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., trade gravity equations). I report additional Monte Carlo results for variations of the model considered in the paper, including unbalanced groups, dependent and heteroskedasctic errors, unit fixed effects, and lower signal-to-noise ratio. Finally, I provide some visualization 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.

*Paris School of Economics, martin.mugnier@psemail.eu

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 even if there are almost as many groups as individuals asymptotically.

Proposition S1.1 *Suppose $\beta^0 = 0$, Assumptions 2–4 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 $1 - \widehat{W}_{ij} = \mathbf{1}\{\widehat{d}^2_\infty(i,j) > c_{NT}\}$ is bounded asymptotically by $o_p(NT^{-\delta})$ uniformly across pairs of units $(i,j) \in \{1,\ldots,N\}^2$ for all $\delta > 0$ as $\min(N,T) \to \infty$. Similarly, outside of an event whose probability tends to zero (that 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 4(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 4(c).

A similar result could plausibly be obtained for the full model with $\beta \neq 0$, 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. Despite the nonlinear multiplicative form, it seems possible, under some assumptions, to follow the same identification strategy as in the linear panel data model: (i) identify β_0 (though identification may deliver slow rates of convergence), (ii) identify the group memberships using tetrad (or triad) pairwise difference comparisons, and (iii) identify the fixed effects.

S2.1.1 Identification of β_0 under serially correlated 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 each $S \subset \mathbb{N}^{*2}$, 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 $S_1, S_2 \subset \{1, \ldots, n\}^2$, $(x_{ij}, \varepsilon_{ij})_{(i,j)\in S_1}$ 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

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

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

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

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

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

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 assume symmetric group-specific effects, a much simpler GMM objective function can be constructed.

Assumption S2 (4-cyclical exogeneity) $(\varepsilon_{ij}, \varepsilon_{jk}, \varepsilon_{kl}, \varepsilon_{li})|x_{12}, \ldots, x_{n(n-1)}$ *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 i to j also affects trade from j to k, from k to l, and from l to i. Permanent unconditional correlation patterns within 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 S2 and equation (B.3) imply that, for any distinct indices i, j, k, l (tetrad),

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

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

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

and

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

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

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

which does not involve any nuisance parameters and holds for all

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

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

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

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, for all distinct indices $(i, j, k, l) \in \{1, \ldots, n\}^4$, define

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

[1] A more complete treatment left for further research should control estimation errors.

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

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

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

$$\begin{aligned} &\bar{\alpha}_{-i} \xrightarrow{p} \bar{\alpha}_{g_i} > 0, \\ &\bar{\gamma}_{-j} \xrightarrow{p} \bar{\gamma}_{g_j} > 0, \\ &\frac{1}{N}\sum_{j=1}^{N} \varepsilon_{ij}\varepsilon_{ji}\varepsilon_{lj}\varepsilon_{jl} - \frac{1}{N}\sum_{j=1}^{N} \varepsilon_{ij}\varepsilon_{ji}\varepsilon_{kj}\varepsilon_{jk} \xrightarrow{p} 0. \end{aligned} \tag{S2.6}$$

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,

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

which converges towards $\rho_{\bar{g},\check{g}} \geq \rho > 0$ uniformly over $i, j, \bar{g}$, and $\check{g}$ under conditions similar to (S2.6), weak group-separation conditions similar to Assumption 4(c), and

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

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, a dissimilarity matrix of tetrad-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

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

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

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

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:

$$y_{it} = x_{it}' \beta + \nu_i + \alpha_{g_i t} + v_{it}. \qquad \text{(S2.7)}$$

Without covariates, the three-step estimation procedure still applies to a within-differenced version of equation (S2.7). With covariates, the nuclear-norm regularized estimator Moon and Weidner (2019) is still theoretically valid under strong factors, and the demeaned first-step residuals can be 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 Supplementary Material.

S2.3 Computational gains in the time-invariant model

If the clustered unobserved heterogeneity is time-invariant, it is possible to relax Assumption 4(f) to Bonhomme and Manresa (2015)'s Assumption 2(e), and a simple within, first-difference, or instrumental differencing estimator can be used as a first step (see, e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Wooldridge, 2010). The method becomes $O(N^2T)$ and enjoy similar oracle asymptotic properties by considering the new distance

$$\widehat{d}^2_{\infty,\text{time-inv}}(i,j) \equiv \left(\frac{1}{T}\sum_{t=1}^{T}(\widehat{v}_{it}-\widehat{v}_{jt})\right)^2. \tag{S2.8}$$

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 1–2 in each case. In Sections S3.4–S3.5, I consider full GFE models similar to that of the paper but introduce heteroskedasticity, serially correlated errors, or unit-specific effects. I replicate Tables 3–4 in each case.

S3.1 Unbalanced groups

For $G = 3$, let

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

For $G = 4$, let

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

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

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

			TPWD			Post-Spectral$^{\bar{G}=2}$		Post-Spectral$^{\bar{G}=3}$		Post-Spectral$^{\bar{G}=4}$		Post-Spectral$^{\bar{G}=10}$		GFE$^{\bar{G}=2}$	GFE$^{\bar{G}=3}$	GFE$^{\bar{G}=4}$	GFE$^{\bar{G}=10}$	Oracle
G	N	T	$\widehat{G}$	RMSE	CPU time	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
3	90	7	13.688	0.210	0.106	1.798	0.217	2.560	0.208	3.100	0.200	·	·	0.107	0.160	0.183	0.236	0.060
		10	8.718	0.153	0.097	1.856	0.209	2.690	0.185	3.210	0.175	7.034	0.160	0.097	0.143	0.156	0.212	0.061
		20	5.066	0.093	0.139	1.900	0.199	2.792	0.145	3.402	0.120	5.222	0.097	0.091	0.103	0.110	0.178	0.060
		40	3.792	0.070	0.301	1.938	0.177	2.858	0.102	3.630	0.071	5.188	0.084	0.091	0.062	0.092	0.153	0.061
3	180	7	27.436	0.234	0.497	1.818	0.201	2.508	0.198	3.018	0.196	·	·	0.089	0.142	0.169	0.221	0.042
		10	16.156	0.168	0.668	1.858	0.197	2.624	0.185	3.222	0.178	6.662	0.163	0.074	0.123	0.148	0.195	0.042
		20	6.478	0.088	1.897	1.872	0.179	2.758	0.133	3.356	0.116	4.792	0.074	0.068	0.103	0.113	0.155	0.043
		40	4.452	0.056	4.023	1.882	0.171	2.814	0.086	3.642	0.049	4.798	0.056	0.068	0.063	0.075	0.128	0.043
4	90	7	6.450	0.158	0.098	1.646	0.280	2.366	0.261	3.076	0.245	·	·	0.148	0.149	0.167	0.234	0.069
		10	5.218	0.131	0.115	1.766	0.263	2.464	0.238	3.076	0.215	7.296	0.173	0.149	0.122	0.141	0.210	0.069
		20	4.218	0.093	0.235	1.860	0.241	2.782	0.165	3.432	0.130	5.932	0.120	0.143	0.100	0.103	0.171	0.070
		40	4.088	0.076	0.334	1.958	0.221	2.894	0.126	3.594	0.102	6.002	0.115	0.144	0.097	0.076	0.145	0.070
4	180	7	8.746	0.156	0.803	1.570	0.275	2.366	0.266	3.134	0.255	·	·	0.121	0.146	0.166	0.220	0.049
		10	6.548	0.125	1.146	1.720	0.254	2.582	0.229	3.152	0.214	5.618	0.205	0.123	0.124	0.139	0.193	0.049
		20	4.570	0.079	1.983	1.810	0.239	2.822	0.160	3.386	0.125	5.444	0.102	0.116	0.085	0.091	0.146	0.050
		40	4.134	0.057	4.125	1.872	0.222	2.928	0.113	3.704	0.092	5.028	0.096	0.116	0.074	0.053	0.117	0.050

Notes: This table reports (i) the estimated number of groups ($\widehat{G}$), the root mean square error (RMSE), and the execution time in seconds (CPU time) for the triad pairwise-difference (TPWD) estimator computed with $\widehat{\beta}^1 = 0$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2 \log(T)/\sqrt{T}$; (ii) $\widehat{G}$ and the RMSE for the post-spectral estimator proposed in Chetverikov and Manresa (2021) with $\tilde{\beta}^0 = \tilde{\beta}^1 = 0$, user-specified number of groups $g \in \{2, 3, 4, 10\}$, and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq g$ (Post-Spectral$^{\bar{G}=g}$); and (iii) the RMSE for Bonhomme and Manresa (2015)'s grouped fixed-effects estimator (their "Algorithm 1") with a user-specified number of groups $g \in \{2, 3, 4, 10\}$ and 500 random initialization points (GFE$^{\bar{G}=g}$) and for the infeasible (Oracle) estimator using the "true" group memberships. A green-shaded cell corresponds to a well-specified estimator. Results for Post-Spectral$^{\bar{G}=10}$ are missing if $T = 7$ since Post-Spectral$^{\bar{G}=g}$ is not properly defined if $g > T$ as it requires computing the g largest eigenvectors of a $T \times T$ matrix. Results are averaged across 500 Monte Carlo samples.

Table S2: Estimation of group memberships (unbalanced groups)

