Optimal Cohort Partitions

Sofoklis Goulas
foundry10 and Yale University
sofoklis.goulas@yale.edu

Faidra Monachou
School of Management, Yale University
faidra.monachou@yale.edu

We study optimal partitions of a heterogeneous population into distinct groups. Through the lens of *majorization theory*, we provide an optimization framework for part-additive objectives that characterizes the fractional and integral partition polytopes, and identifies new classes of optimal policies. Our main result shows that, under edge-quasi-convex objectives, there exists an optimal fractional partition that is *locally uniform*, mixing adjacent types evenly across consecutive sections. Among integral partitions, there exists an optimal *locally extremal* partition, where every pair of sections, holding the others fixed, is maximally balanced or maximally polarized. Under Schur-convex objectives the assortative partition is optimal, while under Schur-concave objectives the uniform fractional and balanced integral partitions are. We next apply our framework to student assignment with peer effects. Using the workhorse *linear-in-means* (LIM) model, we characterize the optimal partitions for *performance*, *inequality*, *welfare*, and *diversity*, which generally conflict. The assortative partition maximizes welfare but produces the greatest inequality and the least diversity. The inequality-minimizing partition changes from balanced to dispersion-maximal as peer effects weaken. The performance-maximizing partition moves from assortative to upper-uniform to uniform as the planner's weights shift from higher- toward lower-performing sections. We further characterize *sorting-neutral* models that, despite being non-linear, exhibit LIM's invariance of total effort. Finally, we complement our theory with an empirical exercise on a novel dataset of quasi-random classroom assignments. We find that greater sorting leaves average performance outcomes broadly unchanged but is associated with sizable gains for lower-achieving students in the most sorted cohorts, while any gains for higher achievers are smaller and less robust.

Key words: partitions; majorization; peer effects; education operations.

1. Introduction

The optimal cohort partition problem involves a planner seeking how to distribute a heterogeneous population of individuals into distinct groups to optimize their objective, and is prevalent across a wide range of applications. Examples include assigning an incoming cohort of students to parallel sections (Carrell et al. 2013), forming teams of experts in organizations (Chade and Eeckhout 2018), and designing contests among teams (Ryvkin 2011), as well as related problems in experiment design (Bertsimas et al. 2015), group testing (Aprahamian et al. 2019), clustering (Boros and Hammer 1989), biodiversity maximization (Perez-Salazar et al. 2021), redistricting (Procaccia et al. 2024), sports drafts (Lee and Liu 2022), and more. Motivated by these applications, in this paper we propose a framework for optimal cohort partition problems under part-additive objectives through

the novel lens of *majorization theory*, and derive structural conditions for when mixing, sorting, or hybrid grouping is optimal. Our main research question is:

How should a planner optimally partition a heterogeneous population into groups under part-additive objectives?

The optimal cohort partition problem has both theoretical and policy appeal. Consider education, our main motivating setting. Understanding peer effects, that is, how the composition of one's peer group shapes individual learning outcomes and group performance, is one of the central questions in the economics of education (Sacerdote 2011, Epple and Romano 2011). The vast majority of research has been almost exclusively empirical, focusing on the empirical identification of peer effects. The literature has paid little attention to the theoretical microfoundations of peer effects and their implications for optimal partition design, even though several empirical works (e.g., Bhattacharya (2009), Carrell et al. (2013), Graham et al. (2014)) have theorized the idea that individuals could be sorted into peer groups to maximize efficiency. Our study helps bridge this theoretical gap by offering optimality guarantees and novel policy insights for when to mix, sort, or partially group individuals. Beyond education, our framework also applies to and extends theoretical results for related problems in team formation (Chade and Eeckhout 2018) and contest design (Ryvkin 2011), providing a common mathematical language for optimal cohort design in different settings.

Apart from its theoretical value, the cohort partition problem also has direct practical and policy relevance. For example, at the U.S. Air Force Academy, Carrell et al. (2013) made a rare empirical-to-design leap. They used experimentally identified peer effects to compute an "optimal" assignment of freshmen to squadrons designed to improve the performance of lower-performing students. The algorithm placed low-performing students with high-performing peers in the same squadrons while middle-performing students formed their own squadrons. Unexpectedly, the new policy harmed the target group—a result that highlights the sensitivity of outcomes to the choice of the partition policy and the planner's objective. Interestingly, this sensitivity extends beyond USAFA. Our Greek high school data also show that quasi-random variation in cohort sortedness generates heterogeneous outcomes for different groups of students (see Section 5).

Similar design questions also arise in educational policy around the world (Brunello and Checchi 2007, Meier and Schütz 2007). In 2021, the Greek Ministry of Education proposed moving from quasi-random class assignment to a "soft sorting" system based on prior achievement, prompting a debate about both equity and effectiveness.[1] Comparable debates over tracking versus detracking

[1] With the current approach in Greece, at each public school, students in first grades of elementary (1^{st} grade), middle (7^{th} grade), and high (10^{th} grade) school are quasi-randomly (alphabetically) divided into parallel groups ("sections") of approximately 20-25 students. The proposed redesign recommended "loosely" sorting students by prior performance while still maintaining a degree of diversity in each class (Hellenic Ministry of Education and Religious Affairs 2021).

policies appear in U.S. school systems (Betts 2011, Truong 2022). Together, these cases highlight the challenges practitioners and policymakers face when designing groups and the need for tools that translate evidence into robust assignment rules.

Our paper contributes to the literature along several dimensions:

A majorization-based framework for cohort partitions. Our first contribution is of conceptual and technical nature. Conceptually, we approach optimal partition problems through the novel perspective of majorization theory. This reduction is based on two key technical observations.

First, we establish a connection between majorization and feasible partitions (Proposition 1). Loosely speaking, we show that every partition policy can be represented by a vector majorized by the assortative partition. This allows us to characterize the geometry of the set of feasible partitions as a majorization polytope. For the fractional case, this polytope is characterized by Dahl (2001). Building upon his characterization, we show that the extreme points of the fractional partition polytope correspond to *locally uniform* partitions. Locally uniform partitions mix adjacent types within a segment and allocate them evenly across consecutive groups associated with that segment.

The integral case, however, requires new analysis, since to the best of our knowledge, the geometry of the integral majorization polytope has not been previously characterized. We establish that the extreme points of the integral partitions polytope belong to a new class of partition policies. *Locally balanced* partitions, the discrete counterpart of locally uniform partitions, are extreme points (Proposition 4), but constitute only a subset of them (Example 1), since integrality generates further extreme points that local balance does not capture. We show that all extreme points must be *locally extremal* (Proposition 5), that is, holding all other section sums fixed, every pair of sections is either as balanced or as polarized as it can feasibly be. Integrality thus pushes each pair to an extreme, and the resulting policies mix the two modes: the assortative partition polarizes every pair, a balanced partition levels every pair, and integrality creates extreme points that do both at once. The latter correspond to locally extremal partitions. This discrete partial analogue of Dahl's characterization of the fractional majorization polytope may be of independent technical interest.

Second, we show that many common optimization objectives admit a *part-additive* form, where each group's contribution is a function of the sum of its members' characteristics. This class is broad and includes canonical measures of performance and equity widely used across education, organizational design, and allocation settings. Together with the fact that under edge-quasi-convex objectives optimal solutions occur at extreme points (Hwang and Rothblum 1996), this observation reduces the original partition problem to characterizing the extreme points of the majorization polytopes above.

Main results: Characterizing optimal partition policies (Theorems 1 and 2). We introduce several novel structural classes of hybrid partition policies, including the families of *balanced*, *locally uniform*, *locally balanced*, and *upper-uniform* partitions, alongside canonical policies such as *assortative* (strict sorting) and *uniform* (full mixing) partitions. Our main results, Theorem 1 and Theorem 2, provide optimality guarantees across these classes under part-additive objectives. For *Schur-convex* objectives, the assortative partition is optimal. For *Schur-concave* objectives, the uniform partition is optimal over the space of fractional partitions, while a balanced partition becomes optimal when restricting attention to integral partitions. We further consider the broader class of edge-quasi-convex objective functions introduced by Hwang and Rothblum (1996), which generalizes quasi-convexity and Schur convexity. For *edge-quasi-convex* objectives, we show that *locally uniform* partitions are optimal among fractional policies, while among integral policies an optimum is always attained at a *locally extremal* partition. Table 1 summarizes these theoretical findings. Taken together, the results formalize when full sorting, full mixing, or hybrid grouping is optimal.

Applications: Characterizing optimal peer-group design in education and beyond. Motivated by open questions in the economics of education (Bhattacharya 2009, Epple and Romano 2011, Sacerdote 2011, Carrell et al. 2013, Graham et al. 2014) and the ongoing educational policy debates discussed earlier, we apply our majorization-theoretic optimization toolkit to study how a planner can strategically shape peer effects through cohort structure. We characterize the optimal partition for four fundamental objectives (*performance*, *inequality*, *welfare*, and *diversity*) under the standard linear-in-means model (Section 4.2), and extend the performance analysis to common non-linear peer-effects models in education (Section 4.3) and other settings (Appendix C), complementing the predominantly empirical literature with theoretical guarantees.

We begin with the workhorse model from the empirical literature on educational peer effects, namely the *linear-in-means* (LIM) model, which posits that each student's action linearly depends on the mean action of her peers. Table 2 summarizes our results on the optimal partitions under the LIM model across objectives. For performance, measured as the weighted sum of total effort across sections, we find that as the planner's priority shifts from the highest- to middle- to lowest-performing sections, the optimal policy moves from assortative to upper-uniform to fully uniform (Proposition 6). With respect to inequality, our results are perhaps surprising. The uniform partition—often deemed the natural egalitarian benchmark—is not optimal in general. Instead, the optimal partition depends on the strength of peer effects. When peer effects are strong, every Gini-minimizing integral partition is balanced (Proposition 7(i)). When they are weak, minimizing inequality is instead equivalent to maximizing the attribute dispersion between classmates

(Lemma B.6); the snake partition used in sports drafts is optimal under certain regimes (Proposition B.14). Furthermore, there exist cohorts for which no fractional partition achieves perfect equality in effort (Proposition 7(iii)). Regarding welfare and diversity, we find that the assortative partition (Proposition B.15), and the uniform partition (or a balanced partition if restricted to integral solutions only) are respectively optimal (Proposition B.16).

All of our core theoretical results above are established in the linear-in-means (LIM) model. A long-standing critique of LIM, dating back to Sacerdote (2011) and others, is that it implies a "net-zero" effect of sorting on aggregate outcomes: swapping a high-ability student with a low-ability student across sections does not change the total unweighted performance when section sizes are equal. A natural question is whether this sorting-neutrality for unweighted objectives is specific to LIM or persists under richer, potentially non-linear peer-effects structures. In Section 4.3, we address this question by studying the class of *sorting-neutral* peer-effects models. Building upon a characterization by Bergstrom and Varian (1985), we show that sorting-neutrality extends beyond LIM to a broader family of models whose equilibrium effort profiles combine a linear individual term with a (possibly non-linear) part-additive component (Proposition 8). We also identify prominent non-linear models that do *not* exhibit sorting-neutrality. Under the quadratic-in-means specification (Garlick 2018), the optimal policy may switch between assortative and uniform (or balanced, in the integral case) grouping depending on the exact parameter values (Proposition 9). Conversely, under a social norm model (Boucher et al. 2024), the assortative partition maximizes total effort for $n \geq 3$, while for $n = 2$ the model is sorting-neutral (Proposition 10). Taken together, these results illustrate that departures from sorting-neutrality can favor either sorting or mixing, depending on the underlying behavioral microfoundations.

While our primary application area is education, in Appendix C we showcase how our framework can be applied more broadly to other partition problems. In particular, Appendix C.1 revisits the optimal assignment of experts to teams, studied by Chade and Eeckhout (2018), while Appendix C.2 studies the (approximately) optimal design of Tullock contests (Tullock 1980) with weakly heterogeneous players (Ryvkin 2011).[2]

Finally, in Section 5 we complement these theoretical results with an empirical exercise using a novel dataset from Greek public high schools. The dataset is unique in its natural variation in classroom sorting at grade 10, generated by the quasi-random (alphabetical) process of classroom assignment. We leverage this quasi-random variation in cohort partition to empirically examine the impact of *"sortedness"*—an empirical measure of the degree to which cohort partitions deviate from a uniform distribution—on students' academic progress. We find that the effects of sorting

[2] Rosenzweig and Xu (2023) recently applied a simplified version of Tullock's model to study educational peer effects.

are heterogeneous and nonlinear. Average outcomes are unaffected once individual and cohort-level prior performance are accounted for, consistent with sorting neutrality. But in the most sorted cohorts, lower-achieving students experience sizable gains—up to 0.14 standard deviations in Modern Greek Language. At the same time, higher-achieving students see no corresponding benefits and, if anything, losses in Modern Greek Language. These findings highlight the heterogeneous consequences of partition design, offering empirical motivation for our theoretical framework.

The remainder of the paper is organized as follows. In Section 1.1, we discuss the related literature. In Section 2, we introduce our baseline model, alongside other technical preliminaries. In Section 3, we conceptualize optimal partitions with the framework of majorization theory and characterize the optimal fractional and integral partition policies under various classes of part-additive objectives. In Section 4, we define the linear-in-means model and then characterize the optimal partitions across four objectives. Furthermore, we characterize sorting-neutral models and study two alternative non-linear peer effects models. In our empirical Section 5, we present our data and empirical findings. Section 6 concludes.

1.1. Related literature

Our work contributes to literature in operations research, economics, and computer science.

Theoretical foundations of optimal partitions. The existing literature on optimal partitions has focused almost exclusively on integral partitions and typically falls into two streams. The first stream (see, e.g., Chakravarty et al. (1982, 1985), Anily and Federgruen (1991), Hwang and Mallows (1995), Hwang and Rothblum (1996)), synthesized in the comprehensive overview of Hwang and Rothblum (2011), examines the properties of part-additive objectives under certain shape constraints that guarantee the existence of an optimal consecutive partition. In consecutive partitions, the indices of the elements of each set are consecutive integers, thus consecutive partitions can be viewed as set-based permutations of the assortative partition under the equal-size constraint in this paper. These results, however, do not deliver a general characterization of optimal partitions for broader families of part-additive objectives.

A second stream derives optimal partitions within application-specific, parametric environments, typically solved by exploiting the structure of each model or via numerical methods (e.g., Chakravarty et al. 1982, Gal and Klots 1995, Boros and Hwang 1996, Schreiber 2014). A prominent example is the standard clustering problem (Boros and Hammer 1989, Boros and Hwang 1996): Boros and Hwang (1996) show that nested partitions optimize variants of sum-of-squares objectives. In economics, Chade and Eeckhout (2018) show that uniform grouping maximizes the total team value; Appendix C.1 revisits their result.

Our work provides a more general characterization of optimal partitions under broad part-additive objectives, yielding structural results for both fractional and integral settings and uncovering new families of optimal policies—such as locally uniform and locally balanced partitions—that do not appear in prior work. Conceptually, we introduce a majorization-theoretic lens to the optimal cohort partition problem.

Peer effects in education and grouping students. A large empirical literature in economics documents the importance of peer effects in shaping student outcomes (Manski 1993, Hoxby 2000, Lazear 2001, Sacerdote 2001, Hoxby and Weingarth 2005, Bhattacharya 2009, Calvó-Armengol et al. 2009, Lavy et al. 2009, Chetty et al. 2011, Lavy et al. 2012, Carrell et al. 2013, Kalogrides and Loeb 2013, Bursztyn et al. 2019, Rosenzweig and Xu 2023, Boucher et al. 2024), including two surveys by Epple and Romano (2011) and Sacerdote (2011). This literature focuses almost entirely on identification rather than on the theoretical microfoundations of peer effects and the design of optimal peer groups. An exception is Boucher et al. (2024), who develop game-theoretic microfoundations for the classic linear-in-means model and introduce a new social norm model; we analyze these models in Section 4. A small number of papers propose potential gains from optimal cohort design (Bhattacharya 2009, Graham et al. 2014, Carrell et al. 2013), but to the best of our knowledge, no existing work provides structural optimality conditions for cohort design under peer effects or characterizes the form that mixing should take. The closest paper in spirit to ours is the USAFA experiment of Carrell et al. (2013), who estimate a model of peer effects, formulate an optimization problem aimed at improving the performance of the bottom one-third of students without harming others, and computationally solve for the optimal assignment of freshmen to squadrons.

Two additional papers consider classroom partitioning without peer effects. Arnott and Rowse (1987) study joint allocation of students and resources, showing numerically that optimal partitions can exhibit irregular mixing patterns, while Agrawal et al. (2014) propose heuristics based on ad hoc gain functions. Our work complements these literatures by providing a general optimization framework for the cohort partition problem across broader classes of objectives and peer effects models, and optimality conditions that characterize when and in what form mixing is optimal.

2. Baseline setting and preliminaries

Section 2.1 introduces the baseline setting. Section 2.2 presents several technical preliminaries from majorization theory. Section 2.3 defines several families and structural types of partition policies. We occasionally borrow terminology from educational settings, such as assigning elementary school students to classrooms or MBA students to parallel sections. While education provides a natural and intuitive interpretation (see Section 4), the same underlying framework applies broadly to

contexts in which a planner partitions heterogeneous agents into groups; we analyze two such applications in Appendix C.

2.1. Main setup

A planner needs to optimally partition a *cohort* $\mathcal{N}$ of $N = |\mathcal{N}|$ heterogeneous *students*, indexed by $i \in [N]$, into $K \geq 2$ *sections*, indexed by $j \in [K]$. Each student i is characterized by an *attribute* $x_i \in \mathbb{R}_+$; x_i could represent, e.g., the student's prior performance, whereas the term cohort represents all students of the same grade at a school. Without loss of generality, we assume that $x_1 \geq x_2 \geq \ldots \geq x_N$. We also rule out the trivial case $x_1 = x_2 = \ldots = x_N$.

In the notation below, we use boldface $\mathbf{x} = (x_1, \ldots, x_N)$ to denote vectors in $\mathbb{R}_+^N$. We denote by $\overline{x}$ the mean value of $\mathbf{x}$. The subscript i typically refers to a student in a given section and j to a section; we sometimes drop j when it is clear from the context. We use the subscript $_{-i}$ to denote all other students in the same section as student i (i.e., her peers).

Partitions. We define partitions of a cohort $\mathcal{N}$ and their properties. A *K-way integral partition*, or simply *integral partition*, of $\mathcal{N}$ is a finite collection of sets $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$ where $\pi_1, \ldots, \pi_K$ are pairwise disjoint, nonempty sets of equal size n, whose union is $\mathcal{N}$. We refer to K as the size of π, and to the sets $\pi_1, ..., \pi_K$ as the parts of π; part π_j corresponds to section j. We require that all $\pi_1, ..., \pi_K$ are of size n thus $N = nK$. For simplicity, we represent each part π_j by the indices of its elements, that is, $\pi_j = \{i_1, \ldots, i_n\}$. Furthermore, the summation vector $\mathbf{X}^\pi$ of an integral partition π is given by

$$\mathbf{X}^\pi = \left(\sum_{i \in \pi_1} x_i, \ldots, \sum_{i \in \pi_K} x_i \right) \in \mathbb{R}_+^K,$$

where its coordinates are denoted by $X_1^\pi, \ldots, X_K^\pi$, respectively.

