	0.12	0.17	0.46	0.07	0.16	0.62
	5	0.14	0.21	0.08	0.08	0.17	0.66
3	2	0.12	0.16	0.22	0.06	0.17	0.52
	3	0.12	0.22	0.18	0.06	0.17	0.52
	5	0.12	0.16	0.04	0.06	0.16	0.56
4	2	0.12	0.15	0.92	0.05	0.15	0.38
	3	0.13	0.13	0.92	0.05	0.12	0.56
	5	0.13	0.14	0.88	0.05	0.12	0.56

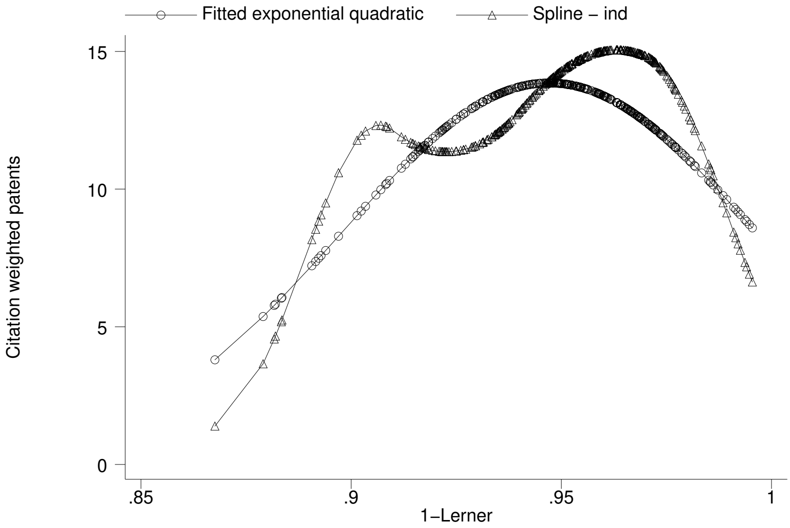
- Small under-coverage in discrete DGP (1, 2, 4), improves as $N, T \to \infty$.

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: $Y_{it}|X_{i1}, ..., X_{it}, \alpha_i, \xi_t \sim \text{Poiss}(X'_{it}\beta + \alpha_i + \xi_t)$, delivers an **"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 (TWFE Poisson)



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: in theory valid to relax this parametric assumption.
  - E.g.: the negative-binomial model verifies our monotonicity and smoothness conditions.

Innovation and Competition Revisited

- I challenge "permanent unobserved technological change" + common trend assumptions.
  - E.g., Telecom/internet revolution might not have affected all industries the same way.

- I substitute ABBGH's **common trend assumption** with that of a finite number of **unobserved clustered trends** and find

  1. Evidence of time-varying UH (low-, increasing-, and high-innovation).
  2. Stable and transitioning data-driven clusters of industries.
  3. **Mildly inverted-U**.

- Clusters effects and cluster memberships can be used as dependent variables for studies aimed at exploring factors which determine the quality or performance of technological change (or the clustering of firms).

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

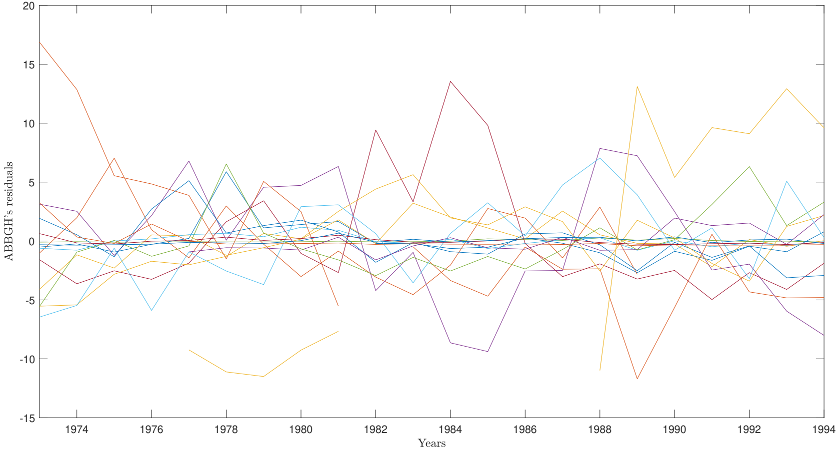
- Suppose we trust ABBGH's estimate $\widehat{\beta}$.
- Do we find evidence of a latent clustering structure in the data? Time-varying UH?
- Apply a smooth exploration method, the **tetrad pairwise distance** (TPWD) estimator, developed in companion paper, to the panel of residuals