			TPWD			Post-Spectral$^{\hat{G}=2}$			Post-Spectral$^{\hat{G}=3}$			Post-Spectral$^{\hat{G}=4}$			Post-Spectral$^{\hat{G}=10}$			GFE$^{\hat{G}=2}$			GFE$^{\hat{G}=3}$			GFE$^{\hat{G}=4}$			GFE$^{\hat{G}=10}$		
G	N	T	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI
3	90	7	0.998	0.248	0.387	0.815	0.874	0.737	0.827	0.776	0.686	0.837	0.733	0.668	·	·	·	0.990	0.985	0.980	0.984	0.508	0.593	0.981	0.365	0.477	0.986	0.134	0.293
		10	0.999	0.418	0.526	0.827	0.887	0.756	0.854	0.783	0.713	0.870	0.733	0.692	0.941	0.703	0.726	0.994	0.996	0.992	0.989	0.519	0.603	0.989	0.401	0.509	0.994	0.138	0.297
		20	1.000	0.769	0.812	0.836	0.841	0.736	0.904	0.785	0.761	0.938	0.811	0.809	0.993	0.918	0.929	0.996	1.000	0.996	0.995	0.685	0.741	0.999	0.495	0.589	0.999	0.147	0.305
		40	1.000	0.965	0.971	0.867	0.897	0.803	0.954	0.912	0.896	0.997	0.987	0.987	1.000	0.973	0.978	0.996	1.000	0.997	1.000	0.991	0.993	1.000	0.516	0.606	1.000	0.159	0.315
3	180	7	0.998	0.134	0.292	0.813	0.909	0.755	0.819	0.833	0.714	0.823	0.777	0.682	·	·	·	0.994	0.980	0.979	0.990	0.501	0.588	0.986	0.339	0.456	0.986	0.132	0.289
		10	1.000	0.289	0.419	0.817	0.877	0.739	0.836	0.763	0.683	0.845	0.700	0.650	0.905	0.733	0.717	0.996	0.996	0.993	0.993	0.507	0.594	0.991	0.343	0.460	0.994	0.134	0.292
		20	1.000	0.616	0.686	0.842	0.897	0.777	0.906	0.824	0.788	0.935	0.781	0.785	0.994	0.941	0.948	0.998	1.000	0.998	0.995	0.510	0.598	0.995	0.389	0.500	1.000	0.142	0.299
		40	1.000	0.949	0.959	0.849	0.908	0.791	0.954	0.930	0.907	1.000	0.998	0.998	1.000	0.984	0.987	0.998	1.000	0.998	0.998	0.795	0.831	1.000	0.513	0.602	1.000	0.152	0.307
4	90	7	0.831	0.613	0.790	0.434	0.957	0.454	0.494	0.900	0.536	0.543	0.852	0.598	·	·	·	0.762	0.951	0.858	0.828	0.806	0.851	0.837	0.630	0.798	0.826	0.221	0.662
		10	0.863	0.763	0.852	0.445	0.950	0.469	0.513	0.903	0.559	0.574	0.873	0.631	0.752	0.803	0.791	0.773	0.967	0.870	0.863	0.920	0.906	0.878	0.687	0.832	0.865	0.238	0.673
		20	0.934	0.942	0.948	0.487	0.939	0.525	0.681	0.930	0.753	0.760	0.940	0.838	0.845	0.925	0.898	0.794	0.995	0.892	0.880	0.991	0.939	0.941	0.806	0.901	0.936	0.275	0.695
		40	0.987	0.983	0.987	0.541	0.943	0.593	0.772	0.967	0.849	0.833	0.985	0.912	0.852	0.927	0.904	0.797	1.000	0.895	0.886	0.999	0.944	0.992	0.972	0.986	0.982	0.306	0.713
4	180	7	0.822	0.522	0.755	0.418	0.985	0.424	0.440	0.913	0.469	0.475	0.854	0.519	·	·	·	0.772	0.959	0.865	0.833	0.696	0.816	0.833	0.556	0.770	0.819	0.214	0.654
		10	0.867	0.700	0.831	0.432	0.971	0.446	0.507	0.922	0.545	0.554	0.892	0.601	0.592	0.853	0.639	0.780	0.969	0.874	0.894	0.801	0.878	0.892	0.635	0.817	0.866	0.232	0.667
		20	0.941	0.930	0.946	0.447	0.950	0.469	0.662	0.921	0.727	0.757	0.932	0.833	0.823	0.960	0.898	0.799	0.995	0.894	0.966	0.965	0.971	0.962	0.793	0.901	0.949	0.270	0.691
		40	0.989	0.984	0.989	0.485	0.943	0.519	0.774	0.966	0.854	0.825	0.983	0.906	0.826	0.970	0.902	0.802	1.000	0.897	0.990	0.997	0.995	0.997	0.989	0.994	0.991	0.296	0.707

Notes: This table reports the precision (P) rate, recall (R) rate, and Rand index (RI) for the triad pairwise-difference (TPWD) estimator with $\hat{\beta}^1 = 0$, average linkage, and cut-off $c_{NT} = 1.5\hat{\sigma}^2 \log(T)/\sqrt{T}$, for the post-spectral estimator proposed in Chetverikov and Manresa (2021) with $\hat{\beta}^0 = \hat{\beta}^1 = 0$, user-specified number of groups $g \in \{2, 3, 4, 10\}$, and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq g$ (Post-Spectral$^{\hat{G}=g}$), and for Bonhomme and Manresa (2015)'s grouped fixed-effects estimator (their "Algorithm 1") with a user-specified number of groups $g \in \{2, 3, 4, 10\}$ and 500 random initialization points (GFE$^{\hat{G}=g}$). A green-shaded cell corresponds to a well-specified estimator. Results for Post-Spectral$^{\hat{G}=10}$ are missing if $T = 7$ since Post-Spectral$^{\hat{G}=g}$ is not properly defined if $g > T$ as it requires computing the g largest eigenvectors of a $T \times T$ matrix. 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–S4 show that results are almost unaffected in comparison with the i.i.d. case.

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

			TPWD			Post-Spectral$^{\tilde{G}=2}$		Post-Spectral$^{\tilde{G}=3}$		Post-Spectral$^{\tilde{G}=4}$		Post-Spectral$^{\tilde{G}=10}$		GFE$^{\tilde{G}=2}$	GFE$^{\tilde{G}=3}$	GFE$^{\tilde{G}=4}$	GFE$^{\tilde{G}=10}$	Oracle
G	N	T	$\widehat{G}$	RMSE	CPU time	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
3	90	7	6.244	0.182	0.098	1.968	0.399	2.714	0.327	3.234	0.302	·	·	0.280	0.120	0.159	0.232	0.060
		10	4.882	0.139	0.114	1.978	0.357	2.804	0.262	3.234	0.241	6.682	0.216	0.269	0.090	0.129	0.207	0.061
		20	3.506	0.078	0.240	1.996	0.309	2.978	0.137	3.254	0.115	6.374	0.118	0.252	0.064	0.096	0.168	0.062
		40	3.102	0.064	0.394	2.000	0.266	3.000	0.066	3.386	0.069	5.724	0.093	0.244	0.063	0.087	0.147	0.063
3	180	7	7.554	0.173	0.612	1.972	0.404	2.712	0.332	3.190	0.309	·	·	0.281	0.113	0.147	0.216	0.043
		10	5.634	0.128	1.100	1.986	0.366	2.820	0.262	3.248	0.243	5.080	0.250	0.269	0.078	0.114	0.186	0.043
		20	3.956	0.064	1.960	1.998	0.314	3.000	0.128	3.184	0.106	5.976	0.090	0.249	0.047	0.078	0.144	0.044
		40	3.246	0.047	3.809	2.000	0.265	3.000	0.045	3.406	0.049	4.264	0.056	0.241	0.044	0.066	0.120	0.044
4	90	7	6.892	0.193	0.080	1.964	0.375	2.594	0.340	3.270	0.313	·	·	0.268	0.149	0.168	0.239	0.071
		10	5.298	0.166	0.109	1.978	0.357	2.764	0.288	3.248	0.275	6.556	0.242	0.252	0.141	0.140	0.215	0.071
		20	4.222	0.122	0.197	1.998	0.331	2.958	0.179	3.364	0.156	6.978	0.158	0.236	0.124	0.099	0.175	0.072
		40	4.056	0.087	0.331	2.000	0.313	3.000	0.125	3.530	0.123	7.122	0.143	0.228	0.118	0.078	0.147	0.072
4	180	7	9.656	0.186	0.484	1.954	0.289	2.544	0.283	3.162	0.281	·	·	0.208	0.157	0.168	0.230	0.050
		10	7.146	0.154	0.635	1.984	0.289	2.662	0.255	3.226	0.240	4.628	0.264	0.190	0.141	0.145	0.205	0.050
		20	4.840	0.104	1.583	2.000	0.280	2.950	0.147	3.354	0.124	6.944	0.126	0.178	0.109	0.102	0.160	0.051
		40	4.220	0.072	3.299	2.000	0.276	3.000	0.103	3.530	0.103	5.660	0.111	0.173	0.101	0.065	0.127	0.051

Notes: This table reports (i) the estimated number of groups ($\widehat{G}$), the root mean square error (RMSE), and the execution time in seconds (CPU time) for the triad pairwise-difference (TPWD) estimator computed with $\widehat{\beta}^1 = 0$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2 \log(T)/\sqrt{T}$; (ii) $\widehat{G}$ and the RMSE for the post-spectral estimator proposed in Chetverikov and Manresa (2021) with $\tilde{\beta}^0 = \tilde{\beta}^1 = 0$, user-specified number of groups $g \in \{2, 3, 4, 10\}$, and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq g$ (Post-Spectral$^{\tilde{G}=g}$); and (iii) the RMSE for Bonhomme and Manresa (2015)'s grouped fixed-effects estimator (their "Algorithm 1") with a user-specified number of groups $g \in \{2, 3, 4, 10\}$ and 500 random initialization points (GFE$^{\tilde{G}=g}$) and for the infeasible (Oracle) estimator using the "true" group memberships. A green-shaded cell corresponds to a well-specified estimator. Results for Post-Spectral$^{\tilde{G}=10}$ are missing if $T = 7$ since Post-Spectral$^{\tilde{G}=g}$ is not properly defined if $g > T$ as it requires computing the g largest eigenvectors of a $T \times T$ matrix. Results are averaged across 500 Monte Carlo samples.

Table S4: Estimation of group memberships (dependent errors)

			TPWD			Post-Spectral$^{\bar{G}=2}$			Post-Spectral$^{\bar{G}=3}$			Post-Spectral$^{\bar{G}=4}$			Post-Spectral$^{\bar{G}=10}$			GFE$^{\bar{G}=2}$			GFE$^{\bar{G}=3}$			GFE$^{\bar{G}=4}$			GFE$^{\bar{G}=10}$		
G	N	T	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI
3	90	7	0.909	0.591	0.848	0.364	0.847	0.455	0.483	0.748	0.629	0.533	0.717	0.683	·	·	·	0.560	0.882	0.736	0.936	0.939	0.959	0.933	0.759	0.904	0.926	0.297	0.763
		10	0.952	0.770	0.913	0.426	0.868	0.543	0.612	0.813	0.752	0.666	0.811	0.787	0.782	0.725	0.824	0.568	0.910	0.746	0.972	0.973	0.982	0.970	0.802	0.928	0.963	0.315	0.773
		20	0.994	0.970	0.988	0.492	0.894	0.647	0.879	0.919	0.924	0.931	0.936	0.954	0.978	0.870	0.951	0.584	0.969	0.765	0.998	0.998	0.999	0.998	0.835	0.946	0.997	0.342	0.785
		40	1.000	0.997	0.999	0.558	0.937	0.734	0.995	0.996	0.997	0.999	0.988	0.996	1.000	0.917	0.973	0.591	0.996	0.774	1.000	1.000	1.000	1.000	0.839	0.947	1.000	0.361	0.792
3	180	7	0.923	0.542	0.834	0.357	0.853	0.439	0.470	0.750	0.613	0.512	0.719	0.663	·	·	·	0.561	0.870	0.733	0.938	0.939	0.959	0.936	0.760	0.904	0.929	0.292	0.759
		10	0.963	0.741	0.905	0.415	0.872	0.520	0.612	0.825	0.749	0.656	0.825	0.780	0.647	0.764	0.762	0.570	0.900	0.744	0.973	0.974	0.983	0.972	0.804	0.928	0.967	0.307	0.768
		20	0.995	0.964	0.987	0.482	0.889	0.633	0.893	0.920	0.932	0.937	0.939	0.958	0.980	0.917	0.966	0.588	0.970	0.767	0.998	0.998	0.998	0.998	0.834	0.945	0.997	0.329	0.778
		40	1.000	0.996	0.999	0.559	0.945	0.731	0.999	0.999	1.000	0.999	0.994	0.998	1.000	0.980	0.993	0.595	0.996	0.775	1.000	1.000	1.000	1.000	0.837	0.946	1.000	0.345	0.784
4	90	7	0.627	0.556	0.812	0.263	0.886	0.365	0.315	0.774	0.512	0.361	0.710	0.606	·	·	·	0.381	0.912	0.620	0.607	0.910	0.836	0.696	0.718	0.856	0.705	0.306	0.801
		10	0.679	0.688	0.845	0.302	0.881	0.454	0.406	0.820	0.640	0.430	0.800	0.675	0.515	0.711	0.744	0.385	0.941	0.622	0.622	0.932	0.847	0.782	0.796	0.897	0.779	0.342	0.817
		20	0.788	0.857	0.905	0.338	0.885	0.542	0.557	0.904	0.793	0.594	0.922	0.824	0.633	0.845	0.841	0.385	0.977	0.617	0.640	0.975	0.861	0.900	0.907	0.953	0.883	0.404	0.843
		40	0.936	0.947	0.971	0.370	0.916	0.599	0.637	0.984	0.860	0.652	0.967	0.866	0.666	0.855	0.860	0.386	0.997	0.616	0.646	0.996	0.867	0.977	0.978	0.989	0.961	0.464	0.866
4	180	7	0.741	0.460	0.690	0.485	0.978	0.526	0.514	0.867	0.570	0.525	0.771	0.584	·	·	·	0.571	0.959	0.663	0.710	0.823	0.774	0.734	0.503	0.699	0.757	0.179	0.611
		10	0.778	0.596	0.746	0.512	0.966	0.569	0.582	0.926	0.663	0.608	0.903	0.691	0.582	0.753	0.648	0.575	0.990	0.671	0.717	0.861	0.789	0.785	0.549	0.734	0.806	0.190	0.621
		20	0.867	0.860	0.877	0.538	0.946	0.613	0.694	0.947	0.791	0.719	0.967	0.818	0.742	0.922	0.823	0.569	0.997	0.664	0.736	0.969	0.833	0.895	0.714	0.837	0.896	0.214	0.641
		40	0.954	0.946	0.955	0.559	0.939	0.645	0.738	0.996	0.842	0.736	0.968	0.833	0.739	0.947	0.828	0.565	1.000	0.659	0.740	0.997	0.844	0.982	0.941	0.966	0.967	0.239	0.659

