Make the Difference! Computationally Trivial Estimators for Grouped Fixed Effects Models*

Martin Mugnier†

March 14, 2022

Abstract

Novel estimators are proposed for linear grouped fixed effects models. Rather than predicting a single grouping of units, they deliver a collection of groupings with the same flavor as the so-called LASSO regularization path. Mild conditions are found that ensure their asymptotic guarantees are the same as the so-called grouped fixed effects and post-spectral estimators (Bonhomme and Manresa, 2015; Chetverikov and Manresa, 2021). In contrast, the new estimators are computationally straightforward and do not require prior knowledge of the number of groups. Monte Carlo simulations suggest good finite sample performance. Applying the approach to real data provides new insights on the potential network structure of the unobserved heterogeneity.

Keywords: panel data, grouped fixed effects, time-varying unobserved heterogeneity, k-means clustering
JEL Codes: C14, C23, C25.

*I thank Stéphane Bonhomme, Xavier D'Haultfœuille, and Ao Wang for helpful advices and discussions. All errors are mine.

†Department of Economics, CREST, ENSAE, Institut Polytechnique de Paris, 5 Av. Le Chatelier, 91120 Palaiseau, France. E-mail: `martin.mugnier@ensae.fr`. Comments are welcome!

1 Introduction

Suppose to observe a sample of longitudinal or panel data $\{(y_{it}, x'_{it})' : 1 \leq i \leq N, 1 \leq t \leq T\}$ and consider the grouped fixed effects model:

$$y_{it} = x'_{it}\theta^0 + \alpha^0_{g^0_i t} + v_{it}, \quad i = 1, ..., N, \ t = 1, ..., T, \tag{1.1}$$

where i denotes cross-sectional units, t denotes time periods, $y_{it} \in \mathbb{R}$ is a dependent variable, $\theta^0 \in \mathbb{R}^p$ is an unknown vector of interest, and $x_{it} \in \mathbb{R}^p$ is a vector of time-varying covariates contemporaneously uncorrelated with the zero-mean random variable $v_{it} \in \mathbb{R}$ but potentially correlated with the group-specific unobservable $\alpha^0_{g^0_i t} \in \mathcal{A} \subset \mathbb{R}$ (which rises the identification challenge). We adopt a "fixed effects" approach, leaving the group membership variable $g^0_i \in \{1, ..., G^0\}$, the number of groups G^0, and the vector of group-specific time effects $(\alpha^0_{1t}, ..., \alpha^0_{G^0 t})' \in \mathcal{A}^{G^0}$ unrestricted and considering them as parameters to be estimated.[1] Under prior restrictions (e.g., a known upper bound) on G^0, Bonhomme and Manresa (2015)'s so-called GFE estimator and Chetverikov and Manresa (2021)'s post-spectral estimator are both consistent for g^0_i (up to group relabeling) and root-NT (resp. root-N) asymptotically normal for θ^0 (resp. α^0_{gt}) as N and T diverge jointly, provided T grows at least as some power of N and G^0 is fixed.

Unfortunately, the GFE estimator cannot be computed in polynomial time in most real-world datasets of interest.[2] While the post-spectral estimator is computationally straightforward, it requires knowledge of G^0, which could be restrictive in practice. This motivates seeking for new identification approaches.

In this paper, we propose a novel identification strategy for all the model parameters, including G^0. The argument is constructive and leads to what we generically call *pairwise differencing* (PWD) estimators. Mild conditions are found that ensure PWD estimators enjoy the same asymptotic guarantees as GFE and post-spectral estimators. In contrast, they have polynomial computational complexity and never

[1] This special case of Bai (2009)'s interactive fixed effects models was introduced in a seminal paper by Bonhomme and Manresa (2015).

[2] This is related to NP-hardness of the k-means problem (see, e.g., Aloise et al., 2009). We note that inference based on heuristic solutions can be misleading (see, e.g., Chen, 2022).

use any prior knowledge on G^0. Instead of predicting a single grouping of units, they provide a collection of groupings indexed by a regularization parameter with the same flavor as the so-called `glmnet`'s LASSO regularization path. This increased flexibility allows for in-depth exploration of the underlying network structure and sensitivity analysis of the results, while delivering an appealing visualization tool. We demonstrate the usefulness of the new approach in a short application to Acemoglu et al. (2008)' panel data.

A quick glimpse at equation (1.1) suffices to see that identification is rendered difficult *only* because the group membership variables g_i^0 are unknown. Otherwise, a pooled ordinary least squares (OLS) regression on the interactions of group and time dummies would deliver, under appropriate regularity conditions, consistent estimates of the remaining structural parameters as N and T diverge (see, e.g., Bonhomme and Manresa, 2015). Our key insight is that identification of the grouping, at least in the special case with no covariates, can be achieved separately from identification of the remaining structural parameters. PWD estimators then proceed in two steps: (i) estimate the grouping, and (ii) project the dependent variable on the interactions of estimated group dummies and time dummies. The approach easily extends to models with covariates if a consistent preliminary estimate $\widehat{\theta}^1$ for θ^0 is available. This is because we can come back to the simple case by considering the sparse "pure factor model" $y_{it} - x_{it}'\widehat{\theta}^1$. We develop two-step PWD estimators whose asymptotic properties are not affected by the first-stage provided $\widehat{\theta}^1$ is $\sqrt{T}$-consistent.[3]

Following the statistical literature on graphon estimation (see, e.g. Klopp et al., 2017), we directly target identification of the adjency matrix $\mathbf{W}^0 := (W_{ij}^0)_{(i,j)\in V^2} = (\mathbb{1}\left\{g_i^0 = g_j^0\right\})_{(i,j)\in V^2}$ of the graph $\mathcal{G}^0$ with vertices $V = \{1, ..., N\}$ spanned by the group membership variables. This approach elegantly incorporates normalization issues since $\mathbf{W}^0$ is invariant to group relabeling. We exploit linearity of eq. (1.1) to build pairwise-level tests $\varphi_{i,j}$ of $H_{0,i,j} : g_i^0 = g_j^0$ (and estimates $\widehat{W}_{ij} = 1 - \varphi_{i,j}$ of W_{ij}^0) based on hard-thresholding of suitable differencing transformations of the dependent

[3] One may use off-the-shelf estimators such as provided in Wooldridge (2010) and Arellano and Bond (1991) (when grouped effects are time-invariant) or the nuclear norm estimator of Moon and Weidner (2019) (for the time varying case). All are computationally straightforward and meet the $\sqrt{T}$-rate restriction under appropriate conditions.

variables.[4] These only require elementary arithmetic operations. To establish asymptotic results, we leverage two other features of the model, which greatly simplify the analysis compared to standard Stochastic Block Models (see, e.g., Gao et al., 2015; Airoldi et al., 2008). First, $\mathcal{G}^0$ is such that there are no edges between units belonging to different groups but all units in a given group are connected. In other words, $\mathcal{G}^0$ consists of perfectly separated clusters or "communities". Second, repeated measurements provide the key identifying variation. Under standard group-separation conditions and weak dependence and tails restrictions on the idiosyncratic shocks, our estimates are sup-norm consistent, i.e., as N and T tend to infinity:

$$\sup_{(i,j)\in\{1,\dots,N\}^2} |\widehat{W}_{ij} - W_{ij}^0| \xrightarrow{p} 0, \tag{1.2}$$

where $\xrightarrow{p}$ denotes convergence in probability. Our second insight is that, given (1.2) and to avoid using k-means clustering of the rows of the matrix $\widehat{\mathbf{W}}$ (a NP-hard problem), we propose to group units with identical rows in $\widehat{\mathbf{W}}$. This procedure always yields a partition of units into $\widehat{G} \in \{1, \dots, N\}$ non-emtpy groups.[5]

A threshold parameter $c_T \in (0, +\infty)$ is needed to calibrate the test $\varphi_{i,j}$ and our asymptotic results hold provided c_T decreases to zero sufficiently slowly as $T \to \infty$. This tuning parameter is the natural counterpart to the "hidden" tuning parameter of GFE and spectral methods, namely, knowledge of the number of groups G^0 (or at least an upper bound). In practice, the researcher may report results for different (incremental) choices of the threshold until it is so large that $\widehat{G} = 1$, or she can choose a value by relying on cross-validation techniques.[6]

[4]In Mugnier (2022), we show how this approach yields nonparametric identification results in nonlinear grouped fixed effects models. It is worth noting that such types of dyad, triad, or tetrad comparisons have proven useful in a variety of different econometric contexts, especially in networks analysis (see, e.g., Graham, 2017; Charbonneau, 2017; Jochmans, 2017; Zeleneev, 2020).

[5]Albeit close in spirit, the procedure is different from the binary segmentation algorithm developed in Wang and Su (2021) or the pairwise comparisons method proposed in Krasnokutskaya et al. (2022). A popular and alternative approach to our hard-thresholding criterion is spectral clustering by considering the Laplacian of a dissimilarity matrix (see, e.g., Ng et al., 2002; von Luxburg, 2007; Chetverikov and Manresa, 2021). Instead, we propose a model-specific solution which yields a regularization path for the number of groups and whose analysis is much simpler.

[6]We find that the path for $\widehat{G}$ as a function of the threshold $c_T \in [0, +\infty)$ exhibits a "roughly

We study four estimators, each of them specifically tailored to accomodate particular submodels of interest such as time-invariant group-specific effects ($\alpha_{gt}^0 = \alpha_g^0$) or models with no covariates ($\theta^0 = 0$). We depart from Pollard (1981) who provides asymptotic theory only for a pseudo-true value in the cross-sectional case (the solution to the population k-means sum of squares problem) since our asymptotic statements hold for the true population parameter. Moreover, our theory covers the estimates effectively reported by the researcher, which fundamentally differs from approaches relying on heurisitic approximating algorithms as proposed in Bonhomme and Manresa (2015).

Monte Carlo simulations in the no covariates case confirm our theoretical results. We find that while the choice of the threshold affects the estimated number of groups in finite samples, it does not harm (up to a certain value) the quality of the predicted clustering in terms of Precision rate and Rand Index.[7] In other words, predicted pairs of units on average effectively belong to the same population group. Hence, the group-specific effects Hausdorff distance to the truth is small on average. The ability of the PWD estimator to regroup these similar predicted pairs in a unique predicted group (Recall rate) is more sensitive to the threshold but quickly improves as T increases. We expect that the same patterns hold true if covariates are included.

The remainder of the paper is organized as follows. Section 2 introduces our main arguments in models without covariates. Section 3 extends the approach to models with covariates. Section 4 gathers results from the Monte Carlo experiments. Section 5 contains the short empirical illustration. Section 6 concludes. All proofs are collected in the Appendix. Matlab code for implementing the methods is provided in Appendix A.9. Throughout the paper, we assume that $\mathcal{A}$ is a compact subset of $\mathbb{R}$.

Notation For any vector-valued sequence $\{z_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$, all $(i, t, s) \in \{1, ..., N\} \times \{1, ..., T\} \times \{2, ..., T\}$, let $\overline{z}_i := T^{-1} \sum_{s=1}^{T} z_{is}$, $\dot{z}_{it} = z_{it} - \overline{z}_i$ and $\ddot{z}_{is} = z_{is} - z_{s-1}$. For any sequences of positive real numbers $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, let $a_n \gtrsim b_n$ if and only if $\exists(C, n_0) \in (0, +\infty) \times \mathbb{N}, \forall n \in \mathbb{N}, n \geq n_0, a_n \geq C b_n$. The indicator function $\mathbb{1}\{\cdot\}$ equals one if $\cdot$ is true, 0 else. For any set A, $|A|$ is the cardinal of A. For any matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$, for all $i \in \{1, ..., n\}$, $\mathbf{M}_{i,\cdot}$ denotes the ith row of $\mathbf{M}$.

decreasing" pattern in all our settings (Monte Carlo and empirical application). Obtaining this path is computationally straightforward given the low CPU time required by the method.

[7] See Section 4 for a formal definition of these metrics.

2 A Simple Case Without Covariates

We present our approach in a simplified version of model (1.1) in which $\theta^0 = 0$ is known (no covariates). Section 2.1 considers a model with time-invariant group-specific effects (i.e., $\alpha^0_{gt} = \alpha^0_g$), introduces the pairwise differencing (PWD) estimator, and establishes its asymptotic properties. Section 2.2 relaxes the time-homogeneity assumption, introduces the tetrad pairwise differencing (TPWD) estimator, and establishes its asymptotic properties.

2.1 Time-Invariant Unobserved Heterogeneity

2.1.1 Model

Consider a specialization of model (1.1) to the following data generating process:

$$y_{it} = \alpha^0_{g^0_i} + v_{it}, \quad i = 1, ..., N,\ t = 1, ..., T. \tag{2.1}$$

Let $\Gamma_{N,G} := \{G\} \times \{1, ..., G\}^N \times \mathcal{A}^G$. Our goal is to estimate a given realization

$$(G^0, g^0_1, ..., g^0_N, \alpha^0_1, ..., \alpha^0_{G^0}) \in \bigcup_{G=1}^{\infty} \Gamma_{N,G}$$

as N and T tend to infinity.

2.1.2 Pairwise Differencing Estimation

The pairwise differencing (PWD) estimator is obtained from the following steps.