We also define *K-way fractional partitions* $M = (\mu_{ij}) \in \mathbb{R}_+^{N \times K}$, or simply *fractional partitions*, where $\mu_{ij} \geq 0$ denotes the fractional assignment of student i with attribute x_i to section j. A fractional partition is *feasible* if $\sum_{i \in [N]} \mu_{ij} = n$ for all $j \in [K]$ and $\sum_{j \in [K]} \mu_{ij} = 1$ for all $i \in [N]$. In settings where multiple individuals share similar attributes, fractional partitions prescribe how (nearly) identical individuals should be split across groups (see Figure 1 that illustrates such a scenario). In educational contexts, such a coarse categorization is particularly common, e.g., categorizing students as low, middle, or high GPA (Carrell et al. 2013, Lavy et al. 2009). Another interpretation of fractional partitions that is more applicable to partitions of experts or workers is time-splitting (Chade and Eeckhout 2018); see Appendix C.1. In this case, μ_{ij} equals the fraction of work time that expert i spends in team j.

As with integral partitions, the summation vector $\mathbf{X}^M$ of a fractional partition M is given by

$$\mathbf{X}^M = \left(\sum_{i \in [N]} \mu_{i1} x_i, \ldots, \sum_{i \in [N]} \mu_{iK} x_i \right) \in \mathbb{R}_+^K,$$

where its coordinates are denoted $X_1^M, \ldots, X_K^M$, respectively. Note that any integral partition π can be written as a fractional partition restricted to $\mu_{ij} \in \{0,1\}$, that is, $\mu_{ij} = \mathbf{1}\{i \in \pi_j\}$.

We denote the set of all feasible integral and fractional partitions of $\mathcal{N}$ by $\Pi_{\mathcal{I}}$ and $\Pi_{\mathcal{F}}$, respectively. Note that $\Pi_{\mathcal{I}} \subset \Pi_{\mathcal{F}}$.

Objective. The planner's problem concerns the selection of a partition $\pi \in \Pi_{\mathcal{I}}$ (respectively, $M \in \Pi_{\mathcal{F}}$) that maximizes an *objective function* $F(\pi)$ (respectively, $F(M)$).

For a vector $\mathbf{X} \in \mathbb{R}_+^K$ and $j = 1, \ldots, K$, let $X_{[j]}$ be the j-th largest coordinate of $\mathbf{X}$. Since section labels are immaterial in our setting, we henceforth relabel sections and, without loss of generality, restrict attention to integral (respectively, fractional) partitions with summation vectors such that $X_{[j]}^{\pi} = X_j^{\pi}$ (respectively, $X_{[j]}^M = X_j^M$). Let $\mathcal{D} = \{\mathbf{X} \in \mathbb{R}_+^K : \ X_1 \geq X_2 \geq \ldots \geq X_K\}$. An objective F is called *part-additive* if there exists $f : \mathcal{D} \to \mathbb{R}$ such that $F(\pi) = f(\mathbf{X}^{\pi})$ for all $\pi \in \Pi_{\mathcal{I}}$ (respectively, $F(M) = f(\mathbf{X}^M)$ for all $M \in \Pi_{\mathcal{F}}$).

2.2. Technical preliminaries

Here we briefly present two standard definitions from majorization theory (Marshall et al. 2011) and refer the reader to Appendix A for additional technical details.

Majorization. A vector $\mathbf{X} \in \mathbb{R}_+^K$ *majorizes* a vector $\mathbf{Y} \in \mathbb{R}_+^K$, namely $\mathbf{X} \succ \mathbf{Y}$, if for all $k = 1, \ldots, K$,

$$\sum_{j \in [k]} X_{[j]} \geq \sum_{j \in [k]} Y_{[j]} \quad \text{and} \quad \sum_{j \in [K]} X_{[j]} = \sum_{j \in [K]} Y_{[j]}.$$

***L*-curves.** Majorization has a useful geometric representation for vectors $\mathbf{X} \in \mathbb{R}_+^K$. For a vector $\mathbf{X} \in \mathbb{R}_+^K$, we define the function $L(z; \mathbf{X}) : [0, K] \to \mathbb{R}$ as the piecewise linear interpolant of the $K+1$ points $(0,0)$ and $(k, \sum_{j \in [k]} X_{[j]})$ for $k = 1, \ldots, K$. The graph of $L(z; \mathbf{X})$ is called the *L-curve of* $\mathbf{X}$, also known as its *Lorenz curve*.

Schur convexity. A function $f : \mathbb{R}_+^K \to \mathbb{R}_+$ is called *Schur convex* (respectively, *Schur concave*) if $f(\mathbf{X}) \geq f(\mathbf{Y})$ (respectively, $f(\mathbf{X}) \leq f(\mathbf{Y})$) for all $\mathbf{X}, \mathbf{Y} \in \mathbb{R}_+^K$ with $\mathbf{X} \succ \mathbf{Y}$.

If f is symmetric and all its first partial derivatives exist, then f is Schur convex if and only if $(x_i - x_j)\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right) \geq 0$ for all $x \in \mathbb{R}_+^K$, $i \neq j$.

2.3. Special classes of partitions

Throughout the paper, we will encounter several families of partition policies. Next, we define three such special partitions—the *assortative*, the *uniform*, and the *t-upper-uniform* partition—and then the class of *balanced*. In Section 3, we will introduce the classes of *locally uniform*, *locally balanced*, and *locally extremal* partitions. Figure 1 illustrates different types of partition policies.

Assortative partition. Let $\widetilde{\mu}_{ij} = \mathbf{1}\{(j-1)n + 1 \leq i \leq jn\}$, $i \in [N]$, $j \in [K]$ denote the *assortative* partition. Partition $\widetilde{M}$ assortatively assigns the group of n students with the highest attributes

$x_1, \ldots, x_n$ to section 1, the group of n students with the next highest attributes $x_{n+1}, \ldots, x_{2n}$ to section 2, and so on. Thus,

$$\widetilde{\mathbf{X}} := \mathbf{X}^{\widetilde{M}} = \left(\sum_{i=1}^{n} x_i, \ldots, \sum_{i=(K-1)n+1}^{N} x_i \right).$$

Of course, $\widetilde{M}$ can also be represented as an integral partition $\widetilde{\boldsymbol{\pi}} \in \Pi_{\mathcal{I}}$, where $\widetilde{\pi}_j = \{i \in [N] : (j-1)n+1 \leq i \leq jn\}$, $j \in [K]$.

Uniform partition. Let $\underline{\mu}_{ij} = 1/K$, $i \in [N]$, $j \in [K]$ denote the *uniform* partition. Intuitively, $\underline{M}$ represents a fully mixed partitioning scheme as it distributes students of different attributes evenly across sections. Thus,

$$\underline{\mathbf{X}} := \mathbf{X}^{\underline{M}} = \left(\sum_{i=1}^{N} \frac{x_i}{K}, \ldots, \sum_{i=1}^{N} \frac{x_i}{K} \right).$$

t-upper-uniform partition. For $t \in [K]$, the *t-upper-uniform* M evenly distributes the top tn students $x_1, \ldots, x_{tn}$ to sections 1 to t and then assortatively assigns the remaining $N - tn$ students $x_{tn+1}, \ldots, x_N$ to sections $t+1$ to K. Formally, $\mu_{ij} = \frac{1}{t}$ for $1 \leq i \leq tn$, $1 \leq j \leq t$, $\mu_{ij} = \mathbf{1}\{j = \lceil \frac{i-tn}{n} \rceil + t\}$ for $tn+1 \leq i \leq N$, $t+1 \leq j \leq K$, and $\mu_{ij} = 0$ otherwise.

Balanced partitions. A partition $\pi \in \Pi_{\mathcal{I}}$ is called *balanced* if for every pair of sections $j \neq k$, the pair (π_j, π_k) minimizes $|X_j^{\pi} - X_k^{\pi}|$ over all 2-way partitions of $\pi_j \cup \pi_k$ into parts of size n. Equivalently, for every $j \neq k$ there is no $\pi' \in \Pi_{\mathcal{I}}$ with $\pi'_\ell = \pi_\ell$ for all $\ell \notin \{j, k\}$ and $\left|X_j^{\pi'} - X_k^{\pi'}\right| < \left|X_j^{\pi} - X_k^{\pi}\right|$. Intuitively, a partition is balanced if no two sections can be reshuffled, by re-partitioning their combined set of students into two groups of size n, in a manner that brings their respective sums closer together.

3. A majorization-theoretic framework of optimal partitions

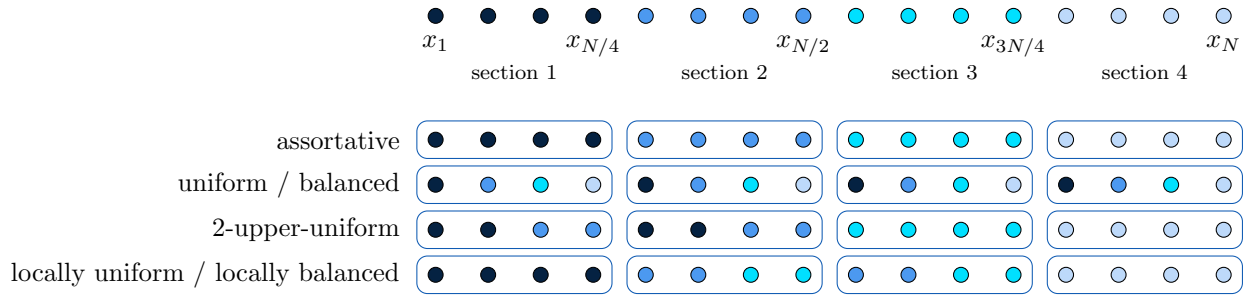
This section presents a majorization-theoretic framework for optimal fractional and integral partitions under part-additive objectives. In particular, Section 3.1 establishes the connection between feasible fractional partitions and the majorization polytope and characterizes optimal policies under Schur convex and Schur concave objectives. Section 3.2 introduces the class of locally uniform partitions and provides sufficient conditions for their optimality. Finally, Section 3.3 extends the results to integral partitions, characterizing the optimality of locally extremal partitions.

3.1. Majorization polytope equivalence and optimal partitions for Schur convex and Schur concave objectives

As we will see in Sections 4 and Appendix C, for many common objective functions $F : \mathbb{R}_+^{N \times K} \to \mathbb{R}$, our original optimization problem

$$\max_{M \in \Pi_{\mathcal{F}}} F(M)$$

Figure 1 **Illustration of common partition policies**



Notes: This example splits $N = 16$ students into $K = 4$ sections. In this example, we consider for simplicity that, for every fixed $1 \le j \le 4$, students $n(j-1)+1 \le i \le nj$ (indicated by the same shade of blue) share identical attributes, e.g., for $j = 1$, we have $x_1 = x_2 = x_3 = x_4$. We present five partition policies: **(i)** The *assortative* partition greedily assigns students to sections in order of their attributes: students $n(j-1)+1 \le i \le nj$ of the same color are assigned to section j. **(ii)** The fractional *uniform* partition splits every student evenly across all K sections ($\mu_{ij} = 1/K$). Because the n students within each shade share the same attribute, this is depicted here as one whole student of each shade per section, thus coinciding with the integral *balanced* partition with the same summation vector. **(iii)** The *2-upper-uniform* partition is identical to the assortative partition for sections 3 and 4 but sections 1 and 2 mix the top half of students (marked by the two darkest blue shades) evenly. **(iv)** In the depicted example, this *locally uniform* partition coincides with the assortative partition except for sections 2 and 3 that evenly mix the middle half of students. This particular locally uniform partition happens to coincide with its integral counterpart, a *locally balanced* partition. Refer to Sections 2.3, 3.2, and 3.3 for further details.

reduces to maximizing a part-additive function $f(\mathbf{X}^M)$:

$$\max_{M \in \Pi_{\mathcal{F}}} f(\mathbf{X}^M). \tag{3.1}$$

When $\pi \in \Pi_{\mathcal{I}}$ is an integral partition and $F(\pi) = f(\mathbf{X}^\pi)$, the analogous reduced problem is:

$$\max_{\pi \in \Pi_{\mathcal{I}}} f(\mathbf{X}^\pi). \tag{3.2}$$

These reductions motivate the main question of this section: *If $F(M) = f(\mathbf{X}^M)$ for all $M \in \Pi_{\mathcal{F}}$ (respectively, $F(\pi) = f(\mathbf{X}^\pi)$ for all $\pi \in \Pi_{\mathcal{I}}$), under what conditions on f can we characterize the optimal fractional partitions in (3.1) (respectively, the optimal integral partitions in (3.2))?*

To answer this question, we first establish a connection between majorization and feasible fractional partitions. More specifically, we show that the summation vector $\mathbf{X}^M$ of any feasible fractional partition M must be majorized by the summation vector $\widetilde{\mathbf{X}}$ of the assortative partition. Conversely, we also show that for any $\mathbf{X} \prec \widetilde{\mathbf{X}}$, there exists a feasible fractional partition M with summation vector $\mathbf{X}^M = \mathbf{X}$. This allows us to focus on the polytope $P(\widetilde{\mathbf{X}})$ of all nonincreasing vectors $\mathbf{X} \in \mathbb{R}_+^K$ that are majorized by $\widetilde{\mathbf{X}}$. The next proposition formalizes the connection between the set of fractional partitions $\Pi_{\mathcal{F}}$ and the majorization polytope $P(\widetilde{\mathbf{X}})$.

PROPOSITION 1. *Let* $\mathbf{X}^M$ *be the summation vector of a feasible fractional partition* $M \in \Pi_{\mathcal{F}}$. *Then,* $\mathbf{X}^M \in P(\widetilde{\mathbf{X}})$. *Conversely, for any* $\mathbf{X} \in P(\widetilde{\mathbf{X}})$, *there exists a fractional partition* $M \in \Pi_{\mathcal{F}}$ *with summation vector* $\mathbf{X}^M = \mathbf{X}$.

The above result already allows the characterization of the optimal partition in the case of Schur convex or Schur concave functions in the following two propositions.

PROPOSITION 2 **(Schur convex objective)**. *Consider the optimization problems* (3.1) *and* (3.2) *with* f *being Schur convex. Then the assortative partition is optimal over both the set of fractional partitions* $\Pi_{\mathcal{F}}$ *and the set of integral partitions* $\Pi_{\mathcal{I}}$. *If* f *is strictly Schur convex, then every optimal partition has summation vector* $\widetilde{\mathbf{X}}$.

PROPOSITION 3 **(Schur concave objective)**. *Consider the optimization problems* (3.1) *and* (3.2) *with* f *being Schur concave. Then the uniform partition is optimal over the set of fractional partitions* $\Pi_{\mathcal{F}}$. *Furthermore, there exists a balanced partition that is optimal over the set of integral partitions* $\Pi_{\mathcal{I}}$. *If* f *is strictly Schur concave, then every optimal fractional partition has summation vector* $\underline{\mathbf{X}}$, *and every optimal integral partition is balanced.*

While the optimality of the assortative and uniform partitions under Schur convex and Schur concave objectives, respectively, follows directly from the definition of majorization, the optimality of balanced partitions over $\Pi_{\mathcal{I}}$ is not as straightforward. We use a constructive approach to prove their optimality. All the proofs can be found in Appendix B.

Additionally, a direct corollary of Proposition 3 is that for any feasible integral partition π, there always exists a balanced partition π_{bal} whose summation vector satisfies $\mathbf{X}^{\pi_{\text{bal}}} \prec \mathbf{X}^{\pi}$. We formally state this fact in Corollary B.4 in Appendix B and use it later in Sections 4.2.1.

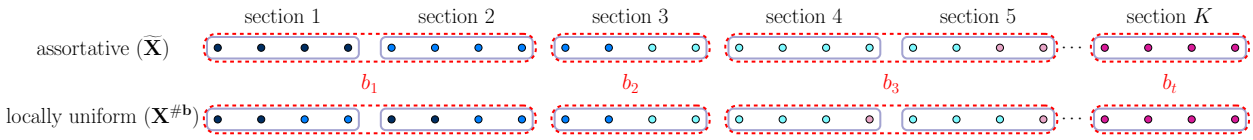
Several of the optimization problems we will see in Section 4 and Appendix C can be successfully reduced to a maximization problem with either a Schur convex or Schur concave f through appropriate transformations. However, what happens if f in (3.1) and (3.2) is neither Schur convex nor Schur concave? We answer this question in Sections 3.2 and 3.3.

3.2. Optimal fractional partitions

We begin by introducing some necessary definitions and formalizing the class of locally uniform partitions. We then establish sufficient conditions under which such partitions are optimal.

Locally uniform partitions. Consider the division of sections $\{1,\dots,K\}$ into t nonempty sets (*blocks*) $\mathbf{b} = (b_1,\dots,b_t)$. We say that $\mathbf{b}$ is a *consecutive division* if there exist integers $0 = k_0 < k_1 < \ldots < k_t = K$ such that $b_i = \{k_{i-1}+1,\dots,k_i\}$ for $i = 1,\dots,t$. (Thus, it is convenient to represent a consecutive division b by its *incidence set* $I(\mathbf{b}) = \{k_0, k_1, \dots, k_t\}$.) We denote the set of all consecutive divisions of $\{1,\dots,K\}$ by $\mathcal{B}$.

Figure 2 Construction of a locally uniform partition given blocks $\mathbf{b}$



Notes: This figure illustrates how a locally uniform partition is constructed from a given block structure $\mathbf{b}$. The top row shows the assortative partition, with consecutive sections that belong to the same block highlighted in red. The bottom row shows the resulting locally uniform partition: within each block, all sections are assigned the same value by replacing their summation-vector entries with the block's average.

For a fixed $\mathbf{X} \in \mathbb{R}_+^K$ and any $\mathbf{b} \in \mathcal{B}$, we define the *interval mean* of $\mathbf{b}$ as $\mathbf{X}^{\#\mathbf{b}} : \mathcal{B} \to \mathbb{R}_+^K$,

$$\mathbf{X}^{\#\mathbf{b}} = \left(\left(\frac{1}{|b_1|} \sum_{j \in b_1} X_j \right) \mathbf{1}_{|b_1|}, \ldots, \left(\frac{1}{|b_t|} \sum_{j \in b_t} X_j \right) \mathbf{1}_{|b_t|} \right),$$

where $\mathbf{1}_k$ denotes the all ones vector of dimension k. In other words, suppose that $i \in b_k$ for some $k \in [t]$. Then, $\mathbf{X}^{\#\mathbf{b}}$ is obtained by replacing each coordinate X_i of $\mathbf{X}$ with the arithmetic mean of all coordinates X_j having indices j in the same block b_k as i.