Notes: This table reports the precision (P) rate, recall (R) rate, and Rand index (RI) for the triad pairwise-difference (TPWD) estimator with $\widehat{\beta}^1 = 0$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2 \log(T)/\sqrt{T}$, for the post-spectral estimator proposed in Chetverikov and Manresa (2021) with $\tilde{\beta}^0 = \tilde{\beta}^1 = 0$, user-specified number of groups $g \in \{2, 3, 4, 10\}$, and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq g$ (Post-Spectral$^{\bar{G}=g}$), and for Bonhomme and Manresa (2015)'s grouped fixed-effects estimator (their "Algorithm 1") with a user-specified number of groups $g \in \{2, 3, 4, 10\}$ and 500 random initialization points (GFE$^{\bar{G}=g}$). A green-shaded cell corresponds to a well-specified estimator. Results for Post-Spectral$^{\bar{G}=10}$ are missing if $T = 7$ since Post-Spectral$^{\bar{G}=g}$ is not properly defined if $g > T$ as it requires computing the g largest eigenvectors of a $T \times T$ matrix. 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–S6 show that the precision of TPWD and GFE deteriorates for small values of T. Both estimators remain comparable.

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

			TPWD			Post-Spectral$^{\bar{G}=2}$		Post-Spectral$^{\bar{G}=3}$		Post-Spectral$^{\bar{G}=4}$		Post-Spectral$^{\bar{G}=10}$		GFE$^{\bar{G}=2}$	GFE$^{\bar{G}=3}$	GFE$^{\bar{G}=4}$	GFE$^{\bar{G}=10}$	Oracle
G	N	T	$\widehat{G}$	RMSE	CPU time	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	$\widehat{G}$	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
3	90	7	12.832	0.450	0.058	1.814	0.446	2.494	0.451	3.116	0.455	·	·	0.330	0.318	0.375	0.494	0.120
		10	7.312	0.336	0.076	1.910	0.434	2.542	0.439	3.230	0.443	7.336	0.456	0.309	0.263	0.319	0.445	0.120
		20	3.456	0.207	0.138	1.960	0.399	2.644	0.387	3.162	0.389	6.294	0.401	0.288	0.175	0.230	0.360	0.121
		40	3.018	0.142	0.292	1.992	0.340	2.754	0.317	3.480	0.315	7.200	0.341	0.272	0.130	0.175	0.296	0.122
3	180	7	23.976	0.473	0.503	1.876	0.441	2.556	0.444	3.056	0.446	·	·	0.320	0.288	0.345	0.463	0.084
		10	12.534	0.340	0.662	1.926	0.431	2.608	0.431	3.184	0.432	6.574	0.439	0.300	0.230	0.286	0.405	0.085
		20	4.020	0.186	1.206	1.990	0.388	2.634	0.375	3.202	0.375	6.070	0.387	0.280	0.144	0.193	0.308	0.087
		40	3.086	0.111	2.470	2.000	0.328	2.850	0.287	3.432	0.285	6.854	0.305	0.262	0.095	0.136	0.238	0.086
4	90	7	14.040	0.468	0.053	1.772	0.410	2.492	0.417	3.084	0.424	·	·	0.343	0.334	0.382	0.498	0.140
		10	7.680	0.352	0.067	1.874	0.407	2.476	0.411	3.112	0.417	7.296	0.443	0.310	0.285	0.330	0.451	0.139
		20	3.542	0.239	0.120	1.950	0.387	2.612	0.386	3.280	0.389	6.436	0.405	0.273	0.217	0.252	0.372	0.140
		40	3.070	0.193	0.256	1.992	0.353	2.718	0.338	3.378	0.344	7.322	0.360	0.254	0.179	0.194	0.312	0.140
4	180	7	26.254	0.492	0.428	1.784	0.311	2.454	0.318	2.902	0.321	·	·	0.321	0.339	0.370	0.473	0.097
		10	15.356	0.371	0.564	1.886	0.314	2.524	0.318	3.060	0.322	6.758	0.346	0.275	0.290	0.318	0.420	0.099
		20	4.502	0.200	1.083	1.976	0.303	2.592	0.305	3.156	0.308	6.202	0.324	0.207	0.212	0.237	0.331	0.099
		40	3.220	0.154	2.375	2.000	0.291	2.732	0.283	3.342	0.282	6.912	0.310	0.190	0.161	0.179	0.266	0.099

Notes: This table reports (i) the estimated number of groups ($\widehat{G}$), the root mean square error (RMSE), and the execution time in seconds (CPU time) for the triad pairwise-difference (TPWD) estimator computed with $\widehat{\beta}^1 = 0$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2 \log(T)/\sqrt{T}$; (ii) $\widehat{G}$ and the RMSE for the post-spectral estimator proposed in Chetverikov and Manresa (2021) with $\tilde{\beta}^0 = \tilde{\beta}^1 = 0$, user-specified number of groups $g \in \{2, 3, 4, 10\}$, and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq g$ (Post-Spectral$^{\bar{G}=g}$); and (iii) the RMSE for Bonhomme and Manresa (2015)'s grouped fixed-effects estimator (their "Algorithm 1") with a user-specified number of groups $g \in \{2, 3, 4, 10\}$ and 500 random initialization points (GFE$^{\bar{G}=g}$) and for the infeasible (Oracle) estimator using the "true" group memberships. A green-shaded cell corresponds to a well-specified estimator. Results for Post-Spectral$^{\bar{G}=10}$ are missing if $T = 7$ since Post-Spectral$^{\bar{G}=g}$ is not properly defined if $g > T$ as it requires computing the g largest eigenvectors of a $T \times T$ matrix. Results are averaged across 500 Monte Carlo samples.

Table S6: Estimation of group memberships (lower signal-to-noise ratio)

			TPWD			Post-Spectral$^{\bar{G}=2}$			Post-Spectral$^{\bar{G}=3}$			Post-Spectral$^{\bar{G}=4}$			Post-Spectral$^{\bar{G}=10}$			GFE$^{\bar{G}=2}$			GFE$^{\bar{G}=3}$			GFE$^{\bar{G}=4}$			GFE$^{\bar{G}=10}$		
G	N	T	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI
3	90	7	0.662	0.282	0.719	0.329	0.934	0.357	0.332	0.883	0.381	0.336	0.850	0.398	·	·	·	0.525	0.812	0.699	0.674	0.692	0.790	0.671	0.529	0.762	0.667	0.211	0.709
		10	0.711	0.482	0.767	0.337	0.905	0.383	0.337	0.864	0.397	0.338	0.832	0.410	0.373	0.655	0.520	0.541	0.837	0.715	0.756	0.768	0.843	0.752	0.595	0.804	0.728	0.236	0.722
		20	0.847	0.838	0.898	0.373	0.863	0.462	0.396	0.804	0.517	0.398	0.777	0.531	0.420	0.686	0.575	0.559	0.880	0.735	0.911	0.915	0.943	0.906	0.738	0.890	0.869	0.298	0.757
		40	0.966	0.968	0.978	0.460	0.874	0.599	0.510	0.818	0.671	0.523	0.787	0.689	0.538	0.679	0.701	0.576	0.937	0.755	0.988	0.988	0.992	0.987	0.824	0.939	0.969	0.354	0.786
3	180	7	0.687	0.186	0.704	0.332	0.939	0.356	0.334	0.906	0.372	0.335	0.894	0.377	·	·	·	0.531	0.811	0.702	0.703	0.712	0.806	0.697	0.542	0.772	0.680	0.211	0.707
		10	0.744	0.385	0.754	0.338	0.914	0.378	0.341	0.875	0.396	0.341	0.849	0.406	0.358	0.767	0.458	0.546	0.833	0.717	0.787	0.793	0.861	0.781	0.615	0.816	0.755	0.237	0.723
		20	0.867	0.821	0.899	0.385	0.861	0.479	0.404	0.809	0.526	0.405	0.785	0.536	0.407	0.740	0.545	0.562	0.871	0.734	0.926	0.927	0.951	0.921	0.752	0.897	0.893	0.293	0.755
		40	0.970	0.970	0.980	0.470	0.879	0.611	0.547	0.799	0.709	0.559	0.780	0.719	0.555	0.753	0.717	0.581	0.939	0.757	0.990	0.990	0.993	0.989	0.826	0.940	0.981	0.339	0.780
4	90	7	0.448	0.253	0.743	0.244	0.950	0.276	0.246	0.910	0.302	0.248	0.877	0.324	·	·	·	0.355	0.774	0.606	0.460	0.651	0.730	0.476	0.499	0.746	0.484	0.207	0.755
		10	0.479	0.464	0.747	0.249	0.923	0.302	0.251	0.886	0.328	0.252	0.856	0.346	0.274	0.682	0.477	0.373	0.833	0.621	0.508	0.719	0.763	0.533	0.559	0.775	0.535	0.232	0.765
		20	0.529	0.771	0.776	0.269	0.882	0.376	0.277	0.836	0.418	0.282	0.808	0.443	0.296	0.713	0.506	0.386	0.908	0.628	0.579	0.837	0.813	0.645	0.667	0.830	0.633	0.290	0.788
		40	0.591	0.902	0.823	0.322	0.869	0.506	0.349	0.805	0.574	0.350	0.773	0.588	0.375	0.669	0.640	0.386	0.957	0.621	0.624	0.930	0.848	0.781	0.797	0.896	0.745	0.365	0.816
4	180	7	0.627	0.144	0.583	0.453	0.990	0.466	0.457	0.979	0.475	0.460	0.974	0.481	·	·	·	0.508	0.630	0.568	0.563	0.453	0.603	0.574	0.336	0.596	0.589	0.137	0.575
		10	0.659	0.326	0.627	0.461	0.984	0.482	0.466	0.973	0.491	0.468	0.967	0.497	0.493	0.901	0.539	0.557	0.785	0.630	0.607	0.516	0.638	0.616	0.370	0.619	0.632	0.147	0.584
		20	0.689	0.810	0.752	0.495	0.956	0.543	0.507	0.936	0.562	0.507	0.923	0.565	0.517	0.878	0.577	0.584	0.971	0.680	0.659	0.639	0.695	0.683	0.434	0.660	0.709	0.170	0.602
		40	0.705	0.953	0.798	0.534	0.926	0.606	0.546	0.822	0.620	0.563	0.827	0.638	0.557	0.747	0.625	0.575	0.994	0.672	0.710	0.847	0.781	0.776	0.542	0.729	0.797	0.202	0.624