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

$$\widehat{W}^{PWD}_{ij} = \mathbb{1}\left\{(\overline{y}_i - \overline{y}_j)^2 \leq c_T\right\}, \quad i = 1, ..., N,\ j = 1, ..., N. \tag{2.2}$$

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

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

2. Compute $\widehat{\alpha}^{PWD}_1, ..., \widehat{\alpha}^{PWD}_{\widehat{G}^{PWD}}$ from:

$$(\widehat{\alpha}^{PWD}_1, ..., \widehat{\alpha}^{PWD}_{\widehat{G}^{PWD}}) = \underset{(\alpha_1, ..., \alpha_{\widehat{G}^{PWD}}) \in \mathcal{A}^{\widehat{G}^{PWD}}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \alpha_{\widehat{g}^{PWD}_i}\right)^2. \tag{2.3}$$

The PWD estimator involves two steps. The "assignment step" (1.) decomposes into a "testing substep" (1.a.), where the squared distance between time-averaged outcomes for each pair of units is compared to the user-specified threshold c_T in order to determine whether or not such units should be assigned an edge in the estimated adjency matrix; and a "grouping substep" (1.b.), where units with the same neighboors are grouped together. It is easy to see that the procedure always yields a partition of $\{1, ..., N\}$ into $\widehat{G}^{PWD} \in \{1, ..., N\}$ non-empty groups.[8] In the "estimation step" (2.), estimates for the group-specific effects are obtained by running a pooled OLS regression on the predicted group dummies. We point out that no optimization is needed as problem (2.3) admits a simple closed form solution.

2.1.3 Asymptotic Properties

We characterize the asymptotic properties of the PWD estimator as N and T tend to infinity in model (2.1). Consider the following assumption.

Assumption 1 *Equation* (2.1) *holds and there exist constants* $a, b, d_1, d_2 > 0$ *and a sequence* $\alpha[t] \leq e^{-at^{d_1}}$ *such that:*

(a) For all $(i, t) \in \{1, ..., N\} \times \{1, ..., T\}$*:* $\mathbb{E}(v_{it}) = 0$*,* $\Pr(|v_{it}| > m) \leq e^{1-(m/b)^{d_2}}$ *for all* $m > 0$*.*

(b) For all $(i, j, g, \widetilde{g}) \in \{1, ..., N\}^2 \times \{1, ..., G^0\}^2$ *such that* $g \neq \widetilde{g}$*, almost surely:* $|\alpha_g^0 - \alpha_{\widetilde{g}}^0| \geq c_{g,\widetilde{g}} > 0$*. Moreover,* $\{(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt})\}_t$ *is a strongly mixing process with mixing coefficients* $\alpha[t]$ *and* $\mathbb{E}((\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt})) = 0$*.*

Assumption 1 provides a set of identication conditions for the group memberships. It imposes moment and tail conditions on the error term, well-separation of groups and vanishing (spatial) dependence between products of differences of unobservables over time. It is akin to Bonhomme and Manresa (2015)'s Assumptions 1(c) and 2(b)-(d), but otherwise allows for asymptotically negligible groups. Below is our main result.

Proposition 2.1 *Let Assumption 1 hold and* $c_T = o(1)$ *such that* $c_T \gtrsim \log(T)/\sqrt{T}$ *as* $T \to \infty$*. Then, for all* $\delta > 0$ *and as* N *and* T *tend to infinity:*

$$\Pr\left(\sup_{(i,j)\in\{1,...,N\}^2} \left|\widehat{W}_{ij}^{PWD} - W_{ij}^0\right| > 0\right) = o(N^2 T^{-\delta}). \tag{2.4}$$

[8] The choice of group labels is innocuous since identification may hold only up to group relabeling.

The following result is a direct consequence of Proposition 2.1.

Corollary 2.2 (Sup-Norm Classification Consistency) *Let Proposition 2.1's Assumptions hold, and let N and T tend to infinity such that, for some $\nu > 0$, $N/T^\nu \to 0$. Then, there exists a permutation $\sigma^\star : \{1, ..., G^0\} \to \{1, ..., G^0\}$ such that, as N and T tend to infinity, we have*

$$\sup_{(i,j)\in\{1,...,N\}^2} \left|\widehat{W}_{ij}^{PWD} - W_{ij}^0\right| \xrightarrow{p} 0, \tag{2.5}$$

$$\widehat{G}^{PWD} \xrightarrow{p} G^0, \tag{2.6}$$

and

$$\sup_{i\in\{1,...,N\}} |\widehat{g}_i^{PWD} - \sigma^\star(g_i^0)| \xrightarrow{p} 0. \tag{2.7}$$

The following assumption is useful to establish limiting distributions.

Assumption 2 *For all $g \in \{1, ..., G^0\}$:*

(a) $\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{g_i^0 = g\} = \pi_g > 0$.

(b)

$$\lim_{N\to\infty} \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}\left(\mathbb{1}\left\{g_i^0 = g\right\} \mathbb{1}\left\{g_j^0 = g\right\} v_{it}v_{jt}v_{is}v_{js}\right) = \omega_g > 0.$$

(c) As N and T tend to infinity: $\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \mathbb{1}\{g_i^0 = g\} v_{it} \xrightarrow{d} \mathcal{N}(0, \omega_g)$.

Assumption 2 is akin to Bonhomme and Manresa (2015)'s Assumptions 2(a) and 3(d)-(e). In particular, Assumption 2(a) rules out asymptotically negligible group.

Corollary 2.3 (Asymptotic Distribution) *Let Assumptions 1 and 2 hold, $c_T = o(1)$ with $c_T \gtrsim \log(T)/\sqrt{T}$ as $T \to \infty$, and let N and T tend to infinity such that, for some $\nu > 0$, $N/T^\nu \to 0$. Then, there exists a permutation $\sigma^\star : \{1, ..., G^0\} \to \{1, ..., G^0\}$ such that, letting $\widehat{\alpha}_g^{PWD} := 0$ if $g > \widehat{G}^{PWD}$, we have*

$$\sqrt{NT}(\widehat{\alpha}_g^{PWD} - \alpha_{\sigma^\star(g)}^0) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_g}{\pi_g^2}\right), \quad g = 1, ..., G^0, \tag{2.8}$$

where ω_g and π_g are defined in Assumption 2.

Consistent plug-in estimates for the asymptotic variances can be easily constructed under cross-sectional independence and weak time-dependence.

2.2 Time-Varying Unobserved Heterogeneity

In this section, we generalize our approach to allow for time-varying grouped effects.

2.2.1 Model

Consider the following specialization of model (1.1):[9]

$$y_{it} = \alpha^0_{g^0_i t} + v_{it}, \quad i = 1, ..., N,\ t = 1, ..., T. \tag{2.9}$$

We now let $\Gamma_{N,T,G} := \{G\} \times \{1, ..., G\}^N \times \mathcal{A}^{GT}$ and seek to estimate

$$(G^0, g^0_1, ..., g^0_N, \alpha^0_{11}, ..., \alpha^0_{1T}, ..., \alpha^0_{G^0 1}, ..., \alpha^0_{G^0 T}) \in \bigcup_{G=1}^{\infty} \Gamma_{N,T,G}.$$

2.2.2 Tetrad Pairwise Differencing Estimation

From now on, we assume that $N \geq 4$. For any set $\mathcal{I} \subset \mathbb{N}^*$, for all $k \in \mathbb{N}^*$, let $\mathscr{P}_k(\mathcal{I}) := \{I \subset \mathcal{I} : |I| = k\}$ denote the set of subsets of $\mathcal{I}$ with cardinal k. For all tetrad $(i, j, k, l) \in \mathscr{P}_4(\{1, ..., N\})$, let

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

The tetrad pairwise differencing (TPWD) estimator is obtained from the following steps.

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

$$\widehat{W}^{TPWD}_{ij} = \mathbb{1}\left\{\max_{(k,l)\in\mathscr{P}_2(\{1,...,N\}\setminus\{i,j\})} |S_{NT}(i, j, k, l)| \leq c_T\right\}, \quad i = 1, ..., N,\ j = 1, ..., N. \tag{2.10}$$

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

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

[9] Since $\alpha^0_{g^0_i t} = \lambda'_{i,0} f_{t,0}$ with $\lambda_{i,0} = (\mathbb{1}\{g^0_i = 1\}, ..., \mathbb{1}\{g^0_i = G^0\})'$ and $f_{t,0} = (\alpha^0_{1t}, ..., \alpha^0_{G^0 t})'$, it is worth noting model (2.9) is a special case of the class of interactive fixed effects panel data models considered in Bai (2003), where the number of groups is unknown. Here, we exploit the extreme sparsity of $\lambda_{i,0}$ (which belongs to the set of vertices of the unit simplex of $\mathbb{R}^{G^0}$) to build an estimator that does *not* use any prior information of G^0 (e.g., its exact value or an upper bound).

2. Compute $\widehat{\alpha} := (\widehat{\alpha}_{11}^{TPWD}, ..., \widehat{\alpha}_{1T}^{TPWD}, ..., \widehat{\alpha}_{\widehat{G}^{TPWD}1}^{TPWD}, ..., \widehat{\alpha}_{\widehat{G}^{TPWD}T}^{TPWD})$ from:[10]

$$\widehat{\alpha} = \underset{\alpha \in \mathcal{A}^{\widehat{G}^{TPWD}T}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \alpha_{\widehat{g}_i^{TPWD}t} \right)^2 . \tag{2.11}$$

The TPWD estimator is based on the intuition that, under weak group-separability assumptions, the maximum of $|S_{NT}(i, j, k, l)|$ over $(k, l) \in \mathscr{P}_2(\{1, ..., N\} \backslash \{i, j\})$ is likely to be large if and only if i and j do not belong to the same group since it is lower bounded by $|S_{NT}(i, j, k, l)|$ evaluated at indices (k, l) such that $g_k^0 = g_i^0$ and $g_l^0 = g_j^0$, and such indices exist with probability approaching one under weak assumptions.

2.2.3 Asymptotic Properties

Consider the following assumption.

Assumption 3 *Equation* (2.9) *holds and there exist constants* $a, b, d_1, d_2 > 0$ *and a sequence* $\alpha[t] \leq e^{-at^{d_1}}$ *such that:*

(a) For all $(i, j, t) \in \{1, ..., N\}^2 \times \{1, ..., T\}$*:* $\mathbb{E}(v_{it}) = \mathbb{E}(v_{it}v_{jt}) = 0$ *and* $\Pr(|v_{it}| > m) \leq e^{1-(m/b)^{d_2}}$ *for all* $m > 0$*.*

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

(c) For all $(i, j, k, l, g, \widetilde{g}) \in \{1, ..., N\}^4 \times \{1, ..., G^0\}^2$ *such that* $g \neq \widetilde{g}$*,* $\{(v_{it} - v_{jt})(v_{kt} - v_{lt})\}_t$*,* $\{\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0\}_t$*, and* $\{(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)(v_{it} - v_{jt})\}_t$ *are strongly mixing processes with mixing coefficients* $\alpha[t]$*. Moreover,* $\mathbb{E}((\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)v_{it}) = 0$*.*

(d) $\lim_{N\to\infty} \Pr\left(\min_{g\in\{1,...,G^0\}} \sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\} \geq 4\right) = 1$*.*

Assumptions 3(a)-(c) are natural extensions of Assumption 1. Assumption 3(d) requires that each group has at least four members with probability approaching one.

Proposition 2.4 *Let Assumption 3 hold and* $c_T = o(1)$ *such that* $c_T \gtrsim \log(T)/\sqrt{T}$ *as* $T \to \infty$*. Then, for all* $\delta > 0$ *and as* N *and* T *tend to infinity:*

$$\Pr\left(\sup_{(i,j)\in\{1,...,N\}^2} \left| \widehat{W}_{ij}^{TPWD} - W_{ij}^0 \right| > 0 \right) = o(1) + o(N^4 T^{-\delta}). \tag{2.12}$$

[10] If necessary, re-allocate randomly units from highly populated groups to low populated groups until the Gram matrix becomes invertible.

The following result is a direct consequence of Proposition 2.4.

Corollary 2.5 (Sup-Norm Classification Consistency) *Let Proposition 2.4's Assumptions hold, and let N and T tend to infinity such that, for some $\nu > 0$, $N/T^{\nu} \to 0$. Then there exists a permutation $\sigma^{\star} : \{1, ..., G^0\} \to \{1, ..., G^0\}$ such that, as N and T tend to infinity, we have*

$$\sup_{(i,j)\in\{1,...,N\}^2} \left|\widehat{W}_{ij}^{TPWD} - W_{ij}^{0}\right| \xrightarrow{p} 0, \tag{2.13}$$

$$\widehat{G}^{TPWD} \xrightarrow{p} G^0, \tag{2.14}$$

and

$$\sup_{i\in\{1,...,N\}} |\widehat{g}_i^{TPWD} - \sigma^{\star}(g_i^0)| \xrightarrow{p} 0. \tag{2.15}$$

The following assumption is useful to establish limiting distributions.

Assumption 4 *For all $(g, t) \in \{1, ..., G^0\} \times \{1, ..., T\}$:*

(a) $\text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\} = \pi_g > 0$.

(b) $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{E}\left(\mathbb{1}\{g_i^0 = g\} \mathbb{1}\left\{g_j^0 = g\right\} v_{it} v_{jt}\right) = \omega_{gt} > 0$.

(c) As N and T tend to infinity: $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\} v_{it} \xrightarrow{d} \mathcal{N}(0, \omega_{gt})$.

Corollary 2.6 (Asymptotic Distribution) *Let Assumptions 3 and 4 hold, $c_T = o(1)$ such that $c_T \gtrsim \log(T)/\sqrt{T}$ as $T \to \infty$, and let N and T tend to infinity such that, for some $\nu > 0$, $N/T^{\nu} \to 0$. Then, there exists a permutation $\sigma^{\star} : \{1, ..., G^0\} \to \{1, ..., G^0\}$ such that, letting $\widehat{\alpha}_{gt}^{TPWD} := 0$ if $g > \widehat{G}^{TPWD}$, we have, for all t,*

$$\sqrt{N}(\widehat{\alpha}_{gt}^{TPWD} - \alpha_{\sigma^{\star}(g)t}^{0}) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_{gt}}{\pi_g^2}\right), \ g = 1, ..., G^0, \tag{2.16}$$

where ω_{gt} and π_g are defined in Assumption 4.