Building on the notion of interval means, we next define the class of *locally uniform* partitions.

Definition 1 **(Locally uniform partitions)**. *A partition $M \in \Pi_{\mathcal{F}}$ is called* locally uniform *if there exists a $\mathbf{b} \in \mathcal{B}$ such that $\mathbf{X}^M = \widetilde{\mathbf{X}}^{\#\mathbf{b}}$.*

Intuitively, locally uniform partitions mix students of adjacent attribute values. More precisely, for a $\mathbf{b} \in \mathcal{B}$ with $I(\mathbf{b}) = \{k_0, k_1, \ldots, k_t\}$, the corresponding locally uniform partition mixes students with attributes $nk_{i-1} + 1 \leq i \leq nk_i$ and then distributes them evenly to sections $k_{i-1} + 1$ to k_i. (Figure 2 illustrates the construction of a locally uniform partition for a given $\mathbf{b} \in B$.) Finally, note that the uniform and t-upper-uniform partitions are special cases of locally uniform partitions.

Directional quasi-convexity. Our first main result (Theorem 1) relies on *edge-quasi-convex* functions. The formal definition appears below; additional discussion can be found in Appendix A.2.

Definition 2 **(Edge-quasi-convexity, Hwang and Rothblum (1996))**. *Let C be a convex subset of $\mathbb{R}^K$ and $\mathbf{d} \in \mathbb{R}^K \setminus \{\mathbf{0}\}$. A function $f : C \to \mathbb{R}$ is said to be* $\mathbf{d}$-quasi-convex*, or* quasi-convex along $\mathbf{d}$*, if the maximum of f over every line-segment in C with direction $\mathbf{d}$ is attained at one of the endpoints of that line-segment. Furthermore, f is* edge-quasi-convex *on C if it is $\mathbf{d}$-quasi-convex for any nonzero $\mathbf{d} \in \mathbb{R}^K$ which is parallel to one of C's edges.*

Intuitively, f is *edge-quasi-convex* on a convex set $C \subseteq \mathbb{R}^K$ if and only if the maximum of f over any line-segment in C which is parallel to one of C's edges is attained at one of its two endpoints.

Characterizing optimal fractional partitions. We are now ready to present our first main result in Theorem 1. Theorem 1 provides sufficient conditions for the optimality of locally uniform partitions (in addition to the optimality of the assortative and uniform partitions under Schur convex and Schur concave objectives established in Section 3.1). We show that locally uniform partitions are optimal whenever the objective is edge-quasi-convex on $P(\tilde{\mathbf{X}})$.

THEOREM 1 **(Characterization of optimal fractional partitions)**. *Consider the optimization problem in* (3.1) *over* $\Pi_{\mathcal{F}}$.

(i) If f is ***edge-quasi-convex*** *on $P(\tilde{\mathbf{X}})$, then there exists an optimal* ***locally uniform*** *partition. If f is strictly edge-quasi-convex on $P(\tilde{\mathbf{X}})$, then every optimal fractional partition must be locally uniform.*

(ii) If f is ***Schur convex****, then the* ***assortative*** *partition $\widetilde{M}$ is optimal.*

(iii) If f is ***Schur concave****, then the* ***uniform*** *partition $\underline{M}$ is optimal.*

We established parts (ii) and (iii) in Propositions 2 and 3, respectively. The result in part (i) follows from Proposition 1 and two additional key technical observations. First, Dahl (2001) implies that the vertices of a majorization polytope $P(\widetilde{\mathbf{X}})$ are the interval means of $\widetilde{\mathbf{X}}$ (Proposition A.13 in Appendix A.3). Second, by Proposition A.11 (Hwang and Rothblum 1996) in Appendix A.2, a maximum of an edge-quasi-convex function over the polytope $P(\widetilde{\mathbf{X}})$ is attained at a vertex of $P(\widetilde{\mathbf{X}})$. If the objective function is strictly edge-quasi-convex on $P(\tilde{\mathbf{X}})$, then all of its maximizers are attained at an extreme point of $P(\widetilde{\mathbf{X}})$. Thus, the proof of Theorem 1 (i) follows from combining Propositions 1, A.13, and A.11.

The result in Proposition A.11—thus Theorem 1 as well—holds for two special cases of directional quasi-convexity: quadratic and quasi-convex objectives in Corollary 1 and 2, respectively.

COROLLARY 1. *Consider the optimization problem in* (3.1). *If*

$$f(\mathbf{X}) = \mathbf{X}^T A\mathbf{X} + \mathbf{c}^T\mathbf{X}$$

for a positive semidefinite matrix $A \in \mathbb{R}^{K\times K}$, then there exists an optimal locally uniform partition.

COROLLARY 2. *Consider the optimization problem in* (3.1). *If f is quasi-convex, then there exists an optimal locally uniform partition.*

3.3. Optimal integral partitions

We now turn to integral partitions $\Pi_{\mathcal{I}}$, where each student must be assigned to exactly one section.

Feasible integral partitions via the majorization polytope $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$. In the fractional setting, Proposition 1 establishes that every point in the majorization polytope $P(\widetilde{\mathbf{X}})$ is attainable, and the optimization problem reduces to selecting an extreme point of this polytope (Proposition A.11).

For integral partitions, this connection breaks: the summation vectors generated by $\Pi_{\mathcal{I}}$ form only a strict subset of $P(\widetilde{\mathbf{X}})$, and some extreme points of $P(\widetilde{\mathbf{X}})$ are not implementable by any integral partition. This motivates our key geometric contribution: we introduce the polytope

$$P_{\mathcal{I}}(\widetilde{\mathbf{X}}) = \operatorname{conv}\left\{ \mathbf{X}^{\pi} \in \mathbb{R}_{+}^{K} : \pi \in \Pi_{\mathcal{I}},\ \mathbf{X}^{\pi} \prec \widetilde{\mathbf{X}} \right\},$$

defined as the convex hull of summation vectors of integral partitions, which serves as the geometric foundation for our analysis in the integral setting.

Locally balanced partitions. In analogy to locally uniform partitions in the fractional setting, we introduce the class of *locally balanced* partitions as the natural candidates for the extreme points of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$. A locally balanced partition exhausts the majorization constraints at specific cut points and, within each resulting block, achieves the most balanced summation vector attainable, since every alternative reassignment of that block's students majorizes it. Intuitively, such partitions behave like the assortative partition at block boundaries, while within each block they are already "as balanced as possible" given integrality constraints. Thus, locally balanced partitions serve as the integral analogue of locally uniform partitions.

Definition 3 **(Locally balanced partitions)**. *An integral partition $\pi \in \Pi_{\mathcal{I}}$ with summation vector $\mathbf{X}^{\pi}$ is* locally balanced *if there exists a consecutive division $\mathbf{b} \in \mathcal{B}$ with incidence set $I(\mathbf{b}) = \{k_0, k_1, \ldots, k_t\}$ such that the following hold:*

(i) The majorization constraint binds at each $k_i \in I(\mathbf{b})$, i.e., $\sum_{j\in[k_i]} X^{\pi}_{[j]} = \sum_{j\in[k_i]} \widetilde{X}_{[j]}$.

(ii) For each block $b_i = \{k_{i-1}+1, \ldots, k_i\}$ with $|b_i| > 1$, and every feasible integral partition $\pi' \in \Pi_{\mathcal{I}}$ such that $X^{\pi'}_{[j]} = X^{\pi}_{[j]}$ for all $j \in [K] \setminus b_i$, $\mathbf{X}^{\pi}_{b_i} \prec \mathbf{X}^{\pi'}_{b_i}$, where $\mathbf{X}^{\pi}_{b_i} = (X^{\pi}_{[k_{i-1}+1]}, \ldots, X^{\pi}_{[k_i]})$ is the block b_i-constrained *summation vector, i.e., the restriction of $\mathbf{X}^{\pi}$ to the coordinates indexed by block b_i.*

Within each block of a locally balanced partition, no further integral redistribution can produce a more balanced summation vector. Proposition 4 formalizes this by establishing that every locally balanced partition corresponds to a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$, similarly to the role locally uniform partitions play in the fractional polytope $P(\widetilde{\mathbf{X}})$.

Proposition 4 **(All locally balanced partitions are vertices of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$)**. *If $\pi \in \Pi_{\mathcal{I}}$ is locally balanced, then $\mathbf{X}^{\pi}$ is a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$.*

Our proof argument shows that local balance is a sufficient condition that prevents any feasible "micro-reallocation" within a block of sections: by definition, there is no other integral partition that keeps all section sums outside a block fixed while strictly improving the distribution inside that block. In the proof, this rules out the possibility that the summation vector could be written as

a convex combination of two other feasible points, since any such decomposition would necessarily identify another integral partition that improves the block, contradicting local balance.

The converse, however, does not hold in general: not every vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$ corresponds to a locally balanced partition when $K \geq 3$. Intuitively, the summation vector of such a partition lies in the interior of the fractional polytope $P(\widetilde{\mathbf{X}})$, but the integrality constraints render it a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$. The following example illustrates this.

EXAMPLE 1. Let $K = 3$, $n = 2$, and $\mathbf{x} = (84, 51, 42, 20, 10, 7)$. Consider the partition π^* given by $\pi_1^* = \{i_1, i_4\}, \pi_2^* = \{i_2, i_6\}, \pi_3^* = \{i_3, i_5\}$. Its summation vector is $\mathbf{X}^{\pi^*} = (104, 58, 52)$. Observe that $\mathbf{X}^{\pi^*}$ is a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$; for example, it is the unique maximizer of the linear objective $2X_{[1]} - 5X_{[2]} + 5X_{[3]}$ over all feasible $\mathbf{X} \in P_{\mathcal{I}}(\widetilde{\mathbf{X}})$. However, π^* is not locally balanced. Indeed, the majorization constraint does not bind at $k = 1$ or $k = 2$, so the only possible consecutive division is the single block. But the feasible partition π' with $\pi'_1 = \{i_1, i_6\}$, $\pi'_2 = \{i_3, i_4\}$, $\pi'_3 = \{i_2, i_5\}$, satisfies $\mathbf{X}^{\pi'} = (91, 62, 61) \prec (104, 58, 52) = \mathbf{X}^{\pi^*}$, so condition (ii) fails. This shows that the converse of Proposition 4 does not hold for $K \geq 3$: while every locally balanced partition corresponds to a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$, there also exist vertices that do not belong to the class of locally balanced partitions.

Locally extremal partitions. Since Example 1 shows that there may exist vertices of $P_{\mathcal{I}}(\widetilde{X})$ that are not locally balanced, we introduce a new class of partitions that broadens the notion of local balance to include all vertices of $P_{\mathcal{I}}(\widetilde{X})$.

For sections $1 \leq r < s \leq K$, let

$$\mathcal{T}_{rs}(\pi) = \left\{\pi' \in \Pi_{\mathcal{I}} : X^{\pi'}_{[j]} = X^{\pi}_{[j]} \text{ for all } j \notin \{r, s\}\right\}$$

be the set of feasible partitions agreeing with π on all section sums except for r and s, and let $\Delta_{rs}(\pi) = X^{\pi}_{[r]} - X^{\pi}_{[s]}$ denote the gap between the two sums of sections r and s.

DEFINITION 4 **(Locally extremal partitions)**. *An integral partition* $\pi \in \Pi_{\mathcal{I}}$ *is* locally extremal *if for every* $1 \leq r < s \leq K$,

$$\Delta_{rs}(\pi) \in \left\{\min_{\pi' \in \mathcal{T}_{rs}(\pi)} \Delta_{rs}(\pi'), \max_{\pi' \in \mathcal{T}_{rs}(\pi)} \Delta_{rs}(\pi')\right\}.$$

Conditional on all other section sums, Definition 4 requires that every pair of sections attains one of the two extremes of its feasible range: the pair is either *balanced* (i.e., its gap is the smallest feasible) or *polarized* (i.e., its gap is the largest feasible). Every locally balanced partition is locally extremal. The assortative partition $\widetilde{\pi}$ is also locally extremal, with every pair polarized. Mixed combinations capture the additional vertices that arise from integrality; in Example 1, π^* has one polarized pair (sections 1 and 2), and two balanced pairs.

Characterizing optimal integral partitions. The following proposition shows that the extreme points of polytope $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$ correspond to locally extremal partitions.

PROPOSITION 5 **(Characterization of extreme points of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$).** *Every vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$ corresponds to a locally extremal partition.*

The converse of Proposition 5 does not hold in general: not every locally extremal partition corresponds to a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$. The proposition provides a necessary local structure that every vertex must satisfy, which suffices for proving the equivalent of Theorem 1 for integral partitions. An exception is $K=2$, for which the converses of Propositions 4 and 5 hold. In particular, when $K=2$, there are at most two vertices, corresponding to the assortative and a balanced partition.

THEOREM 2 **(Characterization of optimal integral partitions).** *Consider the optimization problem in (3.2) over $\Pi_{\mathcal{I}}$.*

*(i) If f is **edge-quasi-convex** on $P_I(\tilde{\mathbf{X}})$, then there exists an optimal **locally extremal** partition. If f is strictly edge-quasi-convex, then every optimal integral partition must be locally extremal.*

*(ii) If f is **Schur convex**, then the **assortative** partition $\widetilde{\pi}$ is optimal.*

*(iii) If f is **Schur concave**, then there exists an optimal **balanced** partition.*

Table 1 summarizes the main theoretical results of this section.

Table 1 Optimal Partitions in $\Pi_{\mathcal{F}}$ and $\Pi_{\mathcal{I}}$.

Objective Function	**Optimal Fractional Partitions** (Theorem 1)	**Optimal Integral Partitions** (Theorem 2)
edge-quasi-convex	locally uniform	locally extremal
Schur convex	assortative	assortative
Schur concave	uniform	balanced

4. An application to education: Characterizing optimal cohort partitions in the presence of peer effects

A large empirical literature in economics documents the importance of peer effects in shaping student outcomes (Sacerdote 2011, Epple and Romano 2011). While several papers have theorized the idea that individuals could be sorted into peer groups to maximize efficiency (Bhattacharya 2009, Carrell et al. 2013, Graham et al. 2014), the existing literature has been vastly empirical. This gap can be consequential: as the USAFA experiment (Carrell et al. 2013) illustrates, assignments without structural optimality guarantees may lead to partition policies that backfire. Our empirical exercise in Section 5 reinforces that partition design matters: comparing hundreds of quasi-random classroom assignments, we find that different partition policies can shift who gains and who loses.

In this section, we apply the majorization-theoretic framework of Section 3 to characterize optimal cohort partitions for four common objectives: performance, inequality, welfare, and diversity.

We begin with the *linear-in-means (LIM) model*, which stands as the cornerstone of the empirical education research on peer effects (Manski 1993), and later study richer non-linear peer effects models (Garlick 2018, Boucher et al. 2024).

4.1. Preliminaries

Peer effects. In a section $j \in [K]$, we consider that students' outcomes (e.g., their final GPA) are determined via a standard effort-based game played among all students in that class. Note that any peer effects from students outside this particular section are deemed negligible. Thus, for simplicity, when referring to a section j, we may omit the superscript j for section j students.

In a section j, each individual student i chooses their *effort* y_i as a function of their own characteristics q_i plus a function of the effort of their classmates, $h(\mathbf{y}_{-i})$. Here, $q_i = \gamma x_i + \epsilon_i$ captures the observable (x_i) and unobservable (ϵ_i) attributes of individual i. We also assume that the error term ϵ_i is such that $\mathbb{E}[\epsilon_i \mid \mathbf{x}] = 0$ and $\mathrm{Var}[\epsilon_i \mid \mathbf{x}] = \sigma^2$ across students, and $\mathbb{E}[\epsilon_i \epsilon_k \mid \mathbf{x}] = 0$ for all $i \neq k$.

Each student i chooses effort y_i that maximizes her utility $U_i(\mathbf{y})$ at a Nash equilibrium. An effort profile $\mathbf{y}$ constitutes a *Nash equilibrium*, if for every student i in the section,

$$U_i(y_i, \mathbf{y}_{-i}) \geq U_i(y_i', \mathbf{y}_{-i}), \forall y_i' \neq y_i.$$

Let also $\mathbf{U} = (U_1, \ldots, U_n)$. Then, a *peer effects model* is specified by $(\mathbf{U}, h, n, K, \mathbf{x})$. In Section 4.2, we review the classic peer effects model used in the education literature, the *linear-in-means* model (Manski 1993, Boucher et al. 2024). Two alternative non-linear models are the *quadratic-in-means* (QIM) model (Garlick 2018) in Section 4.3.2 and the *social norm* model (Boucher et al. 2024) in Section 4.3.3. Additionally, *Tullock's competition* model has been recently adopted to model peer effects in the classroom (Rosenzweig and Xu 2023); we study a related model in Appendix C.2.

Metrics. We are interested in various metrics, defined globally (at the school level, π) and potentially locally (at a section level, π_j). Below, we give the definitions for integral partitions; however, it is straightforward to extend them to fractional partitions M.

- The *section effort* is defined as the (expected) sum of students' achievements, i.e.,

$$e(\pi_j) = \mathbb{E}\left[\sum_{i \in \pi_j} y_i\right]. \tag{4.1}$$

  At the school level, given known weights $\mathbf{w} \in \mathbb{R}_+^K$, we define the *total weighted effort* as $E(\pi) = \sum_{j \in [K]} w_j e(\pi_{[j]})$.
- At the school level, we define the *Gini index of effort* as follows:

$$G_E(\pi) = \frac{1}{2NE(\pi)} \sum_{i \in [N]} \sum_{j \in [N]} |\mathbb{E}[y_i] - \mathbb{E}[y_j]|. \tag{4.2}$$

- The *section utility* or *surplus* is defined as the (expected) sum of students' utilities, i.e.,

$$s(\pi_j) = \mathbb{E}\left[\sum_{i\in\pi_j} U_i(\mathbf{y})\right]. \tag{4.3}$$

  At the school level, the *total utility* of a partition π is defined as $S(\pi) = \sum_{j\in[K]} s(\pi_j)$.

- The *section variance* equals

$$v(\pi_j) = \sum_{i\in\pi_j} (x_i - \overline{x}_j)^2. \tag{4.4}$$

  At the school level, the *total variance* of a partition π is defined as $V(\pi) = \sum_{j\in[K]} v(\pi_j)$.

The above four metrics—total effort, Gini index, total utility, total variance—serve as measures of *performance, inequality, welfare,* and *diversity*, respectively. More specifically, the total effort E can be interpreted as a measure of *performance*. In fact, the empirical literature on peer effects often equates student outcomes, such as grades or test scores, with individual effort (see, e.g., Boucher et al. (2024)). The different weights naturally reflect the planner's attitude towards sections of varying average performance. For example, if $w_1 > w_2 > \ldots > w_K$, then the planner puts the highest value in the effort of the top-performing section and a progressively lower value on the effort of the lower-performing sections. In contrast, a planner placing the highest weight on, e.g., $w_{\lceil K/2\rceil}$ is more concerned about the performance of middle-performing sections.