Notes: This table reports the precision (P) rate, recall (R) rate, and Rand index (RI) for the triad pairwise-difference (TPWD) estimator with $\widehat{\beta}^1 = 0$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2 \log(T)/\sqrt{T}$, for the post-spectral estimator proposed in Chetverikov and Manresa (2021) with $\tilde{\beta}^0 = \tilde{\beta}^1 = 0$, user-specified number of groups $g \in \{2, 3, 4, 10\}$, and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq g$ (Post-Spectral$^{\bar{G}=g}$), and for Bonhomme and Manresa (2015)'s grouped fixed-effects estimator (their "Algorithm 1") with a user-specified number of groups $g \in \{2, 3, 4, 10\}$ and 500 random initialization points (GFE$^{\bar{G}=g}$). A green-shaded cell corresponds to a well-specified estimator. Results for Post-Spectral$^{\bar{G}=10}$ are missing if $T = 7$ since Post-Spectral$^{\bar{G}=g}$ is not properly defined if $g > T$ as it requires computing the g largest eigenvectors of a $T \times T$ matrix. Results are averaged across 500 Monte Carlo samples.

S3.4 Heteroscedasticity and serially correlated errors

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

$$\begin{cases} y_{it} = x_{it}\beta + \alpha_{g_i t} + (x_{it}\beta)v_{it}, \\ v_{it} = 0.2v_{it-1} + \varepsilon_{it}, \\ \mathbb{E}[\varepsilon_{it}] = \mathbb{E}[x_{it}\varepsilon_{it}] = \mathbb{E}[v_{i0}] = \mathbb{E}[x_{it}v_{i0}] = 0, \quad \mathbb{E}[\varepsilon_{it}^2] = 1/9, \quad i = 1, \dots, N, \ t = 1, \dots, T. \end{cases}$$

Tables S7–S8 display excellent results for all clustering methods, though Post-Spectral estimators are still dominated for small values of T. Interactive fixed effects methods have lower coverage.

Table S7: Estimation of the slope coefficient and grouped fixed effects (heteroscedastic and serially correlated errors)

			TPWD					Iterated TPWD					NNR		NN		Spectral		Post-Spectral					GFE			
G	N	T	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$
3	90	7	0.024	0.042	0.810	0.066	4.492	0.000	0.029	0.928	0.044	3.708	0.334	0.336	0.324	0.327	0.040	0.106	0.141	0.216	0.492	0.156	2.776	0.000	0.029	0.930	0.031
		10	0.007	0.028	0.884	0.043	3.648	0.001	0.026	0.904	0.034	3.272	0.290	0.292	0.236	0.245	0.024	0.050	0.010	0.053	0.884	0.041	2.972	0.001	0.026	0.908	0.029
		20	0.000	0.016	0.952	0.030	3.078	0.000	0.016	0.954	0.029	3.018	0.240	0.247	0.148	0.149	0.012	0.023	0.001	0.023	0.954	0.030	2.998	0.000	0.016	0.958	0.029
		40	0.000	0.012	0.946	0.029	3.004	0.000	0.012	0.946	0.029	3.002	0.193	0.201	0.099	0.100	0.008	0.016	0.000	0.012	0.946	0.029	3.000	0.000	0.012	0.946	0.029
3	180	7	0.021	0.032	0.768	0.062	5.300	0.000	0.020	0.930	0.040	4.092	0.319	0.331	0.320	0.321	0.047	0.088	0.165	0.220	0.372	0.176	2.746	0.000	0.021	0.940	0.023
		10	0.006	0.019	0.900	0.037	4.020	0.000	0.017	0.936	0.028	3.456	0.283	0.284	0.237	0.238	0.028	0.041	0.012	0.053	0.898	0.036	2.966	0.001	0.017	0.938	0.021
		20	0.000	0.012	0.946	0.022	3.132	0.000	0.012	0.950	0.021	3.040	0.239	0.239	0.145	0.146	0.013	0.019	0.002	0.024	0.944	0.022	3.000	0.000	0.012	0.950	0.021
		40	0.000	0.009	0.934	0.021	3.012	0.000	0.009	0.934	0.021	3.002	0.192	0.199	0.097	0.098	0.007	0.012	0.000	0.009	0.934	0.021	3.000	0.000	0.009	0.934	0.021
4	90	7	0.050	0.062	0.580	0.097	4.886	0.021	0.035	0.864	0.081	3.892	0.339	0.342	0.328	0.331	−0.532	4.533	0.198	0.283	0.450	0.219	3.636	0.005	0.028	0.932	0.045
		10	0.032	0.044	0.676	0.080	4.242	0.012	0.031	0.840	0.059	4.008	0.296	0.298	0.242	0.244	0.037	0.070	0.046	0.067	0.624	0.104	3.400	0.002	0.024	0.932	0.037
		20	0.003	0.017	0.934	0.038	4.062	0.001	0.017	0.944	0.034	4.024	0.246	0.247	0.147	0.149	0.019	0.030	0.039	0.044	0.388	0.097	3.246	0.000	0.016	0.948	0.033
		40	0.000	0.012	0.942	0.033	4.004	0.000	0.012	0.942	0.033	4.000	0.199	0.199	0.098	0.099	0.012	0.018	0.026	0.036	0.430	0.075	3.538	0.000	0.012	0.942	0.032
4	180	7	0.036	0.043	0.510	0.083	6.380	0.013	0.024	0.846	0.070	5.014	0.282	0.306	0.278	0.280	−2.549	37.246	0.163	0.211	0.348	0.196	4.222	0.004	0.022	0.896	0.049
		10	0.019	0.028	0.742	0.060	5.094	0.003	0.018	0.920	0.041	4.570	0.259	0.260	0.199	0.200	0.036	0.056	0.031	0.039	0.616	0.085	3.394	0.002	0.017	0.918	0.030
		20	0.002	0.012	0.934	0.030	4.236	0.001	0.012	0.942	0.026	4.114	0.215	0.222	0.120	0.121	0.019	0.024	0.032	0.035	0.286	0.087	3.294	0.000	0.012	0.948	0.023
		40	0.001	0.008	0.950	0.023	4.036	0.000	0.008	0.944	0.023	4.014	0.176	0.176	0.080	0.080	0.012	0.015	0.022	0.028	0.374	0.068	3.788	0.000	0.008	0.944	0.023

			GFE with BIC selection					IFE R			IFE R-BC			IFE NNR			IFE NNR-BC			Oracle			
G	N	T	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$
3	90	7	−0.001	0.030	0.908	0.068	4.490	−0.002	0.035	0.682	−0.002	0.036	0.670	−0.002	0.035	0.682	−0.002	0.036	0.670	−0.001	0.029	0.938	0.029
		10	0.001	0.026	0.888	0.054	3.936	0.001	0.027	0.712	0.001	0.028	0.706	0.001	0.027	0.712	0.001	0.028	0.706	0.000	0.026	0.912	0.029
		20	0.000	0.016	0.958	0.030	3.020	−0.001	0.017	0.764	0.000	0.017	0.770	−0.001	0.017	0.764	0.000	0.017	0.770	0.000	0.016	0.958	0.029
		40	0.000	0.012	0.946	0.029	3.000	0.000	0.012	0.810	0.000	0.012	0.798	0.000	0.012	0.810	0.000	0.012	0.798	0.000	0.012	0.946	0.029
3	180	7	0.000	0.022	0.918	0.065	5.000	−0.001	0.024	0.704	−0.001	0.025	0.702	−0.001	0.024	0.704	−0.001	0.025	0.702	−0.001	0.020	0.942	0.020
		10	0.001	0.018	0.930	0.057	4.858	0.001	0.019	0.746	0.001	0.019	0.730	0.001	0.019	0.746	0.001	0.019	0.730	0.000	0.017	0.940	0.021
		20	0.000	0.012	0.940	0.032	3.556	0.000	0.012	0.788	0.000	0.012	0.788	0.000	0.012	0.788	0.000	0.012	0.788	0.000	0.012	0.950	0.021
		40	0.000	0.009	0.934	0.021	3.000	0.000	0.009	0.790	0.000	0.009	0.786	0.000	0.009	0.790	0.000	0.009	0.786	0.000	0.009	0.934	0.021
4	90	7	0.005	0.028	0.914	0.063	4.836	0.004	0.036	0.594	0.004	0.037	0.578	0.004	0.036	0.594	0.004	0.037	0.578	0.001	0.027	0.946	0.032
		10	0.003	0.024	0.928	0.047	4.430	0.004	0.028	0.656	0.004	0.029	0.654	0.004	0.028	0.656	0.004	0.029	0.652	0.001	0.024	0.942	0.033
		20	0.000	0.016	0.948	0.033	4.012	0.001	0.018	0.736	0.001	0.018	0.732	0.001	0.018	0.736	0.001	0.018	0.732	0.000	0.016	0.948	0.033
		40	0.000	0.012	0.942	0.032	4.000	0.000	0.012	0.756	0.000	0.012	0.756	0.000	0.012	0.756	0.000	0.012	0.756	0.000	0.012	0.942	0.032
4	180	7	0.003	0.023	0.868	0.064	5.000	0.003	0.027	0.602	0.005	0.028	0.556	0.003	0.027	0.602	0.005	0.028	0.556	0.000	0.020	0.948	0.023
		10	0.001	0.018	0.898	0.050	5.000	0.002	0.019	0.634	0.002	0.019	0.640	0.002	0.019	0.634	0.002	0.019	0.640	0.001	0.017	0.936	0.023
		20	0.000	0.012	0.932	0.034	4.646	0.001	0.012	0.722	0.001	0.012	0.722	0.001	0.012	0.722	0.001	0.012	0.722	0.000	0.012	0.948	0.023
		40	0.000	0.008	0.944	0.023	4.000	0.001	0.008	0.756	0.001	0.008	0.760	0.001	0.008	0.756	0.001	0.008	0.760	0.000	0.008	0.944	0.023

Notes: This table reports the bias and root mean square error (RMSE) of each estimator of β. When relevant, it also reports the coverage rate of a 95% confidence interval based on large-N, large-T consistent estimates of analytical standard errors (.95 $\widehat{\beta}$), the RMSE of the grouped fixed-effects estimator, and the estimated number of groups ($\widehat{G}$). The triad pairwise-difference (TPWD) estimator is computed with $\widehat{\beta}^1$ set to the nuclear norm regularized (NNR) estimator with $\psi_{NT} = \log(\log(T))/\sqrt{16\min(N,T)}$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2\log(T)/\sqrt{T}$. Iterated TPWD is computed by iterating TPWD 4 times, replacing 3 times the preliminary estimator $\widehat{\beta}^1$ by the TPWD estimate obtained at the previous iteration. NNR is obtained by concentrating out the optimization with respect to the unobserved effects and solving the convex optimization problem with MATLAB `fminsearch` routine. The nuclear norm (NN) estimator is obtained by solving a convex optimization problem with MATLAB `fminsearch` routine. The Spectral estimator is implemented as in Chetverikov and Manresa (2021), with user-specified number of groups for the outcome and regressor equations set to the true number of groups. The Post-Spectral estimator is implemented as in Chetverikov and Manresa (2021), with user-specified number of groups set to the true number of groups and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq G$. The grouped fixed-effects (GFE) estimator proposed in Bonhomme and Manresa (2015) is implemented as in Bonhomme and Manresa (2015)'s Algorithm 1, with a user-specified number of groups set to the true number of groups and 100 random initialization points following the method described in Section S.1.1 Bonhomme and Manresa (2015)'s Supplementary Material with $\theta^{(0)} \sim \mathcal{N}(0,1)$. The GFE with BIC selection of the number of groups is implemented as in Section S.3.2 of Bonhomme and Manresa (2015), with $G_{max} = 5$. Bai (2009)'s interactive fixed effects (IFE) estimator is implemented using 100 random initialization points (IFE R) or taking one optimization step starting from NNR (IFE NNR) and applying bias-correction based on large-N, large-T approximations (IFE R-BC, IFE NNR-BC). The infeasible (Oracle) estimator is obtained from a pooled OLS regression of the outcome on the covariates controlling for the interactions of time and "true" group dummies. Results are averaged across 500 Monte Carlo samples.

