Supplement to “A Simple and Computationally Trivial Estimator for Grouped Fixed Effects Models”

Martin Mugnier*

Abstract

This supplement contains additional material to accompany the main text. First, I show that asymptotic results can be extended to allow for an increasing number of groups. Second, I discuss how the overall estimation strategy allows for individual-specific effects and extends to a class of nonlinear multiplicative models, e.g., for trade gravity equations. I report additional Monte Carlo results for variations of the model considered in the paper, including unbalanced groups, dependent errors, lagged outcomes, unit fixed effects, and lower signal-to-noise ratio. Finally, I provide some visualisation of the regularization path by estimating a pure grouped fixed effects model with time-constant unobserved heterogeneity using Acemoglu, Johnson, Robinson, and Yared (2008)'s panel data set of countries.

*Department of Economics, University of Oxford, martin.mugnier@economics.ox.ac.uk

S1 Large number of groups

Consider a simpler version of the model without covariates ($\beta^0 = 0$ is known). Proposition S1.1 shows that uniformly consistent classification of cross-sectional units remains possible when there are almost as many groups as individuals asymptotically.

Proposition S1.1 *Suppose $\beta^0 = 0$, Assumptions 1-3 hold, $G^0 \equiv G^0_{NT} \leq N/2$, and $\widehat{\beta}^1 = 0$. Then, as N and T tend to infinity,*

$$\left\| \widehat{W} - W^0 \right\|_{\max} = o_p(1).$$

Proof of Proposition S1.1. Similar arguments as in the proof of Proposition 3.1 show that the probability of false-positive for the test $H_{0,ij} : g_i^0 = g_j^0$ based on the statistics $1 - \widehat{W}_{ij} = \mathbf{1}\left\{ \widehat{d}^2_\infty(i,j) > c_{NT} \right\}$ is bounded asymptotically by $o_p(NT^{-\delta})$ uniformly across pairs of units $(i,j) \in \{1,...,N\}^2$ for all $\delta > 0$. Similarly, outside of an event whose probability tends to zero (when one group has less than two units), the probability of false-negative can be shown to be bounded asymptotically as $\min(N,T) \to \infty$ by

$$G^0_{NT}(G^0_{NT} - 1)o_p(T^{-\delta}) = o_p(N^2 T^{-\delta}), \quad \forall \delta > 0.$$

$\square$

The restriction $G^0_{NT} \leq N/2$ is necessary for Assumption 3(e) to hold. The maximal possible rate of divergence of G^0_{NT} depends on the extent to which the bounded support of α_{gt}, $\mathcal{A} \equiv \mathcal{A}_{NT}$, can enlarge as $\min(N,T) \to \infty$ while meeting the limiting group separation restriction imposed by Assumption 3(c).
A similar result can be obtained for the full model where $\beta \neq 0$, but conditions for preliminary consistent estimation may differ (e.g., Beyhum and Gautier, 2023).

S2 Extension of the baseline model

S2.1 Nonlinear multiplicative models for networks

Recall model (B.1) in the paper.

S2.1.1 Identification of β_0 with serial correlation in shocks

In this section, I show that β_0 can be identified by deriving a conditional moment restriction in the spirit of Jochmans (2017). I start by introducing a sampling assumption. For any $S \subset \mathbb{N}^{*2}$, let $\Pi_1 S \equiv \{i \in \mathbb{N}^* : \exists j \in \mathbb{N}^*, (i,j) \in S\}$ and $\Pi_2 S \equiv \{j \in \mathbb{N}^* : \exists i \in \mathbb{N}^*, (i,j) \in S\}$.

Assumption S1 *For any two disjoint subsets $S_1, S_2 \subset \{1,\dots,n\}^2$, $(x_{ij}, \varepsilon_{ij})_{(i,j)\in S_1}$ is independent of $(x_{kl}, \varepsilon_{kl})_{(k,l)\in S_2}$ if and only if $(\Pi_1 S_1 \cup \Pi_2 S_1) \cap (\Pi_1 S_2 \cup \Pi_2 S_2) = \emptyset$.*

Assumption S1 allows for dependence between dyads that have observations in common, which is important in applications. A similar sampling assumption is made in Tabord-Meehan (2019) and Jochmans (2017). Under Assumption S1, (B.3) implies

$$\mathbb{E}[\varepsilon_{ij}\varepsilon_{kl} | x_{12}, \dots, x_{n(n-1)}] = 1, \tag{S2.1}$$

for any distinct indices i, j, k, l. By combining (B.2) and (S2.1), I further obtain

$$\begin{aligned}
\mathbb{E}[u_{ij}u_{kl} | x_{12}, \dots, x_{n(n-1)}] &= (\alpha_i \gamma_j \omega_{g_i,h_j})(\alpha_k \gamma_l \omega_{g_k,h_l}) = \alpha_i\alpha_k\gamma_j\gamma_l\omega_{g_i,h_j}\omega_{g_k,h_l}, \\
\mathbb{E}[u_{il}u_{kj} | x_{12}, \dots, x_{n(n-1)}] &= (\alpha_i \gamma_l \omega_{g_i,h_l})(\alpha_k \gamma_j \omega_{g_k,h_j}) = \alpha_i\alpha_k\gamma_j\gamma_l\omega_{g_i,h_l}\omega_{g_k,h_j}.
\end{aligned}$$

If $g_i = g_k$ or $h_j = h_l$, the difference of the above equations is zero. Hence, the following "infeasible" moment condition holds:

$$\mathbb{E}\left[\sum_{i<k} \mathbf{1}\{g_i = g_k\} \sum_{l<j:\{i,k\}\cap\{j,l\}=\emptyset} m_{ijkl}(\beta_0) + \sum_{j<l} \mathbf{1}\{h_j = h_l\} \sum_{k<i:\{i,k\}\cap\{j,l\}=\emptyset} m_{ijkl}(\beta_0)\right] = 0,$$

where $m_{ijkl}(\beta_0) \equiv \Phi(x_{ij}, x_{kl}, x_{il}, x_{kj})(u_{ij}u_{kl} - u_{il}u_{kj})$ and $\Phi(x_{ij}, x_{kl}, x_{il}, x_{kj})$ is a chosen vector function.

Let $\widehat{\phi}_{ik}(\beta) = \frac{1}{\binom{n-2}{2}} \sum_{l<j:\{i,k\}\cap\{j,l\}=\emptyset} \widehat{m}_{ijkl}(\beta)$, $\widehat{\psi}_{jl}(\beta) = \frac{1}{\binom{n-2}{2}} \sum_{k<i:\{i,k\}\cap\{j,l\}=\emptyset} \widehat{m}_{ijkl}(\beta)$, and $\widehat{m}_{ijkl}(\beta) \equiv \Phi(x_{ij}, x_{kl}, x_{il}, x_{kj})(\widehat{u}_{ij}\widehat{u}_{kl} - \widehat{u}_{il}\widehat{u}_{kj})$, where $\widehat{u}_{ij} \equiv y_{ij}/\varphi(x_{ij};\beta)$. Finally, let

$$\widehat{s}(\beta) = \binom{n}{2}^{-1} \sum_{i<k} \exp\left(-\frac{\widehat{\phi}_{ik}(\beta)^2}{\kappa_n}\right)\widehat{\phi}_{ik}(\beta) + \binom{n}{2}^{-1} \sum_{j<l} \exp\left(-\frac{\widehat{\psi}_{jl}(\beta)^2}{\kappa_n}\right)\widehat{\psi}_{jl}(\beta),$$

for some $\kappa_n \to 0$. A feasible GMM regularized estimator of β_0 is

$$\widehat{\beta} = \underset{\beta\in\mathcal{B}}{\arg\min}\, \widehat{s}(\beta)'\Sigma_n \widehat{s}(\beta),$$

where $\mathcal{B}$ is the parameter space and Σ_n is a chosen positive-definite weight matrix. The intuition behind this estimator is that $\widehat{\phi}_{ik}(\beta_0)^2$ will be strictly positive whenever $g_i \neq g_k$ provided h_n does not decrease too fast and, for all $g \neq g'$ and $h \neq h'$,

$$\det \begin{pmatrix} \omega_{g,h} & \omega_{g,h'} \\ \omega_{g',h} & \omega_{g',h'} \end{pmatrix} \neq 0.$$

I leave the formal statistical analysis of this estimator for further research.

S2.1.2 Identification of β_0 under limited serial correlation in shocks and symmetric groups

If the researcher is willing to impose a slight reinforcement of Assumption S1 and symmetrical group-specific effects, a much simpler GMM objective can be constructed.

Assumption S2 (4-cyclical exogeneity) $(\varepsilon_{ij}, \varepsilon_{jk}, \varepsilon_{kl}, \varepsilon_{li}) \mid x_{12}, \ldots, x_{n(n-1)}$ *are mean independent for all distinct* i, j, k, l.

Assumption S2 allows for arbitrary correlation between pair-specific idiosyncratic shocks $(\varepsilon_{ij}, \varepsilon_{ji})$ but rules out cyclic patterns, i.e., the possibility that shocks that idiosyncratically affect trade from j to i also affects trade from i to l, from l to k and from k to i. Permanent unconditional correlation patterns between tetrads i, j, k, l can be captured by the unobserved effects α_i, γ_j and ω_{g_i,g_j}. This is a rather standard assumption in matching models with transferable utility (see, e.g., Graham, 2017).

Assumption S3 $G_0 = H_0$ *,* $g_i = h_i$*, and* Ω *is symmetric.*

Assumption S1 and equation (B.3) imply that, for any distinct indices i, j, k, l (tetrad),

$$\begin{aligned}\mathbb{E}\left[\varepsilon_{ij}\varepsilon_{kl}\varepsilon_{li}\varepsilon_{jk} \mid x_{12}, ..., x_{n(n-1)}\right] &= \mathbb{E}\left[\varepsilon_{ij} \mid x_{12}, ..., x_{n(n-1)}\right] \times \mathbb{E}\left[\varepsilon_{kl} \mid x_{12}, ..., x_{n(n-1)}\right] \\ &\quad \times \mathbb{E}\left[\varepsilon_{li} \mid x_{12}, ..., x_{n(n-1)}\right] \times \mathbb{E}\left[\varepsilon_{jk} \mid x_{12}, ..., x_{n(n-1)}\right] \\ &= 1. \end{aligned} \tag{S2.2}$$

Combining (B.2) and (S2.1) yields

$$\begin{aligned}\mathbb{E}\left[u_{ij}u_{kl}u_{li}u_{jk} \mid x_{12}, ..., x_{n(n-1)}\right] &= (\alpha_i\gamma_j\omega_{g_i,g_j})(\alpha_k\gamma_l\omega_{g_k,g_l})(\alpha_l\gamma_i\omega_{g_l,g_i})(\alpha_j\gamma_k\omega_{g_j,g_k}), \\ &= \alpha_i\alpha_k\alpha_j\alpha_l\gamma_i\gamma_k\gamma_j\gamma_l\omega_{g_i,g_j}\omega_{g_k,g_l}\omega_{g_l,g_i}\omega_{g_j,g_k},\end{aligned}$$

and

$$\mathbb{E}\left[u_{ji}u_{lk}u_{il}u_{kj} \mid x_{12},...,x_{n(n-1)}\right] = (\alpha_j\gamma_i\omega_{g_j,g_i})(\alpha_l\gamma_k\omega_{g_l,g_k})(\alpha_i\gamma_l\omega_{g_i,g_l})(\alpha_k\gamma_j\omega_{g_k,g_j}),$$
$$= \alpha_i\alpha_k\alpha_j\alpha_l\gamma_i\gamma_k\gamma_j\gamma_l\omega_{g_j,g_i}\omega_{g_l,g_k}\omega_{g_i,g_l}\omega_{g_k,g_j},$$

for any tetrad i, j, k, l. By differencing these equations and using Assumption S3, I obtain the conditional moment condition:

$$\mathbb{E}\left[u_{ij}u_{kl}u_{li}u_{jk} - u_{ji}u_{lk}u_{il}u_{kj} \mid x_{12},...,x_{n(n-1)}\right] = 0, \tag{S2.3}$$

which does not involve any nuisance parameters and holds for all

$$\mathcal{T}_n := \binom{n}{4} = \frac{n(n-1)(n-2)(n-3)}{24},$$

distinct tetrads i, j, k, l.