We adapt the Gini index, a popular measure of *inequality*, to measure the inequality of effort y_i among individuals with varying attributes x_i within a given partition. The Gini index equals the area between the school-level Lorenz curve and the egalitarian line, where all individuals exert equal effort. Total utility is the standard measure of aggregate *welfare*. Similarly, section variance is perhaps the most common measure of *diversity* for non-categorical attributes, reflecting the representativeness of the section in terms of student ability levels. Higher variance suggests greater diversity within each class.

Finally, we note that our analysis is not confined to the specific metrics and objectives presented above or the peer effects models we study in this paper, but can extend to any given metric that can be reduced to a class of part-additive problems (for technical details, see Section 3; for additional examples, see Appendix C).

4.2. Linear-in-means peer effect model

In the empirical literature, the *linear-in-means model* (LIM) takes the following form:

$$y_i = \gamma x_i + \theta \overline{y}_{-i} + \epsilon_i, \tag{4.5}$$

where $\gamma > 0$. Below we present a microfoundational framework of (4.5) based on spillover effects.[3]

[3] The microfoundation in (4.6) is referred as the *LIM spillover model*. An alternative microfoundational formulation is the *LIM conformist* model. In the conformist version, each student i chooses y_i to maximize her utility of the form $U_i^C(y_i, \mathbf{y}_{-i}) = q_i y_i - \frac{1}{2}y_i^2 - \frac{\theta_2}{2}(y_i - \overline{y}_{-i})^2$, where $0 \le \frac{\theta_2}{2} \le 1$ is the *taste for conformity*. The Nash equilibrium and theoretical results are analogous to the LIM spillover model that we present here.

Each student i chooses effort that maximizes the following utility function:

$$U_i^S(y_i, \mathbf{y}_{-i}) = q_i y_i - \frac{1}{2} y_i^2 + \theta_1 y_i \overline{y}_{-i}, \tag{4.6}$$

where $0 < \theta_1 < 1$ is the intensity of the spillover effect and $\overline{y}_{-i}$ denotes the mean of the efforts of student i's classmates. The first term $q_i y_i - \frac{1}{2} y_i^2$ is student i's utility of exerting effort y_i, independently of peer effects. The second term $\theta_1 y_i \overline{y}_{-i}$ is the *spillover effects* of the social norm on student i's own utility.

Observe that the first-order condition corresponds to $y_i = q_i + \theta_1 \overline{y}_{-i}$, or the LIM model in (4.5):

$$y_i = \gamma x_i + \theta_1 \overline{y}_{-i} + \epsilon_i. \tag{4.7}$$

Thanks to the linearity of the system in (4.7), the unique Nash equilibrium $\mathbf{y}$ can be written in a closed form, i.e.,

$$y_i = \frac{(n-1)\, q_i}{n - 1 + \theta_1} + \frac{\theta_1 \left(\gamma X_j^{\pi} + \sum_{k \in \pi_j} \epsilon_k \right)}{(1 - \theta_1)(n - 1 + \theta_1)}, \quad i \in \pi_j. \tag{4.8}$$

Equation (4.8) shows that the equilibrium effort y_i factors in the student's own observed and unobserved attributes (x_i, ϵ_i), the influence of her peers' average attributes $(\overline{x}_{-i}, \overline{\epsilon}_{-i})$, and the parameter θ_1 which adjusts the weight of peer-group pressure.

For any partition π and an arbitrary section j, (4.8) allows us to compute the resulting section effort $e(\pi_j)$ and section utility $s(\pi_j)$ at equilibrium. As shown in Lemma B.5 in Appendix B.2, both metrics are functions of the sum of section j's characteristics, $X_j^{\pi} = \sum_{i \in \pi_j} x_i$. This suffices to ensure that, at a school level, our objectives, total effort $E(\pi) = \sum_{j \in [K]} e(\pi_j)$ and total utility $S(\pi) = \sum_{j \in [K]} s(\pi_j)$, can be rewritten as a part-additive optimization problem.

4.2.1. Optimization and policy implications under the LIM model We now proceed to finding the optimal partition under the LIM model for each of our four objectives: performance, inequality, welfare, and diversity. We discuss our findings and potential policy implications.

Performance. In Proposition 6 below, we focus on optimizing performance, as measured by the total weighted effort. Recall that in a vector of weights $\mathbf{w} \in \mathbb{R}_+^K$, w_j signifies the priority the planner assigns to the effort $e(\pi_{[j]})$ exerted by the j-th highest-performing class. For example, the USAFA experiment aimed at improving the performance of the least-performing students. In our model, this approach could correspond to assigning a significantly higher weight w_K. We say that $\mathbf{w} \in \mathbb{R}_+^K$ is *single-peaked at* t^* if there exists $t^* \in [K]$ such that $w_1 \leq \ldots \leq w_{t^*}$ and $w_{t^*} \geq \ldots \geq w_K$.

Proposition 6 **(LIM, Performance)**. *Fix* $\mathbf{w} \in \mathbb{R}_+^K$. *Under the LIM model, there exists a locally uniform partition that attains the optimal total weighted effort over* $\Pi_{\mathcal{F}}$. *Additionally, if*

$\mathbf{w} \in \mathbb{R}_+^K$ *is single-peaked at* t^**, then the total weighted effort is maximized for a* $\underline{t}$*-upper-uniform partition, where* $\underline{t} \geq t^*$*. If* $t^* = 1$*, then the assortative partition is optimal. If* $t^* = K$*, then the uniform partition is optimal.*

Proposition 6 shows that the structure of the optimal partition heavily depends on the planner's prioritization of different sections. One high-level way to interpret Proposition 6 is as follows: As the planner's highest priority w_{t^*} shifts from the top-performing section $j = 1$ towards the middle-performing sections $1 < j < K$ and eventually to the least-performing section $j = K$, the optimal partition becomes progressively *less assortative* and each section becomes *more diverse*.

For an arbitrary $\mathbf{w} \in \mathbb{R}_+^K$, the optimality of locally uniform partitions follows directly from Theorem 1.[4] Similarly, for $t^* = 1$ (respectively, $t^* = K$), the total weighted effort becomes Schur convex (respectively, Schur concave). Accordingly, Proposition 2 (respectively, Proposition 3) applies. Note that analogous guarantees hold over $\Pi_{\mathcal{I}}$ thanks to Theorem 2.

For the case of a single-peaked $\mathbf{w}$ at $1 < t^* < K$, the objective of total weighted effort is neither Schur convex nor Schur concave. Theorem 1 still applies here; however, given the special structure of $\mathbf{w}$, we can obtain a stronger characterization, showing that the optimal solution is the $\underline{t}$-upper-uniform partition for $\underline{t} \geq t^*$.

To achieve this characterization, we employ a technique mirroring Myerson's "ironing" process (Myerson 1981), in order to *Schur-convexify* objectives of the form $\mathbf{w}^T\mathbf{X}$. Ironing is a common technique in mechanism design problems (see, e.g., Hartline and Roughgarden (2008), Ashlagi et al. (2025)) used to convexify linear functionals of one variable. We develop a counterpart of ironing applicable to vectors in $\mathbb{R}_+^K$, by ironing the L-curve of $\mathbf{w}$. The main idea is to reduce the original problem with single-peaked weights $\mathbf{w}$ to a Schur convex problem with (weakly) decreasing weights by replacing the weights $w_1, \ldots, w_{\underline{t}}$ with their local average $\frac{1}{\underline{t}} \sum_{k \leq \underline{t}} w_k$ (for an appropriately chosen $\underline{t} \geq t^*$). The proof, together with a graphical illustration of the ironing process in Figure B.6, can be found in Appendix B.2.

Note that it is also possible to construct the optimal partition for an arbitrary $\mathbf{w} \in \mathbb{R}_+^K$. One off-the-shelf approach is to apply Dahl's simple and efficient dynamic programming algorithm (Dahl 2001) (see Appendix A.4). Alternatively, one can apply a modified version of our ironing process (see the proof of Claim 2 in Appendix B.2) to construct the optimal solution for any $\mathbf{w} \in \mathbb{R}_+^K$.

Last, it is important to acknowledge that when the planner's goal is to maximize the total unweighted effort, the problem becomes degenerate, as any partition is optimal. This highlights a well-known criticism of the LIM model despite its widespread adoption in the literature: its inherent

[4] In fact, the result can be strengthened further, showing that given any locally uniform partition $M \in \Pi_{\mathcal{F}}$, there exists a weight vector $\mathbf{w} \in \mathbb{R}_+^K$ that induces M as its optimal solution.

Table 2 Summary of Optimal Partitions Under the LIM Model

Objective	Metric	Optimal partition (under LIM)	Reference
Performance	total weighted effort	locally uniform (general $\mathbf{w}$) $\underline{t}$-upper-uniform (single-peaked $\mathbf{w}$) assortative (decreasing $\mathbf{w}$) uniform (increasing $\mathbf{w}$)	Proposition 6
	total effort	any (constant $\mathbf{w}$)	Corollary 3, Proposition 8
Inequality	Gini index of effort	balanced (strong θ_1) dispersion-maximal (weak θ_1) snake (weak θ_1, linear $\mathbf{x}$, even n)	Proposition 7 (i) Lemma B.6 Proposition B.14
Welfare	total utility	assortative	Proposition B.15
Diversity	total variance	uniform (balanced over $\Pi_{\mathcal{I}}$)	Proposition B.16

linearity constrains the net mean effort effect from the reassignment of peers to different sections to be zero (Sacerdote 2011). Interestingly, as we see in Section 4.3.1, the LIM is not the only model with this property and even non-linear models can exhibit the same behavior. Proposition 8 characterizes the class of sorting-neutral peer effects models.

COROLLARY 3. *Let $w_j = 1$ for all $j \in [K]$. Then any feasible partition attains the optimal total effort under the LIM model.*

Inequality. For any partition $\pi \in \Pi_{\mathcal{I}}$, denote (with a slight abuse of notation) its corresponding equilibrium vector of (expected) effort by $\mathbb{E}[\mathbf{y}^\pi] = (\mathbb{E}[y_1^\pi], \ldots, \mathbb{E}[y_N^\pi])$. Similarly, define $\mathbb{E}[\widetilde{\mathbf{y}}]$ as the vector of equilibrium efforts under the assortative partition. The following lemma establishes that $\mathbb{E}[\widetilde{\mathbf{y}}]$ majorizes $\mathbb{E}[\mathbf{y}^\pi]$ of any integral partition π.

LEMMA 1. *Under the LIM model and for any partition $\pi \in \Pi_{\mathcal{I}}$, $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\widetilde{\mathbf{y}}]$. Furthermore, there is no partition $\pi \in \Pi_{\mathcal{I}}$ with $\mathbf{X}^\pi \neq \underline{\mathbf{X}}$ such that $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\underline{\mathbf{y}}]$.*

As a consequence of Lemma 1, it is perhaps unsurprising that the assortative partition exhibits the highest inequality in effort among all integral partitions (Proposition 7 (ii)). On the other hand, Lemma 1 also establishes that no integral partition's effort vector is strictly majorized by that of the uniform partition. Note that this does not imply that the uniform partition always achieves the minimum inequality, as effort vectors incomparable to the one of uniform in a majorization sense may still attain a lower Gini index.

Proposition 7 (i) below shows that, when peer effects are strong, every optimal integral partition is balanced. When peer effects are weak, minimizing inequality is instead equivalent to maximizing the within-section dispersion $W(\pi)$, that is, the total attribute spread between classmates weighted by their distance in rank (Lemma B.6). In this case, the planner minimizes inequality by grouping

dissimilar students, so that weaker students share sections with high-attribute peers. In particular, when $n = 2$, this is achieved uniquely by the snake partition (Proposition B.14 (i)), which pairs the k-th strongest with the k-th weakest student (Definition B.10) and is standard practice in sports and fantasy-sports drafts (Lee and Liu 2022). It is also optimal when n is even and $\mathbf{x}$ is a linear function of the student's rank i (Proposition B.14 (ii)). Interestingly, the uniform partition can be suboptimal for inequality (see Example B.2).

We say that a partition $M \in \Pi_{\mathcal{F}}$ with corresponding equilibrium vector $\mathbf{y}^M$ *achieves perfect equality in effort* if $G_E(\mathbf{y}^M) = 0$. Intuitively, perfect equality is reached when all students exert equal effort at equilibrium. Proposition 7 (iii) shows that the uniform partition, perhaps contrary to the common expectation, does not achieve perfect equality in effort.

PROPOSITION 7 **(LIM, Inequality)**. *Under the LIM model, the following hold:*

(i) There exists threshold $\underline{\theta} \in (0,1)$ such that if peer effects are strong ($\theta_1 > \underline{\theta}$), then every partition minimizing the Gini index of effort over $\Pi_{\mathcal{I}}$ is balanced.

(ii) The assortative partition achieves the maximum Gini index of effort over $\Pi_{\mathcal{I}}$. In particular, if $x_1 > \ldots > x_N$, the assortative partition is the only partition with this property.

(iii) There exist cohorts where no (fractional) partition in $\Pi_{\mathcal{F}}$ can achieve perfect equality in effort.

Part (iii) might seem counterintuitive at first glance. The concept of perfect equality in effort essentially requires that all students exert the same level of effort in equilibrium. Part (iii) shows that no partition can achieve that in the general case. Using a simple counterexample with two types of students, we show that no fractional partition is guaranteed to achieve perfect equality in effort. This highlights that the inherent distributional differences in the student population set a fundamental limit on the planner's ability to enforce perfect equality at an individual level.

Welfare. We measure welfare via total utility. By Lemma B.5, the problem of maximizing total utility reduces to maximizing a Schur convex objective. Proposition 2 then guarantees the optimality of the assortative partition. Proposition B.15 formalizes this result.

Diversity. We measure diversity via total variance. Maximizing total variance directly reduces to maximizing a Schur concave function. Consequently, Proposition 3 implies that the uniform partition is optimal among all feasible fractional partitions while there exists an optimal balanced partition among the set of integral partitions. Proposition B.16 formalizes this result.

Discussion and policy implications. For comparison, the results of this section are summarized in Table 2. Several observations are in order.

One high-level take-away from the study of the LIM model is that the four objectives are not always aligned. On the one hand, optimal welfare is achieved under the assortative partition. On

the other hand, the assortative partition would maximize inequality and minimize diversity, which instead favor various forms of mixing (such as balanced, uniform, snake partitions). The planner's priority weights play a central role in navigating this trade-off. When priorities favor the least-performing students, a balanced partition may simultaneously align performance, diversity, and inequality objectives. More generally, as priorities shift toward higher-performing sections, the performance-optimal partition becomes progressively more segregated; conversely, more redistributive priorities lead to greater mixing. In this regard, the locally uniform partition emerges as a flexible instrument, interpolating between these extremes.

Our findings may offer valuable perspectives on the Greek educational policy debate in Section 1, further complemented by our empirical results using data from Greek public high schools in Section 5. One argument raised against the proposed policy was that it may introduce inequality, contrasting with the perceived fairness of the current quasi-random assignment system (uniform policy). However, as we establish in Proposition 7 (iii), inequality remains a concern even under the uniform policy: no fractional partition, including the uniform itself, achieves perfect equality in effort in general. Heterogeneity in the student population thus sets a fundamental limit on the planner's ability to equalize individual student effort. In fact, the uniform policy may not even minimize the Gini index of effort (Example B.2). The strength of peer effects determines which partition does: when peer effects are strong, the inequality-minimizing integral partition is balanced (Proposition 7 (i)), whereas when they are weak, it is one that maximizes within-section dispersion (Lemma B.6). In either case, the planner may reduce inequality through deliberate mixing rather than through quasi-random assignment.

Let us also revisit the USAFA experiment (Carrell et al. 2013) from Section 1 in light of our results for the LIM model. Suppose first that the planner maximizes total weighted effort with increasing section weights, thus prioritizing the bottom 1/3 of sections. Then, by Proposition 6, the uniform partition is optimal over $\Pi_{\mathcal{F}}$ (resp., a balanced partition over $\Pi_{\mathcal{I}}$). This would imply that USAFA's original policy might have been optimal in the first place, explaining the negative treatment effect observed in practice. Alternatively, suppose that the planner assigns individual weights β_i to students, placing a higher weight on lower-achieving students, and maximizes the weighted total effort $\sum_{i\in[N]} \beta_i \mathbb{E}[y_i]$. This objective is no longer sorting-neutral, but when the weights decline linearly in ability the uniform partition (resp., a balanced partition) remains optimal over $\Pi_{\mathcal{F}}$ (resp., $\Pi_{\mathcal{I}}$), while when they increase linearly in ability the assortative partition is optimal. Appendix B.2.5 examines several formulations.

Nevertheless, Corollary 3 highlights a well-known limitation of the LIM model: sorting neutrality. Next, we explore the class of sorting-neutral peer effects models and some non-linear alternatives.

4.3. Sorting neutrality and non-linear peer effect models

In this section, we study non-linear peer effects. In Section 4.3.1, we characterize *sorting-neutral* models beyond LIM. In Sections 4.3.2 and 4.3.3, we study performance-optimal partitions for two non-linear models. Our results can be of independent technical interest. For example, in Section 4.3.1, we build upon Bergstrom and Varian (1985) and Aczél (1966) to characterize which peer effects models—potentially non-linear—are oblivious to the structure of the underlying partition.

4.3.1. Sorting-neutral peer effects models In Section 4.2.1, we saw that, under the standard LIM model, any partition achieves the optimal total unweighted effort. In this section, we give a characterization of part-additive peer effects models with this property. Formally, a peer effects model $(\mathbf{U}, h, n, K)$ is called *sorting-neutral* if, for every $\mathbf{x} \in \mathbb{R}^{nK}$, each partition $\pi \in \Pi_{\mathcal{I}}$ induces a Nash equilibrium and the value of the total unweighted effort $E(\pi)$ is independent of π.