▸ Details TPWD

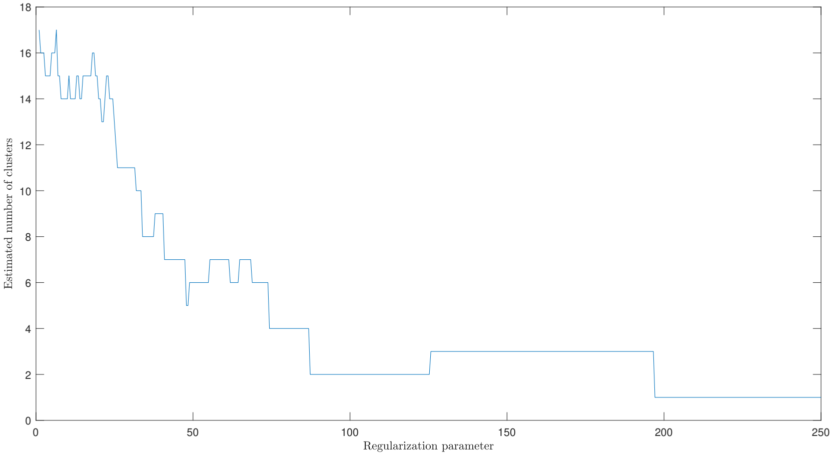
$$Y_{it} - \exp(X_{it}'\widehat{\beta} + \widehat{\alpha}_i + \widehat{\xi}_t).$$

- Unconstrained number of clusters, estimate $\widehat{G}$.
- 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.

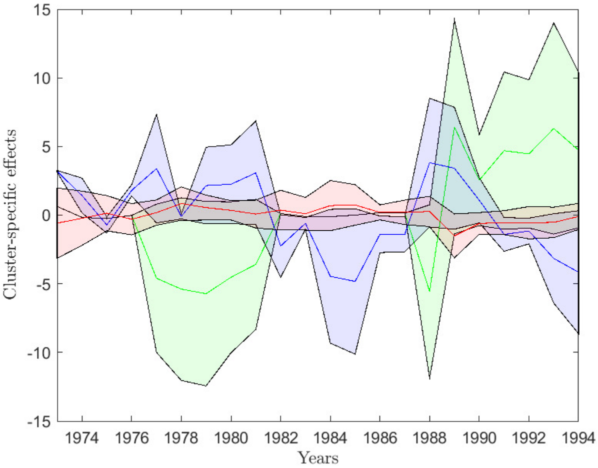
Exercise #1: ABBGH's Residuals



Exercise #1: TPWD Regularization Path



Exercise #1: Three Clusters



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

2 clusters 4 clusters

Exercise #2: Fitting a NGFE Poisson model

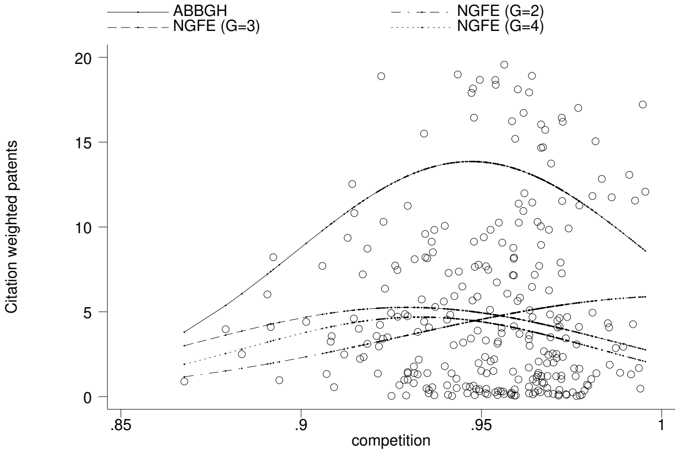
- Suppose we do not trust ABBGH's $\widehat{\beta}$ anymore.
- Allow for unobserved clusters of time-varying heterogeneity, estimate:

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

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

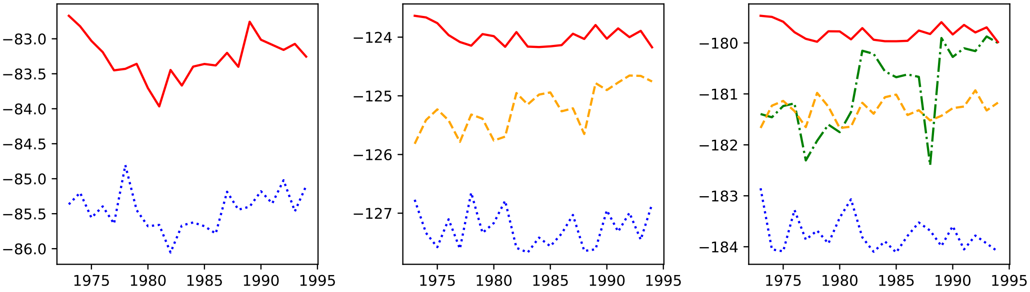
- We obtain a *mildly* inverted-U. ▸ Table

Excercise #2: A mildly Inverted-U



Exercise #2: Unobserved Clustered Dynamics of UH

Figure: Time Effects

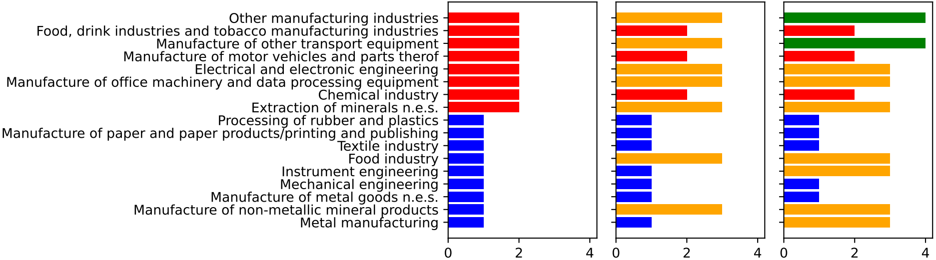


Notes: Solid line=High-Innovation, dotted line=Low-Innovation, dashed line=Steady-Catchers, dashdotted line=Noisy-Catchers.

▸ Visiualize Clusters

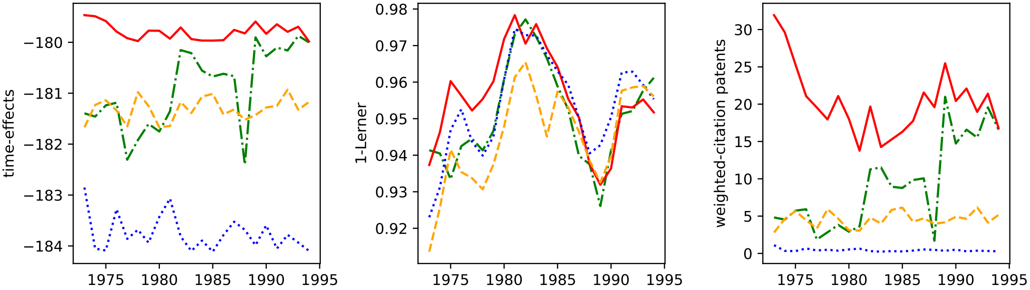
Exercise #2: Data-driven Industries Clustering

Go back



Notes: High-Innovation, Low-Innovation, Steady-Catchers, Noisy-Catchers. From left to right: $G^0 = 2, 3, 4$.

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



Notes: Solid line=High-Innovation, dotted line=Low-Innovation, dashed line=Steady-Catchers, dashdotted line=Noisy-Catchers.

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: The Effect of Competition on Innovation (Control Function Approach)

	TWFE Poisson			NGFE Poisson		
Dependent Variable: Citation-weighted patents$_{it}$	Annual	Before 1983	After 1983	Annual	Before 1983	After 1983
Competition$_{it}$	386.59***	229.18*	113.42	394.23***	265.86***	9.69
	(67.61)	(122.68)	(100.73)	(77.10)	(128.18)	(124.73)
Competition squared$_{it}$	-205.32***	-114.89*	-60.85	-212.35***	-144.18***	-9.41
	(36.11)	(66.49)	(53.37)	(41.14)	(67.95)	(67.46)
Relationship	steep inv-U	increasing		mildly inv-U	mildly inv-U	

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

Outline

Conclusion

- 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... Extension to heterogeneous slope, multinomial choice in the paper.
  - Nonparametric estimation of the link function is theoretically justified.
- Propose and study **semiparametric NGFE estimators**.
  - Consistent & asymptotically normal under regularity conditions.
  - Accurate large-N, T inference in small sample.
- Revisit ABBGH's "inverted-U" relationship between **innovation and competition**.
  - Evidence of **significant unobserved time-varying heterogeneity**, **mildly** inverted-U curve.
- Future research:
  - Estimation of h^0, G^0.
  - Reducing computational burden.
  - Robustness to increasing sample size in empirical application (e.g., using U.S. data).