A n-asymptotically normal GMM estimator $\widehat{\beta}$ for β_0 can be constructed by adapting the analysis in Jochmans (2017). Specifically, Equation (S2.3) implies that the unconditional moment condition

$$\mathbb{E}\left[g(x_{ij}, x_{kl}, x_{li}, x_{jk}, x_{ji}, x_{lk}, x_{il}, x_{kj}; \beta_0)(u_{ij}u_{kl}u_{li}u_{jk} - u_{ji}u_{lk}u_{il}u_{kj})\right] = 0, \tag{S2.4}$$

where g is a chosen (vector) function, holds for all $\mathcal{T}_n$ choices of i, j, k, l. An intuitive way of obtaining an estimating equation for β_0, then, is to work with the empirical counterpart of the average of equation (S2.4) over all $\mathcal{T}_n$ choices. By letting $u_{ij}(\beta) = y_{ij}/\varphi(x_{ij};\beta)$, this empirical moment at a given value β is the U-statistic

$$Q_n(\beta) = \frac{1}{\mathcal{T}_n}\sum_{1\le i<j<k<l\le n} g(x_{ij}, x_{kl}, x_{li}, x_{jk}, x_{ji}, x_{lk}, x_{il}, x_{kj}; \beta)$$
$$\times\,(u_{ij}(\beta)u_{kl}(\beta)u_{li}(\beta)u_{jk}(\beta) - u_{ji}(\beta)u_{lk}(\beta)u_{il}(\beta)u_{kj}(\beta)),$$

where, without loss of generality, I have assumed that the kernel function,

$$g(x_{ij}, x_{kl}, x_{li}, x_{jk}, x_{ji}, x_{lk}, x_{il}, x_{kj}; \beta)(u_{ij}(\beta)u_{kl}(\beta)u_{li}(\beta)u_{jk}(\beta) - u_{ji}(\beta)u_{lk}(\beta)u_{il}(\beta)u_{kj}(\beta)),$$

is permutation invariant in both (i, k) and (j, l). A GMM estimator of β_0 is

$$\widehat{\beta} = \arg\min_{\beta\in\mathcal{B}} Q_n(\beta)'\Sigma_n Q_n(\beta),$$

where $\mathcal{B}$ is the parameter space searched over and Σ_n is a chosen positive-definite weight matrix.

S2.1.3 Identification of group memberships

Given the identification of β_0 and by dividing equation (B.1) by $\varphi(x_{ij};\beta_0)$, the analysis can be restricted to the "pure" gravity model without covariates ($\varphi = 1$):[1]

$$y_{ij} = \alpha_i \gamma_j \omega_{g_i,g_j} \varepsilon_{ij}, \quad i \neq j. \tag{S2.5}$$

Assume that $n \geq 4$ and define, for all distinct indices $(i,j,k,l) \in \{1,...,n\}^4$,

$$\begin{aligned}
\bar{y}_{i.} &\equiv \frac{1}{N-1}\sum_{\ell\neq i} y_{i\ell} = \alpha_i \frac{1}{N-1}\sum_{\ell\neq i}\gamma_\ell \omega_{g_i,g_\ell}\varepsilon_{i\ell} =: \alpha_i \bar{\gamma}_{-i},\\
\bar{y}_{.j} &\equiv \frac{1}{N-1}\sum_{\ell\neq j} y_{\ell j} = \gamma_j \frac{1}{N-1}\sum_{\ell\neq j}\alpha_\ell \omega_{g_\ell,g_j}\varepsilon_{\ell j} =: \gamma_j \bar{\alpha}_{-j},\\
\tilde{y}_{(i,j)} &\equiv \frac{y_{ij}y_{ji}}{\bar{y}_{i.}\bar{y}_{.j}\bar{y}_{j.}\bar{y}_{.i}} = \frac{\omega^2_{g_i,g_j}\varepsilon_{ij}\varepsilon_{ji}}{\bar{\alpha}_{-i}\bar{\alpha}_{-j}\bar{\gamma}_{-i}\bar{\gamma}_{-j}},\\
S_n(i,j,k,l) &\equiv \frac{1}{N-4}\sum_{s\notin\{i,j,k,l\}}(\tilde{y}_{(i,s)} - \tilde{y}_{(j,s)})(\tilde{y}_{(l,s)} - \tilde{y}_{(k,s)}),
\end{aligned}$$

where $\tilde{y}_{(i,j)}$ and S_n are set to 0 if undefined (if a country does not import or export to any other country).[2] If $g_i = g_j = \bar{g}$, then

$$\begin{aligned}
&S_n(i,j,k,l)\\
&= \frac{1}{(N-4)\bar{\alpha}_{-i}\bar{\gamma}_{-i}\bar{\alpha}_{-j}\bar{\gamma}_{-j}\alpha_{-l}\bar{\gamma}_{-l}\bar{\alpha}_{-k}\bar{\gamma}_{-k}}\times\\
&\quad\sum_{s\notin\{i,j,k,l\}}\Bigg[\frac{\omega^2_{\bar{g},g_s}}{\bar{\alpha}_{-s}\bar{\gamma}_{-s}}(\bar{\alpha}_{-j}\bar{\gamma}_{-j}\varepsilon_{is}\varepsilon_{si} - \bar{\alpha}_{-i}\bar{\gamma}_{-i}\varepsilon_{js}\varepsilon_{sj})\\
&\qquad\qquad \times(\omega^2_{g_l,g_s}\bar{\alpha}_{-k}\bar{\gamma}_{-k}\varepsilon_{ls}\varepsilon_{sl} - \omega^2_{g_k,g_s}\bar{\alpha}_{-l}\bar{\gamma}_{-l}\varepsilon_{ks}\varepsilon_{sk})\Bigg],
\end{aligned}$$

converges in probability to 0 (exponentially fast uniformly over i,j,k,l under appropriate regularity conditions) under the mild conditions that for all i,j, $\bar{\alpha}_{-i}, \bar{\gamma}_{-j}$ are bounded below and

$$\begin{aligned}
\bar{\alpha}_i &\to \bar{\alpha}_{g_i} > 0,\\
\bar{\gamma}_j &\to \bar{\gamma}_{g_j} > 0.
\end{aligned}$$

[1] In a more complete treatment left for further research, one should of course control estimation errors.

[2] The reasoning still work for the alternative definitions $\tilde{y}_{(i,j)} := y_{ij}/(\bar{y}_{i.}\bar{y}_{.j})$ or $\tilde{y}_{(i,j)} := y_{ji}/(\bar{y}_{j.}\bar{y}_{.i})$.

Now, if $g_i \neq g_j$, choose k^*, l^* such that $g_{k^*} = g_i = \bar{g}$ and $g_{l^*} = g_j = \check{g}$. This is possible under the assumption that each group has at least two members. Then, for n sufficiently large,

$$\begin{aligned}
&\max_{(k,l)} |S_n(i,j,k,l)| \\
&\quad \geq |S_n(i,j,k^*,l^*)| \\
&\quad = \frac{1}{\bar{\alpha}_{-i}\bar{\gamma}_{-i}\bar{\alpha}_{-j}\bar{\gamma}_{-j}\alpha_{-l^*}\bar{\gamma}_{-l^*}\bar{\alpha}_{-k^*}\bar{\gamma}_{-k^*}} \times \\
&\qquad \Bigg| \frac{1}{(N-4)} \sum_{s \notin \{i,j,k^*,l^*\}} \frac{1}{\bar{\alpha}_{-s}\bar{\gamma}_{-s}} \left(\bar{\alpha}_{-j}\bar{\gamma}_{-j}\omega^2_{\bar{g},g_s}\varepsilon_{is}\varepsilon_{si} - \bar{\alpha}_{-i}\bar{\gamma}_{-i}\omega^2_{\check{g},g_s}\varepsilon_{js}\varepsilon_{sj} \right) \times \\
&\qquad\quad \left(\bar{\alpha}_{-k^*}\bar{\gamma}_{-k^*}\omega^2_{\bar{g},g_s}\varepsilon_{l^*s}\varepsilon_{sl^*} - \bar{\alpha}_{-l^*}\bar{\gamma}_{-l^*}\omega^2_{\check{g},g_s}\varepsilon_{k^*s}\varepsilon_{sk^*} \right) \Bigg|,
\end{aligned}$$

which converges towards $\rho_{\bar{g},\check{g}} \geq \rho > 0$ uniformly over $i, j, \bar{g}$, and $\check{g}$ under weak group-separation conditions similar to Assumption 3(c) and non negligible groups: for all $g \in \{1, \ldots, G_0\}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{g_i = g\} = \pi_g > 0.$$

Hence, pairwise distance can be constructed and an agglomerative clustering algorithm can be applied to the dissimilarity matrix.

S2.1.4 Identification of α, γ, and Ω

Given identification of the g_i's, for any pair (i, i'), under regularity conditions and symmetry,

$$\frac{\sum_{j=1}^n \mathbf{1}\{g_j = g_{i'}\} y_{ij}}{\sum_{j=1}^n \mathbf{1}\{g_j = g_i\} y_{i'j}} \xrightarrow{p} \frac{\alpha_i \omega_{g_i, g_{i'}}}{\alpha_{i'} \omega_{g_{i'}, g_i}} = \frac{\alpha_i}{\alpha_{i'}}.$$

Similarly, for any pair (j, j'),

$$\frac{\sum_{i=1}^n \mathbf{1}\{g_i = g_{j'}\} y_{ij}}{\sum_{i=1}^n \mathbf{1}\{g_i = g_j\} y_{ij'}} \xrightarrow{p} \frac{\gamma_j \omega_{g_{j'}, g_j}}{\gamma_{j'} \omega_{g_j, g_{j'}}} = \frac{\gamma_j}{\gamma_{j'}}.$$

Taking the product of each equality over i', j' and using the normalization (B.4) yields

$$\begin{aligned}
\prod_{\ell=1}^{n} \frac{\sum_{j=1}^n \mathbf{1}\{g_j = g_\ell\} y_{ij}}{\sum_{j=1}^n \mathbf{1}\{g_j = g_i\} y_{\ell j}} &\xrightarrow{p} \alpha_i, \quad \forall i = 1, ..., n, \\
\prod_{\ell=1}^{n} \frac{\sum_{i=1}^n \mathbf{1}\{g_i = g_\ell\} y_{ij}}{\sum_{i=1}^n \mathbf{1}\{g_i = g_j\} y_{i\ell}} &\xrightarrow{p} \gamma_j, \quad \forall j = 1, ..., n.
\end{aligned}$$

Since the left-hand side of each equation is observed, α_i and γ_j are identified. Finally, for all $(g, g') \in \{1, ..., G^0\}$,

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}\{g_i = g\}\mathbf{1}\{g_j = g'\}(y_{ij}/\alpha_i\gamma_j + y_{ji}/\alpha_j\gamma_i)/2}{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{1}\{g_i = g\}\mathbf{1}\{g_j = g'\}} \xrightarrow{p} \omega_{g,g'},$$

which implies that Ω is identified. Other nonlinear models are considered in Mugnier (2022), and I leave the extension to three-way models for further research.