PROPOSITION 8. *Let $n \geq 3$. Consider a peer effects model $\mathcal{Q} = (\mathbf{U}, h, n, K)$, with h and each U_i being continuous, and assume that, for every $\mathbf{x} \in \mathbb{R}^{nK}$, each partition $\pi \in \Pi_{\mathcal{I}}$ induces a unique Nash equilibrium $\mathbf{y} \in \mathbb{R}^{nK}$ that satisfies $y_i = f\left(x_i, \sum_{k \in \pi_j} y_k\right)$, $i \in \pi_j$, $j \in [K]$, for some continuous function f. Then, the section effort $e(\pi_j)$ depends on π_j only through $\sum_{k \in \pi_j} x_k$ and $\mathcal{Q}$ is sorting-neutral, if and only if $\mathbf{y}$ satisfies:*

$$y_i = r\Big(c \sum_{k \in \pi_j} x_k + a\Big) + v\Big(c \sum_{k \in \pi_j} x_k + a\Big) x_i, \qquad i \in \pi_j,\ j \in [K], \tag{4.9}$$

for some constants $c, a \in \mathbb{R}$, and functions $v, r : \mathbb{R} \to \mathbb{R}$ such that, when $c \neq 0$,

$$z = n\, r(z) + v(z)\, \frac{z - a}{c}, \qquad z \in \mathbb{R}. \tag{4.10}$$

The proof is based on two observations. First, in a sorting-neutral model, we show that each section j effort must be affine in $\sum_{k \in \pi_j} x_k$, which follows from an application of a generalized Pexider equation (Aczél 1966). Second, this allows us to utilize the characterization by Bergstrom and Varian (1985) (Proposition B.17) to show that y_i must take the form given in (4.9).

The proposition shows that sorting-neutrality in peer effects models is not exclusive to the linear-in-means (LIM) model. It can also apply to any model where the equilibrium effort y_i induced by any partition exhibits a linear term in x_i and a potentially non-linear, part-additive component that depends on $\sum_{k \in \pi_j} x_k$. A key property of sorting-neutral models is therefore that even though the effort at the *individual* level can be *non-linear*, the aggregate effort at the *class* level eventually reduces to a *linear* form. Sorting-neutrality thus reflects a structural degeneracy of such peer effects models: the partition affects the distribution of effort across individuals, but all such distributional variation washes out in the aggregate.

Beyond the LIM model that is sorting-neutral, in the next two subsections we provide additional non-linear effort-based models that are *not* sorting-neutral: the quadratic-in-means model (Garlick 2018) and the social norm model for $\beta = 1/2$ (Boucher et al. 2024).

4.3.2. The Quadratic-In-Means (QIM) Model Inspired by the quadratic production function of non-linear peer effects in Garlick (2018) (see his equation (4)), we consider the following effort profile of student i in section j:

$$y_i = \gamma_0 + \gamma_1 x_i + \gamma_2 \overline{x}_j + \gamma_{11} x_i^2 + \gamma_{22} \overline{x}^2 + \gamma_{12} x_i \overline{x}_j + \epsilon_i, \tag{4.11}$$

where ϵ_i is a mean-zero idiosyncratic noise term. Then, the section effort equals

$$e(\pi_j) = \gamma_0 n + (\gamma_1 + \gamma_2) \sum_{i \in \pi_j} x_i + \gamma_{11} \sum_{i \in \pi_j} x_i^2 + \frac{1}{n} (\gamma_{22} + \gamma_{12}) \left(\sum_{i \in \pi_j} x_i \right)^2. \tag{4.12}$$

Only the last quadratic term influences the value of the total unweighted effort. When $\gamma_{22} + \gamma_{12} > 0$ (resp., when $\gamma_{22} + \gamma_{12} < 0$), Proposition 2 (resp., Proposition 3) applies, giving the following result.

PROPOSITION 9. *In the quadratic-in-means model, the assortative partition maximizes total effort if $\gamma_{22} + \gamma_{12} > 0$, whereas the uniform partition maximizes total effort over $\Pi_{\mathcal{F}}$ and a balanced partition maximizes total effort over $\Pi_{\mathcal{I}}$ if $\gamma_{22} + \gamma_{12} < 0$.*

4.3.3. Peer effects under a non-linear social norm In recent work, Boucher et al. (2024) generalize the LIM model in (4.5) to include more general peer effects functions $h(\mathbf{y}_{-i})$. They prove uniqueness and existence of a Nash equilibrium (see Lemma B.7) but do not provide a closed-form solution of $\mathbf{y}$ or the section effort due to the intractability of the model. We provide an analytical solution for $\mathbf{y}$ for the common value $\beta = 1/2$. Specifically, we replace the mean $\overline{y}_{-i}$ in (4.5) with the *social norm*

$$\widetilde{y}_{-i} = \left(\frac{1}{n-1} \sum_{j \neq i} \sqrt{y_j} \right)^2, \tag{4.13}$$

which corresponds to the general formulation $\tilde{y}_{-i} = \left(\frac{1}{n-1} \sum_{j \neq i} y_j^{\beta} \right)^{1/\beta}$ evaluated at $\beta = 1/2$. We assume that the utility $U_i(y_i, \mathbf{y}_{-i})$ equals

$$U_i(y_i, \mathbf{y}_{-i}) = q_i y_i + \theta_1 y_i \widetilde{y}_{-i} - \frac{1}{2} \left[y_i^2 + \theta_2 (y_i - \widetilde{y}_{-i})^2 \right], \tag{4.14}$$

where $q_i = q(x_i) > 0$ for a nondecreasing function q.

Solving the social norm model with $\beta = 1/2$, we find that the assortative partition maximizes total effort for $n \geq 3$, while for $n = 2$ the model is sorting-neutral.

PROPOSITION 10. *Consider the social norm model with $\beta = 1/2$. For $n \geq 3$, the assortative partition $\widetilde{\pi}$ maximizes total effort over $\Pi_{\mathcal{I}}$. For $n = 2$, the model is sorting-neutral.*

The key technical difficulty is that the first-order condition pins down y_i as a function of $\hat{S}_j = \sum_{i\in\pi_j}\sqrt{y_i}$ rather than of $S_j = \sum_{i\in\pi_j} y_i$ directly, where $a_i = \frac{q_i}{1+\theta_2}$ and $b = \frac{\theta_1+\theta_2}{(1+\theta_2)(n-1)^2}$. Summing over $i \in \pi_j$ yields the following fixed-point equation

$$\hat{S}_j = \frac{\sum_{i\in\pi_j}\sqrt{b\hat{S}_j^2 + (1-b)a_i}}{1+(n-1)b},$$

which involves generally intractable cross terms. By managing to eliminate these, we obtain $e(\pi_j) = \left(\sum_{i\in\pi_j} a_i + b(n-2)\hat{S}_j^2\right)/(1-b)$ (see Lemma B.8). For $n \geq 3$, maximizing total effort is thus equivalent to maximizing $\sum_{j\in[K]}\hat{S}_j^2$, and Lemmas B.9–B.11 give the result.

5. Empirical insights from Greek public schools

This empirical exercise leverages the quasi-random variation in cohort partition sortedness across Greek public high schools to document how the degree of sorting is associated with student performance, motivating and contextualizing our theoretical framework.

Data. We use data from the *myschool* administrative database of the Hellenic Ministry of Education, which covers the universe of middle and high school students in Greece (grades 7 through 12). Each student record includes an individual identifier, school and classroom identifiers, gender, a complete history of section and track enrollment, high school graduation status and year, and test scores in every subject and grade. Our analysis relies on prior performance from grade 9 (before the quasi-random classroom assignment at the beginning of grade 10) and outcomes from grade 10 for all public school students between 2023 and 2025. The analytic dataset includes more than 200,000 students enrolled in more than 900 public high schools.

Quasi-random classroom assignment and cohort partition sortedness. The Greek educational system is highly centralized (OECD 2018). Students are assigned to public schools through a zoning process based on their residential address. Upon enrollment in a high school, each student is placed in a classroom where they attend all courses. In accordance with a strictly enforced law, students are allocated to classrooms in alphabetical order based on their last name[5] and are not permitted to switch classrooms. This alphabetical assignment rule essentially generates quasi-random variation in classroom peer composition (see Table D.3 for supporting evidence).

We exploit this quasi-random variation to construct a measure of cohort partition *sortedness*: the percentage difference between the Lorenz-curve area of the empirical classroom partition and that of a theoretically uniform partition within each school-grade cohort. This measure captures the degree to which a cohort's actual classroom partition deviates from the uniform partition benchmark. Higher values indicate greater sorting by prior performance, with the maximum achieved by the

[5] Government Gazette of the Hellenic Republic 167 A/1566/1985, Ministry of Education Bulletin 100749/Γ2/17-09-07.

assortative partition. Figure D.7 in Appendix D presents its empirical distribution across cohorts for Mathematics and Modern Greek Language. Both are centered around moderate levels of sortedness, with Mathematics exhibiting greater dispersion, indicating that sorting varies substantially across school cohorts.

Empirical strategy. Figure 3 plots the relationship between each cohort's partition sortedness in grade 10 and its average academic growth between grades 9 and 10. The estimated correlations are $\rho = 0.06$ for Mathematics and $\rho = 0.07$ for Modern Greek Language and statistically significant, reflecting a modest unconditional positive association.

Our main specification estimates nonlinear heterogeneous associations, allowing the estimated association between sortedness and performance to vary flexibly both along the sortedness distribution and across students' positions in the prior-achievement distribution:

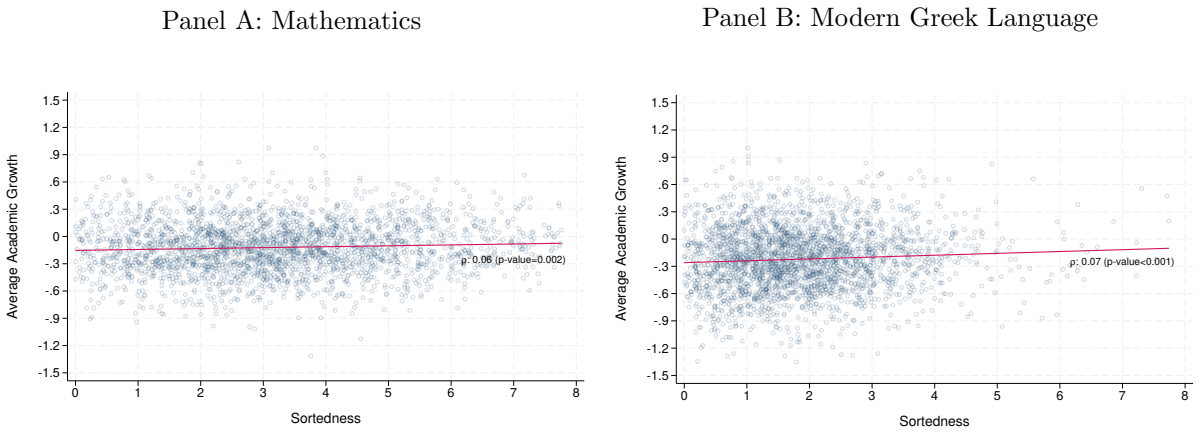
$$Y_{isct} = \alpha + \sum_{g=1}^{3}\sum_{q=2}^{5} \beta_{gq}\, \mathbf{1}\{\text{Sortedness}_{sct} \in Q_q\} \times \mathbf{1}\{i \in T_g\} + \gamma X_{isct} + \delta_c + \tau_t + \varepsilon_{isct}, \tag{5.1}$$

where Y_{isct} is the standardized grade-10 exam score of student i in subject s, school c, and year t; Q_q denotes the q-th quintile of cohort partition sortedness in the school-year-subject cohort, with the least-sorted quintile (Q_1) as the omitted category within each tertile; and T_g denotes the bottom, middle, and top tertile of students' own grade-9 achievement, with the tertile indicators included as controls. The vector X_{isct} includes individual grade-9 subject performance, gender, and the mean and standard deviation of prior performance in the school-year-subject cohort. School and year fixed effects, δ_c and τ_t, absorb time-invariant differences across schools and common year shocks. In richer specifications, we also include teacher fixed effects. Standard errors are clustered at school level. The model is estimated separately by subject. The coefficients β_{gq} trace out, for students at each part of the prior-achievement distribution, how performance in more sorted cohorts compares with performance in the least sorted cohorts.

Identification relies on the institutional feature that students are alphabetically assigned to classrooms within each grade. This process generates quasi-random variation in classroom composition with respect to prior academic performance, yielding exogenous differences in cohort partition sortedness across school-grade cohorts. Because sorting is mechanically induced by the random distribution of surnames rather than academic criteria, the resulting variation can be interpreted as plausibly exogenous to unobserved determinants of student achievement. Figure D.8 illustrates this variation by plotting cohort partition sortedness against cohort-level mean prior performance (from grade 9). Table D.3 provides evidence supporting the identifying assumption that classroom assignment within school cohorts is effectively quasi-random.

We also estimate overall associations in Appendix D: linear specifications that replace the quintile-tertile interactions with the continuous sortedness measure, both pooled and interacted with prior-achievement tertiles (Table D.4), and a pooled nonlinear specification with quintile indicators alone (Table D.5).

Figure 3 Cohort Partition Sortedness and Average Academic Growth



Notes: Each marker is a school × year cohort in grade 10 per subject. *Sortedness* is measured as the percentage difference in Lorenz-curve areas between the empirical partition and a theoretically uniform partition within a cohort; higher values indicate greater sorting of students. *Average Academic Growth* is the mean change in standardized final exam scores from grade 9 to grade 10. The red line is the fitted OLS trend. The figure reports the Pearson correlation coefficient and the p-value from a test of significance. To limit the influence of extreme values, we exclude observations falling outside the 1^{st}–99^{th} percentile range of the distribution, leaving 2,642 observations in Mathematics and 2,645 in Modern Greek.

Results. Figure 4 presents our main results. Each panel plots the estimated coefficients β_{gq} from specification (5.1), comparing students in each sortedness quintile to students in the least sorted quintile (Q_1), separately for the bottom, middle, and top tertiles of prior achievement; Table D.6 in Appendix D reports the point estimates across specifications.[6]

Three patterns stand out. First, the positive estimates are concentrated among lower-achieving students. Students in the bottom tertile of prior achievement score significantly higher in the most sorted cohorts than in the least sorted ones. In the specification with teacher fixed effects, being in a cohort of top-quintile sortedness is associated with scores that are 0.08 standard deviations higher

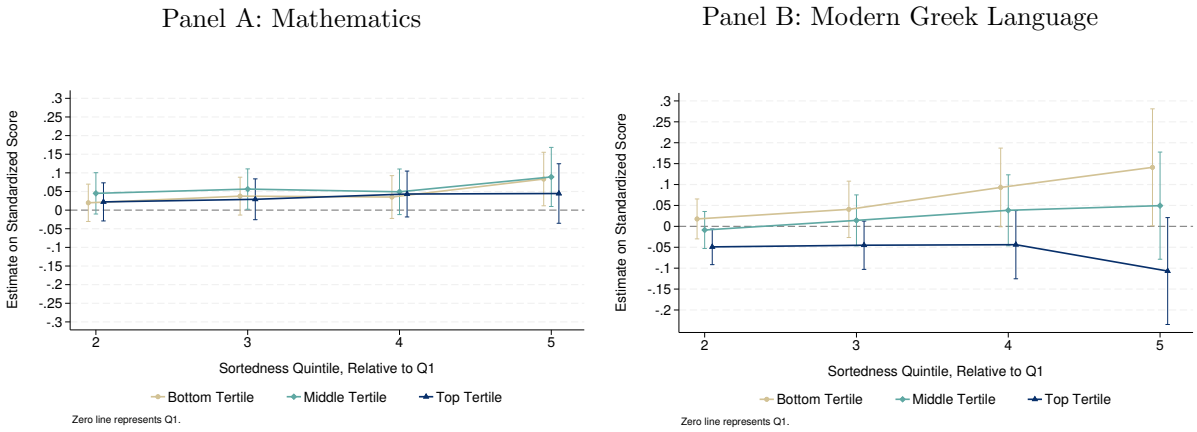
[6] Controlling for the distribution of cohort prior achievement is important because the extent of feasible sorting depends on the underlying achievement distribution of the cohort. Alphabetical assignment randomizes sortedness only within that feasible range, so controlling for the mean and standard deviation of cohort prior achievement is what makes the variation comparable across cohorts. Table D.7 reports estimates that omit these controls. In Mathematics the positive associations disappear once cohort prior achievement is accounted for, while in Modern Greek Language the bottom-tertile association is strengthened.

in Mathematics and 0.14 standard deviations higher in Modern Greek Language. In Modern Greek Language, these differences rise roughly monotonically across sortedness quintiles. Second, there is little evidence of corresponding gains for high-achieving students. For the top tertile, estimates are small and statistically insignificant in Mathematics once teacher fixed effects are included, and in Modern Greek Language the point estimates are negative throughout, significantly so only for the second quintile in the specification with teacher fixed effects. Middle-tertile students fall in between, with positive and sometimes significant estimates in Mathematics and no discernible differences in Modern Greek Language. Third, the estimates are nonlinear: in the specifications with teacher fixed effects, contrasts at the bottom of the sortedness distribution are generally small and imprecise, and the largest differences appear in the most sorted cohorts, consistent with our motivation for a specification that does not impose linearity.[7]

Discussion. The incidence pattern in Figure 4 is informative about which peer effects models can rationalize the data, and it points in different directions in the two subjects. In Modern Greek Language, the pattern resembles a redistribution: students in the bottom tertile gain in the most sorted cohorts, top-tertile point estimates are negative, and pooled estimates are statistically indistinguishable from zero (Tables D.4 and D.5). This is consistent with sorting neutrality—recall that under sorting-neutral peer effects models, any partition yields the same total unweighted effort (Section 4.3.1)—but not with the distributional prediction of LIM. LIM is sorting-neutral (Corollary 3) and thus a natural candidate for this dataset. Lemma 1 further suggests that moving to the fully assortative benchmark maximizes the dispersion of expected efforts, so that in the extreme, sorting raises the peer component for high-ability students and lowers it for low-ability ones. Partitions ordered by our sortedness measure need not be comparable in the majorization order, but this endpoint comparison suggests that the same incidence should hold as sortedness rises; empirically, we observe the opposite. Which sorting-neutral model best accounts for this pattern remains an open question. In Mathematics, by contrast, the gains at high sortedness are significant for the bottom and middle tertiles and positive, though insignificant, for the top tertile, and pooled nonlinear estimates are significantly positive in the most sorted cohorts (Table D.5). This pattern points to a departure from sorting neutrality under which sorting increases aggregate

[7] The overall effects reported in Appendix D are consistent with, but far less sharp than, these patterns. In linear specifications (Table D.4), cohort sortedness has no statistically significant effect on average grade-10 performance in either subject; interacting the continuous measure with prior-achievement tertiles yields marginally significant positive effects for bottom-tertile students in Modern Greek Language and marginally significant negative effects for top-tertile students in Mathematics, with the remaining estimates imprecise. This is expected: because the effects in Figure 4 are concentrated in the most sorted cohorts, a linear specification averages them with the near-zero contrasts at lower sortedness levels and understates the incidence of sorting. Similarly, the pooled nonlinear specification (Table D.5) shows achievement gains concentrated in the most sorted quintiles in Mathematics but smaller, insignificant estimates in Modern Greek Language, because pooling across students averages the bottom-tertile gains with the null or negative estimates for top-tertile students visible in Figure 4.