3 Including Covariates

In this section, we show how our approach easily extends to models with covariates. We propose two-step estimators based on a preliminary estimate for θ^0. The 2PWD (resp. 2TPWD) estimator is obtained by applying the PWD (resp. TPWD) estimator to the residuals $y_{it} - x_{it}'\widehat{\theta}^1$ obtained from a preliminary $\sqrt{T}$-consistent estimate $\widehat{\theta}^1$ for θ^0, and running a pooled OLS regression of y_{it} on x_{it} and the (interactions of time and) estimated group dummies. From now on, we assume that eq. (1.1) holds.

Assumption 5

(a) $||\widehat{\theta}^1 - \theta^0|| = O_p(T^{-1/2})$.

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

$$\sup_{i \in \{1,\ldots,N\}} \Pr\left(\frac{1}{T}\sum_{t=1}^{T} ||x_{it}|| \geq M\right) = o(T^{-\delta}) \text{ for all } \delta > 0.$$

Estimators verifying Assumption 5(a) can be found in, e.g., Wooldridge (2010), Arellano and Bond (1991), and Moon and Weidner (2019). Assumption 5(b) is identical to Bonhomme and Manresa (2015)'s Assumption 2(e). It holds if covariates have bounded support or if they satisfy dependence and tail conditions similar to v_{it} (see, e.g., the discussion made by the authors).

Proposition 3.1 *Let Assumption 5 hold and* $c_T = o(1)$ *with* $c_T \gtrsim \log(T)/\sqrt{T}$ *as* $T \to \infty$*, and let* N *and* T *tend to infinity such that, for some* $\nu > 0$, $N/T^\nu \to 0$.

- *If* $\alpha_{gt}^0 = \alpha_g^0$ *and Assumptions 1(a)-(b) hold, then as* N *and* T *tend to infinity:*

$$\Pr\left(\sup_{(i,j) \in \{1,\ldots,N\}^2} \left|\widehat{W}_{ij}^{2PWD} - W_{ij}^0\right| > 0\right) = o(1). \tag{3.1}$$

- *If Assumptions 3(a)-(d) hold, then as* N *and* T *tend to infinity:*

$$\Pr\left(\sup_{(i,j) \in \{1,\ldots,N\}^2} \left|\widehat{W}_{ij}^{2TPWD} - W_{ij}^0\right| > 0\right) = o(1). \tag{3.2}$$

Results similar to Corollaries 2.2 and 2.5 immediately follow from Proposition 3.1. The following assumptions are useful to establish limiting distributions and are reminiscent of Bonhomme and Manresa (2015).

Assumption 6 (Time-Invariant Unobserved Heterogeneity)

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

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

$$\Sigma_\theta = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x}_{g_i^0})(x_{it} - \overline{x}_{g_i^0})',$$
$$\Omega_\theta = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}\left[v_{it} v_{js} (x_{it} - \overline{x}_{g_i^0})(x_{js} - \overline{x}_{g_j^0})'\right],$$

where $\overline{x}_g := \left(T \sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\}\right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{1}\{g_i^0 = g\} x_{it}$.

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

Assumption 7 (Time-Varying Unobserved Heterogeneity)

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

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

$$\Sigma_\theta = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \overline{x}_{g_i^0 t})(x_{it} - \overline{x}_{g_i^0 t})',$$
$$\Omega_\theta = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}\left[v_{it} v_{js} (x_{it} - \overline{x}_{g_i^0 t})(x_{js} - \overline{x}_{g_j^0 s})'\right],$$

where $\overline{x}_{gt} := \left(\sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\}\right)^{-1} \sum_{i=1}^{N} \mathbb{1}\{g_i^0 = g\} x_{it}$.

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

Corollary 3.2 (Asymptotic Distribution) *Let Assumption 5 hold and $c_T = o(1)$ with $c_T \gtrsim \log(T)/\sqrt{T}$ as $T \to \infty$, and let N and T tend to infinity such that, for some $\nu > 0$, $N/T^\nu \to 0$.*

- *If $\alpha_{gt}^0 = \alpha_g^0$ and Assumptions 1(a)-(b), 2 and 6 hold, then there exists a permutation $\sigma^\star : \{1, ..., G^0\} \to \{1, ..., G^0\}$ such that, letting $\widehat{\alpha}_g^{2PWD} = 0$ if $g > \widehat{G}^{2PWD}$, we have*

$$\sqrt{NT}(\widehat{\theta}^{2PWD} - \theta^0) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_\theta^{-1} \Omega_\theta \Sigma_\theta^{-1}\right), \tag{3.3}$$

  and

$$\sqrt{NT}(\widehat{\alpha}_g^{2PWD} - \alpha_{\sigma^\star(g)}^0) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_g}{\pi_g^2}\right), \quad g = 1, ..., G^0, \tag{3.4}$$

  where $\Sigma_\theta, \Omega_\theta, \omega_g$, and π_g are defined in Assumptions 2 and 6.

- *If Assumptions 3(a)-(d), 4 and 7 hold, then there exists a permutation* $\sigma^\star : \{1, ..., G^0\} \to \{1, ..., G^0\}$ *such that, letting* $\widehat{\alpha}_g^{2TPWD} = 0$ *if* $g > \widehat{G}^{2TPWD}$, *we have*

$$\sqrt{NT}(\widehat{\theta}^{2TPWD} - \theta^0) \xrightarrow{d} \mathcal{N}\left(0, \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1}\right), \tag{3.5}$$

and, for all t:

$$\sqrt{N}(\widehat{\alpha}_{gt}^{2TPWD} - \alpha^0_{\sigma^\star(g)t}) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_{gt}}{\pi_g^2}\right), \quad g = 1, ..., G^0, \tag{3.6}$$

where $\Sigma_\theta, \Omega_\theta, \omega_{gt}$, *and* π_g *are defined in Assumptions 4 and 7.*

4 Monte Carlo Simulations

In this section, we investigate (i) the finite sample performance of the PWD estimator in correctly specified models, and (ii) the finite sample sensitivity to the choice of the threshold c_T.

First, we assess the consistency of the PWD estimator for $c_T = 2\log(T)/\sqrt{T}$.[11] For each $(G^0, N) \in \{2, 5, 10, 50\} \times \{50, 100, 200, 500\}$ and $T \in \text{ceil}(\text{linspace}(\sqrt{N}, N, 4))$, we draw $1,000$ Monte Carlo samples generated according to model (2.1), where $(\alpha_1^0, ..., \alpha_{G^0}^0)' = \text{linspace}(-G^0/2, G^0/2, G^0)$ and $(g_1^0, ..., g_N^0) = \text{repelem}([1 : G^0], N/G^0)$ are deterministic (each group has equal size N/G^0), and we consider three DGPs for the noise random variable v_{it} as summarized in Table 1.

Table 1: 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.
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.

Tables 2-4 report Monte Carlo averages of the estimated number of groups $\widehat{G}^{PWD}$, the Hausdorff distance between estimated effects $\{\widehat{\alpha}_1^{PWD}, ..., \widehat{\alpha}_{\widehat{G}^{PWD}}^{PWD}\}$ and true effects

[11] We study sensitivity to this choice later.

$\{\alpha_1^0, ..., \alpha_{G^0}^0\}$, Rand Index (RI), and CPU time. The Rand Index (RI) is the proportion of correctly predicted pair (true or false) returned by the PWD estimator, i.e.:

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

where

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

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

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

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

Results suggest good finite sample performance, although deteriorating with the degree of time-dependence of the idiosyncratic shocks. In the most favorable case of independent normal errors (Tables 2 and 4), it is remarkable how perfect or almost perfect classification is achieved for moderate sample sizes 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, we investigate finite sample sensitivity of the results to the choice of the thresholding parameter c_T. We consider DGP 1 only, fix $N = 120$, and let $(G^0, T) \in \{2, 3, 4\} \times \{11, 66, 120\}$. Figures 2-6 plot Monte Carlo averages of $\widehat{G}$, 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 colored 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 individuals, what is the proportion of correct ones? Both formally write

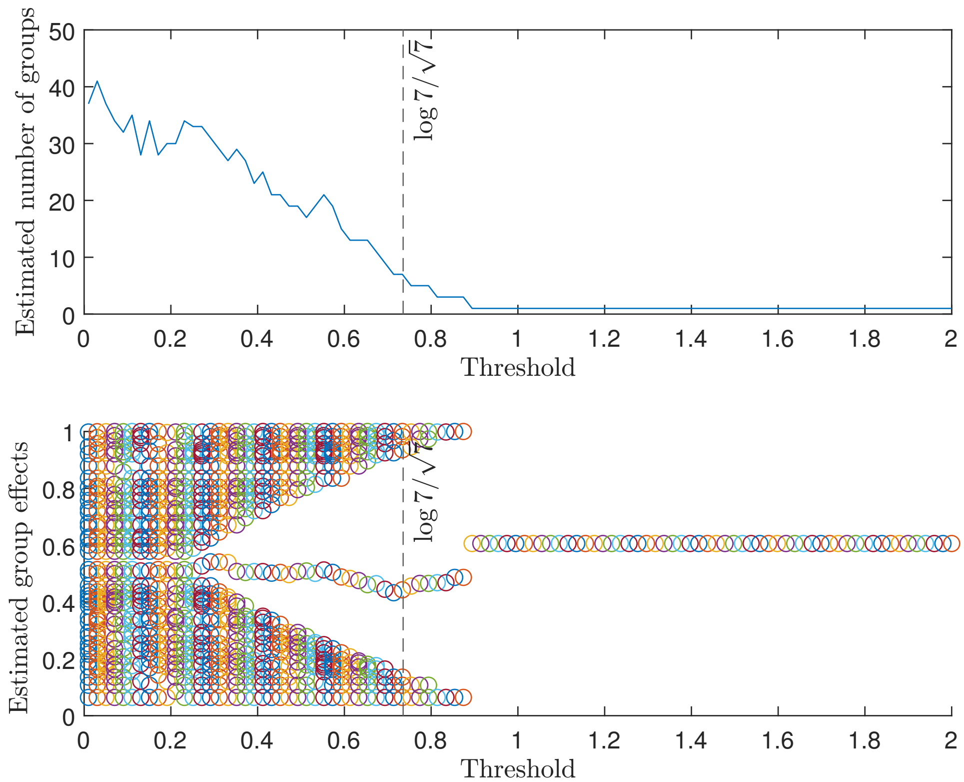
$$\text{R} = \frac{TP}{TP + FN}, \quad \text{P} = \frac{TP}{TP + FP}.$$

Figure 2 suggests that the larger the T, the larger the range of values for c for which $\hat{G}^{PWD} = G^0$. The two bumps can be explained by looking at Figures 3-5, which inform about the grouping composition. By definition of the PWD estimator, low values of c are necessarily associated with almost perfect precision (see Fig. 4) and low recall (see Fig.5), but the former compensates at first the latter, wich justifies the high Rand Index value (see Fig. 3). Such a high value of RI is not incompatible with a high $\widehat{G}^{PWD}$; it merely tells that the PWD estimator is so precise (or "conservative") that it is not yet able to merge all the similar pairs into G^0 groups (i.e., to detect all similar pairs) but that predicted similar pairs tend to be predicted together. The recall rate then increases steeply with c until experiencing a sudden drop (whereas the precision rate remains almost constant), and then increases again. This explains the first bump in $\widehat{G}^{PWD}$ and the fact that the decrease in RI is more tempered: with finite T, there exists a range for c which is "non-optimal" for high recall. Before the second bump, the range is optimal as $\widehat{G}^{PWD} = G^0$ across all DGPs and the Rand Index is very close to one. Figures 4 and 5 show that the second bump in $\widehat{G}^{PWD}$ is driven by a simultaneously drop in precision and recall: as c continues to increase, the hard-thresholding procedure becomes too coarse and the grouping quickly ends up in a degenerated prediction with a unique group containing all individuals. A key insight is that the Rand Index never drops too much before the second bump (because of the good precision rate). Hence, we can be confident that for a large range of values of c (at least $[0.1, 5]$ if $G^0 = 2$ and $[0.1, 10]$ if $G^0 = 4$), we will estimate precisely the group-specific effects. This is exactly what the small Hausdorff distances corresponding to these ranges suggest in Figure 6.

5 Empirical Illustration

In this section, we give a short illustration of the PWD estimator which allows to visualize the grouping path. We use the balanced subsample of Acemoglu et al. (2008) which contains the Freedom House Index of democracy for $N = 74$ countries (after dropping missing values) observed during $T = 7$ periods over the time span 1970-2000. We estimate model (2.1) using the PWD estimator for different values of $c_T \in (0, 2)$. Figure 1 reports the estimated number of groups and group-specific effects as a function of the threshold. The top-panel shows that the "regularization path" for $\widehat{G}^{PWD}(c_T)$ is surprisingly smooth and exhibits a general decreasing pattern

Figure 1: Panel of Countries (Freedom House Democracy Index)



from $\widehat{G}(0.01) \approx 40$ to $\widehat{G}^{PWD}(2) = 1$.[12] The bottom-panel suggests a convergence toward $\widehat{G}^{PWD} = 3$ groups before a sudden phase-transition to $\widehat{G}^{PWD} = 1$.

Finally, we investigate the performance of the PWD estimator through a specific Monte-Carlo simulation calibrated to the application. We consider the same sample size as in the empirical application: $N = 74$ and $T = 7$. The goal is to study the effective performance in a similar setting. We consider a well-specified case of discrete (grouped) unobserved heterogeneity. We compute the PWD estimates corresponding to $c_T \in \{0,8341, 0,7738, 0,6733\}$, which respectively yield $\widehat{G}^{PWD} \in \{3, 5, 11\}$. For a given number of estimated groups, we draw $1,000$ Monte Carlo samples under a data generating process following model (2.1), where the parameters α_g^0 and g_i^0 are fixed to their estimated PWD values. The error terms are generated as i.i.d. normal draws across units and time periods, with variance equal to the mean of squared

[12] The exact same pattern is observed when using the TPWD estimator.