S2.2 Individual-specific unobserved heterogeneity

Model (1.1) can be augmented with a unit-specific ν_i^0:

$$y_{it} = x_{it}'\theta^0 + \nu_i^0 + \alpha^0_{g_i^0 t} + v_{it}. \tag{S2.6}$$

Without covariates, the two-step estimator can still be applied to a within differenced version of equation (S2.6). With covariates, Moon and Weidner (2019) still apply under strong factors and the demeaned first-step residuals are used to estimate the groups. The projection step is performed with demeaned outcomes and covariates. From a theoretical standpoint, one needs to redefine separation assumptions as in Bonhomme and Manresa (2015)'s Supplemental Material.

S2.3 Computational gains in the time-invariant model

If the clustered unobserved heterogeneity is time-invariant, one can relax Assumption 2(b) to Bonhomme and Manresa (2015)'s Assumption 2(e). Also, it is recommended to use a simple within, first-difference, or instrumental differencing estimator as a first step in this case (see, e.g., Arellano and Bonhomme, 2011; Wooldridge, 2010). The method becomes $O(N^2T)$ without changing the asymptotic properties by considering the new distance

$$\widehat{d}^2_{\infty,\text{time-inv}}(i, j) \equiv \left(\frac{1}{T}\sum_{t=1}^{T}(\widehat{v}_{it} - \widehat{v}_{jt})\right)^2. \tag{S2.7}$$

I call the estimator based upon $\widehat{d}^2_{\infty,\text{time-inv}}(i, j)$ the pairwise-distance (PWD) estimator. Its asymptotic properties are studied in detail in the previous arXiv version of this paper.

S3 Additional Monte Carlo simulations

In this section, I report the results of additional Monte Carlo simulations. In Sections S3.1-S3.3, I consider the same pure GFE model as in the paper but varying one dimension: unbalanced groups, dependent errors, and signal-to-noise ratio. I replicate Tables 1 and 2 in each case. In Sections S3.5 and S3.4, I consider full GFE models similar to that of the paper but introduce unit-specific effects or a lagged outcome. I replicate Tables 3 and 4 in each case.

S3.1 Unbalanced groups

For $G = 3$, let

$$g_i = 1 + \mathbf{1}\{i > 2\} + \mathbf{1}\{i > N/10\}, \quad i = 1, \dots, N,$$

For $G = 4$, let

$$g_i = 1 + \mathbf{1}\{i > 2\} + \mathbf{1}\{i > N/10\} + \mathbf{1}\{i > N/2\}, \quad i = 1, \dots, N.$$

In each setting, Group 1 has only two units. Tables S1 and S2 suggest that results are almost unaffected in comparison to the balanced case.

Table S1: Estimation of the grouped fixed effects (unbalanced groups)

				TPWD		$\text{GFE}^{G=2}$	$\text{GFE}^{G=3}$	$\text{GFE}^{G=10}$
G	N	T	$\widehat{G}$	RMSE	CPU time	RMSE	RMSE	RMSE
3	90	7	4.822	0.124	0.076	0.122	0.159	0.233
		10	4.164	0.097	0.101	0.100	0.138	0.210
		20	3.448	0.069	0.209	0.094	0.118	0.176
		40	3.062	0.061	0.422	0.093	0.107	0.154
3	180	7	5.818	0.116	0.694	0.095	0.140	0.220
		10	4.688	0.084	0.934	0.075	0.119	0.194
		20	3.756	0.054	1.693	0.068	0.098	0.156
		40	3.178	0.044	3.281	0.068	0.087	0.130
4	90	7	4.1700	0.1680	0.0830	0.1500	0.1630	0.2300
		10	3.7120	0.1770	0.1100	0.1540	0.1530	0.2080
		20	3.5900	0.1790	0.2080	0.1510	0.1280	0.1730
		40	3.8360	0.1610	0.4230	0.1530	0.1160	0.1500
4	180	7	4.9600	0.1550	0.6930	0.1230	0.1470	0.2190
		10	4.0940	0.1620	0.9310	0.1230	0.1330	0.1920
		20	3.7840	0.1600	1.6950	0.1180	0.1030	0.1500
		40	3.9220	0.1440	3.2670	0.1200	0.0890	0.1210

Notes: This table reports the estimated number of group ($\widehat{G}$), the root mean square errors (RMSE) and the execution time in seconds (CPU time) for the triad pairwise distance estimator (TPWD) with $\widehat{\beta}^1 = 0$ and $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$. It reports the RMSE for a Lloyd's approximation to the grouped fixed effects estimator with user-specified number of groups $g \in \{2, 3, 10\}$ and 10,000 initializers ($\text{GFE}^{G=g}$). Results averaged across 500 Monte Carlo samples.

Table S2: Classification accuracy (unbalanced groups)

G	N	T	TPWD P	TPWD R	TPWD RI	GFE$^{G=2}$ P	GFE$^{G=2}$ R	GFE$^{G=2}$ RI	GFE$^{G=3}$ P	GFE$^{G=3}$ R	GFE$^{G=3}$ RI	GFE$^{G=10}$ P	GFE$^{G=10}$ R	GFE$^{G=10}$ RI
3	90	7	0.990	0.796	0.827	0.983	0.945	0.944	0.980	0.540	0.617	0.977	0.149	0.304
		10	0.995	0.897	0.913	0.991	0.989	0.984	0.988	0.562	0.638	0.986	0.157	0.312
		20	1.000	0.977	0.981	0.992	0.994	0.989	0.992	0.586	0.659	0.991	0.176	0.328
		40	1.000	0.998	0.999	0.992	0.999	0.992	0.993	0.627	0.693	0.991	0.191	0.340
3	180	7	0.993	0.771	0.809	0.992	0.970	0.970	0.988	0.517	0.600	0.980	0.140	0.295
		10	0.996	0.882	0.901	0.996	0.995	0.993	0.993	0.527	0.611	0.988	0.144	0.299
		20	0.999	0.972	0.977	0.997	1.000	0.998	0.996	0.547	0.628	0.992	0.155	0.308
		40	1.000	0.997	0.998	0.997	1.000	0.997	0.996	0.565	0.643	0.993	0.177	0.327
4	90	7	0.7120	0.8100	0.7840	0.7570	0.9510	0.8530	0.7980	0.7520	0.8190	0.8120	0.2480	0.6680
		10	0.6530	0.8110	0.7360	0.7560	0.9640	0.8500	0.8230	0.7970	0.8450	0.8520	0.2720	0.6820
		20	0.6090	0.8130	0.6980	0.7650	0.9940	0.8600	0.8710	0.8890	0.8990	0.9110	0.3160	0.7070
		40	0.6650	0.8290	0.7480	0.7580	0.9990	0.8520	0.8900	0.9430	0.9270	0.9470	0.3480	0.7250
4	180	7	0.7250	0.8060	0.7910	0.7690	0.9560	0.8620	0.8000	0.7170	0.8070	0.8100	0.2280	0.6580
		10	0.6740	0.8250	0.7550	0.7780	0.9700	0.8710	0.8290	0.7740	0.8380	0.8610	0.2520	0.6730
		20	0.6510	0.8260	0.7350	0.7940	0.9950	0.8880	0.8960	0.8960	0.9120	0.9370	0.3050	0.7030
		40	0.6990	0.8440	0.7780	0.7910	1.0000	0.8850	0.9150	0.9510	0.9410	0.9790	0.3530	0.7290

Notes: This table reports the precision rate (P), recall rate (R), and Rand Index (RI) for the triad pairwise distance estimator (TPWD) with $\widehat{\beta}^1 = 0$ and $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$, as well as for a Lloyd's approximation to the grouped fixed effects estimator with user-specified number of groups $g \in \{2, 3, 10\}$ and 10,000 initializers (GFE$^{G=g}$). Results are averaged across 500 Monte Carlo samples.

S3.2 Dependent errors

Consider an AR(1) model for the error process:

$$v_{it} = 0.2v_{it-1} + \varepsilon_{it}, \quad \mathbb{E}[\varepsilon_{it}] = 0, \quad \mathbb{E}[\varepsilon_{it}^2] = 1/9.$$

The variance is set to maintain a signal-to-noise ratio of 1. Tables S3 and S4 show that results are almost unaffected in comparison to the i.i.d. case.

Table S3: Estimation of the grouped fixed effects (dependent errors)

				TPWD		$\text{GFE}^{G=2}$	$\text{GFE}^{G=3}$	$\text{GFE}^{G=10}$
G	N	T	$\widehat{G}$	RMSE	CPU time	RMSE	RMSE	RMSE
3	90	7	5.2340	0.1920	0.0810	0.2800	0.1260	0.2250
		10	4.4600	0.1600	0.1090	0.2720	0.1010	0.1970
		20	3.4720	0.0970	0.2090	0.2550	0.0820	0.1580
		40	3.0660	0.0650	0.4190	0.2460	0.0880	0.1390
3	180	7	6.3760	0.1840	0.6760	0.2800	0.1130	0.2110
		10	5.1760	0.1440	0.8930	0.2710	0.0790	0.1780
		20	3.8920	0.0830	1.6340	0.2530	0.0560	0.1360
		40	3.1840	0.0470	3.2110	0.2420	0.0750	0.1120
4	90	7	5.5280	0.2060	0.0840	0.2760	0.1540	0.2320
		10	4.5540	0.1950	0.1090	0.2590	0.1500	0.2090
		20	3.8660	0.1720	0.2100	0.2410	0.1320	0.1690
		40	3.7500	0.1560	0.4250	0.2320	0.1270	0.1430
4	180	7	6.6840	0.1770	0.6830	0.2200	0.1640	0.2240
		10	5.2200	0.1580	0.9010	0.1940	0.1470	0.1990
		20	4.2500	0.1380	1.6580	0.1860	0.1160	0.1560
		40	4.0520	0.1210	3.1920	0.1900	0.1100	0.1210

Notes: This table reports the estimated number of group ($\widehat{G}$), the root mean square errors (RMSE) and the execution time in seconds (CPU time) for the triad pairwise distance estimator (TPWD) with $\widehat{\beta}^1 = 0$ and $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$. It reports the RMSE for a Lloyd's approximation to the grouped fixed effects estimator with user-specified number of groups $g \in \{2, 3, 10\}$ and 10,000 initializers ($\text{GFE}^{G=g}$). Results averaged across 500 Monte Carlo samples.