Thank you!

martin.mugnier@ensae.fr

Examples ▸ Go Back

- **Binary outcome**: $Y_{it} = \mathbf{1}\left\{X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} \geq 0\right\}$, 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,

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

- **Ordered outcome**:

$$Y_{it} = \begin{cases} 0 & \text{if } X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} < d_1^0. \\ 1 & \text{if } d_1^0 \leq X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} < d_2^0. \\ 2 & \text{if } X_{it}'\beta^0 + \alpha^0_{g_i^0 t} - \varepsilon_{it} \geq d_2^0, \end{cases} \tag{3}$$

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

Estimation: Computation ▸ Go Back

An Iterative (Heuristic) Algorithm

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

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

3. Compute:

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

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

Go Back

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

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

$\Longrightarrow$ MLE with known group dummies.

Assumption (A.N. 1)

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

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

(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

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!

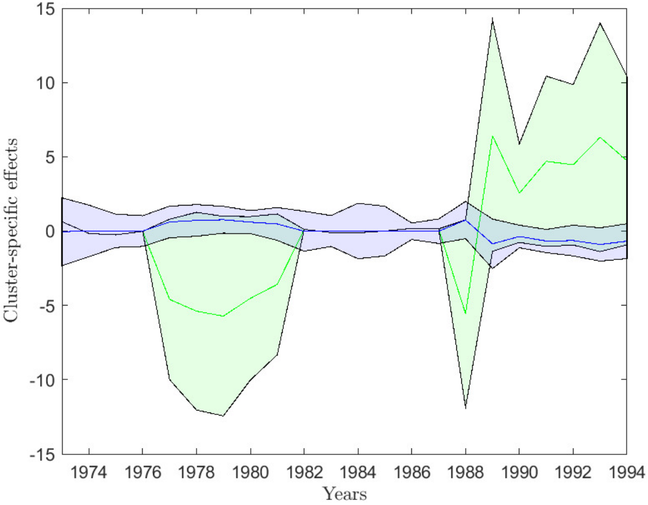
Industries at the 2 Digits

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

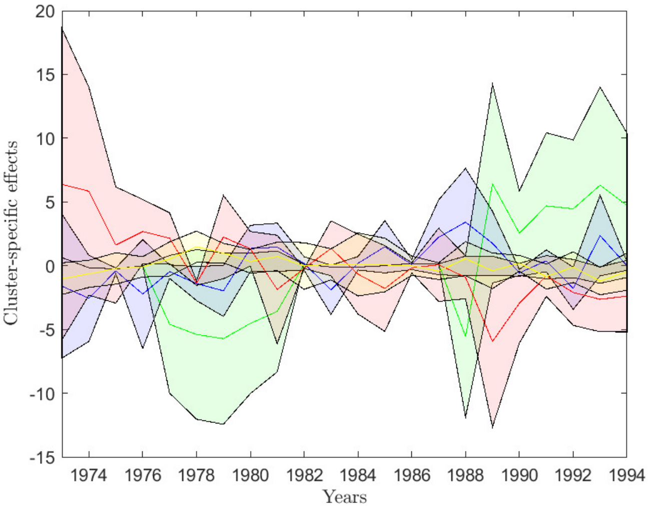
Two Clusters

Go Back



Four Clusters

Go Back



Mildly Inverted-U

Go Back

Table: THE EFFECT OF COMPETITION ON INNOVATION

	TWFE Poisson		NGFE Poisson		
Dependent variable: Citation-weighted patents$_{it}$	(1)	(2)	(3)	(4)	(5)
Competition$_{it}$	152.80***	387.46***	171.28***	273.62***	392.23***
	(55.74)	(67.74)	(71.51)	(70.21)	(70.35)
Competition squared$_{it}$	-80.99***	-204.55***	-85.15***	-147.21***	-210.19***
	(29.61)	(36.17)	(38.18)	(37.62)	(37.73)
Year effects	Yes	Yes			
Industry effects		Yes			
Time-varying clustered effects			Yes	Yes	Yes
Number of clusters			2	3	4

Notes: Analytical standard errors are under parentheses. The sample includes 354 observations from an unbalanced panel of 17 industries over the period 1973-1994. Competition$_{it}$ is measured by (1-Lerner index)$_{it}$ in the industry-year. NGFE estimates are computed using Lloyd's algorithm with 2,000 random initializers. ***, **, * denote statistical significance at 1, 5, and 10% respectively.

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:

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

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

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

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).