PWD residuals. Figures 7-12 plot Monte Carlo averages of the PWD estimator of the number of groups, probability to identifiy the correct number of groups, HD between estimated/true group effects, RI, R, and P as a function of c_T. As expected from the empirical application, the predicted number of group is a perfectly continuous function of the threshold parameter and the Hausdorff distance remains moderately low given the scale of the estimated effects.

6 Conclusion

Grouped fixed effects models are plagued with an underlying NP-hard combinatorial problem, rendering estimation and inference difficult. In this paper, we propose a novel constructive identification argument for all the model parameters including the number of groups. The corresponding estimation methods have polynomial computational cost and are straightforward to implement (only basic arithmetic operations are required). They are based on thresholding suitable pairwise differencing transformation of the regression equation. Mild conditions are given under which our estimators are uniformly consistent for the underlying grouping structure and asymptotically normal as both dimensions diverge jointly, where the time-dimension can grow much more slowly than the cross-sectional dimension. Monte Carlo simulations suggest good finite sample performance but also leave many interesting questions unanswered. First, could the approach be fruitful to build a test for the grouping assumption? Second, how do the new estimators compare to existing methods requiring the number of groups to be known or when covariates are included? We leave these questions for further research.

References

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

Airoldi, E. M., Blei, D. M., Fienberg, S. E. and Xing, E. P. (2008), 'Mixed membership stochastic blockmodels', *Journal of Machine Learning Research* **9**(65), 1981–2014.

Aloise, D., Deshpande, A., Hansen, P. and Popat, P. (2009), 'Np-hardness of euclidean sum-of-squares clustering', *Machine Learning* **75**, 245–248.

Arellano, M. and Bond, S. (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.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–171.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

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

Charbonneau, K. B. (2017), 'Multiple fixed effects in binary response panel data models', *The Econometrics Journal* **20**(3), S1–S13.

Chen, Y.-C. (2022), 'Statistical inference with local optima'.

Chetverikov, D. and Manresa, E. (2021), Spectral and post-spectral estimators for grouped panel data models, Technical report.

Gao, C., Lu, Y. and Zhou, H. H. (2015), 'Rate-optimal graphon estimation', *The Annals of Statistics* **43**(6).

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.

Klopp, O., Tsybakov, A. B. and Verzelen, N. (2017), ‘Oracle inequalities for network models and sparse graphon estimation’, *The Annals of Statistics* **45**(1), 316–354.

Krasnokutskaya, E., Song, K. and Tang, X. (2022), ‘Estimating unobserved individual heterogeneity using pairwise comparisons’, *Journal of Econometrics* **226**(2), 477–497.

Moon, H. R. and Weidner, M. (2019), ‘Nuclear norm regularized estimation of panel regression models’.

Mugnier, M. (2022), Unobserved clusters of time-varying heterogeneity in nonlinear panel data models, Technical report.

Ng, A., Jordan, M. and Weiss, Y. (2002), On spectral clustering: Analysis and an algorithm, *in* T. Dietterich, S. Becker and Z. Ghahramani, eds, ‘Advances in Neural Information Processing Systems’, Vol. 14, MIT Press.

Pollard, D. (1981), ‘Strong Consistency of K-Means Clustering’, *The Annals of Statistics* **9**(1), 135 – 140.

von Luxburg, U. (2007), ‘A tutorial on spectral clustering’.

Wang, W. and Su, L. (2021), ‘Identifying latent group structures in nonlinear panels’, *Journal of Econometrics* **220**(2), 272–295. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.

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

Zeleneev, A. (2020), Identification and estimation of network models with nonparametric unobserved heterogeneity, Technical report.

A Appendix

A.1 Proof of Proposition 2.1

Let $\delta > 0$ and define $Z_{1NT}(i,j) = \widehat{W}_{ij}^{PWD}(1 - W_{ij}^0)$ and $Z_{2NT}(i,j) = (1 - \widehat{W}_{ij}^{PWD})W_{ij}^0$. By the union bound,

$$\begin{aligned}
&\Pr\left(\sup_{(i,j)\in\{1,\dots,N\}^2} |\widehat{W}_{ij}^{PWD} - W_{ij}^0| > 0\right) \\
&\quad \le \sum_{(i,j)\in\{1,\dots,N\}^2} \Pr\left(\widehat{W}_{ij}^{PWD} \neq W_{ij}^0\right) \\
&\quad = \sum_{(i,j)\in\{1,\dots,N\}^2} \Pr\left(Z_{1NT}(i,j) = 1\right) + \Pr\left(Z_{2NT}(i,j) = 1\right). \qquad \text{(A.1)}
\end{aligned}$$

We show below that, as N and T tend to infinity,

$$\sup_{(i,j)\in\{1,\dots,N\}^2} \Pr(Z_{1NT}(i,j) = 1) = o(T^{-\delta}), \qquad \text{(A.2)}$$

and

$$\sup_{(i,j)\in\{1,\dots,N\}^2} \Pr\left(Z_{2NT}(i,j) = 1\right) = o(T^{-\delta}). \qquad \text{(A.3)}$$

The result follows by combining (A.1)-(A.3).

Step 1: (A.2) **holds.** Let $(i,j) \in \{1,\dots,N^2\}$. If $G^0 = 1$, almost surely $Z_{1NT}(i,j) = 0$. Assume without loss that $G^0 \ge 2$. By definition of $\widehat{W}_{ij}^{PWD}$ and W_{ij}^0, we have

$$\begin{aligned}
Z_{1NT}(i,j) &= \sum_{\substack{(g,\widetilde{g})\in\{1,\dots,G^0\}^2 \\ g\neq\widetilde{g}}} \mathbb{1}\left\{g_i^0 = g\right\}\mathbb{1}\left\{g_j^0 = \widetilde{g}\right\}\mathbb{1}\left\{(\overline{y}_i - \overline{y}_j)^2 \le c_T\right\} \\
&\le \sum_{\substack{(g,\widetilde{g})\in\{1,\dots,G^0\}^2 \\ g\neq\widetilde{g}}} \mathbb{1}\left\{g_i^0 = g\right\}\mathbb{1}\left\{g_j^0 = \widetilde{g}\right\}\mathbb{1}\left\{\sum_{t=1}^T(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \le -\frac{T}{2}(c_{g,\widetilde{g}}^2 - c_T)\right\} \\
&\le \max_{\substack{(g,\widetilde{g})\in\{1,\dots,G^0\}^2 \\ g\neq\widetilde{g}}} \mathbb{1}\left\{\sum_{t=1}^T(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \le -\frac{T}{2}(c_{g,\widetilde{g}}^2 - c_T)\right\},
\end{aligned}$$

where the first inequality is obtained by developing the square and using Assumption 1(b), and the second inequality follows because there is at most a unique $(g,\widetilde{g}) \in$

$\{1,...,G^0\}^2$ such that $\mathbb{1}\{g_i^0 = g\}\mathbb{1}\left\{g_j^0 = \widetilde{g}\right\} = 1$. By the union bound,

$$\Pr(Z_{1NT}(i,j) = 1) \leq \sum_{\substack{(g,\widetilde{g})\in\{1,...,G^0\}^2 \\ g\neq\widetilde{g}}} \Pr\left(\sum_{t=1}^{T}(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \leq -\frac{T}{2}(c_{g,\widetilde{g}}^2 - c_T)\right). \tag{A.4}$$

Let $(g,\widetilde{g}) \in \{1,...,G^0\}^2$, $g \neq \widetilde{g}$. For T sufficiently large, $c_T \leq \min_{g\neq\widetilde{g}} c_{g,\widetilde{g}}^2/2$. Hence,

$$\begin{aligned}
&\Pr\left(\sum_{t=1}^{T}(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \leq -\frac{T}{2}(c_{g,\widetilde{g}}^2 - c_T)\right) \\
&\leq \Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \leq -\frac{c_{g,\widetilde{g}}^2}{4}\right).
\end{aligned} \tag{A.5}$$

By Assumption 1(b), the process $\{(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt})\}_t$ has zero mean and is strongly mixing with faster-than-polynomial decay rate. Moreover, for all i, t, and $m > 0$,

$$\Pr\left(|(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt})| > m\right) \leq \Pr\left(|v_{it} - v_{jt}| > \frac{m}{2\sup_{\alpha\in\mathcal{A}}|\alpha|}\right),$$

so $\{(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt})\}_t$ also satisfies the tail condition of Assumption 1(a), albeit with a different constant $\widetilde{b} > 0$ instead of $b > 0$. Applying Lemma B.5 in Bonhomme and Manresa (2015) with $z_t = (\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt})$ and taking $z = c_{g,\widetilde{g}}^2/4$ yields

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \leq -\frac{c_{g,\widetilde{g}}^2}{4}\right) = o(T^{-\delta}). \tag{A.6}$$

Note that the above upper bound on the probability does not depend on i, j, g, and $\widetilde{g}$. Combining (A.4)-(A.6) and taking the supremum over $(i,j) \in \{1,...,N\}^2$ yields

$$\sup_{(i,j)\in\{1,...,N\}^2} \Pr(Z_{1NT}(i,j) = 1) \leq G^0(G^0 - 1)o(T^{-\delta}) = o(T^{-\delta}).$$

i.e., (A.2) holds.

Step 2: (A.3) **holds.** Similarly, we have

$$\begin{aligned}
Z_{2NT}(i,j) &= \sum_{g=1}^{G^0} \mathbb{1}\left\{g_i^0 = g\right\}\mathbb{1}\left\{g_j^0 = g\right\}(1 - \widehat{W}_{ij}^{PWD}) \\
&\leq \mathbb{1}\left\{\left|\frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})\right| > \sqrt{c_T}\right\}.
\end{aligned} \tag{A.7}$$

A close inspection of the proof of Lemma B.5 in Bonhomme and Manresa (2015) reveals that, by taking $z = \sqrt{c_T}$ and because $c_T \gtrsim \log(T)/T^{1/2}$,

$$\begin{aligned}\Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}(v_{it}-v_{jt})\right| \geq \sqrt{c_T}\right) &\leq 4\left(1+\frac{\log(T)}{C_1}\right)^{-(1/2)T^{1/2}} \\ &\quad + C_2\sqrt{\frac{T^{1/2}}{\log(T)}}\exp\left(-C_3\left(T^{1/2}\sqrt{C_4\frac{\log(T)}{T^{1/2}}}\right)^{C_5}\right) \\ &= o(T^{-\delta}),\end{aligned}$$

where C_1, C_2, C_3, C_4, and C_5 are positive constants that do not depend on i and j. Thus, (A.3) holds.

The proof of Proposition 2.1 is complete.

A.2 Proof of Corollary 2.2

Let $\delta > 2\nu$. Then, eq. (2.5) follows from an application of Proposition 2.1. Next,

$$\Pr(\widehat{G}^{PWD} \neq G^0) \leq \Pr\left(\sup_{(i,j)\in\{1,...,N\}^2} |\widehat{W}_{ij}^{PWD} - W_{ij}^0| > 0\right) = o(1),$$

which proves (2.6). Define the probability events $\mathcal{E}_{1NT} = \{\forall (i,j) \in \{1,...,N\}^2 : \widehat{W}_{ij}^{PWD} = W_{ij}^0\}$ and $\mathcal{E}_{2NT} = \{\widehat{G}^{PWD} = G^0\}$. Then, there must exist a permutation $\sigma^\star : \{1,...,G^0\} \to \{1,...,G^0\}$ such that, on $\mathcal{E}_{1NT} \cap \mathcal{E}_{2NT}$, for all $i \in \{1,...,N\}$, $\widehat{g}_i^{PWD} = \sigma^\star(g_i^0)$. Given (2.5) and (2.6), deduce that

$$\Pr\left(\sup_{i\in\{1,...,N\}} |\widehat{g}_i^{PWD} - \sigma^\star(g_i^0)| > 0\right) \leq \Pr\left(\mathcal{E}_{1NT}^c \cup \mathcal{E}_{2NT}^c\right) \leq \Pr\left(\mathcal{E}_{1NT}^c\right) + \Pr\left(\mathcal{E}_{2NT}^c\right) = o(1),$$

which proves (2.7).

A.3 Proof of Corollary 2.3

Let $(\widetilde{\alpha}_1, ..., \widetilde{\alpha}_{G^0})' \in \mathbb{R}^{G^0}$ denote the infeasible "oracle" estimates obtained from the pooled OLS regression of y_{it} on $\mathbb{1}\{g_i^0 = 1\}, ..., \mathbb{1}\{g_i^0 = G^0\}$. It can be checked that

$$\sqrt{NT}(\widetilde{\alpha}_g - \alpha_g^0) = \frac{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{1}\{g_i^0 = g\}\sum_{t=1}^{T} v_{it}}{\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{g_i^0 = g\}}, \quad g = 1, ..., G^0.$$

Under Assumption 2, we have

$$\sqrt{NT}(\widetilde{\alpha}_g - \alpha_g^0) \xrightarrow{d} \mathcal{N}\left(0, \frac{\omega_g}{\pi_g^2}\right), \quad g = 1, ..., G^0.$$

By Corollary 2.2, there exists a permutation $\sigma^\star : \{1, ..., G^0\} \to \{1, ..., G^0\}$ such that

$$\Pr(\widehat{\alpha}_g^{PWD} \neq \widetilde{\alpha}_{\sigma^\star(g)}) \leq \Pr(\widehat{G}^{PWD} \neq G^0) + \Pr\left(\sup_{i\in\{1,...,N\}} |\widehat{g}_i^{PWD} - \sigma^\star(g_i^0)| > 0\right) = o(1).$$

The result follows as

$$\begin{aligned}
&\left|\Pr\left(\sqrt{NT}(\widehat{\alpha}_g^{PWD} - \alpha_{\sigma^\star(g)}^0) \leq c\right) - \Pr\left(\sqrt{NT}(\widetilde{\alpha}_{\sigma^\star(g)} - \alpha_{\sigma^\star(g)}^0) \leq c\right)\right| \\
&\quad \leq \left|\Pr\left(\sqrt{NT}(\widehat{\alpha}_g^{PWD} - \alpha_{\sigma^\star(g)}^0) \leq c, \sqrt{NT}(\widetilde{\alpha}_{\sigma^\star(g)} - \alpha_{\sigma^\star(g)}^0) > c\right)\right| \\
&\qquad + \left|\Pr\left(\sqrt{NT}(\widehat{\alpha}_g^{PWD} - \alpha_{\sigma^\star(g)}^0) > c, \sqrt{NT}(\widetilde{\alpha}_{\sigma^\star(g)} - \alpha_{\sigma^\star(g)}^0) \leq c\right)\right| \\
&\quad \leq \quad \Pr(\widehat{\alpha}_g^{PWD} \neq \widetilde{\alpha}_{\sigma^\star(g)}) + \Pr(\widehat{\alpha}_g^{PWD} \neq \widetilde{\alpha}_{\sigma^\star(g)})
\end{aligned}$$

for any c.