Table S4: Classification accuracy (dependent errors)

G	N	T	TPWD P	TPWD R	TPWD RI	GFE$^{G=2}$ P	GFE$^{G=2}$ R	GFE$^{G=2}$ RI	GFE$^{G=3}$ P	GFE$^{G=3}$ R	GFE$^{G=3}$ RI	GFE$^{G=10}$ P	GFE$^{G=10}$ R	GFE$^{G=10}$ RI
3	90	7	0.8300	0.7230	0.8620	0.5600	0.8820	0.7360	0.9250	0.9330	0.9520	0.9240	0.3430	0.7770
		10	0.8770	0.8200	0.9040	0.5650	0.8990	0.7420	0.9490	0.9630	0.9680	0.9610	0.3730	0.7910
		20	0.9630	0.9520	0.9720	0.5790	0.9520	0.7590	0.9560	0.9830	0.9730	0.9960	0.4140	0.8090
		40	0.9980	0.9960	0.9980	0.5880	0.9860	0.7710	0.9380	0.9800	0.9620	1.0000	0.4310	0.8150
3	180	7	0.8500	0.6970	0.8600	0.5620	0.8740	0.7340	0.9380	0.9390	0.9600	0.9270	0.3290	0.7700
		10	0.9030	0.8230	0.9130	0.5680	0.8920	0.7410	0.9720	0.9730	0.9810	0.9680	0.3590	0.7850
		20	0.9700	0.9490	0.9730	0.5840	0.9530	0.7610	0.9790	0.9910	0.9870	0.9970	0.3990	0.8020
		40	0.9990	0.9960	0.9980	0.5940	0.9930	0.7740	0.9320	0.9770	0.9580	1.0000	0.4350	0.8140
4	90	7	0.539	0.692	0.782	0.381	0.895	0.622	0.601	0.904	0.829	0.695	0.341	0.804
		10	0.545	0.760	0.787	0.387	0.924	0.626	0.610	0.920	0.835	0.758	0.385	0.821
		20	0.573	0.829	0.806	0.389	0.961	0.624	0.622	0.966	0.844	0.864	0.456	0.851
		40	0.613	0.843	0.831	0.393	0.989	0.625	0.627	0.987	0.848	0.943	0.517	0.876
4	180	7	0.699	0.711	0.738	0.564	0.918	0.651	0.704	0.795	0.762	0.748	0.198	0.615
		10	0.708	0.829	0.774	0.577	0.985	0.673	0.710	0.834	0.777	0.799	0.215	0.628
		20	0.716	0.918	0.802	0.569	0.991	0.664	0.727	0.950	0.819	0.884	0.251	0.654
		40	0.753	0.951	0.840	0.565	0.994	0.658	0.727	0.978	0.826	0.966	0.304	0.687

Notes: This table reports the precision rate (P), recall rate (R), and Rand Index (RI) for the triad pairwise distance estimator (TPWD) with $\widehat{\beta}^1 = 0$ and $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$, as well as for a Lloyd's approximation to the grouped fixed effects estimator with user-specified number of groups $g \in \{2, 3, 10\}$ and 10,000 initializers (GFE$^{G=g}$). Results are averaged across 500 Monte Carlo samples.

S3.3 Lower signal-to-noise ratio

The standard deviation of the error term is set to 2/3, which corresponds to a signal-to-noise ratio of 0.5. Tables S5 and S6 show that the precision of TPWD and GFE deteriorates for small values of T. Both estimators remain comparable.

Table S5: Estimation of the grouped fixed effects (lower signal-to-noise ratio)

				TPWD		$\text{GFE}^{G=2}$	$\text{GFE}^{G=3}$	$\text{GFE}^{G=10}$
G	N	T	$\widehat{G}$	RMSE	CPU time	RMSE	RMSE	RMSE
3	90	7	22.382	0.508	0.091	0.331	0.337	0.482
		10	15.212	0.416	0.120	0.309	0.285	0.433
		20	6.830	0.263	0.219	0.289	0.200	0.348
		40	4.502	0.180	0.431	0.275	0.157	0.288
3	180	7	41.006	0.518	0.705	0.321	0.300	0.453
		10	28.068	0.423	0.935	0.300	0.236	0.395
		20	10.286	0.248	1.691	0.282	0.151	0.298
		40	5.638	0.151	3.206	0.266	0.113	0.228
4	90	7	22.788	0.515	0.090	0.348	0.352	0.487
		10	15.606	0.426	0.119	0.317	0.305	0.439
		20	6.950	0.284	0.220	0.280	0.236	0.362
		40	4.586	0.224	0.427	0.264	0.197	0.307
4	180	7	39.740	0.516	0.711	0.327	0.346	0.463
		10	27.678	0.422	0.934	0.287	0.296	0.410
		20	10.490	0.254	1.682	0.221	0.223	0.324
		40	5.292	0.179	3.211	0.196	0.182	0.262

Notes: This table reports the estimated number of group ($\widehat{G}$), the root mean square errors (RMSE) and the execution time in seconds (CPU time) for the triad pairwise distance estimator (TPWD) with $\widehat{\beta}^1 = 0$ and $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$. It reports the RMSE for a Lloyd's approximation to the grouped fixed effects estimator with user-specified number of groups $g \in \{2, 3, 10\}$ and 10,000 initializers ($\text{GFE}^{G=g}$). Results averaged across 500 Monte Carlo samples.

Table S6: Classification accuracy (lower signal-to-noise ratio)

G	N	T	TPWD			GFE$^{G=2}$			GFE$^{G=3}$			GFE$^{G=10}$		
			P	R	RI	P	R	RI	P	R	RI	P	R	RI
3	90	7	0.645	0.204	0.704	0.524	0.812	0.699	0.631	0.668	0.762	0.657	0.232	0.710
		10	0.693	0.333	0.734	0.541	0.838	0.716	0.702	0.735	0.809	0.715	0.266	0.726
		20	0.823	0.651	0.841	0.557	0.872	0.733	0.847	0.876	0.903	0.856	0.342	0.767
		40	0.940	0.880	0.943	0.571	0.921	0.749	0.915	0.953	0.947	0.954	0.408	0.801
3	180	7	0.656	0.145	0.693	0.530	0.810	0.701	0.682	0.695	0.792	0.676	0.226	0.709
		10	0.719	0.255	0.721	0.546	0.834	0.716	0.775	0.783	0.853	0.750	0.261	0.728
		20	0.848	0.592	0.831	0.560	0.865	0.731	0.910	0.917	0.942	0.889	0.345	0.770
		40	0.959	0.855	0.940	0.575	0.919	0.750	0.947	0.970	0.966	0.979	0.419	0.806
4	90	7	0.433	0.193	0.743	0.354	0.761	0.605	0.431	0.622	0.709	0.465	0.223	0.750
		10	0.461	0.322	0.743	0.372	0.817	0.621	0.471	0.684	0.735	0.514	0.257	0.761
		20	0.523	0.613	0.769	0.388	0.889	0.633	0.544	0.807	0.784	0.602	0.322	0.784
		40	0.572	0.776	0.803	0.392	0.930	0.632	0.587	0.894	0.815	0.689	0.387	0.809
4	180	7	0.621	0.144	0.582	0.501	0.607	0.560	0.549	0.437	0.593	0.584	0.145	0.576
		10	0.656	0.268	0.614	0.545	0.730	0.614	0.594	0.502	0.629	0.625	0.160	0.586
		20	0.693	0.614	0.709	0.588	0.934	0.681	0.647	0.616	0.682	0.691	0.190	0.604
		40	0.722	0.851	0.789	0.582	0.988	0.680	0.667	0.718	0.718	0.758	0.234	0.628

Notes: This table reports the precision rate (P), recall rate (R), and Rand Index (RI) for the triad pairwise distance estimator (TPWD) with $\widehat{\beta}^1 = 0$ and $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$, as well as for a Lloyd's approximation to the grouped fixed effects estimator with user-specified number of groups $g \in \{2, 3, 10\}$ and 10,000 initializers (GFE$^{G=g}$). Results are averaged across 500 Monte Carlo samples.

S3.4 Unit-specific effects

Consider the model

$$y_{it} = x_{it}\beta + \alpha_{g_i t} + \nu_i + v_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T,$$

where $\nu_i \overset{iid}{\sim} \mathcal{N}(0, 1)$, $x_{it} = 0.5\alpha_{g_i t} + 0.2\nu_i + u_{it}$, and the $(\nu_i)_i$ are mutually independent from the other variables. To ensure time-demeaned group separation, the time effects of Group 1 is now $\alpha_{1t} = 1 - \alpha_{2t}$. The clustering step of the TPWD estimator is applied to demeaned residuals, and the projection step to demeaned outcomes and covariates. Tables S7 and S8 show the results. The TPWD estimator has less bias than the NN-reg estimator, but the bias dominates the variance. Clustering consistency is much slower, which results in poor coverage of asymptotic confidence intervals.

Table S7: Estimation of the slope coefficient and grouped fixed effects (unit-specific FE)

			NN-reg		TPWD					Infeasible OLS			
G	N	T	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95	RMSE $\widehat{\alpha}_{gt}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95	RMSE $\widehat{\alpha}_{gt}$
3	90	7	0.450	0.453	0.197	0.217	0.102	0.220	2.650	0.001	0.002	0.973	0.055
		10	0.360	0.362	0.212	0.228	0.068	0.225	2.096	0.002	0.001	0.964	0.058
		20	0.244	0.247	0.210	0.219	0.010	0.228	1.638	-0.001	0.001	0.970	0.059
		40	0.189	0.190	0.195	0.203	0.004	0.219	1.670	0.001	0.000	0.966	0.060
3	180	7	0.442	0.453	0.174	0.187	0.030	0.210	3.334	0.000	0.001	0.977	0.040
		10	0.357	0.363	0.185	0.200	0.018	0.209	2.422	0.002	0.001	0.982	0.041
		20	0.232	0.245	0.196	0.207	0.002	0.217	1.796	-0.001	0.000	0.981	0.042
		40	0.171	0.187	0.177	0.187	0.000	0.206	1.710	0.000	0.000	0.966	0.042
4	90	7	0.471	0.474	0.192	0.215	0.164	0.216	2.564	0.001	0.002	0.970	0.064
		10	0.376	0.378	0.210	0.228	0.052	0.223	2.088	0.001	0.001	0.964	0.066
		20	0.255	0.257	0.218	0.228	0.010	0.232	1.788	-0.001	0.001	0.966	0.068
		40	0.197	0.198	0.209	0.217	0.000	0.228	1.758	0.001	0.000	0.965	0.069
4	180	7	0.457	0.468	0.086	0.100	0.384	0.151	2.824	0.000	0.001	0.973	0.046
		10	0.359	0.366	0.105	0.118	0.166	0.157	2.164	0.002	0.001	0.985	0.047
		20	0.230	0.243	0.114	0.127	0.034	0.165	1.970	-0.001	0.000	0.980	0.049
		40	0.170	0.184	0.109	0.118	0.000	0.161	1.954	0.000	0.000	0.969	0.049

Notes: This table reports results for the nuclear norm regularized (NN-reg) estimator with regularization parameter $\psi = \log(\log(T))\min(N,T)^{-1/2}/4$; triad pairwise differencing estimator (TPWD) obtained after four iterations starting at NN-reg and setting $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$; and infeasible pooled OLS regression with known group memberships (Infeasible OLS). .95 denotes coverage rate of a 95%-level confidence interval for β based on large N,T approximations and an estimator of the asymptotic variance clustered at the unit level. Results are averaged across 500 Monte Carlo samples.

Table S8: Classification accuracy of the TPWD estimator (unit-specific FE)

G	N	T	Precision	Recall	Rand Index
3	90	7	0.428	0.811	0.563
		10	0.396	0.881	0.491
		20	0.367	0.932	0.427
		40	0.372	0.925	0.440
3	180	7	0.446	0.784	0.591
		10	0.428	0.860	0.545
		20	0.390	0.928	0.465
		40	0.398	0.922	0.483
4	90	7	0.325	0.832	0.524
		10	0.304	0.877	0.460
		20	0.277	0.917	0.377
		40	0.275	0.927	0.366
4	180	7	0.583	0.939	0.672
		10	0.550	0.965	0.631
		20	0.521	0.978	0.586
		40	0.517	0.982	0.581

Notes: This table reports results for the triad pairwise differencing estimator (TPWD) obtained after four iterations starting at the nuclear norm regularized estimator $\widehat{\beta}^1(\psi)$ with $\psi = \log(\log(T)) \min(N,T)^{-1/2}/4$ and setting $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$. Results are averaged across 500 Monte Carlo samples.