Table S8: Classification accuracy (heteroscedastic and serially correlated errors)

			TPWD			Iterated TPWD			Post-Spectral			GFE			GFE with BIC selection		
G	N	T	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI
3	90	7	0.991	0.903	0.966	1.000	0.973	0.991	0.784	0.930	0.863	0.999	0.999	0.999	0.999	0.820	0.941
		10	0.998	0.970	0.990	1.000	0.992	0.997	0.984	0.996	0.991	1.000	1.000	1.000	1.000	0.878	0.960
		20	1.000	0.998	0.999	1.000	1.000	1.000	0.998	0.999	0.999	1.000	1.000	1.000	1.000	0.998	0.999
		40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	180	7	0.994	0.894	0.964	0.999	0.970	0.990	0.744	0.925	0.843	0.999	0.999	0.999	0.999	0.751	0.918
		10	0.999	0.968	0.989	1.000	0.991	0.997	0.980	0.997	0.989	1.000	1.000	1.000	1.000	0.767	0.923
		20	1.000	0.998	0.999	1.000	1.000	1.000	0.997	0.999	0.998	1.000	1.000	1.000	1.000	0.913	0.971
		40	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	90	7	0.662	0.860	0.857	0.651	0.961	0.865	0.476	0.883	0.684	0.952	0.955	0.977	0.951	0.880	0.960
		10	0.822	0.920	0.923	0.891	0.973	0.956	0.638	0.978	0.858	0.991	0.991	0.996	0.990	0.954	0.987
		20	0.987	0.988	0.994	0.998	0.997	0.999	0.672	0.992	0.876	0.999	0.999	1.000	0.999	0.999	1.000
		40	1.000	0.999	1.000	1.000	1.000	1.000	0.795	0.990	0.922	1.000	1.000	1.000	1.000	1.000	1.000
4	180	7	0.796	0.872	0.841	0.790	0.956	0.863	0.610	0.926	0.688	0.954	0.885	0.931	0.961	0.585	0.806
		10	0.940	0.938	0.945	0.978	0.975	0.979	0.739	0.992	0.841	0.995	0.990	0.993	0.994	0.590	0.817
		20	0.993	0.993	0.994	0.998	0.998	0.998	0.747	0.942	0.832	1.000	1.000	1.000	1.000	0.733	0.882
		40	1.000	1.000	1.000	1.000	1.000	1.000	0.843	0.849	0.859	1.000	1.000	1.000	1.000	1.000	1.000

Notes: This table reports the precision (P) rate, recall (R) rate, and Rand index (RI) for each estimator. The triad pairwise-difference (TPWD) estimator is computed with $\widehat{\beta}^1$ set to the nuclear norm regularized (NNR) estimator with $\psi_{NT} = \log(\log(T))/\sqrt{16\min(N,T)}$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2\log(T)/\sqrt{T}$. Iterated TPWD is computed by iterating TPWD 4 times, replacing 3 times the preliminary estimator $\widehat{\beta}^1$ by the TPWD estimate obtained at the previous iteration. The Post-Spectral estimator is implemented as in Chetverikov and Manresa (2021), with user-specified number of groups set to the true number of groups and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq G$. The grouped fixed-effects (GFE) estimator proposed in Bonhomme and Manresa (2015) is implemented as in Bonhomme and Manresa (2015)'s Algorithm 1, with a user-specified number of groups set to the true number of groups and 100 random initialization points following the method described in Section S.1.1 Bonhomme and Manresa (2015)'s Supplementary Material with $\theta^{(0)} \sim \mathcal{N}(0,1)$. The GFE with BIC selection of the number of groups is implemented as in Section S.3.2 of Bonhomme and Manresa (2015), with $G_{max} = 5$. Results are averaged across 500 Monte Carlo samples.

S3.5 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 effect 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 S9–S10 show the results. The TPWD estimator has less bias than the NNR estimator, but the bias dominates the variance. Clustering consistency is much slower, which results in poor coverage of asymptotic confidence intervals in some settings. TPWD coverage uniformly dominates that of the post-spectral estimator.

Table S9: Estimation of the slope coefficient and grouped fixed effects (unit-specific effects)

			TPWD					Iterated TPWD					NNR		NN		Spectral		Post-Spectral					GFE			
G	N	T	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$
3	90	7	0.321	0.332	0.002	0.489	14.848	0.047	0.078	0.668	0.449	7.474	0.580	0.589	0.583	0.588	Inf	Inf	−0.280	0.299	0.150	0.567	2.624	0.041	0.069	0.736	0.435
		10	0.142	0.153	0.082	0.453	7.406	0.029	0.052	0.794	0.432	4.244	0.453	0.456	0.453	0.456	0.049	0.112	−0.269	0.285	0.110	0.560	2.784	0.024	0.049	0.818	0.427
		20	0.066	0.077	0.318	0.430	2.476	0.019	0.040	0.806	0.420	2.900	0.302	0.304	0.294	0.296	0.021	0.046	−0.263	0.267	0.000	0.552	2.712	0.008	0.027	0.918	0.417
		40	0.063	0.068	0.098	0.429	2.070	0.020	0.040	0.652	0.418	2.694	0.224	0.225	0.201	0.202	0.011	0.024	−0.220	0.225	0.008	0.562	2.846	0.000	0.017	0.938	0.413
3	180	7	0.345	0.351	0.000	0.495	27.920	0.049	0.066	0.518	0.452	13.226	0.574	0.580	0.577	0.579	−0.660	2.885	−0.269	0.285	0.058	0.568	3.414	0.034	0.049	0.728	0.430
		10	0.150	0.156	0.002	0.455	14.214	0.025	0.037	0.786	0.430	6.298	0.446	0.452	0.449	0.450	0.053	0.084	−0.267	0.282	0.038	0.559	3.368	0.022	0.035	0.814	0.423
		20	0.035	0.042	0.520	0.422	3.238	0.009	0.020	0.910	0.416	3.070	0.285	0.292	0.286	0.287	0.019	0.032	−0.261	0.264	0.000	0.550	2.766	0.005	0.018	0.926	0.414
		40	0.019	0.036	0.706	0.416	2.774	0.001	0.013	0.926	0.411	3.000	0.209	0.217	0.196	0.197	0.008	0.018	−0.206	0.212	0.002	0.567	2.834	0.000	0.013	0.934	0.411
4	90	7	0.328	0.340	0.000	0.478	13.770	0.050	0.073	0.712	0.432	5.980	0.616	0.620	0.616	0.620	−1.888	40.992	−0.228	0.252	0.246	0.578	3.274	0.053	0.077	0.650	0.434
		10	0.149	0.157	0.044	0.429	6.416	0.043	0.059	0.742	0.407	3.640	0.478	0.481	0.478	0.481	0.086	0.252	−0.229	0.252	0.212	0.551	3.096	0.036	0.055	0.774	0.413
		20	0.079	0.086	0.174	0.404	2.272	0.043	0.055	0.576	0.396	2.600	0.315	0.322	0.310	0.312	0.024	0.050	−0.254	0.258	0.000	0.520	3.152	0.014	0.030	0.876	0.394
		40	0.071	0.074	0.044	0.399	2.006	0.058	0.064	0.204	0.395	2.176	0.237	0.245	0.216	0.217	0.016	0.029	−0.175	0.181	0.028	0.540	3.618	0.006	0.019	0.906	0.385
4	180	7	0.322	0.328	0.000	0.459	21.180	0.031	0.047	0.746	0.403	5.774	0.636	0.638	0.636	0.638	−0.292	6.249	−0.203	0.214	0.076	0.566	3.632	0.045	0.059	0.584	0.415
		10	0.117	0.124	0.046	0.384	8.756	0.030	0.041	0.744	0.362	3.746	0.484	0.490	0.487	0.489	0.101	0.178	−0.188	0.198	0.056	0.521	3.594	0.038	0.049	0.614	0.376
		20	0.050	0.056	0.272	0.349	2.280	0.031	0.040	0.558	0.344	2.462	0.309	0.315	0.309	0.310	0.018	0.034	−0.247	0.253	0.000	0.464	3.238	0.014	0.024	0.824	0.347
		40	0.041	0.044	0.130	0.339	2.000	0.036	0.039	0.226	0.337	2.068	0.226	0.227	0.210	0.211	0.011	0.020	−0.135	0.141	0.014	0.494	3.630	0.008	0.015	0.882	0.333