A.4 Proof of Proposition 2.4

Let $\delta > 0$, $Z_{1NT}(i,j) = \widehat{W}_{ij}^{TPWD}(1 - W_{ij}^0)$, $Z_{2NT}(i,j) = (1 - \widehat{W}_{ij}^{TPWD})W_{ij}^0$, and define the probability event $\mathcal{E}_N = \{\min_{g\in\{1,...,G^0\}} \sum_{i=1}^N \mathbb{1}\{g_i^0 = g\} \geq 4\}$. By standard probability algebra and similar arguments as the begining of the proof of Proposition 2.1:

$$\begin{aligned}
&\Pr\left(\sup_{(i,j)\in\{1,...,N\}^2} |\widehat{W}_{ij}^{TPWD} - W_{ij}^0| > 0\right) \\
&\quad \leq \Pr\left(\mathcal{E}_N^c\right) + \sum_{(i,j)\in\{1,...,N\}^2} \Pr\left(Z_{1NT}(i,j) = 1, \mathcal{E}_N\right) + \Pr\left(Z_{2NT}(i,j) = 1, \mathcal{E}_N\right). \quad \text{(A.8)}
\end{aligned}$$

Since Assumption 3(d) ensures $\lim_{N\to\infty} \Pr\left(\mathcal{E}_N^c\right) = 0$, the rest of the proof consists in proving that (A.2) and (A.3) (with $o(N^2T^{-\delta})$ in place of $o(T^{-\delta})$) hold under the new definition of $Z_{1NT}(i,j)$ and $Z_{2NT}(i,j)$ and intersecting each probability event with $\mathcal{E}_N$ (we call them (A.2)' and (A.3)').

Step 1: (A.2)' holds. For all $(i,j) \in \mathscr{P}_2(\{1, ..., N\})$, let $(k^*(i,j,g_i^0), l^*(i,j,g_j^0)) \in \mathscr{P}_2(\{1, ..., N\}\backslash\{i,j\})$ such that, on $\mathcal{E}_N$, $g_{k^*(i,j,g_i^0)}^0 = g_i^0$ and $g_{l^*(i,j,g_j^0)}^0 = g_j^0$. Let $(i,j) \in$

$\{1, ..., N\}^2$. Since almost surely $Z_{1NT}(i,j) = 0$ if $G^0 = 1$, assume without loss that $G^0 \geq 2$. Similar reasoning as Step 1 in Section A.1 yields that, on $\mathcal{E}_N$, we have

$$
\begin{aligned}
Z_{1NT}(i,j) \leq & \max_{\substack{(g,\widetilde{g})\in\{1,...,G^0\}^2 \\ g\neq\widetilde{g}}} \mathbb{1}\left\{g_i^0 = g\right\} \mathbb{1}\left\{g_j^0 = \widetilde{g}\right\} \\
& \times \mathbb{1}\left\{\max_{(k,l)\in\mathscr{P}_2(\{1,...,N\}\setminus\{i,j\})} \left|\frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)(\alpha_{g_k^0 t}^0 - \alpha_{g_l^0 t}^0)\right.\right. \\
& + \frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)(v_{kt} - v_{lt}) + \frac{1}{T}\sum_{t=1}^{T}(\alpha_{g_k^0 t}^0 - \alpha_{g_l^0 t}^0)(v_{it} - v_{jt}) \\
& \left.\left. + \frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})(v_{kt} - v_{lt})\right| \leq c_T\right\} \\
\leq & \max_{\substack{(g,\widetilde{g})\in\{1,...,G^0\}^2 \\ g\neq\widetilde{g}}} \mathbb{1}\left\{g_i^0 = g\right\} \mathbb{1}\left\{g_j^0 = \widetilde{g}\right\} \\
& \times \mathbb{1}\left\{\left|\frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)(\alpha^0_{g^0_{k^*(i,j,g_i^0)} t} - \alpha^0_{g^0_{l^*(i,j,g_j^0)} t})\right.\right. \\
& + \frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)(v_{k^*(i,j,g_i^0)t} - v_{l^*(i,j,g_j^0)t}) + \frac{1}{T}\sum_{t=1}^{T}(\alpha^0_{g^0_{k^*(i,j,g_i^0)} t} - \alpha^0_{g^0_{l^*(i,j,g_j^0)} t})(v_{it} - v_{jt}) \\
& \left.\left. + \frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})(v_{k^*(i,j,g_i^0)t} - v_{l^*(i,j,g_j^0)t})\right| \leq c_T\right\}.
\end{aligned}
$$

where the second inequality is by definition of the max and because $(k^*(i,j,g_i^0), l^*(i,j,g_j^0)) \in \mathscr{P}_2(\{1, ..., N\}\setminus\{i,j\})$. Because $\mathbb{1}\{|a| \leq b\} \leq \mathbb{1}\{a \leq b\}$ for any $(a,b) \in \mathbb{R}\times\mathbb{R}^*$ and using that $g^0_{k^*(i,j,g_i^0)} = g_i^0$ and $g^0_{l^*(i,j,g_j^0)} = g_j^0$, we have

$$
\begin{aligned}
Z_{1NT}(i,j) \leq & \max_{\substack{(g,\widetilde{g})\in\{1,...,G^0\}^2 \\ g\neq\widetilde{g}}} \mathbb{1}\left\{g_i^0 = g\right\} \mathbb{1}\left\{g_j^0 = \widetilde{g}\right\} \mathbb{1}\left\{\frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)^2\right. \\
& + \frac{1}{T}\sum_{t=1}^{T}(\alpha_{gt}^0 - \alpha_{\widetilde{g}t}^0)(v_{it} - v_{jt} + v_{k^*(i,j,g)t} - v_{l^*(i,j,\widetilde{g})t}) \\
& \left. + \frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})(v_{k^*(i,j,g)t} - v_{l^*(i,j,\widetilde{g})t}) \leq c_T\right\}.
\end{aligned}
$$

By the union bound,

$$
\begin{aligned}
&\Pr(Z_{1NT}(i,j)=1,\mathcal{E}_N)\\
&\quad\le \sum_{\substack{(g,\widetilde{g})\in\{1,\ldots,G^0\}^2\\ g\neq\widetilde{g}}}\Bigg[\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})^2\le\frac{c_{g,\widetilde{g}}}{2}\right)\\
&\qquad+\Pr\left(\sum_{t=1}^{T}(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})(v_{it}-v_{jt})\le-\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2}-c_T\right)\right)\\
&\qquad+\Pr\left(\sum_{t=1}^{T}(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})(v_{k^*(i,j,g)t}-v_{l^*(i,j,\widetilde{g})t})\le-\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2}-c_T\right)\right)\\
&\qquad+\Pr\left(\sum_{t=1}^{T}(v_{it}-v_{jt})(v_{k^*(i,j,g)t}-v_{l^*(i,j,\widetilde{g})t})\le-\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2}-c_T\right)\right)\Bigg].
\end{aligned}
\tag{A.9}
$$

Focus on the first term in (A.9). By Assumption 3(b), we have $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})^2]=c_{g,\widetilde{g}}$. So for T large enough, we have

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})^2]\ge\frac{2c_{g,\widetilde{g}}}{3}.$$

Applying Lemma B.5 in Bonhomme and Manresa (2015) to $z_t=(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})^2-\mathbb{E}[(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})^2]$, which satisfies appropriate mixing and tail conditions by Assumption 3(a) and (c), and taking $z=c_{g,\widetilde{g}}/6$ yields, as T tends to infinity,

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})^2\le\frac{c_{g,\widetilde{g}}}{2}\right)=o(T^{-\delta}),\tag{A.10}$$

uniformly over g and $\widetilde{g}$. Now, focus on the second term in (A.9). For T sufficiently large, we have $c_T\le c_{g,\widetilde{g}}/4$. Hence,

$$
\begin{aligned}
&\Pr\left(\sum_{t=1}^{T}(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})(v_{it}-v_{jt})\le-\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2}-c_T\right)\right)\\
&\quad\le\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})(v_{it}-v_{jt})\le-\frac{c_{g,\widetilde{g}}}{24}\right).
\end{aligned}
$$

By Assumption 3(c), the process $\{(\alpha^0_{gt}-\alpha^0_{gt})(v_{it}-v_{jt})\}_t$ has zero mean, and is strongly mixing with faster-than-polynomial decay rate. Moreover, for all i,t and $m>0$,

$$\Pr\Big(|(\alpha^0_{gt}-\alpha^0_{\widetilde{g}t})(v_{it}-v_{jt})|>m\Big)\le\Pr\left(|v_{it}-v_{jt}|>\frac{m}{2\sup_{\alpha_t\in\mathcal{A}}|\alpha_t|}\right),$$

so $\{(\alpha^0_{gt} - \alpha^0_{\widetilde{g}t})(v_{it} - v_{jt})\}_t$ also satisfies the tail condition of Assumption 3(a), albeit with a different constant $\widetilde{b} > 0$ instead of $b > 0$. Lastly, applying Lemma B.5 from Bonhomme and Manresa (2015) again with $z_t = (\alpha^0_{gt} - \alpha^0_{\widetilde{g}t})(v_{it} - v_{jt})$ and taking $z = c_{g,\widetilde{g}}/24$ yields

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha^0_{gt} - \alpha^0_{\widetilde{g}t})(v_{it} - v_{jt}) \leq -\frac{c_{g,\widetilde{g}}}{24}\right) = o(T^{-\delta}). \tag{A.11}$$

Note that the above upper bound does not depend on i, j, and $g, \widetilde{g}$. Hence, we deduce

$$\Pr\left(\sum_{t=1}^{T}(\alpha^0_{gt} - \alpha^0_{\widetilde{g}t})(v_{it} - v_{jt}) \leq -\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2} - c_T\right)\right) = o(T^{-\delta}), \tag{A.12}$$

uniformly over i, j, g and $\widetilde{g}$. Similarly,

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha^0_{gt} - \alpha^0_{\widetilde{g}t})(v_{k^*(i,j,g)t} - v_{l^*(i,j,\widetilde{g})t}) \leq -\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2} - c_T\right)\right) = o(T^{-\delta}), \tag{A.13}$$

uniformly over i, j, g, and $\widetilde{g}$. Similar arguments can be used to show that

$$\Pr\left(\frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})(v_{k^*(i,j,g)t} - v_{l^*(i,j,\widetilde{g})t}) \leq -\frac{T}{6}\left(\frac{c_{g,\widetilde{g}}}{2} - c_T\right)\right) = o(T^{-\delta}). \tag{A.14}$$

uniformly over i, j. Combining results (A.9)-(A.14), and taking supremum over all $(i, j) \in \{1, ..., N\}$ yields

$$\begin{aligned}
&\sup_{(i,j)\in\{1,\dots,N\}^2} \Pr(Z_{1NT}(i,j) = 1, \mathcal{E}_N) \\
&\leq G^0(G^0 - 1)[o(T^{-\delta}) + o(T^{-\delta}) + o(T^{-\delta}) + o(T^{-\delta})] \\
&= o(T^{-\delta}),
\end{aligned}$$

i.e., (A.2)' holds.

Step 2: (A.3)**' holds.** We have

$$\begin{aligned}
Z_{2NT}(i,j) \leq \mathbb{1}\Bigg\{ &\max_{(k,l)\in\mathscr{P}_2(\{1,\dots,N\}\setminus\{i,j\})} \Bigg| \frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})(v_{kt} - v_{lt}) \\
&+ \frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})(\alpha^0_{g^0_k t} - \alpha^0_{g^0_l t}) \Bigg| > c_T \Bigg\}.
\end{aligned}$$

By the union bound and the triangle inequality, we have

$$\Pr\left(Z_{2NT}(i,j)=1,\mathcal{E}_N\right) \leq \binom{N-2}{2} \sup_{(k,l)\in\{1,...,N\}^2} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}(v_{it}-v_{jt})(v_{kt}-v_{lt})\right|\right.$$
$$\left.+\left|\frac{1}{T}\sum_{t=1}^{T}(v_{it}-v_{jt})(\alpha^0_{g^0_k t}-\alpha^0_{g^0_l t})\right| > c_T\right).$$

Under the strong mixing and tail conditions given by Assumptions 3(a) and (c), the upper bound in the expression above can be shown to be $o(N^2T^{-\delta})$ uniformly over i,j by similar arguments as in Step 1 and Section A.1 (Step 2). Hence, (A.3)' holds and the proof of Proposition 2.4 is complete.

A.5 Proof of Corollary 2.5

Given Proposition 2.4, the proof is identical to that of Corollary 2.2.

A.6 Proof of Corollary 2.6

Given Corollary 2.5, the proof follows closely the same lines as that of Corollary 2.3 so we omit it here.