S3.5 Lagged outcome

Consider the model

$$y_{it} = x_{it}\beta_1 + y_{it-1}\beta_2 + \alpha_{g_i t} + v_{it}, \quad i = 1, \dots, N,\ t = 1, \dots, T,$$

where $\beta_1 = 1$, $\beta_2 = 0.1$, $x_{it} = 0.2\alpha_{g_i t} + u_{it}$, and $y_{i0} \sim \mathcal{N}(\mathbf{1}\{g_i = 1\} - \mathbf{1}\{g_i = 2\}/T, (1/6)^2)$. Tables S9 and S10 show the results. The TPWD estimator is less biased than the NN-reg estimator, but the bias dominates the variance. Clustering consistency is much slower, which results in poor coverage of asymptotic confidence intervals (which improves for $T \geq 40$).

Table S9: Estimation of the slope coefficient and grouped fixed effects (lagged outcome)

			NN-reg				TPWD								Infeasible OLS						
G	N	T	Bias $\widehat{\beta}_1$	Bias $\widehat{\beta}_2$	RMSE $\widehat{\beta}_1$	RMSE $\widehat{\beta}_2$	$\widehat{G}$	Bias $\widehat{\beta}_1$	Bias $\widehat{\beta}_2$	RMSE $\widehat{\beta}_1$	RMSE $\widehat{\beta}_2$	RMSE $\widehat{\alpha}_{gt}$	.95 β_1	.95 β_2	Bias $\widehat{\beta}_1$	Bias $\widehat{\beta}_2$	RMSE $\widehat{\beta}_1$	RMSE $\widehat{\beta}_2$	RMSE $\widehat{\alpha}_{gt}$	.95 β_1	.95 β_2
3	90	7	0.151	0.584	0.162	0.584	4.168	0.130	0.370	0.142	0.379	0.449	0.248	0.000	-0.002	-0.002	0.041	0.033	0.065	0.942	0.924
		10	0.131	0.550	0.140	0.550	2.658	0.127	0.374	0.136	0.384	0.454	0.156	0.002	0.000	0.003	0.035	0.025	0.064	0.926	0.950
		20	0.101	0.439	0.105	0.439	1.982	0.115	0.368	0.121	0.382	0.454	0.046	0.008	-0.001	0.000	0.024	0.017	0.062	0.952	0.944
		40	0.089	0.332	0.091	0.332	2.404	0.066	0.199	0.089	0.269	0.280	0.374	0.354	0.001	-0.001	0.017	0.012	0.062	0.946	0.946
3	180	7	0.147	0.585	0.161	0.586	6.360	0.123	0.343	0.130	0.352	0.422	0.078	0.000	-0.002	0.000	0.030	0.022	0.047	0.936	0.948
		10	0.129	0.549	0.133	0.549	3.516	0.120	0.338	0.126	0.349	0.419	0.034	0.000	0.001	0.001	0.024	0.017	0.045	0.944	0.946
		20	0.096	0.431	0.100	0.431	2.402	0.106	0.314	0.112	0.337	0.400	0.056	0.032	0.000	-0.001	0.017	0.012	0.044	0.948	0.936
		40	0.083	0.316	0.084	0.316	2.588	0.050	0.151	0.075	0.232	0.218	0.504	0.450	0.000	0.000	0.012	0.009	0.044	0.942	0.930
4	90	7	0.169	0.562	0.186	0.563	4.004	0.130	0.334	0.141	0.342	0.402	0.250	0.006	-0.002	-0.002	0.041	0.033	0.073	0.944	0.918
		10	0.147	0.526	0.155	0.526	2.630	0.127	0.367	0.134	0.373	0.427	0.150	0.000	0.000	0.003	0.035	0.025	0.072	0.930	0.936
		20	0.107	0.423	0.111	0.423	1.726	0.116	0.371	0.121	0.379	0.435	0.042	0.000	-0.001	0.000	0.024	0.017	0.071	0.950	0.944
		40	0.091	0.324	0.093	0.324	2.344	0.080	0.245	0.089	0.270	0.318	0.128	0.002	0.001	-0.001	0.017	0.012	0.071	0.944	0.952
4	180	7	0.176	0.507	0.180	0.508	5.016	0.048	0.089	0.064	0.114	0.191	0.628	0.304	-0.002	0.000	0.030	0.022	0.051	0.944	0.944
		10	0.146	0.456	0.149	0.456	3.618	0.049	0.109	0.059	0.126	0.190	0.534	0.094	0.000	0.001	0.024	0.017	0.051	0.936	0.948
		20	0.098	0.356	0.100	0.356	3.326	0.042	0.115	0.048	0.127	0.184	0.430	0.004	0.000	-0.001	0.017	0.012	0.050	0.952	0.930
		40	0.077	0.267	0.078	0.267	3.668	0.035	0.103	0.037	0.106	0.170	0.258	0.450	0.000	0.000	0.012	0.009	0.050	0.942	0.936

Notes: This table reports results for the nuclear norm regularized (NN-reg) estimator with regularization parameter $\psi = \log(\log(T)) \min(N,T)^{-1/2}/4$; triad pairwise differencing estimator (TPWD) obtained after four iterations starting at NN-reg and setting $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$; and infeasible pooled OLS regression with known group memberships (Infeasible OLS). .95 denotes coverage rate of a 95%-level confidence interval for β based on large N, T approximations and an estimator of the asymptotic variance clustered at the unit level. Results are averaged across 500 Monte Carlo samples.

Table S10: Classification accuracy of the TPWD estimator (lagged outcome)

G	N	T	Precision	Recall	Rand Index
3	90	7	0.421	0.730	0.559
		10	0.408	0.845	0.508
		20	0.416	0.912	0.488
		40	0.676	0.936	0.731
3	180	7	0.457	0.684	0.609
		10	0.452	0.812	0.577
		20	0.489	0.880	0.580
		40	0.758	0.949	0.796
4	90	7	0.301	0.745	0.500
		10	0.276	0.863	0.396
		20	0.281	0.945	0.354
		40	0.405	0.935	0.583
4	180	7	0.650	0.857	0.730
		10	0.630	0.927	0.723
		20	0.613	0.970	0.710
		40	0.629	0.963	0.728

Notes: This table reports results for the triad pairwise differencing estimator (TPWD) obtained after four iterations starting at the nuclear norm regularized estimator $\widehat{\beta}^1(\psi)$ with $\psi = \log(\log(T)) \min(N, T)^{-1/2}/4$ and setting $c_{NT} = \widehat{\sigma}^1 \log(T)/\sqrt{T}$. Results are averaged across 500 Monte Carlo samples.

S3.6 Time-invariant unobserved heterogeneity

In this section, I investigate (i) the finite sample performance of the PWD estimator in the correctly specified model:

$$y_{it} = \alpha^0_{g^0_i} + v_{it}, \tag{S3.1}$$

and (ii) finite sample sensitivity to the choice of the thresholding parameter $c_T \equiv$. First, I assess the consistency of the PWD estimator for $c_T = 2\log(T)/\sqrt{T}$.[3] For each $(G^0, N) \in \{2, 5, 10, 50\} \times \{50, 100, 200, 500\}$ and T in a linearly equally spaced grid of 4 elements from $\lceil\sqrt{N}\rceil$ to N, I draw $1,000$ Monte Carlo samples from model (S3.1), in which $(\alpha^0_1, ..., \alpha^0_{G^0})' = \text{linspace}(-G^0/2, G^0/2, G^0)$ and $(g^0_1, ..., g^0_N) = (1, ..., 1, ..., G^0, ..., G^0)$ are deterministic and each group has equal size N/G^0. I consider three DGPs for the noise random variable v_{it}, summarized in Table S11. Tables S12-S14 report Monte

Table S11: Data Generating Processes

DGP	Noise
1	$v_{it} \sim \mathcal{N}(0, 1)$ i.i.d. across i and t.
2	$v_{it} = 0.5v_{it-1} + \xi_{it}$ with $\xi_{it} \sim \mathcal{N}(0, 1)$ i.i.d. across i and t, independent of $v_{i0} \sim \mathcal{N}(0, 1)$.
3	$\sigma_i \sim \mathcal{U}[0.5, 1.5]$ and $v_{it} \mid \sigma_i \sim \mathcal{N}(0, \sigma_i)$ independent across i and i.i.d. across t for a given i.

Carlo averages of the estimated number of groups $\widehat{G}^{\text{PWD}}$, the Hausdorff distance between estimated effects $(\widehat{\alpha}^{\text{PWD}}_1, ..., \widehat{\alpha}^{\text{PWD}}_{\widehat{G}^{\text{PWD}}})'$ and true effects $(\alpha^0_1, ..., \alpha^0_{G^0})'$, Rand Index (RI), and CPU time. RI is the proportion of correctly predicted pair (true or false) returned by the PWD estimator:

$$\text{RI} = \frac{TP + TN}{TP + TN + FP + FN},$$

[3] I study sensitivity to this choice later.

where

$$
\begin{aligned}
TP \equiv \text{True Positives} &:= \sum_{i<j} \mathbf{1}\left\{\widehat{g}_i^{\text{PWD}} = \widehat{g}_j^{\text{PWD}}\right\} \mathbf{1}\left\{g_i^0 = g_j^0\right\}, \\
TN \equiv \text{True Negatives} &:= \sum_{i<j} \mathbf{1}\left\{\widehat{g}_i^{\text{PWD}} \neq \widehat{g}_j^{\text{PWD}}\right\} \mathbf{1}\left\{g_i^0 \neq g_j^0\right\}, \\
FP \equiv \text{False Positives} &:= \sum_{i<j} \mathbf{1}\left\{\widehat{g}_i^{\text{PWD}} = \widehat{g}_j^{\text{PWD}}\right\} \mathbf{1}\left\{g_i^0 \neq g_j^0\right\}, \\
FN \equiv \text{False Negatives} &:= \sum_{i<j} \mathbf{1}\left\{\widehat{g}_i^{\text{PWD}} \neq \widehat{g}_j^{\text{PWD}}\right\} \mathbf{1}\left\{g_i^0 = g_j^0\right\}.
\end{aligned}
$$

Results suggest good finite sample performance, although deteriorating with the degree of time dependence of the idiosyncratic shocks. In the most favourable case of independent normal errors (Tables S12 and S14), it is remarkable how perfect or almost perfect classification is achieved for moderate sample sizes and even for a large number of groups (e.g., for $(N, T, G^0) = (50, 36, 2)$ or $(N, T, G^0) = (500, 500, G^0 = 50)$).