			GFE with BIC selection					IFE R			IFE R-BC			IFE NNR			IFE NNR-BC			Oracle			
G	N	T	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$	$\widehat{G}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	Bias $\widehat{\beta}$	RMSE $\widehat{\beta}$	.95 $\widehat{\beta}$	RMSE $\widehat{\alpha}$
3	90	7	0.049	0.076	0.666	0.449	4.514	0.002	0.068	0.762	0.001	0.069	0.764	0.002	0.068	0.764	0.001	0.069	0.766	−0.001	0.046	0.926	0.412
		10	0.026	0.050	0.802	0.429	3.188	0.001	0.049	0.792	0.001	0.049	0.794	0.001	0.049	0.792	0.001	0.049	0.794	−0.001	0.037	0.924	0.412
		20	0.008	0.027	0.916	0.417	2.996	0.001	0.028	0.886	0.001	0.028	0.878	0.001	0.028	0.886	0.001	0.028	0.878	0.001	0.025	0.944	0.413
		40	0.000	0.017	0.938	0.413	3.000	0.000	0.019	0.916	0.000	0.019	0.910	0.000	0.019	0.916	0.000	0.019	0.910	−0.001	0.017	0.946	0.413
3	180	7	0.046	0.060	0.572	0.447	5.000	−0.001	0.047	0.782	−0.001	0.048	0.770	−0.001	0.047	0.784	−0.001	0.047	0.772	0.000	0.030	0.956	0.410
		10	0.028	0.041	0.736	0.435	4.818	0.001	0.032	0.854	0.001	0.032	0.846	0.001	0.032	0.854	0.001	0.032	0.846	0.000	0.025	0.932	0.410
		20	0.005	0.018	0.926	0.414	3.000	0.000	0.019	0.910	0.000	0.019	0.900	0.000	0.019	0.910	0.000	0.019	0.900	−0.001	0.016	0.952	0.410
		40	0.000	0.013	0.934	0.411	3.000	0.000	0.013	0.914	0.000	0.013	0.910	0.000	0.013	0.914	0.000	0.013	0.910	0.000	0.012	0.936	0.410
4	90	7	0.054	0.079	0.642	0.438	4.554	0.007	0.091	0.642	0.008	0.092	0.628	0.008	0.091	0.642	0.009	0.092	0.630	0.000	0.045	0.924	0.401
		10	0.038	0.056	0.772	0.410	3.544	0.006	0.054	0.776	0.007	0.054	0.764	0.006	0.054	0.776	0.007	0.054	0.764	−0.001	0.034	0.954	0.388
		20	0.023	0.037	0.798	0.391	2.922	0.007	0.031	0.854	0.007	0.031	0.848	0.007	0.031	0.854	0.007	0.031	0.848	0.000	0.024	0.946	0.384
		40	0.022	0.034	0.750	0.386	2.846	0.005	0.019	0.894	0.005	0.020	0.882	0.005	0.019	0.894	0.005	0.020	0.882	0.001	0.017	0.932	0.381
4	180	7	0.041	0.057	0.590	0.420	5.000	0.006	0.059	0.700	0.007	0.061	0.676	0.006	0.059	0.702	0.007	0.061	0.674	−0.002	0.031	0.948	0.380
		10	0.031	0.044	0.658	0.380	4.998	0.007	0.038	0.782	0.008	0.039	0.756	0.007	0.038	0.782	0.008	0.039	0.756	0.000	0.025	0.944	0.347
		20	0.017	0.027	0.768	0.347	3.792	0.003	0.020	0.874	0.003	0.021	0.860	0.003	0.020	0.874	0.003	0.021	0.860	0.000	0.017	0.946	0.334
		40	0.018	0.025	0.662	0.332	2.754	0.003	0.014	0.898	0.003	0.014	0.900	0.003	0.014	0.898	0.003	0.014	0.900	0.001	0.012	0.950	0.327

Notes: This table reports the bias and root mean square error (RMSE) of each estimator of β. When relevant, it also reports the coverage rate of a 95% confidence interval based on large-N, large-T consistent estimates of analytical standard errors (.95 $\widehat{\beta}$), the RMSE of the grouped fixed-effects estimator, and the estimated number of groups ($\widehat{G}$). The triad pairwise-difference (TPWD) estimator is computed with $\widehat{\beta}^1$ set to the nuclear norm regularized (NNR) estimator with $\psi_{NT} = \log(\log(T))/\sqrt{16\min(N,T)}$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2\log(T)/\sqrt{T}$. Iterated TPWD is computed by iterating TPWD 4 times, replacing 3 times the preliminary estimator $\widehat{\beta}^1$ by the TPWD estimate obtained at the previous iteration. NNR is obtained by concentrating out the optimization with respect to the unobserved effects and solving the convex optimization problem with MATLAB `fminsearch` routine. The nuclear norm (NN) estimator is obtained by solving a convex optimization problem with MATLAB `fminsearch` routine. The Spectral estimator is implemented as in Chetverikov and Manresa (2021), with user-specified number of groups for the outcome and regressor equations set to the true number of groups. The Post-Spectral estimator is implemented as in Chetverikov and Manresa (2021), with user-specified number of groups set to the true number of groups and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq G$. The grouped fixed-effects (GFE) estimator proposed in Bonhomme and Manresa (2015) is implemented as in Bonhomme and Manresa (2015)'s Algorithm 1, with a user-specified number of groups set to the true number of groups and 100 random initialization points following the method described in Section S.1.1 Bonhomme and Manresa (2015)'s Supplementary Material with $\theta^{(0)} \sim \mathcal{N}(0,1)$. The GFE with BIC selection of the number of groups is implemented as in Section S.3.2 of Bonhomme and Manresa (2015), with $G_{max} = 5$. Bai (2009)'s interactive fixed effects (IFE) estimator is implemented using 100 random initialization points (IFE R) or taking one optimization step starting from NNR (IFE NNR) and applying bias-correction based on large-N, large-T approximations (IFE R-BC, IFE NNR-BC). The infeasible (Oracle) estimator is obtained from a pooled OLS regression of the outcome on the covariates controlling for the interactions of time and "true" group dummies. Results are averaged across 500 Monte Carlo samples.

Table S10: Classification accuracy (unit-specific effects)

			TPWD			Iterated TPWD			Post-Spectral			GFE			GFE with BIC selection		
G	N	T	P	R	RI	P	R	RI	P	R	RI	P	R	RI	P	R	RI
3	90	7	0.523	0.211	0.681	0.701	0.483	0.765	0.440	0.881	0.565	0.716	0.732	0.818	0.715	0.505	0.774
		10	0.591	0.479	0.721	0.752	0.709	0.829	0.476	0.898	0.614	0.793	0.803	0.867	0.791	0.769	0.859
		20	0.653	0.880	0.790	0.860	0.914	0.915	0.564	0.917	0.733	0.925	0.928	0.952	0.923	0.928	0.951
		40	0.611	0.982	0.781	0.863	0.987	0.923	0.539	0.906	0.690	0.989	0.990	0.993	0.989	0.990	0.993
3	180	7	0.542	0.129	0.677	0.720	0.375	0.746	0.431	0.886	0.546	0.746	0.753	0.834	0.730	0.456	0.765
		10	0.623	0.351	0.716	0.782	0.659	0.827	0.475	0.911	0.606	0.814	0.820	0.879	0.800	0.529	0.802
		20	0.805	0.830	0.875	0.907	0.912	0.940	0.555	0.930	0.718	0.934	0.936	0.957	0.934	0.936	0.957
		40	0.879	0.969	0.929	0.984	0.984	0.990	0.479	0.914	0.598	0.992	0.992	0.995	0.992	0.992	0.995
4	90	7	0.389	0.235	0.726	0.500	0.587	0.757	0.328	0.805	0.526	0.533	0.561	0.775	0.538	0.502	0.775
		10	0.427	0.543	0.711	0.517	0.781	0.768	0.353	0.825	0.563	0.594	0.618	0.805	0.571	0.693	0.797
		20	0.443	0.913	0.690	0.534	0.937	0.773	0.441	0.868	0.700	0.728	0.744	0.870	0.598	0.929	0.829
		40	0.450	0.984	0.698	0.491	0.991	0.737	0.446	0.818	0.701	0.836	0.849	0.923	0.615	0.985	0.843
4	180	7	0.594	0.180	0.583	0.686	0.719	0.730	0.488	0.802	0.532	0.652	0.389	0.638	0.661	0.308	0.624
		10	0.645	0.586	0.674	0.694	0.898	0.777	0.493	0.754	0.543	0.683	0.419	0.658	0.697	0.330	0.640
		20	0.656	0.970	0.756	0.694	0.982	0.796	0.637	0.895	0.726	0.766	0.533	0.722	0.758	0.563	0.727
		40	0.680	0.995	0.786	0.695	0.998	0.803	0.542	0.696	0.604	0.824	0.611	0.771	0.731	0.995	0.835

Notes: This table reports the precision (P) rate, recall (R) rate, and Rand index (RI) for each estimator. The triad pairwise-difference (TPWD) estimator is computed with $\widehat{\beta}^1$ set to the nuclear norm regularized (NNR) estimator with $\psi_{NT} = \log(\log(T))/\sqrt{16\min(N,T)}$, average linkage, and cut-off $c_{NT} = 1.5\widehat{\sigma}^2 \log(T)/\sqrt{T}$. Iterated TPWD is computed by iterating TPWD 4 times, replacing 3 times the preliminary estimator $\widehat{\beta}^1$ by the TPWD estimate obtained at the previous iteration. The Post-Spectral estimator is implemented as in Chetverikov and Manresa (2021), with user-specified number of groups set to the true number of groups and smallest λ chosen in the grid $\{1, 1.5, 2, 2.5, \ldots\}$ such that $m(\lambda) \leq G$. The grouped fixed-effects (GFE) estimator proposed in Bonhomme and Manresa (2015) is implemented as in Bonhomme and Manresa (2015)'s Algorithm 1, with a user-specified number of groups set to the true number of groups and 100 random initialization points following the method described in Section S.1.1 Bonhomme and Manresa (2015)'s Supplementary Material with $\theta^{(0)} \sim \mathcal{N}(0,1)$. The GFE with BIC selection of the number of groups is implemented as in Section S.3.2 of Bonhomme and Manresa (2015), with $G_{max} = 5$. Results are averaged across 500 Monte Carlo samples.

S3.6 Time-invariant unobserved heterogeneity

In this section, I investigate 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 the finite sample sensitivity to the choice of the thresholding parameter c_T.

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 $(\alpha^0_1, \ldots, \alpha^0_{G^0})' = \text{linspace}(-G^0/2, G^0/2, G^0)$ and $(g^0_1, \ldots, g^0_N) = (1, \ldots, 1, \ldots, G^0, \ldots, G^0)$ are deterministic so that 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 Carlo averages of the estimated number of groups $\widehat{G}^{\text{PWD}}$, the Hausdorff distance between estimated effects $(\widehat{\alpha}^{\text{PWD}}_1, \ldots, \widehat{\alpha}^{\text{PWD}}_{\widehat{G}^{\text{PWD}}})'$ and true effects $(\alpha^0_1, \ldots, \alpha^0_{G^0})'$, Rand Index (RI), and CPU time. RI is the proportion of correctly

[3] I study sensitivity to this choice later.