A.7 Proof of Proposition 3.1

We only prove the first point since the second follows from similar arguments (the proof is available upon request). The proof is similar to that of Proposition 2.1, in that only the additional control on the residual is required. We therefore keep the same notation for $Z_{1NT}(i,j)$ and $Z_{2NT}(i,j)$ by replacing $\widehat{W}^{PWD}_{ij}$ by $\widehat{W}^{2PWD}_{ij}$. Let $(i,j)\in\{1,...,N\}^2$, $\delta>0$, $\epsilon>0$, and $K>0$ such that $\Pr\left(||\widehat{\theta}^1-\theta^0||>K/\sqrt{T}\right)\leq\epsilon$.

1. We first show that $\sup_{(i,j)\in\{1,...,N\}^2}\Pr(Z_{1NT}(i,j)=1,||\widehat{\theta}^1-\theta^0||\leq K/\sqrt{T})=o(T^{-\delta})$. Following similar arguments than before (assume without loss $G^0\geq 2$), we have

$$Z_{1NT}(i,j)\leq \max_{g\neq\widetilde{g}} \mathbb{1}\left\{\left(\alpha^0_g-\alpha^0_{\widetilde{g}}\right)\left((\overline{x}_i-\overline{x}_j)'(\theta^0-\widehat{\theta}^1)+\overline{v}_i-\overline{v}_j\right)\leq -\frac{1}{2}(c^2_{g,\widetilde{g}}-c_T)\right\}.$$

Letting $\bar{a} = \sup\{a \in \mathcal{A}\}$, the triangle and Cauchy-Schwarz inequalities yield

$$\begin{aligned}(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(\overline{x}_i - \overline{x}_j)'(\theta^0 - \widehat{\theta}^1) &\geq -2\bar{a}||\widehat{\theta}^1 - \theta^0|| \times \left(\frac{1}{T}\sum_{t=1}^{T}||x_{it}|| + \frac{1}{T}\sum_{t=1}^{T}||x_{jt}||\right) \\ &\geq \frac{-2\bar{a}K}{\sqrt{T}} \times \left(\frac{1}{T}\sum_{t=1}^{T}||x_{it}|| + \frac{1}{T}\sum_{t=1}^{T}||x_{jt}||\right).\end{aligned}$$

For T sufficently large, $4\bar{a}KM/\sqrt{T} \leq \min_{g\neq\widetilde{g}} c_{g,\widetilde{g}}^2/4$ and $c_T \leq \min_{g\neq\widetilde{g}} c_{g,\widetilde{g}}^2/4$. Hence, by decompositing the probability event on the left-hand side and applying the union bound, we obtain,

$$\begin{aligned}&\Pr\left(Z_{1NT}(i,j) = 1, ||\widehat{\theta}^1 - \theta^0|| \leq K/\sqrt{T}\right) \\ &\quad \leq \sum_{g\neq\widetilde{g}} \Pr\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha_g^0 - \alpha_{\widetilde{g}}^0)(v_{it} - v_{jt}) \leq -\frac{c_{g,\widetilde{g}}}{4}\right) + G^0(G^0-1)\left[\Pr\left(\frac{1}{T}\sum_{t=1}^{T}||x_{it}|| \geq M\right)\right. \\ &\qquad \left. + \Pr\left(\frac{1}{T}\sum_{t=1}^{T}||x_{jt}|| \geq M\right)\right]. \qquad (A.15)\end{aligned}$$

Combining (A.6), (A.15), and Assumption 5(b) yields

$$\begin{aligned}&\sup_{(i,j)\in\{1,\ldots,N\}^2} \Pr\left(Z_{1NT}(i,j) = 1, ||\widehat{\theta}^1 - \theta^0|| \leq K/\sqrt{T}\right) \\ &\quad \leq G^0(G^0-1)[o(T^{-\delta}) + o(T^{-\delta}) + o(T^{-\delta})] \\ &\quad = o(T^{-\delta}). \qquad (A.16)\end{aligned}$$

2. Next, we can use similar arguments to obtain, for T sufficiently large,

$$\begin{aligned}&\sup_{(i,j)\in\{1,\ldots,N\}^2} \Pr\left(Z_{2NT}(i,j) = 1, ||\widehat{\theta}^1 - \theta^0|| \leq K/\sqrt{T}\right) \\ &\quad \leq G^0(G^0-1)\left[\sup_{(i,j)\in\{1,\ldots,N\}^2} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})\right| > \sqrt{c_T C}\right)\right. \\ &\qquad \left. + 2\sup_{i\in\{1,\ldots,N\}} \Pr\left(\frac{1}{T}\sum_{t=1}^{T}||x_{it}|| \geq M\right)\right], \qquad (A.17)\end{aligned}$$

for some constant $C > 0$. Similar arguments than in step 2. in Section A.1 show that

$$\sup_{(i,j)\in\{1,\ldots,N\}^2} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}(v_{it} - v_{jt})\right| > \sqrt{c_T C}\right) = o(T^{-\delta}). \qquad (A.18)$$

Combining (A.17)-(A.18), and Assumption 5(b) yields

$$\sup_{(i,j)\in\{1,\ldots,N\}^2} \Pr\left(Z_{2NT}(i,j) = 1, ||\widehat{\theta}^1 - \theta^0|| \leq K/\sqrt{T}\right) = o(T^{-\delta}). \qquad (A.19)$$

3. Finally, by the union bound

$$
\begin{aligned}
&\Pr\left(\sup_{(i,j)\in\{1,\ldots,N\}^2}|\widehat{W}_{ij}^{2PWD}-W_{ij}^0|>0\right)\\
&\quad\le \Pr\left(||\widehat{\theta}^1-\theta^0||>K/\sqrt{T}\right)\\
&\qquad+N(N-1)\left[\sup_{(i,j)\in\{1,\ldots,N\}^2}\Pr\left(Z_{1NT}(i,j)=1,||\widehat{\theta}^1-\theta^0||\le K/\sqrt{T}\right)\right.\\
&\qquad\left.+\sup_{(i,j)\in\{1,\ldots,N\}^2}\Pr\left(Z_{2NT}(i,j)=1,||\widehat{\theta}^1-\theta^0||\le K/\sqrt{T}\right)\right]\\
&\quad=\epsilon+o(N^2T^{-\delta}).
\end{aligned}
$$

As ϵ was unrestricted, the proof of Proposition 3.1 is complete.

A.8 Proof of Corollary 3.2

Given Proposition 3.1, the proof follows the same lines as that of Corollary 2.3 and 2.6 so we omit it here (see Bonhomme and Manresa (2015)'s Supplemental Material for details).

A.9 MATLAB Code

Below is some MATLAB code for the PWD and TPWD estimators. Replication codes for the Monte Carlo simulation and the empirical application can be found at martinmugnier.github.io.

```
classdef pwd_estimators

    methods(Static)

        function [G,grp_labels,grp_effects] = pwd(Y,c)
            % This function returns the PWD estimator.
            % INPUTS: Y : NxT array of balanced panel data outcome;
            %         c : scalar threshold.
            [N T] = size(Y);
            Ybar = nanmean(Y,2);
            % compute the W matrix
            W = (bsxfun(@minus,Ybar,Ybar').^2<=c);
```

```
        % obtain estimates for G and group labels
        G = size(unique(W,'rows'),1);
        [~,~,grp_labels] = unique(W,'rows');
        % generate group dummies
        exog = dummyvar(grp_labels);
        exog = repmat(exog',T,1);
        exog = reshape(exog,[],N*T)';
        endog = reshape(Y',N*T,1);
        % run pooled OLS and obtain group-specific effects estimates
        grp_effects = OLS(endog,exog);
    end

    function [G,grp_labels,params] = tpwd(Y,c)
        % This function returns the TPWD estimator.
        % INPUTS: Y : NxT array of balanced panel data outcome;
        %         c : scalar threshold.
        [N T] = size(Y);
        s = permute(Y-permute(Y,[3 2 1]),[1 3 2]);
        S = max(abs(mean(s.*permute(s,[5 4 3 2 1]),3)),[],[4 5]);
        % compute the W matrix
        W = (S<=c);
        % obtain estimates for G and group labels
        G = size(unique(W,'rows'),1);
        [~,~,grp_labels] = unique(W,'rows');
        % generate group dummies
        exog = dummyvar(grp_labels);
        exog = repmat(exog',T,1);
        exog = reshape(exog,[],N*T)';
        timedum = kron(ones(N,1), eye(T));
        % take interactions exog x timedum
        % ...TBD...
        % merge singleton groups
        % ...TBD...
        % run pooled OLS and obtain structural parameters
        endog = reshape(Y',N*T,1);
        params = OLS(endog,exog);
    end
```

```
    end
end

function[theta] = OLS(Y,X)
   % This function returns the OLS estimator.
   % INPUTS:
   % -------
   %          Y : n x 1 array;
   %          X : n x p array.
    theta = inv(X'*X)*X'*Y;
end
```

A.10 Tables and Figures

Table 2: Consistency Under IID Errors (PWD Estimator)

		$G^0 = 2$				$G^0 = 5$				$G^0 = 10$				$G^0 = 50$			
N	T	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	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

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

Table 3: Consistency Under Weakly Dependent Errors (PWD Estimator)

		$G^0 = 2$				$G^0 = 5$				$G^0 = 10$				$G^0 = 50$			
N	T	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	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

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

Table 4: Consistency Under Heteroskedastic Errors (PWD Estimator)

		$G^0=2$				$G^0=5$				$G^0=10$				$G^0=50$			
N	T	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}^{PWD}$	HD	RI	CPU time	$\widehat{G}$	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