Second, I investigate the finite sample sensitivity of the results to the choice of the thresholding parameter c_T. I consider DGP 1 only, fix $N = 120$, and let $(G^0, T) \in \{2, 3, 4\} \times \{11, 66, 120\}$. Figures S2-S6 plot Monte Carlo averages of $\widehat{G}^{\text{PWD}}$, HD, RI, Precision (P) and Recall (R) rates as functions of $c \in \text{linspace}(0.1, 20, 40)$ with $c_T = c \log(T)/\sqrt{T}$, where each coloured line corresponds to $\sigma \in \{0.25, 0.5, 1\}$, where σ is the standard-deviation of the random noise v_{it}. The Recall rate (R) measures the ability of the PWD estimator to identify pairs that truly belong to the same group. The Precision rate (P) measures how precise the pairing prediction is: among all predicted pairs of units, what is the proportion of correct ones? Both formally write

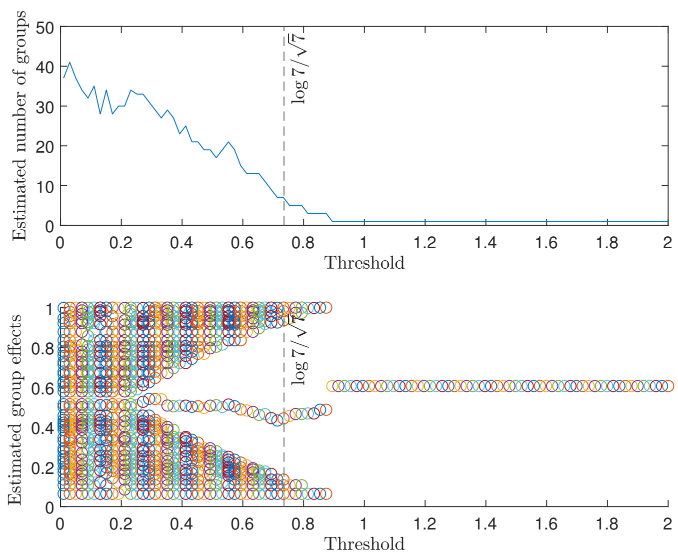
$$
\text{R} = \frac{TP}{TP + FN}, \quad \text{P} = \frac{TP}{TP + FP}.
$$

Figure S2 suggests that the larger the T, the larger the range of values for c for which $\hat{G}^{\text{PWD}} = G^0$. Figures S3-S5, which inform about the grouping composition, confirm that the larger T the better in terms of RI, Recall and Precision and that classification is fast (e.g., compare with $T = 66$). This is also exactly what the small Hausdorff distances suggest in Figure S6.

S4 Empirical illustration: visualisation of the regularization path

In this section, I give a short illustration of the PWD estimator which allows to visualize the grouping "regularization path". I use the balanced subsample of Acemoglu, Johnson, Robinson, and Yared (2008) which contains the Freedom House Index of democracy for $N = 74$ countries (after dropping missing values) observed during $T = 7$ periods over the period 1970-2000. I estimate model (S3.1) using the PWD estimator for different values of $c_T \in (0, 2)$. Figure S1 reports the estimated number of groups and group-specific effects as a function of c_T. The top-panel shows that

Figure S1: PWD Regularization Path



Notes: Acemoglu, Johnson, Robinson, and Yared (2008)'s democracy data. The top panel plots $\widehat{G}^{\mathrm{PWD}}$ *as a function* c_T*. The bottom-panel plots* $\left\{\widehat{\alpha}_g^{\mathrm{PWD}} : g \in \left\{1, ..., \widehat{G}^{\mathrm{PWD}}\right\}\right\}$ *as a function of* c_T*.*

the regularization path for $\widehat{G}^{\mathrm{PWD}}(c_T)$ is smooth and exhibits a general decreasing pattern from $\widehat{G}^{\mathrm{PWD}}(0.01) \approx 40$ to $\widehat{G}^{\mathrm{PWD}}(2) = 1$. The same pattern is observed when estimating model (1.1) under the constraint $\beta^0 = 0$. The bottom panel suggests a convergence toward $\widehat{G}^{\mathrm{PWD}} = 3$ groups before a sudden phase-transition to $\widehat{G}^{\mathrm{PWD}} = 1$.

S5 Additional tables and figures

Table S12: Consistency of the PWD estimator under i.i.d. errors

		$G^0=2$				$G^0=5$				$G^0=10$				$G^0=50$			
N	T	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time
50	8	13.448	0.853	0.689	0.0231	31.569	0.6837	0.8347	0.0245	33.595	0.6025	0.9285	0.0268	41.097	1.0521	0.9918	0.0271
	22	3.12	0.2724	0.9679	0.0197	31.078	0.5124	0.8398	0.028	33.978	0.5033	0.9316	0.033	44.195	0.7252	0.9952	0.038
	36	2.057	0.0523	0.9987	0.0202	29.031	0.4071	0.8472	0.0351	34.036	0.4676	0.9318	0.0385	43.049	0.6741	0.9943	0.0384
	50	2.0	0.0318	1.0	0.0179	25.567	0.3422	0.8598	0.0392	34.006	0.4309	0.9332	0.0388	41.194	0.6594	0.9928	0.0363
100	10	20.265	0.8568	0.6847	0.0356	61.661	0.7014	0.8192	0.0514	66.048	0.6207	0.9167	0.0473	76.09	0.6061	0.9909	0.0468
	40	2.067	0.0431	0.9991	0.0365	54.776	0.4204	0.8283	0.1461	65.684	0.468	0.9184	0.159	77.127	0.521	0.9916	0.186
	70	2.0	0.0195	1.0	0.0366	36.093	0.3167	0.8603	0.1492	62.926	0.3492	0.9211	0.2321	75.669	0.5423	0.9907	0.2714
	100	2.0	0.0164	1.0	0.0464	19.101	0.2653	0.9213	0.1268	53.595	0.2724	0.9283	0.2673	75.805	0.5456	0.9911	0.3696
200	15	20.541	0.7577	0.7738	0.0388	121.746	0.6481	0.8105	0.2366	129.174	0.5911	0.9095	0.2539	142.493	0.518	0.9877	0.2808
	77	2.0	0.0129	1.0	0.0331	61.0	0.3285	0.8491	0.4463	119.47	0.3402	0.9122	0.8661	142.23	0.4934	0.9874	1.058
	139	2.0	0.0098	1.0	0.0387	13.602	0.2447	0.9704	0.2244	66.249	0.2485	0.934	0.8318	133.374	0.375	0.9892	1.719
	200	2.0	0.0082	1.0	0.0377	5.492	0.0657	0.9988	0.1662	26.174	0.2079	0.9774	0.4734	96.023	0.2161	0.9935	1.7135
500	23	17.832	0.6652	0.8941	0.2407	294.97	0.6124	0.8047	1.7975	319.05	0.5614	0.9041	1.9425	338.79	0.5145	0.9837	2.1004
	182	2.0	0.0054	1.0	0.2244	9.838	0.2209	0.993	0.5474	80.128	0.2383	0.9483	3.1459	236.88	0.2404	0.9868	9.3569
	341	2.0	0.0038	1.0	0.309	5.0	0.0085	1.0	0.5465	10.566	0.0589	0.9997	0.9849	60.138	0.1724	0.9992	4.3177
	500	2.0	0.0031	1.0	0.4193	5.0	0.0071	1.0	0.7088	10.0	0.012	1.0	1.2732	50.056	0.039	1.0	5.2066

Notes: Results are averaged over 1,000 Monte Carlo replications. G^0 ≡True number of groups; $\widehat{G}^{\text{PWD}}$ ≡Estimated number of groups; HD≡Hausdorff Distance between estimated and true group effects; RI≡Rand Index; CPU time≡MATLAB's `cputime`.

Table S13: Consistency of the PWD estimator under weakly dependent errors

		$G^0 = 2$				$G^0 = 5$				$G^0 = 10$				$G^0 = 50$			
N	T	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time
50	8	28.595	1.4024	0.5414	0.0216	32.107	1.1536	0.8214	0.0275	33.375	0.9449	0.9176	0.028	35.682	1.3475	0.9855	0.0273
	22	17.803	0.9754	0.6238	0.0267	31.736	0.7796	0.832	0.03	33.706	0.6663	0.9254	0.0315	38.623	1.1209	0.9894	0.0325
	36	10.578	0.7982	0.7543	0.0246	31.378	0.6548	0.8374	0.039	33.919	0.5848	0.9293	0.0382	39.864	0.9718	0.9912	0.0386
	50	6.528	0.6328	0.8646	0.0245	30.584	0.5979	0.8417	0.0412	34.134	0.545	0.9309	0.0418	39.854	0.867	0.9915	0.0368
100	10	54.019	1.4465	0.5244	0.0466	63.822	1.2461	0.8121	0.0525	66.017	1.0834	0.9102	0.0495	69.514	0.9166	0.9864	0.0441
	30	17.864	0.8408	0.7239	0.051	61.334	0.6931	0.8205	0.1498	65.385	0.613	0.917	0.1548	74.838	0.6368	0.9904	0.178
	70	6.447	0.5802	0.9223	0.0457	55.833	0.5859	0.827	0.2099	65.453	0.5384	0.9187	0.2397	75.422	0.5907	0.9906	0.2736
	100	3.413	0.3072	0.9795	0.0419	48.448	0.5243	0.8368	0.2423	63.861	0.504	0.9204	0.3224	75.413	0.5825	0.9907	0.3614
200	15	91.877	1.3407	0.5195	0.1777	125.656	1.1787	0.8074	0.2408	130.224	1.0442	0.9063	0.2491	136.702	0.7222	0.9849	0.2714
	77	10.08	0.6585	0.9134	0.1068	106.55	0.6098	0.816	0.7937	127.67	0.5544	0.9102	0.9589	142.40	0.5066	0.9873	1.0767
	139	2.965	0.2254	0.9924	0.0463	71.937	0.4896	0.8375	0.9182	115.835	0.4826	0.9131	1.4926	142.416	0.4901	0.9878	1.8468
	200	2.166	0.0551	0.999	0.0502	41.683	0.4085	0.8811	0.778	92.895	0.4117	0.92	1.676	136.125	0.4528	0.9889	2.469
500	23	170.186	1.2131	0.5171	1.057	310.33	1.1007	0.8035	1.869	322.65	1.0035	0.9031	1.9837	337.404	0.7414	0.9827	2.0861
	182	3.232	0.2498	0.9958	0.2965	122.31	0.4737	0.8396	4.8164	247.13	0.4717	0.9081	9.8664	326.424	0.469	0.9841	13.09
	341	2.016	0.0115	1.0	0.3321	22.837	0.3443	0.9669	1.8112	96.738	0.3481	0.9392	6.9564	215.331	0.3505	0.9878	15.6834
	500	2.0	0.0063	1.0	0.4076	6.274	0.1528	0.9985	0.8521	30.89	0.2868	0.985	3.3249	108.39	0.2895	0.9949	11.2575