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 $v_{it}\vert\sigma_i \sim \mathcal{N}(0,\sigma_i)$ independent across i and i.i.d. across t for a given i.

predicted pair (true or false) returned by the PWD estimator:

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

where

$$TP \equiv \text{True Positives} \equiv \sum_{i<j} \mathbf{1}\{\widehat{g}_i^{\text{PWD}} = \widehat{g}_j^{\text{PWD}}\}\mathbf{1}\{g_i^0 = g_j^0\},$$

$$TN \equiv \text{True Negatives} \equiv \sum_{i<j} \mathbf{1}\{\widehat{g}_i^{\text{PWD}} \neq \widehat{g}_j^{\text{PWD}}\}\mathbf{1}\{g_i^0 \neq g_j^0\},$$

$$FP \equiv \text{False Positives} \equiv \sum_{i<j} \mathbf{1}\{\widehat{g}_i^{\text{PWD}} = \widehat{g}_j^{\text{PWD}}\}\mathbf{1}\{g_i^0 \neq g_j^0\},$$

$$FN \equiv \text{False Negatives} \equiv \sum_{i<j} \mathbf{1}\{\widehat{g}_i^{\text{PWD}} \neq \widehat{g}_j^{\text{PWD}}\}\mathbf{1}\{g_i^0 = g_j^0\}.$$

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 measures the ability of the PWD estimator to identify pairs that truly belong to the same group. The Precision rate measures how precise the pairing prediction is: among all predicted pairs of units, what is the proportion of correct ones? Both formally write

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

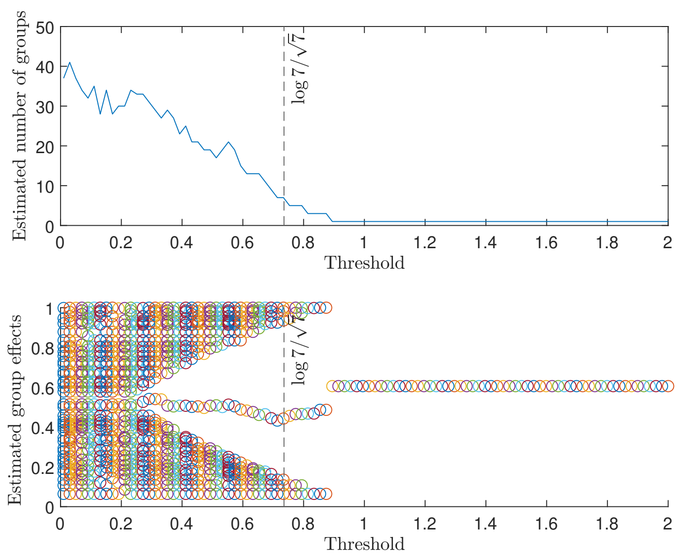
Figure S2 suggests that the larger the T, the larger the range of values for c for which $\widehat{G}^{\mathrm{PWD}} = G^0$. Figures S3–S5, which inform about the grouping composition, confirm that the larger T the better in terms of RI, Recall and Precision the classification – which remains quite is fast (e.g., compare with $T = 66$). This is confirmed by the small Hausdorff distances displayed 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 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

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* $\left\{\widehat{\alpha}_g^{\mathrm{PWD}} : g \in \left\{1, \ldots, \widehat{G}^{\mathrm{PWD}}\right\}\right\}$ *as a function of* c_T*.*

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	$\widehat{G}^{\mathrm{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\mathrm{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\mathrm{PWD}}$	HD	RI	CPU time	$\widehat{G}^{\mathrm{PWD}}$	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

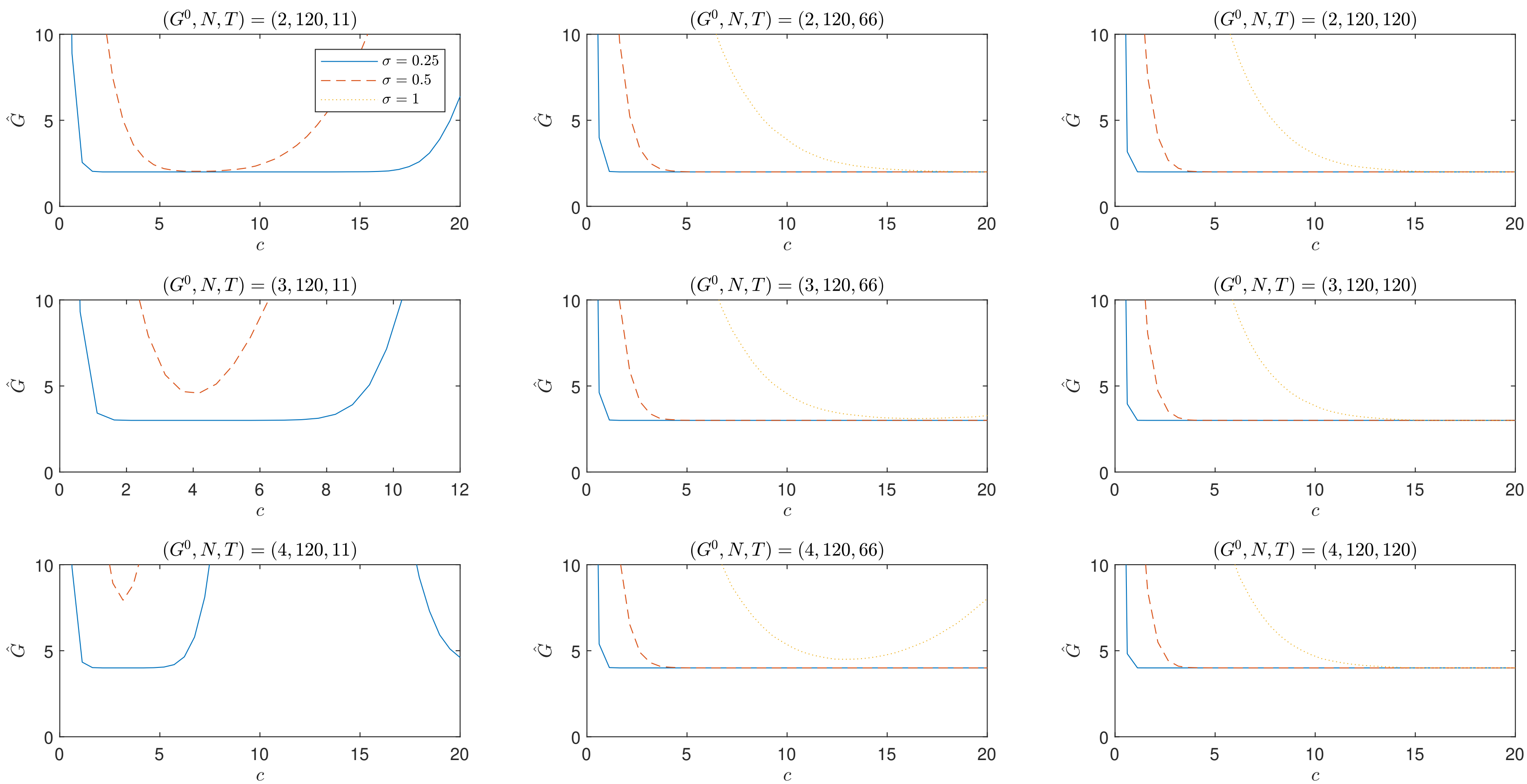
Notes: Results are averaged over 1,000 Monte Carlo replications. G^0 ≡True number of groups; $\widehat{G}^{\mathrm{PWD}}$ ≡Estimated number of groups; HD≡Hausdorff Distance between estimated and true group effects; RI≡Rand Index; CPU time≡MATLAB's `cputime`.

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

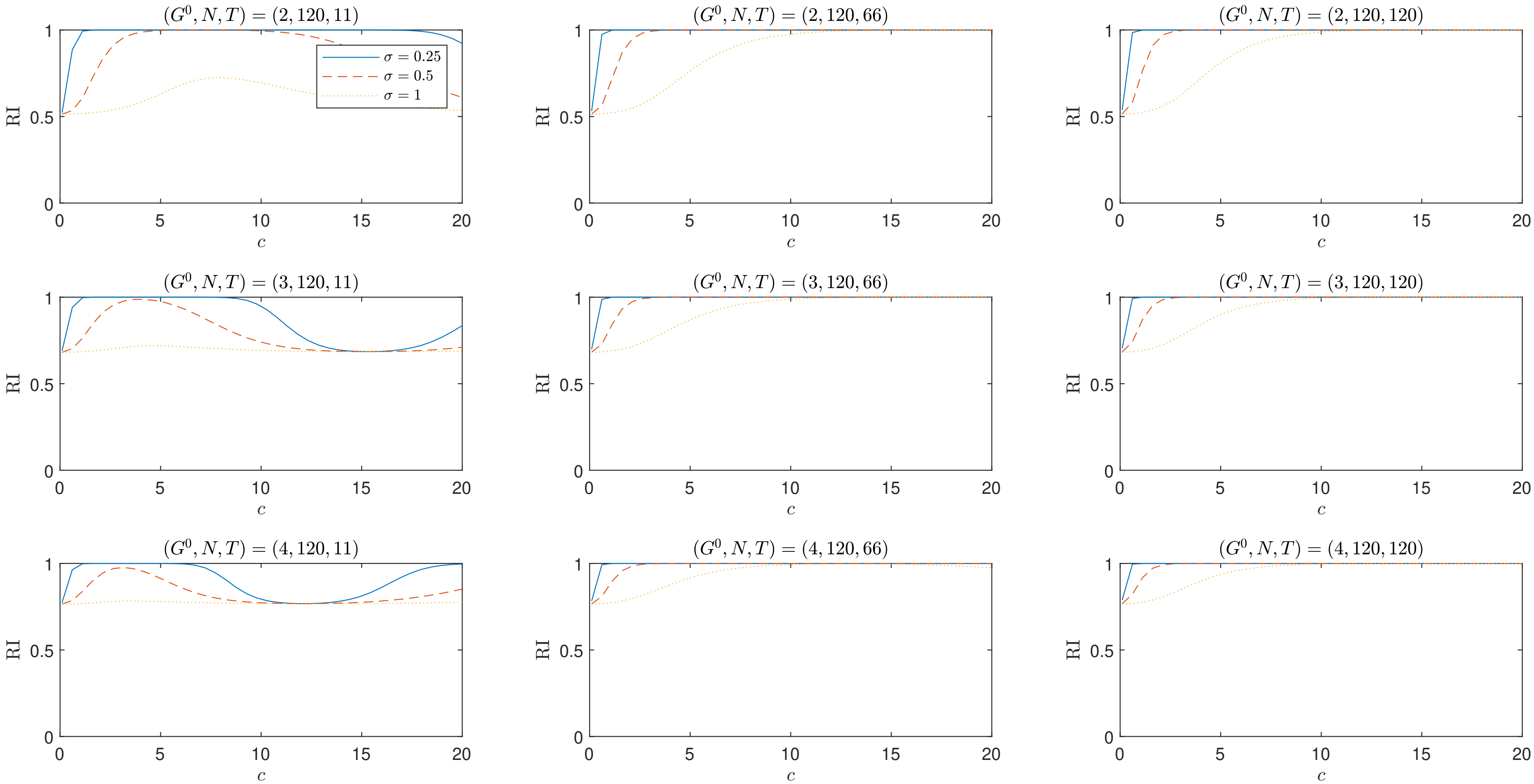
Notes: Results are averaged over 1,000 Monte Carlo replications. G^0 ≡True number of groups; $\widehat{G}^{\text{PWD}}$ ≡Estimated number of groups; HD≡Hausdorff Distance between estimated and true group effects; RI≡Rand Index; CPU time≡MATLAB's `cputime`.