Figure 4 Nonlinear Heterogeneous Effects of Cohort Partition Sortedness by Prior Student Performance



Notes: Academic performance is measured by standardized grade 10 final examination scores. Prior performance is measured by standardized grade 9 final examination scores. Sortedness is defined as the percentage difference in Lorenz curve areas between the empirical partition and a theoretically uniform partition. The first quintile of sortedness represents school cohorts with the least homogeneous classroom assignments, whereas the fifth quintile represents cohorts with the most internally homogeneous classrooms (greatest between-classroom separation by prior achievement. Coefficients represent differences relative to the bottom quintile of cohort partition sortedness within each tertile of prior student performance. Standard errors are clustered at school level. Vertical lines denote 95 percent confidence intervals.

performance, in line with non-linear specifications such as the quadratic-in-means model, where the assortative partition can be optimal (Proposition 9). Regardless of which microfoundation ultimately fits best, the evidence indicates that the cohort partition structure is systematically associated with heterogeneous and nonlinear differences in student outcomes. This is precisely the setting in which the choice of partition and of the planner's objective matters, motivating the optimal cohort partition problem studied in this paper.

6. Conclusion and future directions

In this paper, we initiated a theoretical study of optimal cohort partitions in education and related contexts. Our results suggest that the structure of the objective—whether it is Schur-convex, Schur-concave, or edge-quasi-convex—shapes the optimal partition rule in a tractable way, potentially offering guidance to administrators and policymakers who face these decisions in practice. Our empirical findings from Greek public schools, where greater sortedness is associated with gains for lower-achieving students and, once teacher fixed effects are included, no measurable gains for high achievers, underscore that the choice of partition has real distributional consequences. Several questions remain open, both regarding policy implications (e.g., the individual fairness of different partitions) and theoretical extensions (e.g., extending our results to multi-parameter partitions

and networks). We hope that the theoretical framework in Section 3, similar to its applications to education (Section 4) as well as team formation and contests (Appendix C), will prove useful in various other contexts.

References

Aczél, János. 1966. *Lectures on Functional Equations and Their Applications*. Academic Press.

Agrawal, Rakesh, Behzad Golshan, Evimaria Terzi. 2014. Grouping Students in Educational Settings. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1017–1026.

Anily, Shoshana, Awi Federgruen. 1991. Structured Partitioning Problems. *Operations Research* **39**(1) 130–149.

Aprahamian, Hrayer, Douglas R Bish, Ebru K Bish. 2019. Optimal Risk-based Group Testing. *Management Science* **65**(9) 4365–4384.

Arnold, Barry C, José María Sarabia. 2018. *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*, vol. 7. Springer.

Arnott, Richard, John Rowse. 1987. Peer Group Effects and Educational Attainment. *Journal of Public Economics* **32**(3) 287–305.

Ashlagi, Itai, Faidra Monachou, Afshin Nikzad. 2025. Optimal Allocation via Waitlists: Simplicity Through Information Design. *Review of Economic Studies* **92**(1) 40–68.

Bergstrom, Theodore C, Hal R Varian. 1985. When Are Nash Equilibria Independent of the Distribution of Agents' Characteristics? *Review of Economic Studies* **52**(4) 715–718.

Bertsimas, Dimitris, Mac Johnson, Nathan Kallus. 2015. The Power of Optimization over Randomization in Designing Experiments Involving Small Samples. *Operations Research* **63**(4) 868–876.

Betts, Julian R. 2011. The Economics of Tracking in Education. *Handbook of the Economics of Education*, vol. 3. Elsevier, 341–381.

Bhattacharya, Debopam. 2009. Inferring Optimal Peer Assignment from Experimental Data. *Journal of the American Statistical Association* **104**(486) 486–500.

Boros, Endre, Peter L Hammer. 1989. On Clustering Problems with Connected Optima in Euclidean Spaces. *Discrete Mathematics* **75**(1-3) 81–88.

Boros, Endre, F.K. Hwang. 1996. Optimality of Nested Partitions and Its Application to Cluster Analysis. *SIAM Journal on Optimization* **6**(4) 1153–1162.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, Yves Zenou. 2024. Toward a General Theory of Peer Effects. *Econometrica* **92**(2) 543–565.

Brunello, Giorgio, Daniele Checchi. 2007. Does School Tracking Affect Equality of Opportunity? New International Evidence. *Economic Policy* **22**(52) 782–861.

Bursztyn, Leonardo, Georgy Egorov, Robert Jensen. 2019. Cool to Be Smart or Smart to Be Cool? Understanding Peer Pressure in Education. *Review of Economic Studies* **86**(4) 1487–1526.

Calvó-Armengol, Antoni, Eleonora Patacchini, Yves Zenou. 2009. Peer Effects and Social Networks in Education. *Review of Economic Studies* **76**(4) 1239–1267.

Carrell, Scott E, Bruce I Sacerdote, James E West. 2013. From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica* **81**(3) 855–882.

Chade, Hector, Jan Eeckhout. 2018. Matching Information. *Theoretical Economics* **13**(1) 377–414.

Chakravarty, AK, JB Orlin, UG Rothblum. 1982. A partitioning problem with additive objective with an application to optimal inventory groupings for joint replenishment. *Operations Research* **30**(5) 1018–1022.

Chakravarty, Amiya K, James B Orlin, Uriel G Rothblum. 1985. Consecutive Optimizers for a Partitioning Problem with Applications to Optimal Inventory Groupings for Joint Replenishment. *Operations Research* **33**(4) 820–834.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, Danny Yagan. 2011. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics* **126**(4) 1593–1660.

Dahl, Geir. 2001. Principal Majorization Ideals and Optimization. *Linear Algebra and Its Applications* **331**(1-3) 113–130.

Dalton, Hugh. 1920. The Measurement of the Inequality of Incomes. *Economic Journal* **30**(119) 348–361.

Epple, Dennis, Richard E Romano. 2011. Peer Effects in Education: A Survey of the Theory and Evidence. *Handbook of Social Economics*, vol. 1. Elsevier, 1053–1163.

Gal, Shmuel, Boris Klots. 1995. Optimal Partitioning Which Maximizes the Sum of the Weighted Averages. *Operations Research* **43**(3) 500–508.

Garlick, Robert. 2018. Academic Peer Effects with Different Group Assignment Policies: Residential Tracking Versus Random Assignment. *American Economic Journal: Applied Economics* **10**(3) 345–369.

Graham, Bryan S, Guido W Imbens, Geert Ridder. 2014. Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis. *Quantitative Economics* **5**(1) 29–66.

Hardy, Godfrey Harold, John Edensor Littlewood, George Pólya. 1929. Some Simple Inequalities Satisfied by Convex Functions. *Messenger Math* **58** 145–152.

Hartline, Jason D, Tim Roughgarden. 2008. Optimal Mechanism Design and Money Burning. *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*. 75–84.

Hellenic Ministry of Education and Religious Affairs. 2021. School Reform, Empowerment of Teachers, and Other Provisions (Article 85: Distribution of Students into Classes). Draft Law, Public Consultation, Open Government Initiative, `http://www.opengov.gr/ypepth/?p=5915`. URL `http://www.opengov.gr/ypepth/?p=5915`.

Hoxby, Caroline M. 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation.

Hoxby, Caroline M, Gretchen Weingarth. 2005. Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Tech. rep., Citeseer.

Hwang, Frank K, Colin L Mallows. 1995. Enumerating Nested and Consecutive Partitions. *Journal of Combinatorial Theory, Series A* **70**(2) 323–333.

Hwang, Frank K, Uriel G Rothblum. 1996. Directional-Quasi-Convexity, Asymmetric Schur-Convexity and Optimality of Consecutive Partitions. *Mathematics of Operations Research* **21**(3) 540–554.

Hwang, Frank Kwang-ming, Uriel R Rothblum. 2011. *Partitions: Optimality and Clustering - Volume I: Single-Parameter*, vol. 19. World Scientific.

Kalogrides, Demetra, Susanna Loeb. 2013. Different Teachers, Different Peers: The Magnitude of Student Sorting Within Schools. *Educational Researcher* **42**(6) 304–316.

Lavy, Victor, Olmo Silva, Felix Weinhardt. 2009. The Good, the Bad, and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools. Tech. rep., National Bureau of Economic Research.

Lavy, Victor, Olmo Silva, Felix Weinhardt. 2012. Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *Review of Economics and Statistics* **94**(4) 1079–1096.

Lazear, Edward P. 2001. Educational Production. *Quarterly Journal of Economics* **116**(3) 777–803.

Lee, Michael D, Siqi Liu. 2022. Drafting Strategies in Fantasy Football: A Study of Competitive Sequential Human Decision Making. *Judgment and Decision Making* **17**(4) 691–719.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* **60**(3) 531–542.

Marshall, Albert W, Ingram Olkin, Barry C Arnold. 2011. Inequalities: Theory of Majorization and Its Applications .

Meier, Volker, Gabriela Schütz. 2007. The Economics of Tracking and Non-Tracking. Tech. rep., Ifo Institute.

Myerson, Roger B. 1981. Optimal Auction Design. *Mathematics of Operations Research* **6**(1) 58–73.

OECD. 2018. *Education for a Bright Future in Greece*. OECD Publishing, Paris. doi:https://doi.org/10.1787/9789264298750-en. URL `https://www.oecd-ilibrary.org/content/publication/9789264298750-en`.

Perez-Salazar, Sebastian, Alfredo Torrico, Victor Verdugo. 2021. Preserving Diversity When Partitioning: A Geometric Approach. *Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–11.

Procaccia, Ariel, Isaac Robinson, Jamie Tucker-Foltz. 2024. School Redistricting: Wiping Unfairness Off the Map. *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2704–2724.

Rosenzweig, Mark R, Bing Xu. 2023. Classroom Competition, Student Effort, and Peer Effects. Tech. rep., National Bureau of Economic Research.

Ryvkin, Dmitry. 2011. The Optimal Sorting of Players in Contests Between Groups. *Games and Economic Behavior* **73**(2) 564–572.

Sacerdote, Bruce. 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics* **116**(2) 681–704.

Sacerdote, Bruce. 2011. Peer Effects in Education: How Might They Work, How Big Are They, and How Much Do We Know Thus Far? *Handbook of the Economics of Education*, vol. 3. Elsevier, 249–277.

Schreiber, E. L. 2014. Optimal multi-way number partitioning. Phd thesis, UCLA. URL `https://escholarship.org/uc/item/30g6n09q`.

Truong, Debbie. 2022. Detracking in K-12 Classrooms. *U.S. News* URL `https://www.usnews.com/education/k12/articles/detracking-in-k-12-classrooms`.

Tullock, Gordon. 1980. Efficient Rent Seeking. *Toward a Theory of the Rent-Seeking Society*. Texas A&M University Press.

Appendix A: Technical background

In this appendix, we include some technical background on polytopes (Appendix A.1), directional convexity (Appendix A.2), majorization (Appendix A.3), and Dahl's algorithm for linear objectives (Appendix A.4). The material is adapted from Hwang and Rothblum (2011, 1996), Dahl (2001) and Marshall et al. (2011).

A.1. Polytopes

A *polytope* in $\mathbb{R}^K$ is the convex hull of finitely many vectors in $\mathbb{R}^K$. A *polyhedron* in $\mathbb{R}^K$ is the solution set of a finite system of linear inequalities. The Main Theorem of Polytopes asserts that a subset of $\mathbb{R}^K$ is a polytope if and only if it is a bounded polyhedron.

A *face* of a polytope $P \subset \mathbb{R}^K$ is any set of the form $F = P \cap \{\mathbf{x} \in \mathbb{R}^K : \sum_{j\in[K]} c_j x_j = \gamma\}$ where $P \subseteq \{\mathbf{x} \in \mathbb{R}^K : \sum_{j\in[K]} c_j x_j \leq \gamma\}$. Faces of dimension 0 and 1 are respectively called *vertices* and *edges*. If P is a polytope in $\mathbb{R}^K$, then it is the convex hull of its vertices (Proposition 3.1.1 (a), Hwang and Rothblum (2011)).

An *extreme point* of a convex set C is a point $x \in \mathbb{R}^K$ for which $\{x\}$ is a face of C. The extreme points of a polytope are its vertices (Proposition 3.1.2 (l), Hwang and Rothblum (2011))

A.2. Generalized notions of convexity and optimization

When maximizing a function f over a bounded polyhedron, quasi-convexity is usually the standard minimal requirement for f to attain its maximum at an extreme point. Hwang and Rothblum (1996) obtain a weaker property that suffices for optimality of extreme points by introducing the concept of quasi-convexity along sets, namely *directional quasi-convexity*. Under the umbrella of directional quasi-convexity, they include the notions of quasi-convexity, Schur convexity, and asymmetric Schur convexity (see Figure A.5).

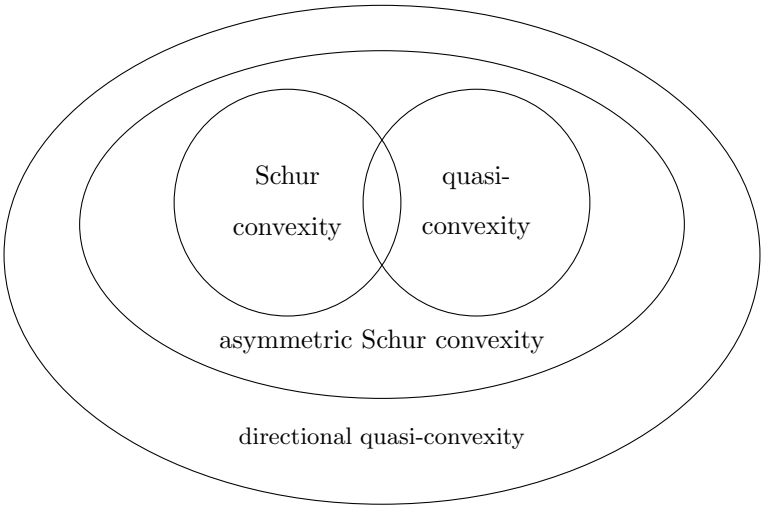


Figure A.5 Relationship of convexity notions (Hwang and Rothblum 1996)

Definition A.5 **(Quasi-convexity, Hwang and Rothblum (2011)).** *For a convex $C \subseteq \mathbb{R}^K$, a function $f : C \to \mathbb{R}$ is called* quasi-convex *if for all $x, x' \in C$ and $\lambda \in [0,1]$, it holds that $f(\lambda x + (1-\lambda)x') \leq \max\{f(x), f(x')\}$. It is* strictly quasi-convex *if the previous inequality is strict for all $x \neq x'$ and $\lambda \in (0,1)$.*

DEFINITION A.6 **(Asymmetric Schur convexity, Hwang and Rothblum (2011))**. *For $k \in [K]$, let $\mathbf{u}^k$ be the k-th unit vector in $\mathbb{R}^K$. Let $\Lambda = \{(i,j) : i,j \in [K], i \neq j\}$ and for $(i,j) \in \Lambda$, $\mathbf{u}^{ij} := \mathbf{u}^i - \mathbf{u}^j$. For a convex $C \subseteq \mathbb{R}^K$, the function $f : C \to \mathbb{R}$ is called* (strictly) asymmetric Schur convex *if it is (strictly) quasi-convex along the set $\{\mathbf{u}^{ij} : (i,j) \in \Lambda\}$.*

LEMMA A.2 **(Lemma 2.1 in Hwang and Rothblum (1996))**. *Let C be a convex subset of $\mathbb{R}^K$, $f : C \to \mathbb{R}$ and $\mathbf{d} \in \mathbb{R}^K$. Then f is $\mathbf{d}$-quasi-convex if and only if for each $\mathbf{x} \in C$, the restriction to $C \cap \{\mathbf{x} + \gamma \mathbf{d} : \gamma \in \mathbb{R}\}$ when viewed as a function of the parameter γ cannot decrease after an increase.*

PROPOSITION A.11 **(Corollary 2.5 in Hwang and Rothblum (1996))**. *Let P be a polytope and let $f : P \to \mathbb{R}$ be edge-quasi-convex on P. Then f attains a maximum over P on one of P's vertices. Further, if f is strictly edge-quasi-convex, then every maximizer of f over P is a vertex of P.*

A.3. Majorization

DEFINITION A.7 **(Weak majorization)**. *For $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^K$, we say that $\mathbf{X}$* weakly majorizes *$\mathbf{Y}$, that is $\mathbf{Y} \prec_w \mathbf{X}$, if $\sum_{j \leq k} Y_{[j]} \leq \sum_{j \leq k} X_{[j]}$ for every $k \in [K]$.*

DEFINITION A.8 **(Doubly stochastic matrix)**. *A $K \times K$ matrix $P = (p_{ij})$ is* doubly stochastic *if $p_{ij} \geq 0$ for all $i,j \in [K]$, $\sum_{i=1}^K p_{ij} = 1$ for all $j \in [K]$, and $\sum_{j \in [K]} p_{ij} = 1$ for all $i \in [K]$.*

LEMMA A.3 **(Hardy, Littlewood, and Pólya (1929))**. *$\mathbf{X} \succ \mathbf{Y}$ if and only if there exists a doubly stochastic matrix Q such that $\mathbf{Y} = Q\mathbf{X}$.*

DEFINITION A.9 **(Supermodular function)**. *A function $\phi : \mathbb{R}^m \to \mathbb{R}$ is called* supermodular *if for any $\mathbf{x} \in \mathbb{R}^m$ and $i \neq j$,*

$$\phi(x_1, \ldots, x_i + \delta_1, \ldots, x_j + \delta_2, \ldots, x_m) + \phi(x_1, \ldots, x_i - \delta_1, \ldots, x_j - \delta_2, \ldots, x_m) \geq$$
$$\phi(x_1, \ldots, x_i + \delta_1, \ldots, x_j - \delta_2, \ldots, x_m) + \phi(x_1, \ldots, x_i - \delta_1, \ldots, x_j + \delta_2, \ldots, x_m),$$

whenever $\delta_1, \delta_2 \geq 0$.

PROPOSITION A.12 **(Karamata Inequality, Arnold and Sarabia (2018))**. *Let $\mathbf{X}, \mathbf{Y} \in \mathbb{R}_+^K$. Then, $\mathbf{X} \prec \mathbf{Y}$ if and only if*

$$\sum_{i \in [K]} f(X_i) \leq \sum_{i \in [K]} f(Y_i) \tag{A.1}$$

for every convex function $f : \mathbb{R} \to \mathbb{R}$. Inequality (A.1) reverses if f is concave.