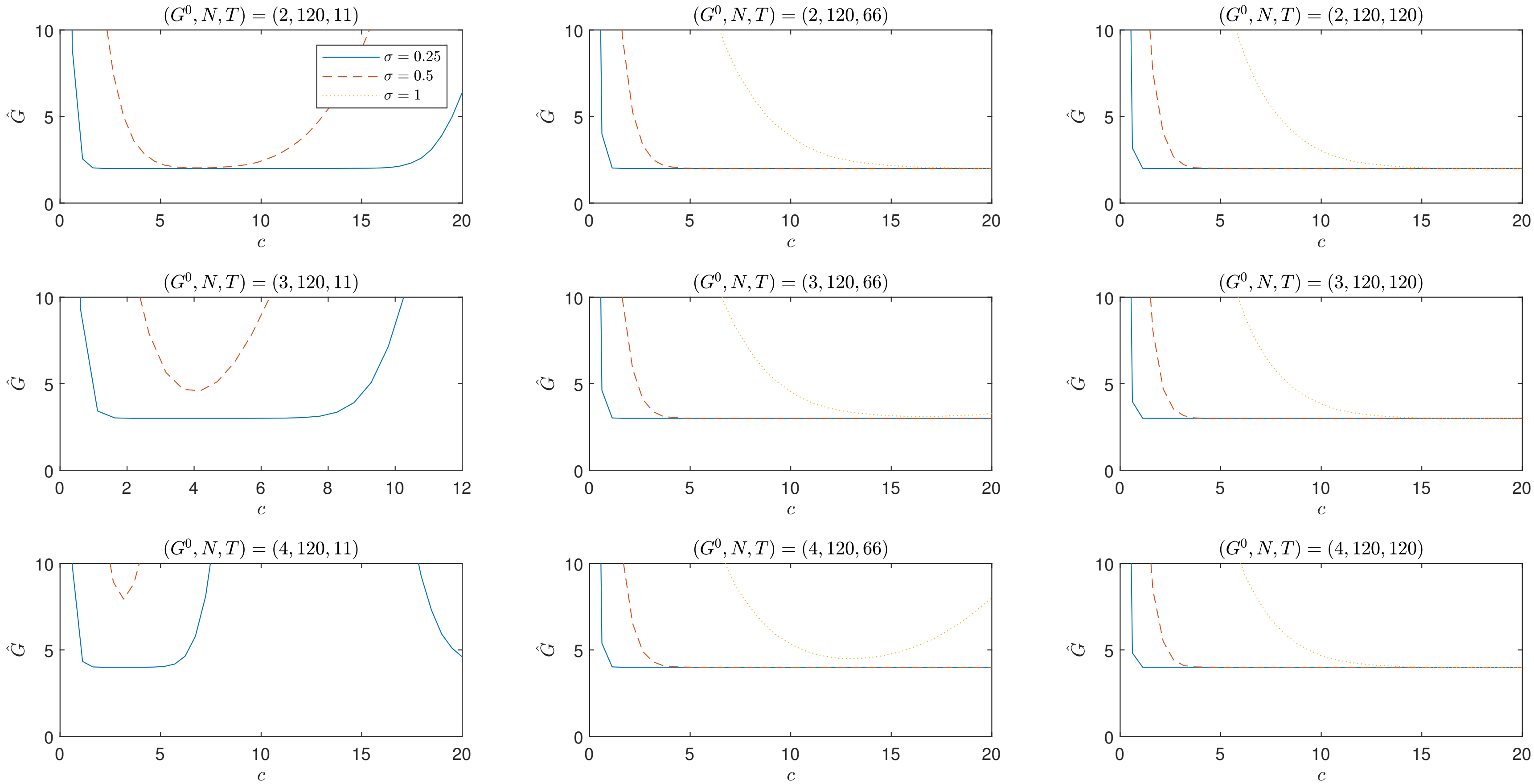
Notes: Results are averaged over 1,000 Monte Carlo replications. G^0 ≡True number of groups; $\widehat{G}^{\text{PWD}}$ ≡Estimated number of groups; HD≡Hausdorff Distance between estimated and true group effects; RI≡Rand Index; CPU time≡MATLAB's `cputime`.

Table S14: Consistency of the PWD estimator under heteroskedastic errors

		$G^0=2$				$G^0=5$				$G^0=10$				$G^0=50$			
N	T	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\text{PWD}}$	HD	RI	CPU time
50	8	15.869	0.9211	0.6471	0.0225	31.651	0.7309	0.8332	0.0255	33.744	0.6399	0.9268	0.026	40.022	1.0997	0.9906	0.0275
	22	4.841	0.4884	0.9134	0.0224	31.374	0.5593	0.8386	0.0311	34.255	0.5243	0.9313	0.0332	43.216	0.8085	0.9943	0.0365
	36	2.257	0.1016	0.9937	0.022	30.082	0.454	0.8442	0.033	34.272	0.4806	0.9316	0.0363	42.551	0.7016	0.9939	0.0443
	50	2.006	0.0381	0.9999	0.0185	26.943	0.3792	0.8545	0.0364	34.135	0.4532	0.9328	0.0355	41.12	0.679	0.9928	0.0415
100	10	29.069	0.9689	0.605	0.0456	62.248	0.7981	0.8179	0.0568	65.794	0.7018	0.9153	0.0474	74.017	0.6838	0.9898	0.0549
	40	2.344	0.1132	0.9955	0.0457	56.172	0.4677	0.8265	0.1434	65.958	0.4917	0.9182	0.159	76.597	0.5338	0.9915	0.1772
	70	2.0	0.0213	1.0	0.0416	39.427	0.3467	0.8523	0.1642	63.817	0.3783	0.9205	0.2393	75.513	0.5473	0.9907	0.2803
	100	2.0	0.0177	1.0	0.0402	22.032	0.2861	0.9072	0.1375	55.448	0.2937	0.9267	0.2832	75.832	0.5458	0.9911	0.3649
200	15	35.896	0.8967	0.6458	0.0561	122.235	0.7455	0.8101	0.226	129.324	0.6703	0.9089	0.2543	142.211	0.5459	0.987	0.2748
	77	2.0	0.0143	1.0	0.0307	68.329	0.3566	0.8408	0.4835	121.611	0.3689	0.9117	0.9058	142.751	0.4924	0.9874	1.0951
	139	2.0	0.0101	1.0	0.0396	16.56	0.2582	0.9591	0.2576	72.042	0.2621	0.9302	0.9223	134.743	0.3999	0.9891	1.8094
	200	2.0	0.0084	1.0	0.0454	5.76	0.087	0.9981	0.1657	28.772	0.2154	0.9736	0.5166	99.223	0.2239	0.9931	1.8396
500	33	37.936	0.7906	0.7672	0.3116	298.003	0.6836	0.8046	1.8629	319.607	0.6228	0.904	1.9931	339.888	0.5361	0.9835	2.1764
	182	2.0	0.0056	1.0	0.2115	11.672	0.2355	0.9897	0.599	88.894	0.2462	0.9435	3.6233	245.004	0.2504	0.9865	10.57
	341	2.0	0.0041	1.0	0.3299	5.0	0.0088	1.0	0.5171	10.966	0.0819	0.9995	0.9751	63.247	0.18	0.9989	4.7129
	500	2.0	0.0033	1.0	0.4094	5.0	0.0073	1.0	0.7156	10.0	0.0124	1.0	1.271	50.098	0.0433	1.0	5.4568

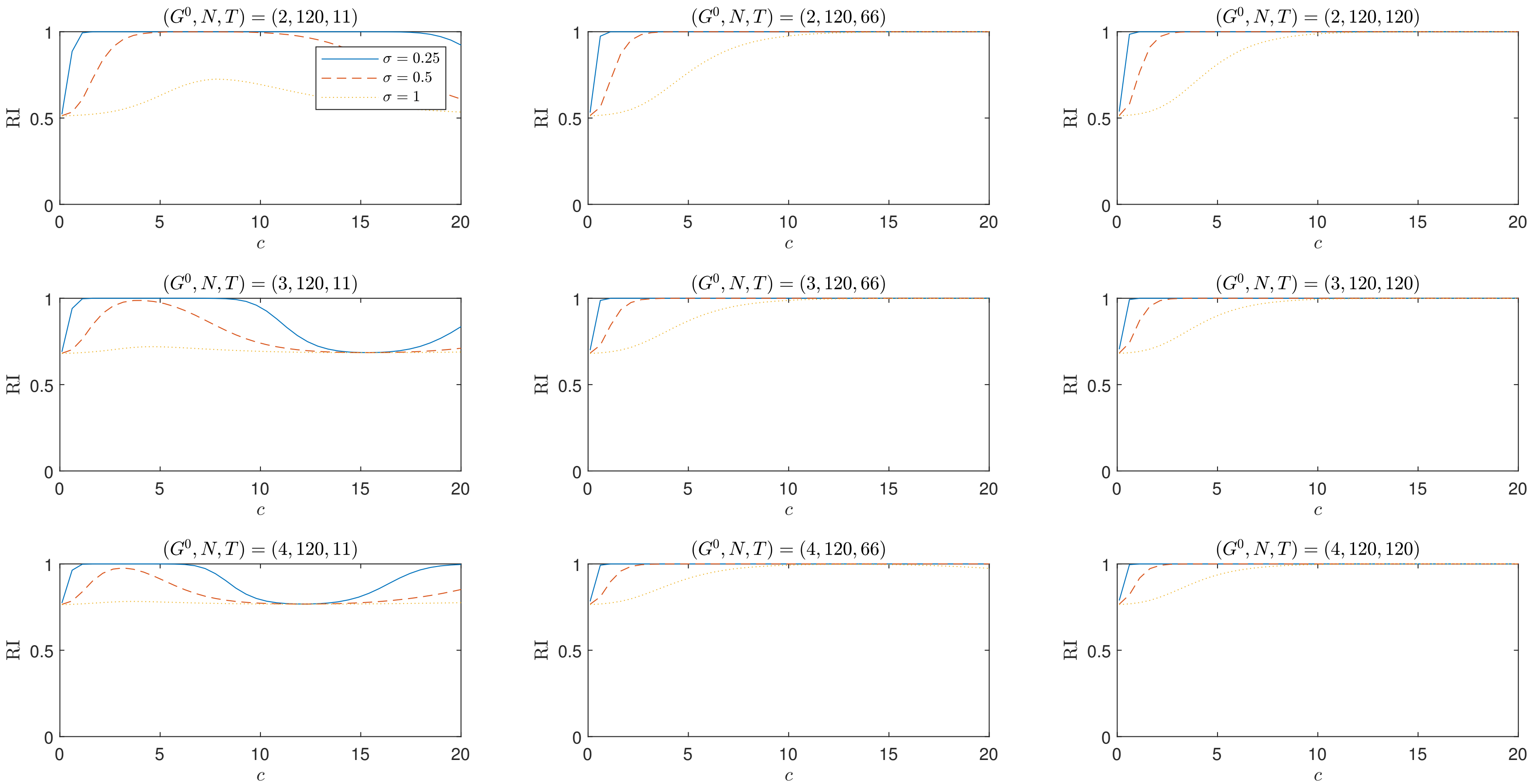
Notes: Results are averaged over 1,000 Monte Carlo replications. G^0 ≡True number of groups; $\widehat{G}^{\text{PWD}}$ ≡Estimated number of groups; HD≡Hausdorff Distance between estimated and true group effects; RI≡Rand Index; CPU time≡MATLAB's `cputime`.

Figure S2: Sensitivity of the estimated number of groups ($\widehat{G}$)



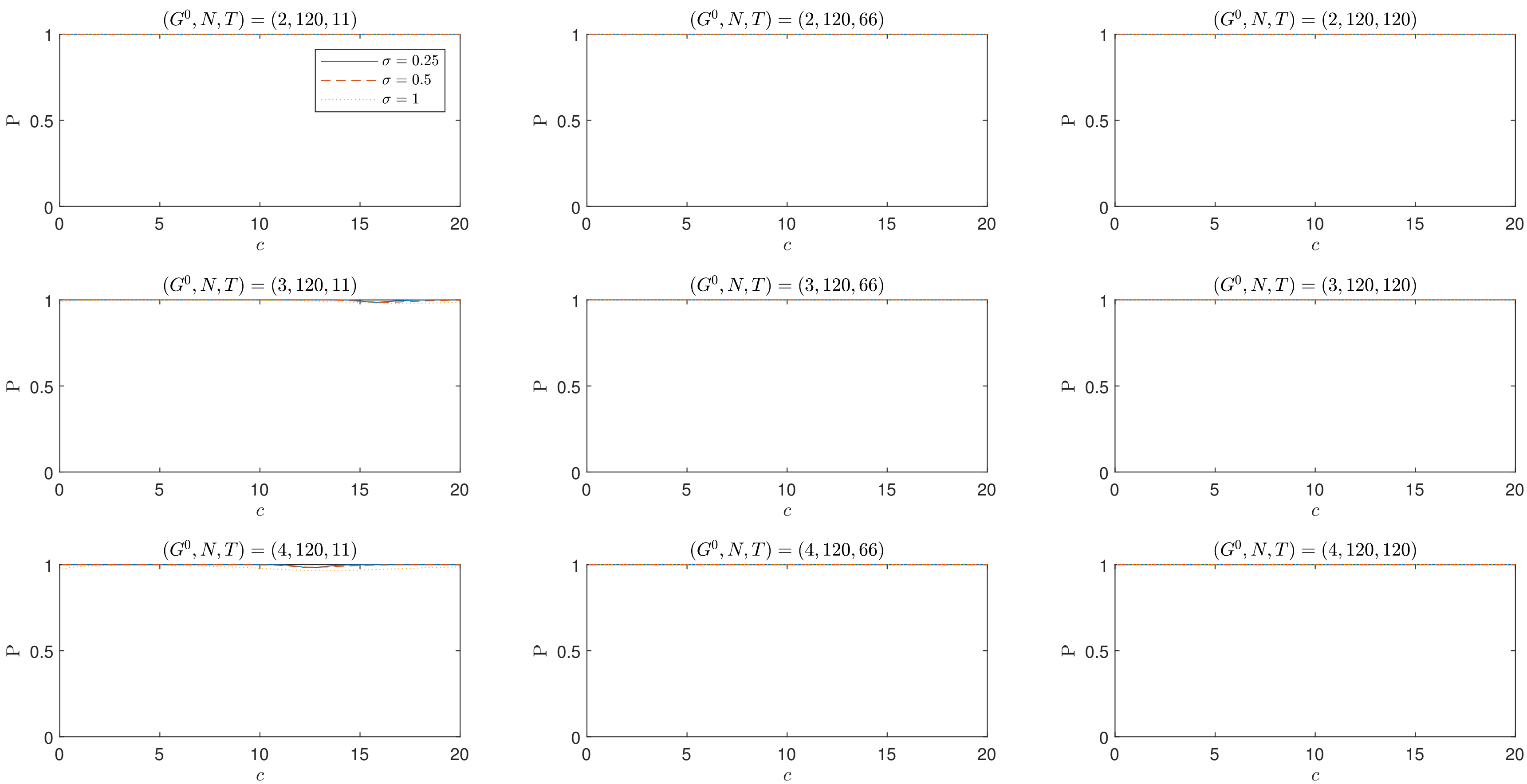
Notes: solid blue, dashed red, and dotted yellow lines report averages of $\widehat{G}$ with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are averaged over $1,000$ Monte Carlo replications.