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

Figure 2: SENSITIVITY OF THE NUMBER OF GROUPS $\widehat{G}$

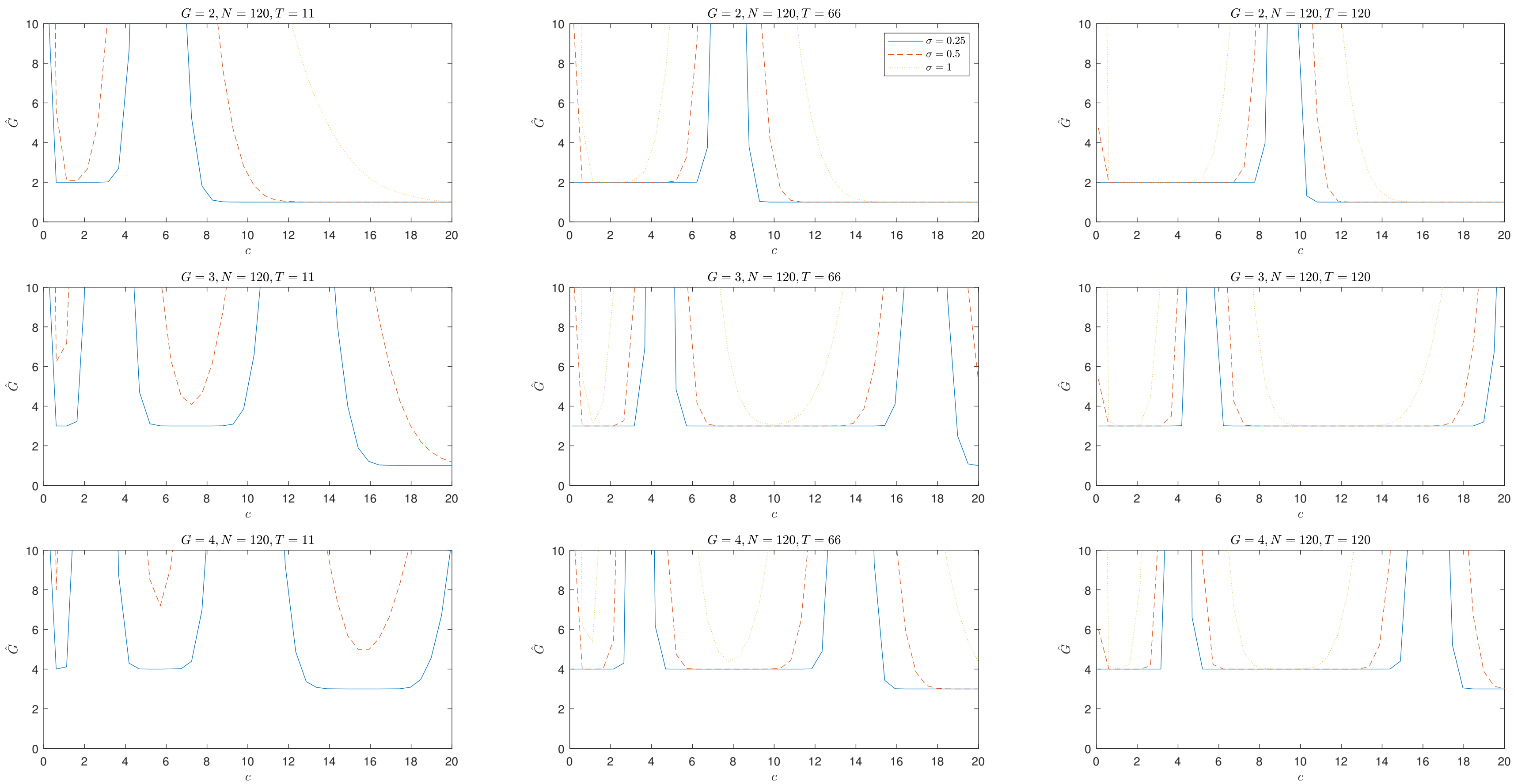


Figure 3: Sensitivity of the Rand Index

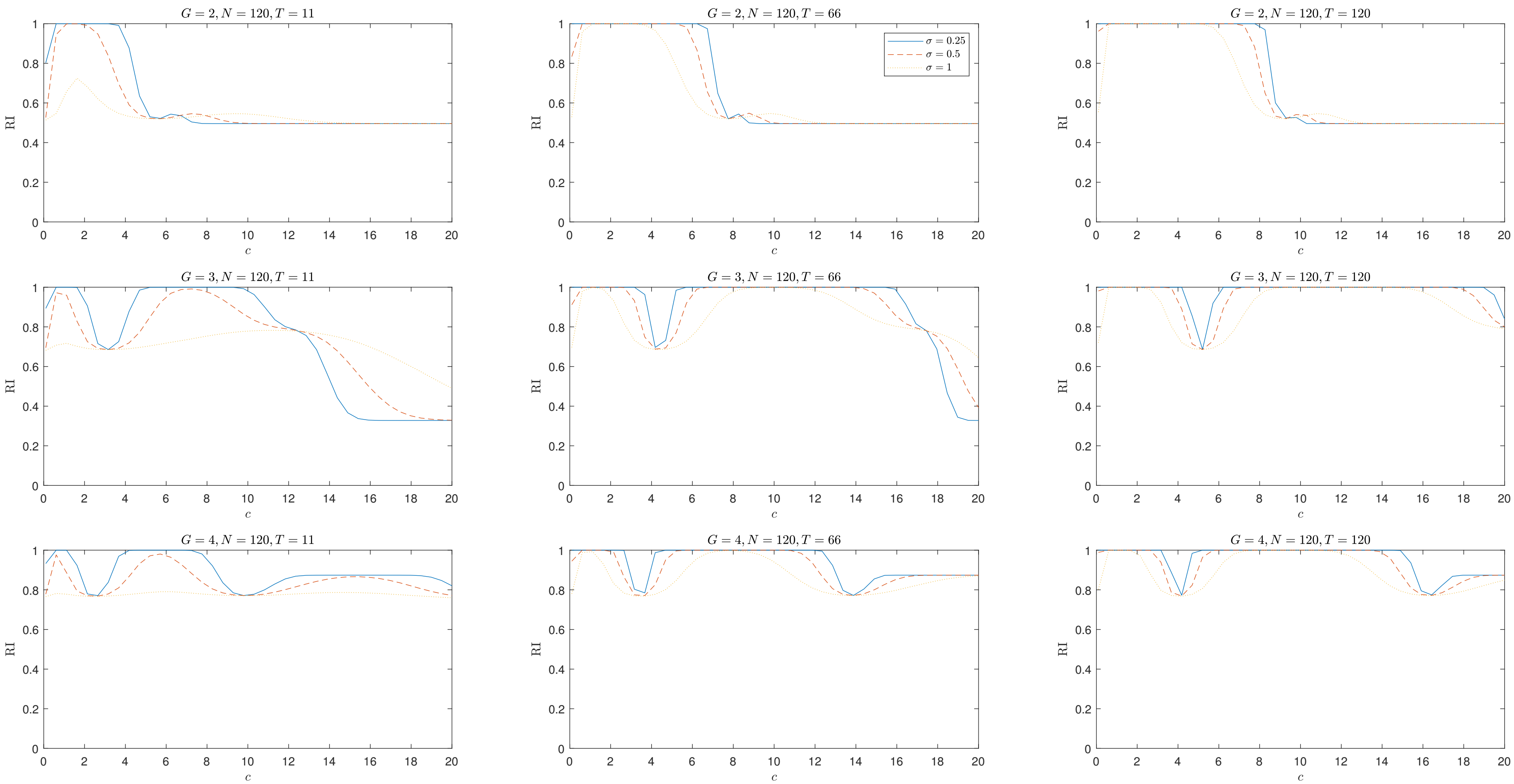


Figure 4: SENSITIVITY OF THE PRECISION RATE

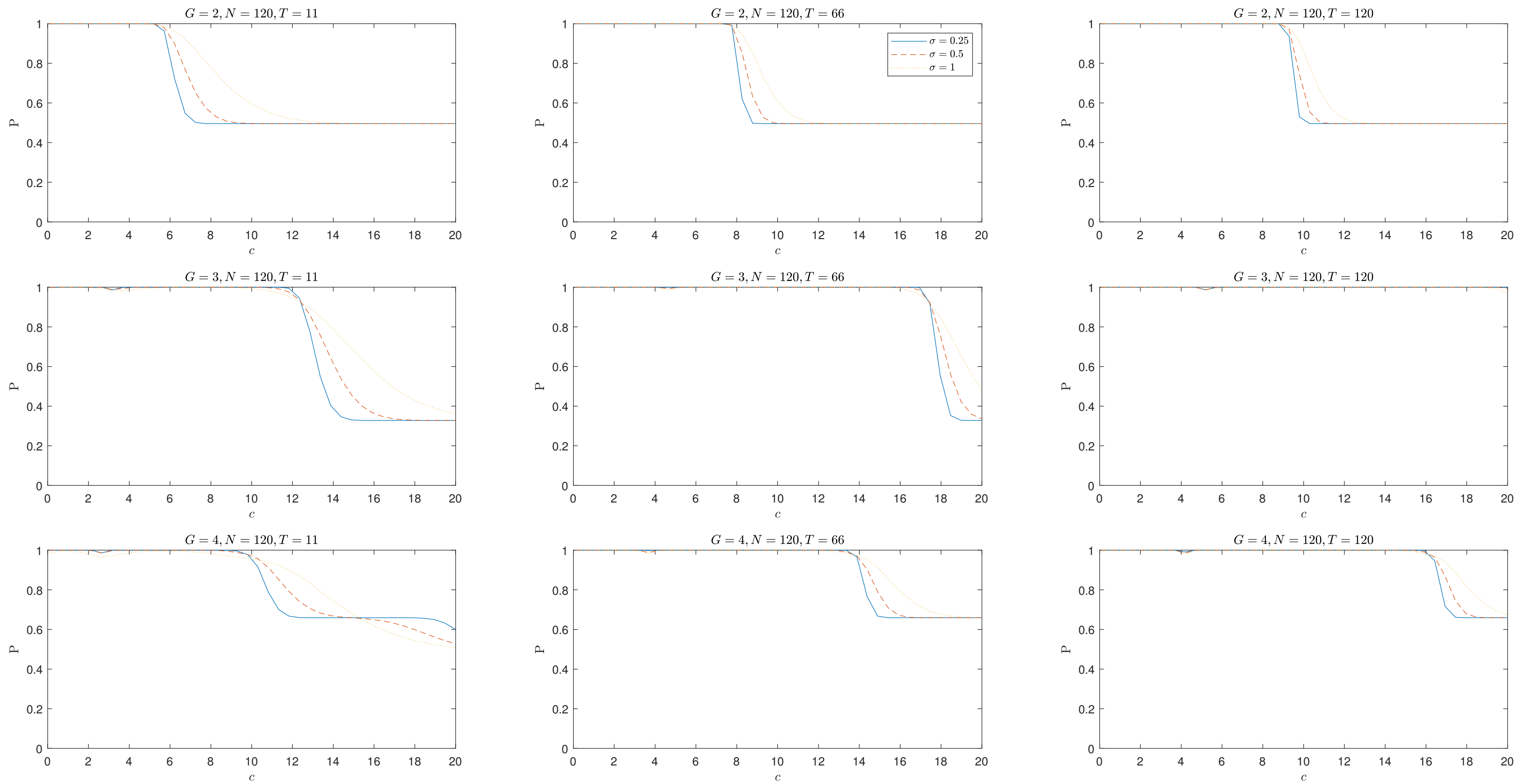


Figure 5: Sensitivity of the Recall Rate

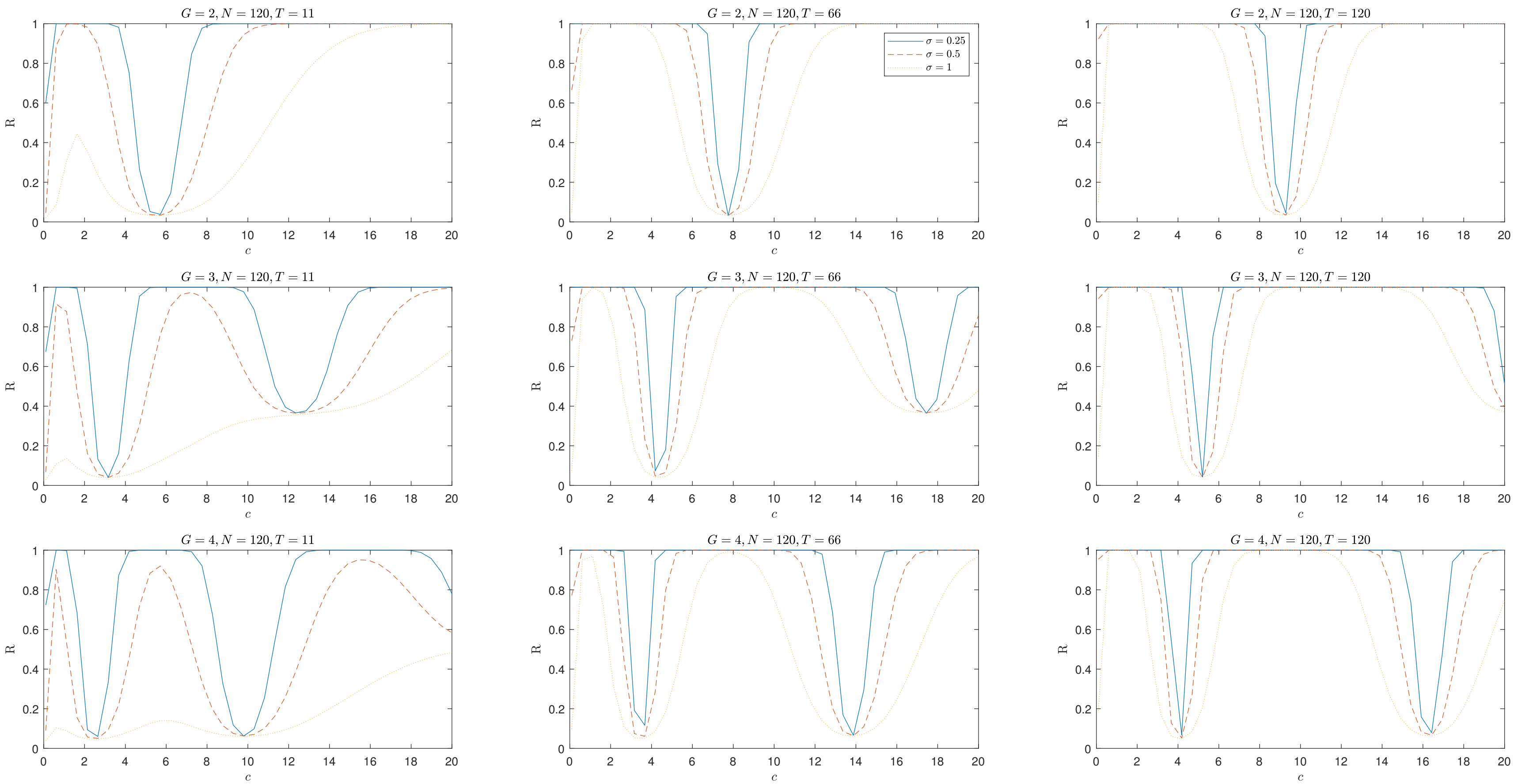


Figure 6: Sensitivity of the Hausdorff Distance

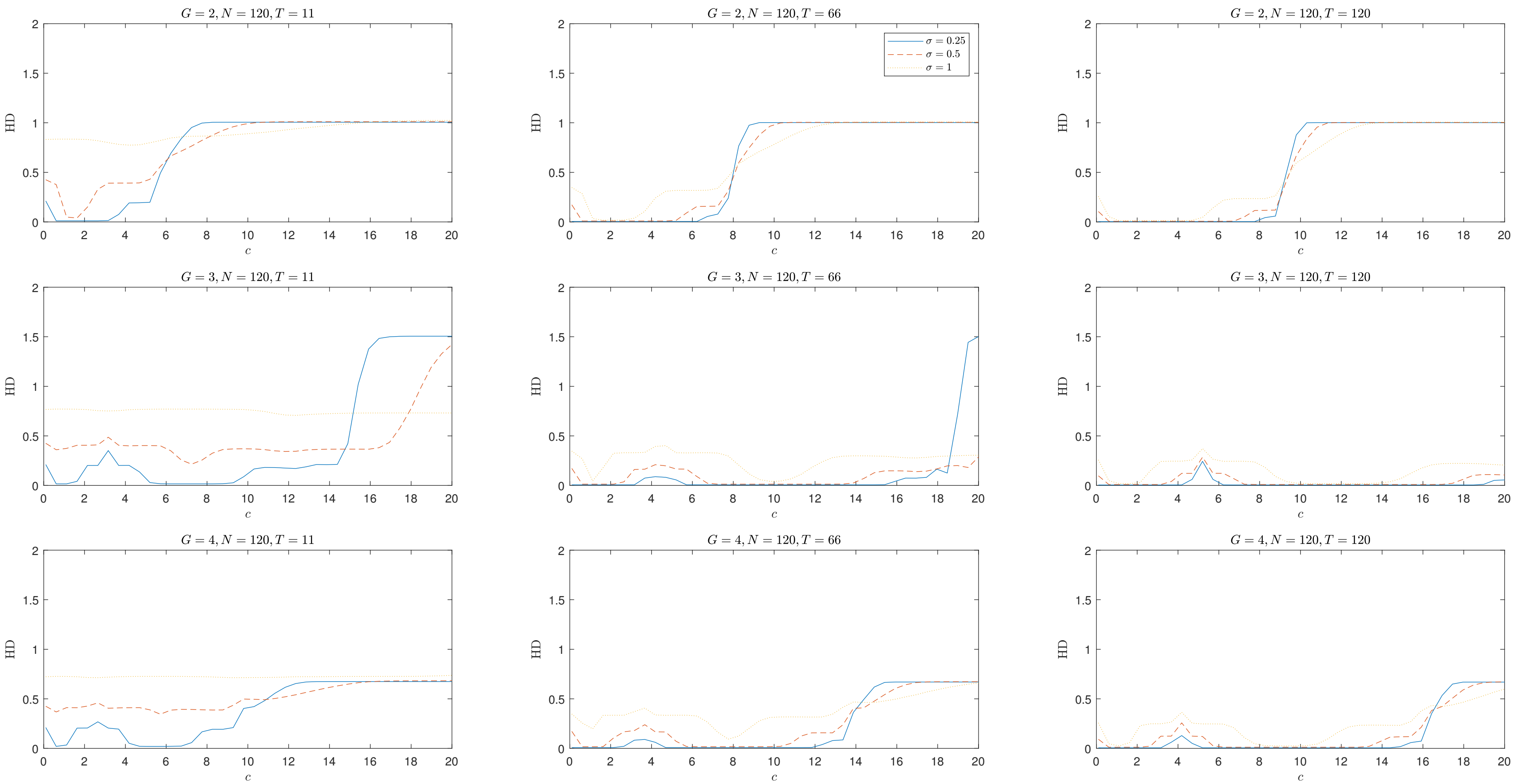