LEMMA A.4. *Let $u = (u_1, u_2), v = (v_1, v_2) \in \mathbb{R}_+^2$ satisfy $u_1 + u_2 = v_1 + v_2$. Then*

$$u \prec v \quad \Longleftrightarrow \quad |u_1 - u_2| \leq |v_1 - v_2|.$$

Proof. Let $u^{\downarrow}$ and $v^{\downarrow}$ denote the vectors obtained by sorting u and v in nonincreasing order. Under the equal-sum assumption $u_1 + u_2 = v_1 + v_2$, the definition of majorization in $\mathbb{R}_+^2$ reduces to

$$u \prec v \quad \Longleftrightarrow \quad u_1^{\downarrow} \leq v_1^{\downarrow}.$$

Let $S = u_1 + u_2 = v_1 + v_2$, and define $d_u = |u_1 - u_2|, d_v = |v_1 - v_2|$. Assume without loss of generality that $u_1 \geq u_2$, so $u_1^{\downarrow} = u_1$. Then

$$u_1^{\downarrow} = u_1 = \frac{(u_1 + u_2) + (u_1 - u_2)}{2} = \frac{S + d_u}{2}.$$

If instead $u_2 \geq u_1$, the same calculation with indices reversed shows that $u_1^{\downarrow} = (S + d_u)/2$. Thus, in all cases,

$$u_1^{\downarrow} = \frac{S + d_u}{2}, \qquad v_1^{\downarrow} = \frac{S + d_v}{2}.$$

Therefore,

$$u \prec v \iff u_1^{\downarrow} \leq v_1^{\downarrow} \iff \frac{S + d_u}{2} \leq \frac{S + d_v}{2} \iff d_u \leq d_v \iff |u_1 - u_2| \leq |v_1 - v_2|.$$

□

PROPOSITION A.13 **(Theorem 2.1 in Dahl (2001))**. *Let* $\mathbf{X} \in \mathbb{R}_+^K$ *with* $X_1 \geq X_2 \ldots \geq X_K$. *The mapping* $\mathbf{X}^{\#b}$ *is a bijection between the set of consecutive divisions* $\mathcal{B}$ *and the vertex set of* $P(\mathbf{X})$. *Thus, the vertices of* $P(\mathbf{X})$ *are the interval means of* $\mathbf{X}$.

A.4. Dahl's algorithm

We can exploit Proposition A.13 to construct the optimal solution for optimization problems of the form:

$$\max \sum_{j \in [K]} w_j X_{[j]} \text{ subject to } \mathbf{X} \prec \widetilde{\mathbf{X}}. \tag{A.2}$$

Dahl (2001) uses the following efficient dynamic programming algorithm to construct the optimal solution.

ALGORITHM 1: Dahl's algorithm

$z_0 \leftarrow 0$;

for $k = 1, \ldots, K$ **do**
  $p(k) \leftarrow \arg\max\{z_t + \hat{w}_{t+1:k} : t = 0, \ldots, k-1\}$;
  $z_k \leftarrow z_{p(k)} + \hat{w}_{p(k)+1:k}$;
end

For a block $b_r = \{i, i+1, \ldots, j\}$, define its contribution by

$$\widehat{w}_{i:j} = \left(\frac{1}{j - i + 1} \sum_{\ell=i}^{j} \widetilde{X}_\ell \right) \left(\sum_{\ell=i}^{j} w_\ell \right). \tag{A.3}$$

Then, the maximum value of (A.2) equals z_K and the optimal solution is determined by

$$I(b) = \{K, p(K), p(p(K)), \ldots, 0\},$$

using Algorithm 1.

Appendix B: Proofs

Note: For the sake of readability, in some of the proofs we work directly with integral partitions (e.g., Lemma B.5) and then adjust the objective for fractional partitions. The form of the objective can be verified through tedious algebra that we omit.

B.1. Proofs from Section 3

PROPOSITION 1. *Let $\mathbf{X}^M$ be the summation vector of a feasible fractional partition $M \in \Pi_{\mathcal{F}}$. Then, $\mathbf{X}^M \in P(\widetilde{\mathbf{X}})$. Conversely, for any $\mathbf{X} \in P(\widetilde{\mathbf{X}})$, there exists a fractional partition $M \in \Pi_{\mathcal{F}}$ with summation vector $\mathbf{X}^M = \mathbf{X}$.*

Proof. We first show that $\mathbf{X}^M \in P(\widetilde{\mathbf{X}})$. Indeed,

$$\sum_{j\in[K]} X^M_{[j]} = \sum_{j\in[K]} \widetilde{X}_{[j]} = \sum_{i\in[N]} x_i.$$

Next we prove that for every $k \in [K]$,

$$\sum_{j\in[k]} X^M_{[j]} \le \sum_{j\in[k]} \widetilde{X}_{[j]}.$$

The proof is by contradiction. Suppose instead that k^* is an arbitrary section at which the majorization condition fails, i.e.,

$$\sum_{j=1}^{k^*} X^M_{[j]} > \sum_{j=1}^{k^*} \widetilde{X}_{[j]}.$$

Note that both $\mathbf{X}^M$ and $\widetilde{\mathbf{X}}$ have nk^* students in sections 1 to k^* but $\mathbf{X}^M$ has a higher aggregate quality of students. This is a contradiction, since the assortative partition $\widetilde{\mathbf{X}}$ assigns students to sections in a greedy manner starting from the student with the highest attribute x_1.

For the other direction, consider any $\mathbf{X} \in P(\widetilde{\mathbf{X}})$. We need to show that there exists a feasible partition $M \in \mathbb{R}_+^{N\times K}$ with $\mathbf{X}^M = \mathbf{X}$. By Lemma A.3, $\mathbf{X} \in P(\widetilde{\mathbf{X}})$ implies that there exists a doubly stochastic matrix $Q \in \mathbb{R}_+^{K\times K}$ such that $\mathbf{X} = Q\widetilde{\mathbf{X}}$. Let $M \in \mathbb{R}_+^{N\times K}$ be a matrix defined as

$$\mu_{ij} = q_{j\lceil i/n\rceil}, \quad i \in [N],\, j \in [K].$$

Then, given that Q is doubly stochastic, $\mu_{ij} \in [0,1]$ and

$$\sum_{j\in[K]} \mu_{ij} = \sum_{j\in[K]} q_{j\lceil i/n\rceil} = 1,$$

for all $i \in [N]$. Similarly,

$$\sum_{i\in[N]} \mu_{ij} = n\sum_{k=1}^{K} q_{jk} = n\cdot 1 = n,$$

for all $j \in [K]$. Consequently, M is a feasible partition, with $\mathbf{X}^M = \mathbf{X}$ following directly by the construction of M. □

PROPOSITION 2 **(Schur convex objective)**. *Consider the optimization problems (3.1) and (3.2) with f being Schur convex. Then the assortative partition is optimal over both the set of fractional partitions $\Pi_{\mathcal{F}}$ and the set of integral partitions $\Pi_{\mathcal{I}}$. If f is strictly Schur convex, then every optimal partition has summation vector $\widetilde{\mathbf{X}}$.*

Proof. For any partition M, $\mathbf{X}^M \prec \widetilde{\mathbf{X}}$. By the definition of Schur convexity, it thus follows that $f(\widetilde{\mathbf{X}}) \ge f(\mathbf{X}^M)$. Therefore, $\widetilde{\mathbf{X}}$ must attain the optimal value in (3.1).

For the second statement, let M be optimal and suppose for the sake of contradiction that $\mathbf{X}^M$ is not a permutation of $\widetilde{\mathbf{X}}$. Since $\mathbf{X}^M \prec \widetilde{\mathbf{X}}$,strict Schur convexity gives $f(\mathbf{X}^M) < f(\widetilde{\mathbf{X}})$, contradicting optimality. □

PROPOSITION 3 **(Schur concave objective)**. *Consider the optimization problems (3.1) and (3.2) with f being Schur concave. Then the uniform partition is optimal over the set of fractional partitions $\Pi_{\mathcal{F}}$. Furthermore, there exists a balanced partition that is optimal over the set of integral partitions $\Pi_{\mathcal{I}}$. If f is strictly Schur concave, then every optimal fractional partition has summation vector $\underline{\mathbf{X}}$, and every optimal integral partition is balanced.*

Proof. For $\Pi_{\mathcal{F}}$, the proof follows as in Proposition 2 from the definition of Schur concavity by observing that $\underline{\mathbf{X}} \prec \mathbf{X}^M$ for any $\mathbf{X}^M \in P(\widetilde{\mathbf{X}})$.

For the second statement, we consider only integral partitions in $\Pi_{\mathcal{I}}$, and show that some balanced partition is optimal. We first prove the following claim.

CLAIM 1. *Suppose that $\pi, \pi' \in \Pi_{\mathcal{I}}$ agree on all sections except for $\{j,k\}$ and satisfy $X_j^{\pi'} + X_k^{\pi'} = X_j^{\pi} + X_k^{\pi}$. If $|X_j^{\pi'} - X_k^{\pi'}| \le |X_j^{\pi} - X_k^{\pi}|$, then $\mathbf{X}^{\pi'} \prec \mathbf{X}^{\pi}$.*

Proof of Claim 1. The vectors $\mathbf{X}^{\pi'}$ and $\mathbf{X}^{\pi}$ coincide on all sections except for $\{j,k\}$, while sections j,k satisfy $X_j^{\pi'} + X_k^{\pi'} = X_j^{\pi} + X_k^{\pi}$, thus $\sum_{\ell\in[K]} X_\ell^{\pi'} = \sum_{\ell\in[K]} X_\ell^{\pi}$. By Lemma A.4, the hypothesis $|X_j^{\pi'} - X_k^{\pi'}| \le |X_j^{\pi} - X_k^{\pi}|$ gives $(X_j^{\pi'}, X_k^{\pi'}) \prec (X_j^{\pi}, X_k^{\pi})$. Since the two vectors agree on the remaining $K-2$ coordinates, every partial sum $\sum_{\ell\in[k']} X_{[\ell]}^{\pi'}$ is at most $\sum_{\ell\in[k']} X_{[\ell]}^{\pi}$ for all $k' \in [K]$, and the total sums coincide. Therefore, $\mathbf{X}^{\pi'} \prec \mathbf{X}^{\pi}$. □

For a partition $\pi \in \Pi_{\mathcal{I}}$, define the $\Phi(\pi) = \sum_{\ell\in[K]} (X_\ell^{\pi})^2$, which takes finitely many values over the finite set $\Pi_{\mathcal{I}}$. Starting from any optimal partition π^0, we construct a sequence of partitions $\pi^0, \pi^1, \pi^2, \dots$ as follows. If π^t is balanced, set $\pi^* := \pi^t$ and stop. Otherwise, pick a pair (j,k) for which (π_j^t, π_k^t) is not a balanced 2-way partition of $\pi_j^t \cup \pi_k^t$, and obtain π^{t+1} from π^t by replacing (π_j^t, π_k^t) with a balanced 2-way partition, keeping all other sections the same.

Each step keeps the sum $X_j^{\pi^t} + X_k^{\pi^t}$ fixed and strictly reduces the gap $|X_j^{\pi^t} - X_k^{\pi^t}|$. Since

$$(X_j^{\pi^t})^2 + (X_k^{\pi^t})^2 = \tfrac{1}{2}\big(X_j^{\pi^t} + X_k^{\pi^t}\big)^2 + \tfrac{1}{2}\big(X_j^{\pi^t} - X_k^{\pi^t}\big)^2$$

is strictly increasing in $|X_j^{\pi^t} - X_k^{\pi^t}|$, and all other section sums are unchanged, we have $\Phi(\pi^{t+1}) < \Phi(\pi^t)$. Thus, the sequence $\Phi(\pi^0) > \Phi(\pi^1) > \cdots$ is strictly decreasing. On top of this, since Φ takes finitely many values over $\Pi_{\mathcal{I}}$, the sequence must be finite and the construction terminates at some balanced π^*.

By Claim 1, each step satisfies $\mathbf{X}^{\pi^{t+1}} \prec \mathbf{X}^{\pi^t}$. Given that f is Schur concave, $f(\mathbf{X}^{\pi^{t+1}}) \ge f(\mathbf{X}^{\pi^t})$ at every step. Chaining these inequalities gives $f(\mathbf{X}^{\pi^*}) \ge f(\mathbf{X}^{\pi^0})$. Since π^0 is optimal, π^* is optimal as well. Therefore, there exists an optimal balanced partition in $\Pi_{\mathcal{I}}$.

For the last statement, the proof is by contradiction. Let $M \in \Pi_{\mathcal{F}}$ be optimal. Since $\underline{\mathbf{X}} \prec \mathbf{X}^M$, if $\mathbf{X}^M \neq \underline{\mathbf{X}}$ then strict Schur concavity gives $f(\underline{\mathbf{X}}) > f(\mathbf{X}^M)$, contradicting optimality.

Now let $\pi \in \Pi_{\mathcal{I}}$ be optimal and suppose it is not balanced. Then there exist sections $j \neq k$ and $\pi' \in \Pi_{\mathcal{I}}$ that agrees with π on all other sections, but $X_j^{\pi'} + X_k^{\pi'} = X_j^{\pi} + X_k^{\pi}$ and $|X_j^{\pi'} - X_k^{\pi'}| < |X_j^{\pi} - X_k^{\pi}|$. By Claim 1, $\mathbf{X}^{\pi'} \prec \mathbf{X}^{\pi}$. Furthermore, $\mathbf{X}^{\pi'}$ is not a permutation of $\mathbf{X}^{\pi}$, since the two vectors agree outside $\{j,k\}$ while on $\{j,k\}$ they have equal sums but different gaps. If f is strictly Schur concave, then $f(\mathbf{X}^{\pi'}) > f(\mathbf{X}^{\pi})$, which is again a contradiction. □

COROLLARY B.4. *For any $\pi \in \Pi_{\mathcal{I}}$, there exists a balanced partition $\pi_{\mathrm{bal}} \in \Pi_{\mathcal{I}}$ such that $\mathbf{X}^{\pi} \succ \mathbf{X}^{\pi_{\mathrm{bal}}}$.*

Proof. Starting from $\pi^0 := \pi$, construct a sequence $\pi^0, \pi^1, \ldots \in \Pi_{\mathcal{I}}$ as in the proof of Proposition 3. If π^t is balanced, stop; otherwise pick a pair (j,k) for which (π^t_j, π^t_k) is not a balanced 2-way partition of $\pi^t_j \cup \pi^t_k$, and obtain π^{t+1} from π^t by replacing (π^t_j, π^t_k) with such a balanced 2-way partition, keeping all other sections the same. Note that this construction does not depend on any objective. However, as shown in the proof, each step strictly decreases $\Phi(\pi) = \sum_{\ell \in [K]} (X^{\pi}_{\ell})^2$, which takes finitely many values over $\Pi_{\mathcal{I}}$, so the procedure terminates at a balanced π_{bal}. By Claim 1, $\mathbf{X}^{\pi^{t+1}} \prec \mathbf{X}^{\pi^t}$ at every step, therefore $\mathbf{X}^{\pi} \succ \mathbf{X}^{\pi_{\mathrm{bal}}}$. □

THEOREM 1 **(Characterization of optimal fractional partitions)**. *Consider the optimization problem in (3.1) over $\Pi_{\mathcal{F}}$.*

*(i) If f is **edge-quasi-convex** on $P(\tilde{\mathbf{X}})$, then there exists an optimal **locally uniform** partition. If f is strictly edge-quasi-convex on $P(\tilde{\mathbf{X}})$, then every optimal fractional partition must be locally uniform.*

*(ii) If f is **Schur convex**, then the **assortative** partition $\widetilde{M}$ is optimal.*

*(iii) If f is **Schur concave**, then the **uniform** partition $\underline{M}$ is optimal.*

Proof. Parts (ii) and (iii) correspond to Propositions 2 and 3, respectively. For part (i), by Proposition 1, the summation vectors of feasible fractional partitions form the majorization polytope $P(\widetilde{\mathbf{X}})$. Since f is edge-quasi-convex on $P(\widetilde{\mathbf{X}})$, Proposition A.11 implies that the maximum of f over $P(\widetilde{\mathbf{X}})$ is attained at an extreme point. By Proposition A.13, every extreme point of $P(\widetilde{\mathbf{X}})$ corresponds to a locally uniform partition, which proves that there exists an optimal locally uniform partition. If f is strictly edge-quasi-convex, then every maximizer of f over $P(\widetilde{\mathbf{X}})$ is an extreme point, so every optimal fractional partition must be locally uniform. □

PROPOSITION 4 **(All locally balanced partitions are vertices of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$)**. *If $\pi \in \Pi_{\mathcal{I}}$ is locally balanced, then $\mathbf{X}^{\pi}$ is a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$.*

Proof. The proof is by contradiction. Let π be a locally balanced partition via consecutive division $\mathbf{b}$ with incidence set $I(\mathbf{b}) = \{k_1, \ldots, k_t\}$. Suppose for the sake of contradiction that $\mathbf{X}^{\pi}$ is not a vertex of $P_{\mathcal{I}}(\widetilde{X})$. Then there exist distinct feasible partitions $\pi^{(1)}, \ldots, \pi^{(r)} \in \Pi_{\mathcal{I}}$, each with $\mathbf{X}^{\pi^{(j)}} \neq \mathbf{X}^{\pi}$, and weights $\alpha_j > 0$ with $\sum_{j \in [r]} \alpha_j = 1$, such that

$$\mathbf{X}^{\pi} = \sum_{j \in [r]} \alpha_j \, \mathbf{X}^{\pi^{(j)}}. \tag{B.1}$$

Step 1: We verify that all $\pi^{(j)}$ satisfy the same binding constraints as π. For each $k_i \in I(\mathbf{b})$, condition (i) gives $\sum_{\ell \in [k_i]} X^{\pi}_{[\ell]} = \sum_{\ell \in [k_i]} \widetilde{X}_{[\ell]}$, which is the maximum achievable value of that partial sum over all feasible partitions. Since every feasible partition satisfies

$$\sum_{\ell \in [k_i]} X^{\pi^{(j)}}_{[\ell]} \leq \sum_{\ell \in [k_i]} \widetilde{X}_{[\ell]},$$

and by (B.1) the convex average of the LHS equals the maximum, we have:

$$\sum_{\ell \in [k_i]} X^{\pi^{(j)}}_{[\ell]} = \sum_{\ell \in [k_i]} \widetilde{X}_{[\ell]} \qquad \text{for all } j \in [r],\ k_i \in I(\mathbf{b}).$$

Step 2: Fix any block $b_i = \{k_{i-1}+1, \dots, k_i\}$. We show that within b_i, $\mathbf{X}_{b_i}^{\pi^{(j)}} = \mathbf{X}_{b_i}^{\pi}$ for all $j \in [r]$. If $|b_i| = 1$, the claim follows from Step 1 applied at k_{i-1} and k_i. Suppose $|b_i| \geq 2$.

We first explain why condition (ii) can be applied to the block vector of each $\pi^{(j)}$. If a partition binds at k_h, then its first k_h sections contain nk_h students and attain the largest possible sum, thus they must contain exactly the attributes $x_1, \dots, x_{nk_h}$. Applying this observation at both k_{i-1} and k_i, it follows that the sections in block b_i must contain exactly the attributes $x_{nk_{i-1}+1}, \dots, x_{nk_i}$. By Step 1, this holds for π and every $\pi^{(j)}$.