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



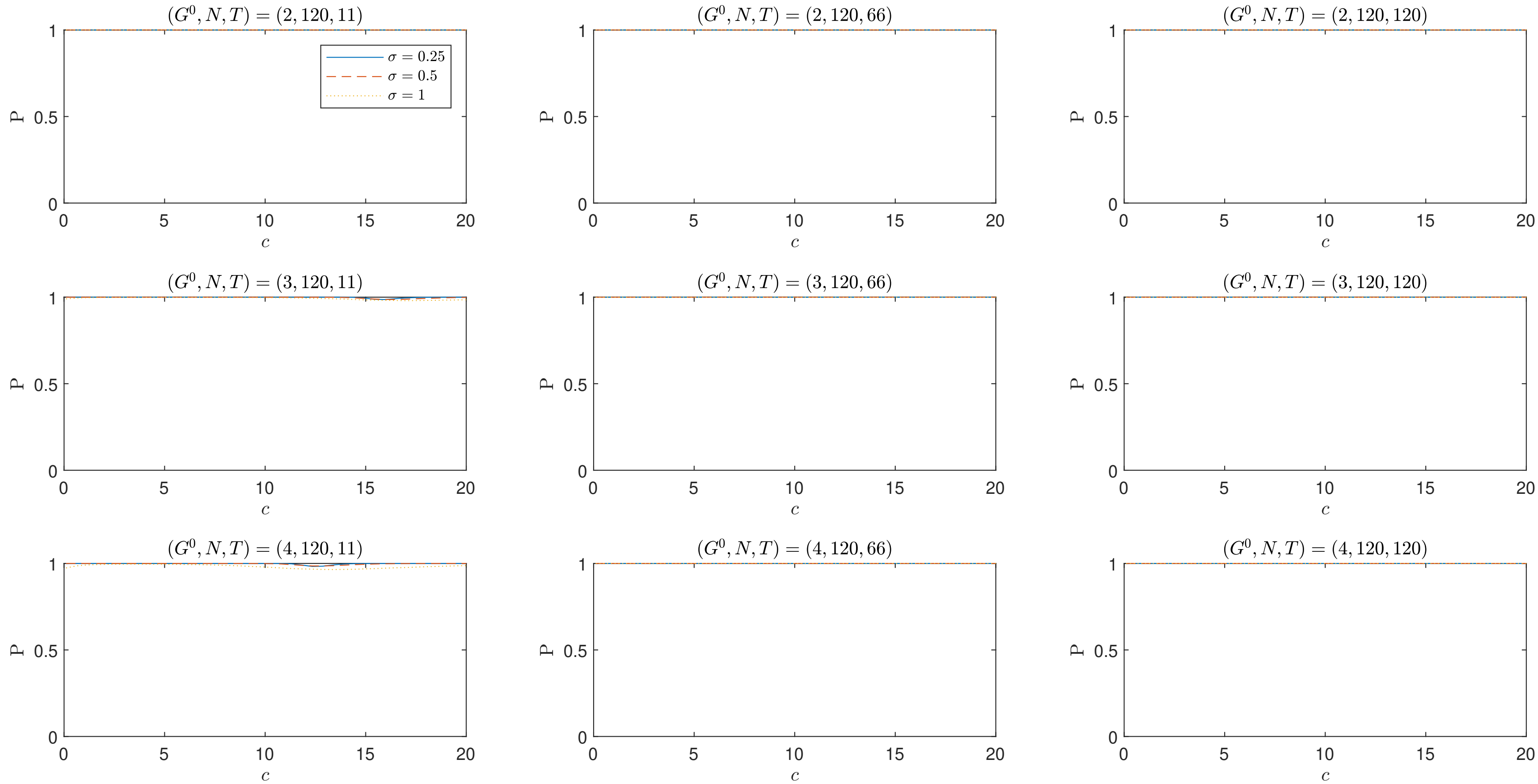
Notes: solid blue, dashed red, and dotted yellow lines report averages of $\widehat{G}$ with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$.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

Figure S3: Sensitivity of the Rand Index (RI)



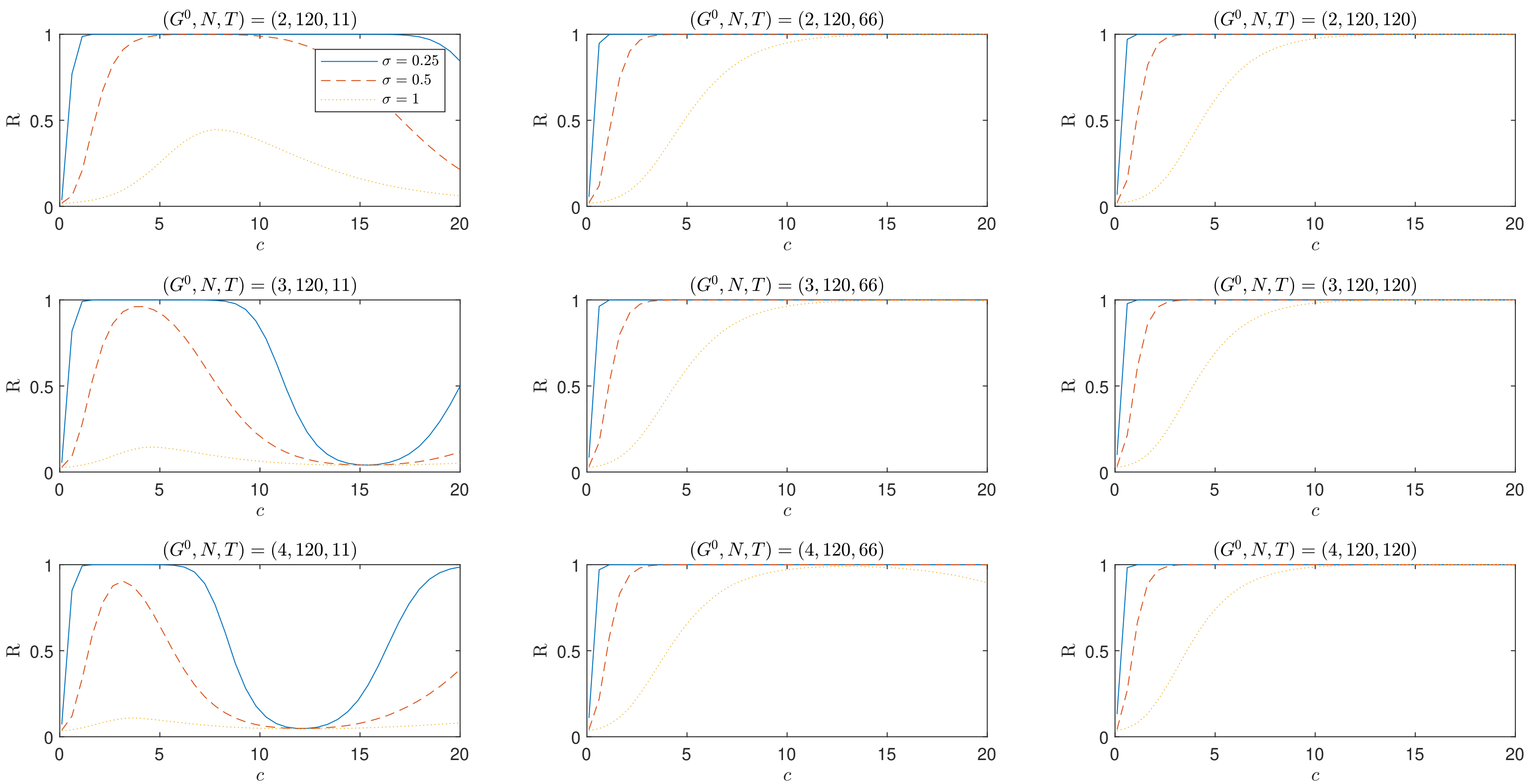
Notes: solid blue, dashed red, and dotted yellow lines report averages of RI with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are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

Figure S4: Sensitivity of the Precision rate (P)



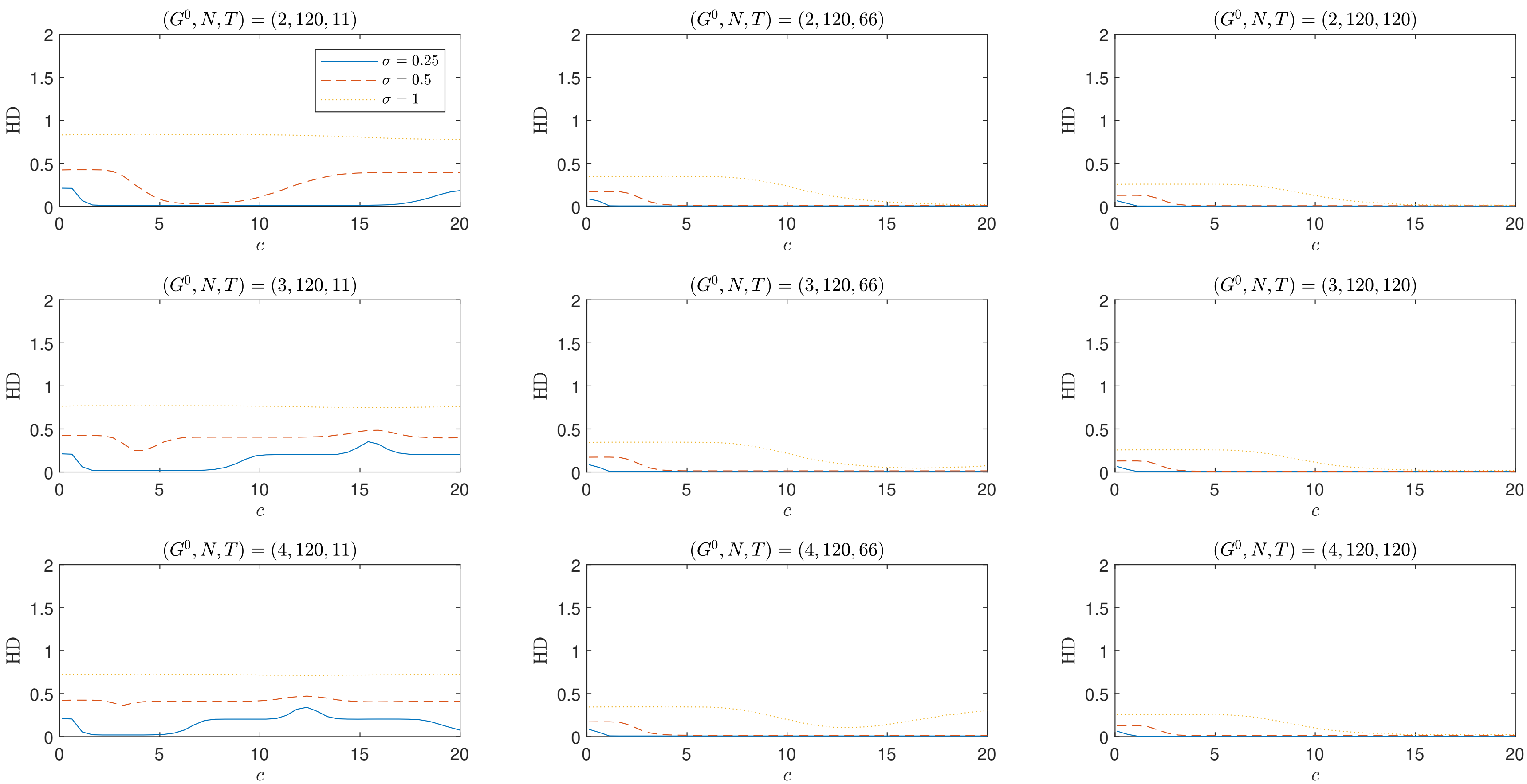
Notes: solid blue, dashed red, and dotted yellow lines report averages of P with $c_T = cT^{-4/5}$ as a function of $c \in (0, 20)$ for noise standard deviation of .25, .5 and 1 respectively. In each plot, $N = 120$; from top to bottom $G^0 = 2, 3, 4$; from left to right $T = 11, 66, 120$. Results are 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. Bond (1991): “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *The Review of Economic Studies*, 58(2), 277–297.

Arellano, M., and O. Bover (1995): “Another look at the instrumental variable estimation of error-components models,” *Journal of Econometrics*, 68(1), 29–51.

Bai, J. (2009): “Panel Data Models With Interactive Fixed Effects,” *Econometrica*, 77(4), 1229–1279.

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.

Chetverikov, D., and E. Manresa (2021): “Spectral and Post-Spectral Estimators for Grouped Panel Data Models,” Discussion paper.

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.