Figure S3: Sensitivity of the Rand Index (RI)



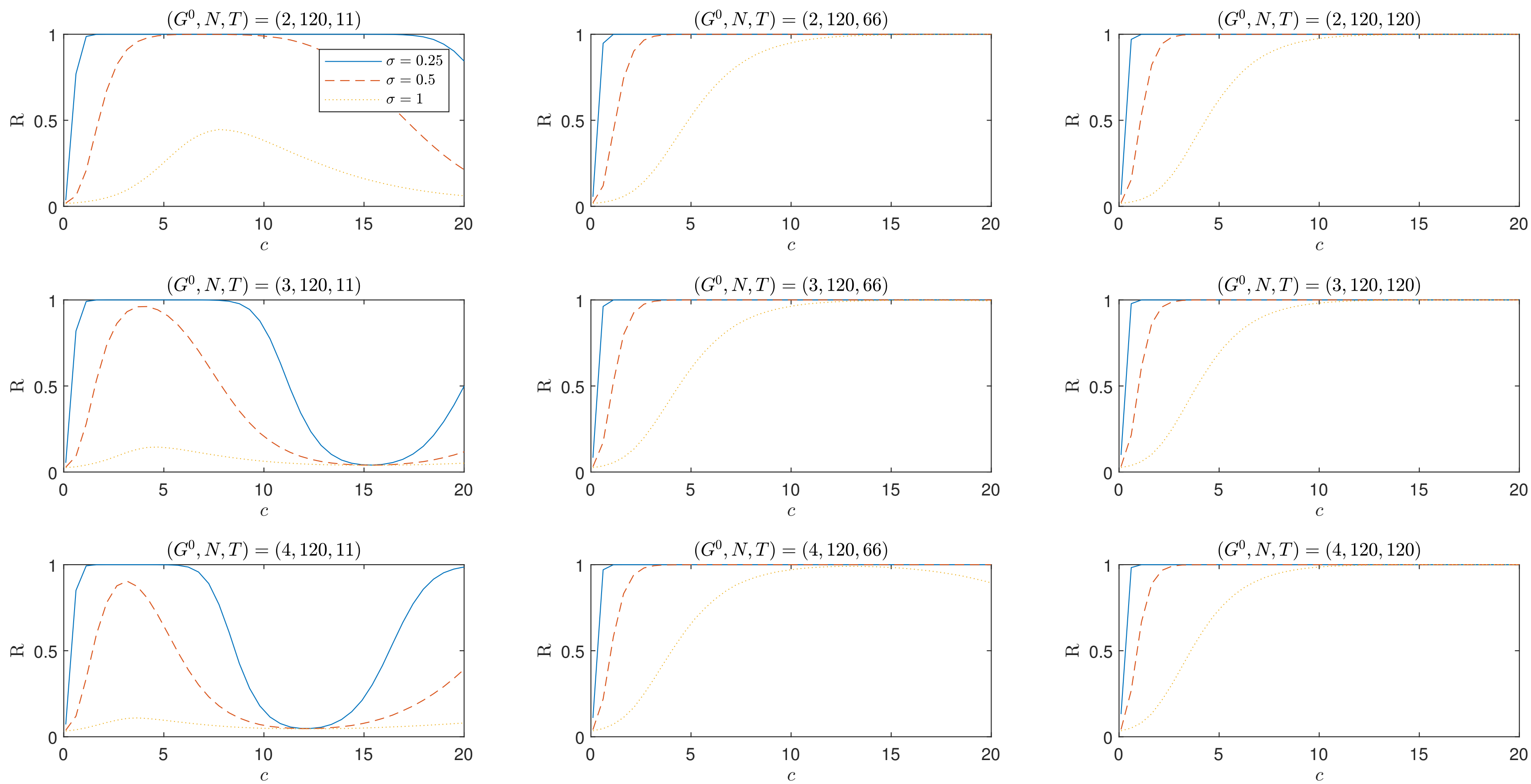
Notes: solid blue, dashed red, and dotted yellow lines report averages of RI with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are averaged over $1,000$ Monte Carlo replications.

Figure S4: Sensitivity of the Precision rate (P)



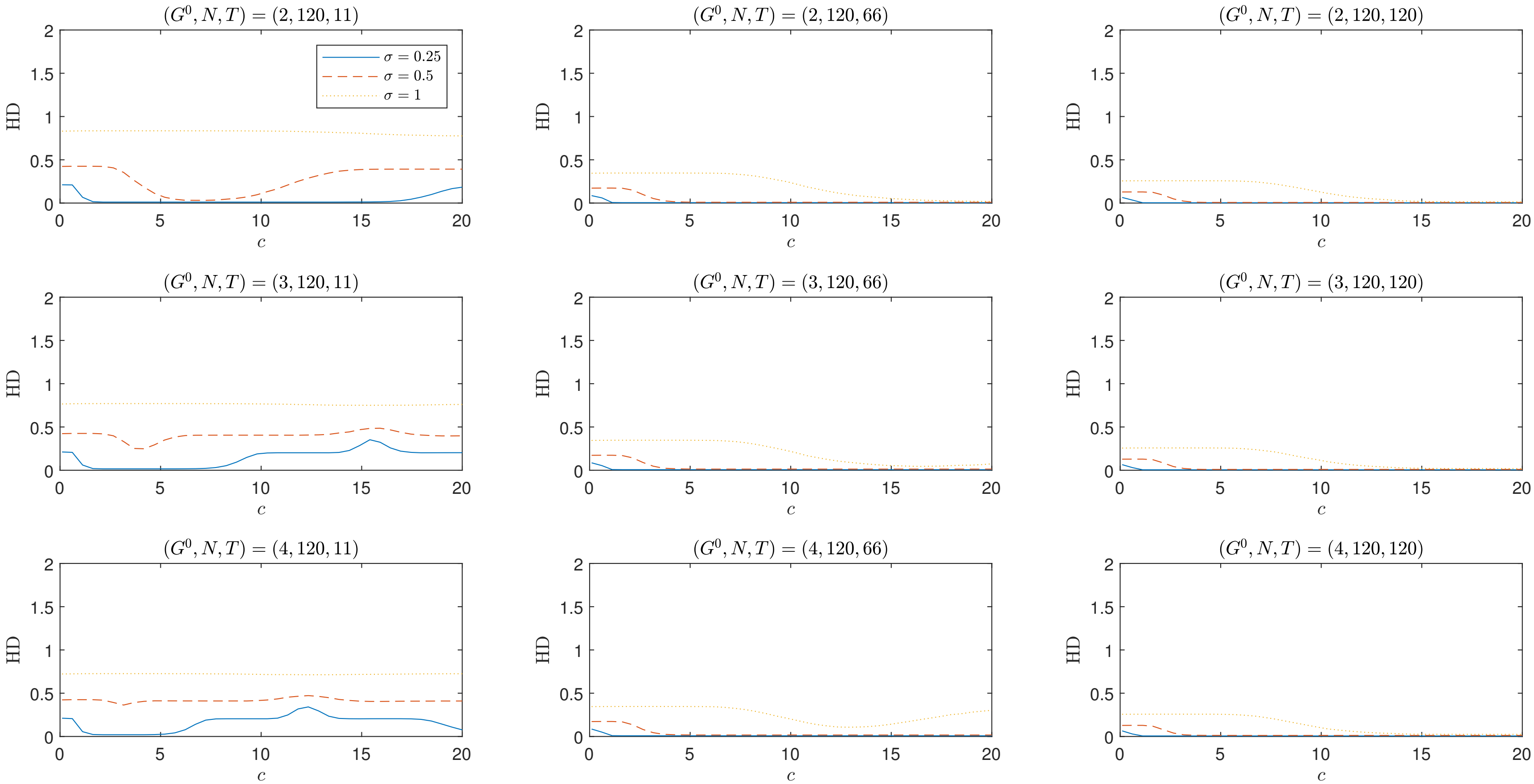
Notes: solid blue, dashed red, and dotted yellow lines report averages of P with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are averaged over $1,000$ Monte Carlo replications.

Figure S5: Sensitivity of the Recall rate (R)



Notes: solid blue, dashed red, and dotted yellow lines report averages of R with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are averaged over $1,000$ Monte Carlo replications.

Figure S6: Sensitivity of the Hausdorff Distance (HD)



Notes: solid blue, dashed red, and dotted yellow lines report averages of HD with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are averaged over $1,000$ Monte Carlo replications.

References

Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008): "Income and Democracy," *American Economic Review*, 98(3), 808–42.

Arellano, M., and S. Bonhomme (2011): "Nonlinear Panel Data Analysis," *Annual Review of Economics*, 3(1), 395–424.

Beyhum, J., and E. Gautier (2023): "Factor and Factor Loading Augmented Estimators for Panel Regression With Possibly Nonstrong Factors," *Journal of Business & Economic Statistics*, 41(1), 270–281.

Bonhomme, S., and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83(3), 1147–1184.

Graham, B. S. (2017): "An Econometric Model of Network Formation With Degree Heterogeneity," *Econometrica*, 85(4), 1033–1063.

Jochmans, K. (2017): "Two-Way Models for Gravity," *The Review of Economics and Statistics*, 99(3), 478–485.

Moon, H. R., and M. Weidner (2019): "Nuclear Norm Regularized Estimation of Panel Regression Models," .

Mugnier, M. (2022): "Unobserved Clusters of Time-Varying Heterogeneity in Nonlinear Panel Data Models," Discussion paper.

Tabord-Meehan, M. (2019): "Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic," *Journal of Business & Economic Statistics*, 37(4), 671–680.

Wooldridge, J. (2010): *Econometric Analysis of Cross Section and Panel Data, second edition*, Econometric Analysis of Cross Section and Panel Data. MIT Press.