Figure 7: Calibrated Monte Carlo - Average $\widehat{G}$

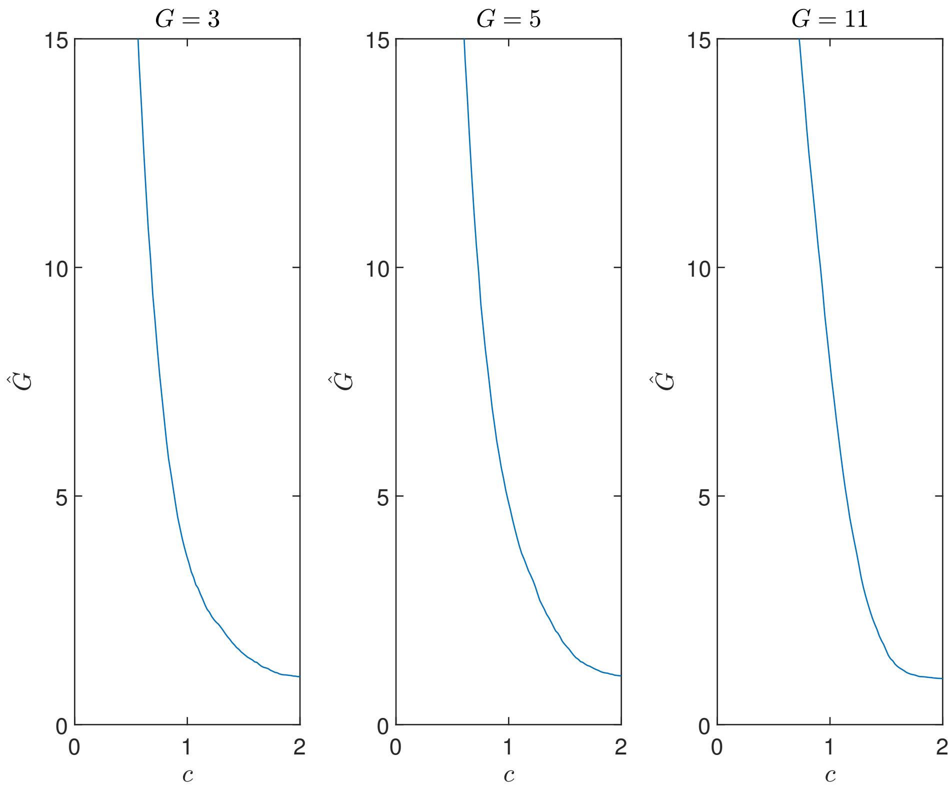


Figure 8: CALIBRATED MONTE CARLO - FREQUENCY OF $\mathbb{1}\left\{\widehat{G}=G\right\}$

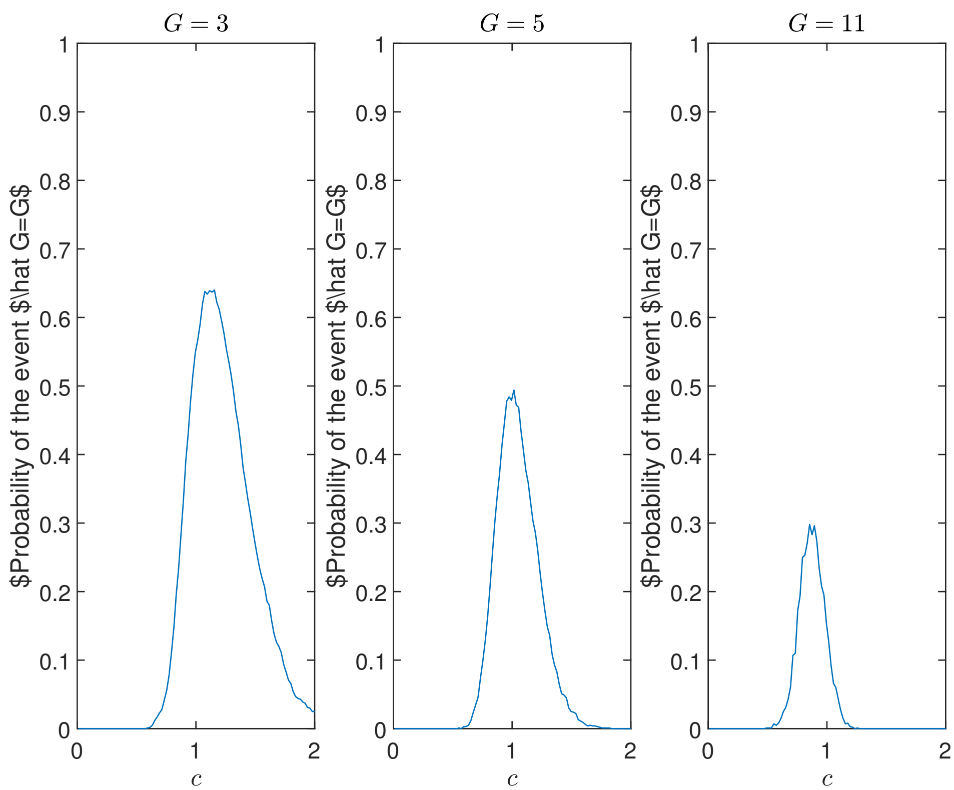


Figure 9: Calibrated Monte Carlo - Hausdorff Distance

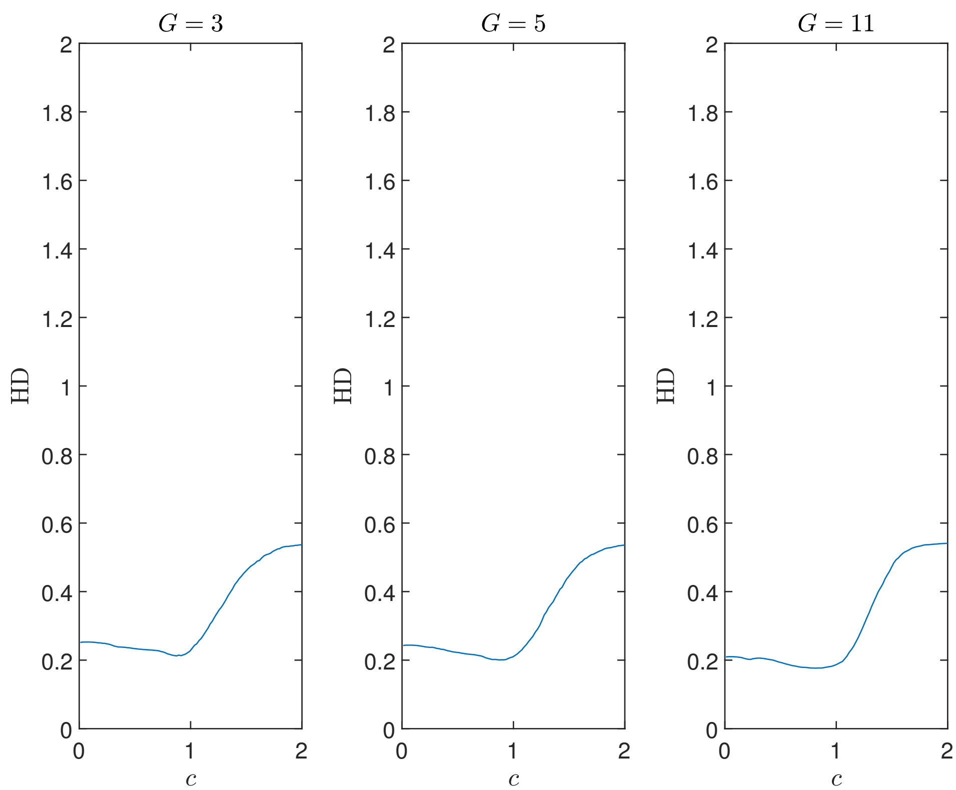


Figure 10: Calibrated Monte Carlo - Rand Index

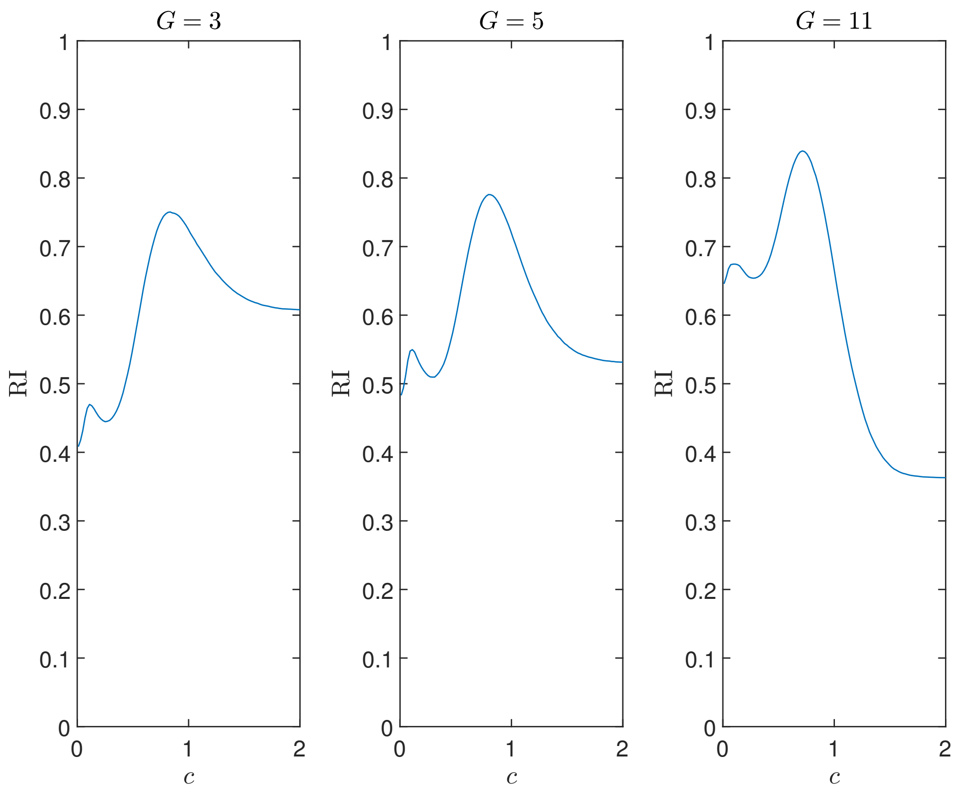


Figure 11: Calibrated Monte Carlo - Precision Rate

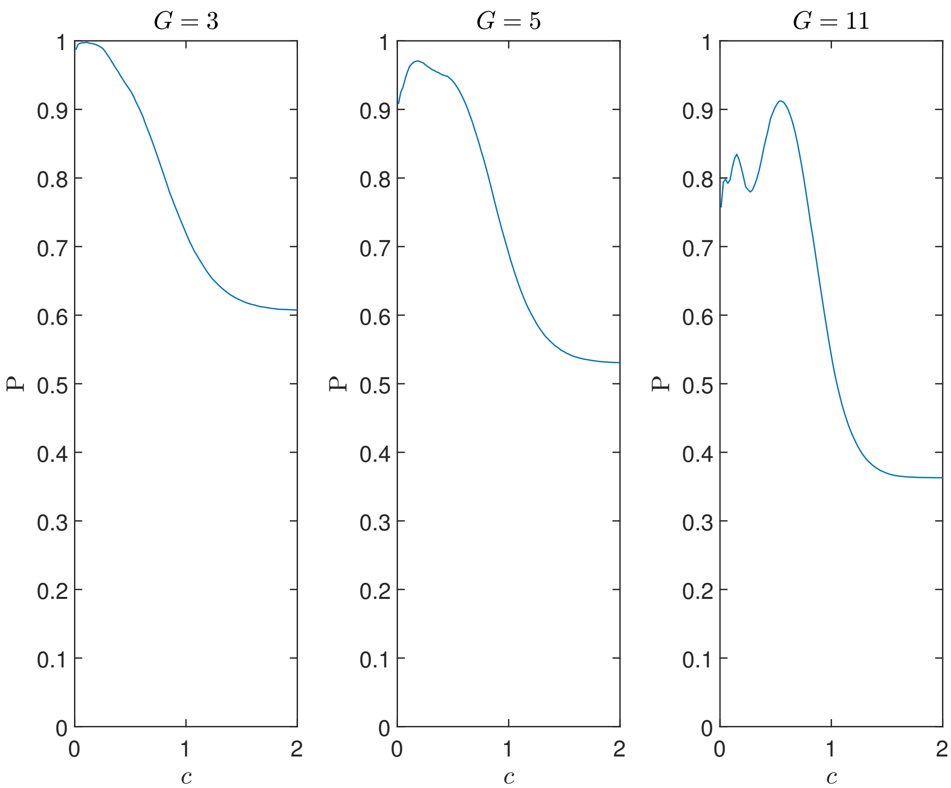


Figure 12: Calibrated Monte Carlo - Recall Rate