Next, fix $j \in [r]$. Since $\pi^{(j)}$ and π contain the same attributes in block b_i, we can construct a new partition $\widehat{\pi}$ as follows:

$$\widehat{\pi}_\ell = \begin{cases} \pi_\ell^{(j)}, & \ell \in b_i, \\ \pi_\ell, & \ell \notin b_i. \end{cases}$$

In other words, $\widehat{\pi}$ keeps the assignment of π outside b_i, but inside b_i it reproduces $\pi^{(j)}$. Consequently, $X_{[\ell]}^{\widehat{\pi}} = X_{[\ell]}^{\pi}$, for all $\ell \notin b_i$, and $\mathbf{X}_{b_i}^{\widehat{\pi}} = \mathbf{X}_{b_i}^{\pi^{(j)}}$.

Therefore we can substitute $\pi' = \hat{\pi}$ in condition (ii), i.e.,

$$\mathbf{X}_{b_i}^{\pi} \prec \mathbf{X}_{b_i}^{\widehat{\pi}} = \mathbf{X}_{b_i}^{\pi^{(j)}}.$$

By the definition of majorization, for every $q = 1, \dots, |b_i|$,

$$\sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi} \leq \sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi^{(j)}}. \tag{B.2}$$

Restricting (B.1) to the coordinates in b_i and applying (B.2) gives

$$\sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi} = \sum_{j \in [r]} \alpha_j \sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi^{(j)}} \geq \sum_{j \in [r]} \alpha_j \sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi}.$$

Since $\alpha_j > 0$, it must hold that

$$\sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi^{(j)}} = \sum_{\ell=k_{i-1}+1}^{k_{i-1}+q} X_{[\ell]}^{\pi},$$

for all $q = 1, \dots, |b_i|$ and $j \in [r]$. Equality of all the partial sums in turn implies equality coordinate by coordinate, i.e.,

$$\mathbf{X}_{b_i}^{\pi^{(j)}} = \mathbf{X}_{b_i}^{\pi} \qquad \text{for all } j \in [r].$$

Step 3: We obtain the contradiction. Steps 1 and 2 imply that $\mathbf{X}^{\pi^{(j)}} = \mathbf{X}^{\pi}$ for all $j \in [r]$, contradicting the initial assumption that $\mathbf{X}^{\pi^{(j)}} \neq \mathbf{X}^{\pi}$. This concludes the proof. □

PROPOSITION 5 **(Characterization of extreme points of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$).** *Every vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$ corresponds to a locally extremal partition.*

Proof. Let $\mathbf{X}^{\pi}$ be a vertex of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$ and fix a pair of sections $1 \leq r < s \leq K$. Note that $\pi \in \mathcal{T}_{rs}(\pi)$, so this set is nonempty, and given that $\Pi_{\mathcal{I}}$ is finite, the minimum and maximum of Δ_{rs} over $\mathcal{T}_{rs}(\pi)$ are attained.

Every feasible summation vector has the same total $\sum_{j=1}^{K} X_{[j]}^{\pi'} = T$. For $\pi' \in \mathcal{T}_{rs}(\pi)$, all coordinates outside $\{r, s\}$ agree with those of $\mathbf{X}^{\pi}$, so

$$X_{[r]}^{\pi'} + X_{[s]}^{\pi'} = T - \sum_{j \notin \{r,s\}} X_{[j]}^{\pi} = X_{[r]}^{\pi} + X_{[s]}^{\pi} = S_{rs}.$$

Thus, $X^{\pi'}_{[r]}$ and $X^{\pi'}_{[s]}$ have fixed sum S_{rs} and difference $\Delta_{rs}(\pi')$, therefore we can write

$$X^{\pi'}_{[r]} = \frac{S_{rs} + \Delta_{rs}(\pi')}{2}, \qquad X^{\pi'}_{[s]} = \frac{S_{rs} - \Delta_{rs}(\pi')}{2},$$

i.e., both coordinates are linear in $\Delta_{rs}(\pi')$.

Next we show that the gap must be extremal. Suppose, for the sake of contradiction, that

$$\min_{\pi' \in \mathcal{T}_{rs}(\pi)} \Delta_{rs}(\pi') < \Delta_{rs}(\pi) < \max_{\pi' \in \mathcal{T}_{rs}(\pi)} \Delta_{rs}(\pi'),$$

and let $\pi^-, \pi^+ \in \mathcal{T}_{rs}(\pi)$ attain the minimum and the maximum, respectively. Set

$$\lambda = \frac{\Delta_{rs}(\pi^+) - \Delta_{rs}(\pi)}{\Delta_{rs}(\pi^+) - \Delta_{rs}(\pi^-)}.$$

The denominator is positive and the strict inequalities above give $\lambda \in (0,1)$ and $\Delta_{rs}(\pi) = \lambda \Delta_{rs}(\pi^-) + (1-\lambda)\Delta_{rs}(\pi^+)$.

We claim that $\mathbf{X}^{\pi} = \lambda \mathbf{X}^{\pi^-} + (1-\lambda)\mathbf{X}^{\pi^+}$. For $j \notin \{r,s\}$, the identity holds trivially by definition. For $j = r$, we have that

$$\lambda X^{\pi^-}_{[r]} + (1-\lambda) X^{\pi^+}_{[r]} = \frac{S_{rs} + \lambda \Delta_{rs}(\pi^-) + (1-\lambda)\Delta_{rs}(\pi^+)}{2} = \frac{S_{rs} + \Delta_{rs}(\pi)}{2} = X^{\pi}_{[r]},$$

and the case $j = s$ is symmetric with $-\Delta_{rs}(\pi)$. Note that $\lambda \in (0,1)$, and since $\Delta_{rs}(\pi^-) < \Delta_{rs}(\pi) < \Delta_{rs}(\pi^+)$. the vectors $\mathbf{X}^{\pi^-}$ and $\mathbf{X}^{\pi^+}$ are distinct from each other and from $\mathbf{X}^{\pi}$. Thus, $\mathbf{X}^{\pi}$ is a nontrivial convex combination of points of $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$, contradicting that it is a vertex.

Therefore, $\Delta_{rs}(\pi)$ attains its minimum or its maximum over $\mathcal{T}_{rs}(\pi)$, and since the pair $r < s$ was arbitrary, this holds for every pair thus π is locally extremal. □

THEOREM 2 **(Characterization of optimal integral partitions)**. *Consider the optimization problem in (3.2) over $\Pi_{\mathcal{I}}$.*

*(i) If f is **edge-quasi-convex** on $P_I(\tilde{\mathbf{X}})$, then there exists an optimal **locally extremal** partition. If f is strictly edge-quasi-convex, then every optimal integral partition must be locally extremal.*

*(ii) If f is **Schur convex**, then the **assortative** partition $\widetilde{\pi}$ is optimal.*

*(iii) If f is **Schur concave**, then there exists an optimal **balanced** partition.*

Proof. Parts (ii) and (iii) are established in Propositions 2 and 3, respectively. For part (i), since f is edge-quasi-convex, Proposition A.11 implies that the maximum of f over $P_{\mathcal{I}}(\widetilde{\mathbf{X}})$ is attained at a vertex. By Proposition 5, every vertex corresponds to a locally extremal partition. Therefore there exists an optimal integral partition that is locally extremal. If f is strictly edge-quasi-convex, every maximizer is a vertex, so it must in turn be locally extremal. □

B.2. Proofs from Section 4

B.2.1. Proofs from Subsection 4.2

LEMMA B.5. *In a partition π, consider a section π_j at equilibrium. In the LIM model, the section effort equals*[8]

$$e(\pi_j) = \frac{\gamma}{1-\theta_1} \sum_{i \in \pi_j} x_i$$

and the section utility equals

$$s(\pi_j) = \frac{\gamma^2}{2} \left(\frac{(n-1)^2}{(n-1+\theta_1)^2} \sum_{i \in \pi_j} x_i^2 + \frac{\theta_1 \big(2(n-1)(1-\theta_1) + n\theta_1\big)}{(1-\theta_1)^2 (n-1+\theta_1)^2} \left(\sum_{i \in \pi_j} x_i \right)^2 \right) + \frac{n}{2} \cdot \frac{(n-1)^2(1-\theta_1)^2 + 2\theta_1(n-1)(1-\theta_1) + n\theta_1^2}{(1-\theta_1)^2(n-1+\theta_1)^2} \sigma^2.$$

Proof. For the section effort, taking the expectation of (4.8) and using $\mathbb{E}[\epsilon_i \mid \mathbf{x}] = 0$ gives us the result for $e(\pi_j)$. For section utility, we replace (4.7) into (4.6):

$$\begin{aligned} U_i^S(y_i, \mathbf{y}_{-i}) &= (\gamma x_i + \epsilon_i)(\gamma x_i + \theta_1 \overline{y}_{-i} + \epsilon_i) - 0.5(\gamma x_i + \theta_1 \overline{y}_{-i} + \epsilon_i)^2 + \theta_1 (\gamma x_i + \theta_1 \overline{y}_{-i} + \epsilon_i) \overline{y}_{-i} \\ &= (\gamma x_i + \theta_1 \overline{y}_{-i} + \epsilon_i)(\gamma x_i + \epsilon_i - 0.5(\gamma x_i + \theta_1 \overline{y}_{-i} + \epsilon_i) + \theta_1 \overline{y}_{-i}) \\ &= 0.5 y_i^2 \end{aligned}$$

Summing over i and taking the expectation, we get:

$$s(\pi_j) = \mathbb{E}\left[0.5 \sum_{i \in \pi_j} y_i^2 \right] = 0.5 \sum_{i \in \pi_j} \mathbb{E}\left[(A\, q_i + B\, Q_j)^2 \right], \tag{B.3}$$

where $Q_j = \sum_{k \in \pi_j} q_k$ and, by (4.8), $A = \frac{n-1}{n-1+\theta_1}$ and $B = \frac{\theta_1}{(1-\theta_1)(n-1+\theta_1)}$. Using $\mathbb{E}[\epsilon_i \mid \mathbf{x}] = 0$, $\mathbb{E}[\epsilon_i^2 \mid \mathbf{x}] = \sigma^2$ and $\mathbb{E}[\epsilon_i \epsilon_k \mid \mathbf{x}] = 0$ for $i \neq k$, we have that

$$\mathbb{E}[q_i^2] = \gamma^2 x_i^2 + \sigma^2, \quad \mathbb{E}[q_i Q_j] = \gamma^2 x_i \sum_{i \in \pi_j} x_i + \sigma^2, \quad \mathbb{E}[Q_j^2] = \gamma^2 \left(\sum_{i \in \pi_j} x_i \right)^2 + n\sigma^2.$$

Therefore,

$$\begin{aligned} &0.5 \sum_{i \in \pi_j} \mathbb{E}\left[(A\, q_i + B\, Q_j)^2 \right] \\ &= 0.5\gamma^2 \left(A^2 \sum_{i \in \pi_j} x_i^2 + \big(2AB + nB^2\big) \left(\sum_{i \in \pi_j} x_i \right)^2 \right) + \frac{n}{2} \left(A^2 + 2AB + nB^2\right) \sigma^2. \end{aligned}$$

Putting everything together gives us the result for $s(\pi_j)$. □

PROPOSITION 6 **(LIM, Performance).** *Fix $\mathbf{w} \in \mathbb{R}_+^K$. Under the LIM model, there exists a locally uniform partition that attains the optimal total weighted effort over $\Pi_{\mathcal{F}}$. Additionally, if $\mathbf{w} \in \mathbb{R}_+^K$ is single-peaked at t^*, then the total weighted effort is maximized for a $\underline{t}$-upper-uniform partition, where $\underline{t} \geq t^*$. If $t^* = 1$, then the assortative partition is optimal. If $t^* = K$, then the uniform partition is optimal.*

[8] For the sake of readability, we have chosen to present our statements in this section using integral partitions. Expressing these results for fractional partitions is straightforward: in Lemma B.5, for a given $M \in \Pi_{\mathcal{F}}$, the sum $\sum_{i \in \pi_j} x_i$ found in both $e(\pi_j)$ and $s(\pi_j)$ is replaced by $\sum_{i \in [N]} \mu_{ij} x_i$.

Proof. By Lemma B.5, the total weighted effort $E(\pi)=\sum_{j\in[K]} w_j e_n(\pi_{[j]})$, is linear in each $X_j^\pi=\sum_{i\in\pi_j} x_i$. By Theorem 1, it follows that there exists a locally uniform partition that attains the optimal total weighted effort.

If $t^*=1$ then $w_1\geq\ldots\geq w_K$. In this case, f is Schur convex and the result follows from Proposition 2. If $t^*=K$ then $w_1\leq\ldots\leq w_K$ and f is Schur concave thus Proposition 3 applies.

For any other $1<t^*<K$, we will show that $E(\pi)$ is maximized for a $\underline{t}$-upper-uniform partition, where $\underline{t}>t^*$.

First, using an "ironing" argument à la Myerson (1981), we prove the following claim:

Claim 2. *There exists a $\underline{t}\geq t^*$ such that for any $\mathbf{X}\in P(\widetilde{X})$,*

$$\sum_{j\in[K]}\hat{w}_j X_j\geq\sum_{j\in[K]} w_j X_j \tag{B.4}$$

where

$$\hat{\mathbf{w}}:=\left(\frac{1}{\underline{t}}\sum_{j\leq\underline{t}} w_j\mathbf{1}_{\underline{t}}, w_{\underline{t}+1},\ldots,w_K\right) \tag{B.5}$$

and (B.4) holds with equality whenever $X_j=c$ for some constant $c\in\mathbb{R}_+$ and all $j\leq\underline{t}$.

Proof of Claim 2. Analogously to L-curves, we define the step function

$$g(x)=\sum_{j\in[K]} w_{K-j+1}\,\mathbf{1}\{j-1\leq x\leq j\}, 0\leq x\leq K.$$

Let $G(x)=\int_0^x g(q)\,\mathrm{d}q$ and find the convex hull of G, which we will call $H(x)$. Then, define $\hat{\mathbf{w}}$ as follows: Let $h(x)$ be the derivative of $H(x)$, wherever defined, and extend it to $[0,K]$ by right continuity.

By construction of g and the single-peakedness of $\mathbf{w}$, it follows that $g(x)$ is unimodal. Thus, $G(x)$ is convex-concave, meaning that there exists a $\underline{x}$ such that G is convex for $x\in[0,\underline{x}]$ and concave for $x\in[\underline{x},K]$. Consequently, $H(x)$, as the convex hull of $G(x)$, coincides with $G(x)$ for $x\in[0,K-\underline{t}]$, while for $x\in[K-\underline{t},K]$ is linear and its slope is equal to $\frac{1}{\underline{t}}\sum_{j\leq\underline{t}} w_j$. Next, we construct $\hat{\mathbf{w}}$. For $j\in[K]$, set

$$\hat{w}_{K-j+1}=h(j-0.5).$$

By the previous arguments, it follows that $\hat{\mathbf{w}}$ takes the form in (B.5).

Thus, we get that for $\mathbf{X}\in P(\widetilde{\mathbf{X}})$,

$$\sum_{j\in[K]}\hat{w}_j X_j=\left(\frac{1}{\underline{t}}\sum_{j\leq\underline{t}} w_j\right)\left(\sum_{1\leq j\leq\underline{t}} X_j\right)+\sum_{\underline{t}<j\leq K} w_j X_j$$

Recall that H is linear on each $[K-\underline{t},K]$ with slope $\frac{1}{\underline{t}}\sum_{j\leq\underline{t}} w_j$. Since $H(x)\leq G(x)$ with $H(K-\underline{t})=G(K-\underline{t})$ and $G(K-\underline{t}+m)-G(K-\underline{t})=\sum_{j=\underline{t}-m+1}^{\underline{t}} w_j$, we obtain

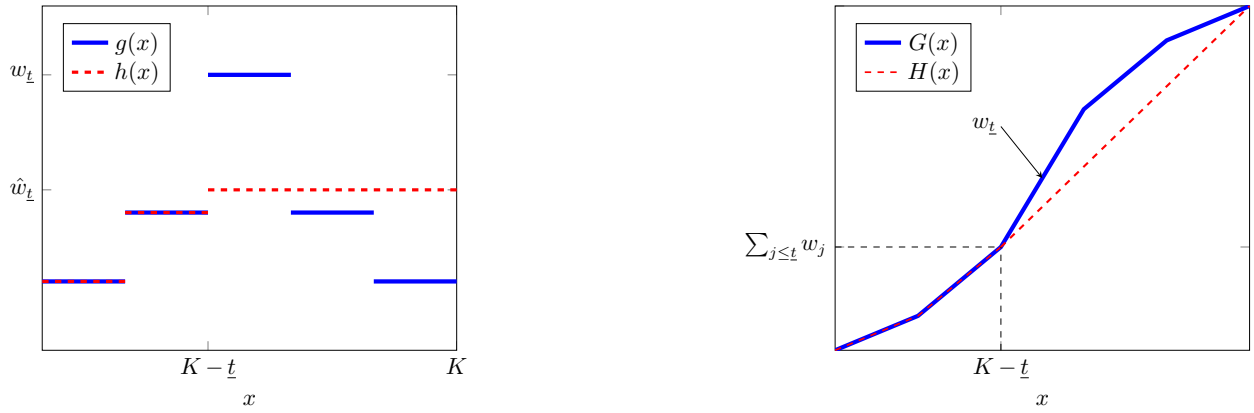
$$\sum_{j\in[k]} w_j\ \leq\ \frac{k}{\underline{t}}\sum_{j\leq\underline{t}} w_j\qquad\text{for every } k\leq\underline{t}.$$

Since $X_j\geq 0$, for all $j\in[\underline{t}]$, by summing by parts and then rearranging, we get that

$$\sum_{1\leq j\leq\underline{t}} w_j X_j\leq\frac{1}{\underline{t}}\left(\sum_{1\leq j\leq\underline{t}} w_j\right)\left(\sum_{1\leq j\leq\underline{t}} X_j\right). \tag{B.6}$$

If $X_j=c$ for all $j\leq\underline{t}$, (B.6) holds with equality. This concludes the proof of the claim. □

Figure B.6 "Ironing" of $\mathbf{w} \in \mathbb{R}_+^K$ in Claim 2.



Notes: **Left**: With the solid blue line we depict $g(x)$; with the dashed red line we depict its ironed counterpart $h(x)$. The value of j-th step of g corresponds to w_{K-j+1}; observe that the vector $\mathbf{w}$ is single-peaked at $t^* = \underline{t}$. Since h is "ironed" only for $x \in (K-\underline{t}, K]$, h coincides with g for $x \in [0, K-\underline{t}]$, but equals the local average $\hat{w}_{\underline{t}} = \frac{1}{\underline{t}} \sum_{j\le\underline{t}} w_j$ for $x \in (K-\underline{t}, K]$. **Right**: With the solid blue line we depict $G(x)$; with the dashed red line we depict its convex hull $H(x)$. Observe that, due to the form of g, $G(x) = \int_0^x g(q)dq$ consists of K line segments; the derivative of the j-th line segment equals w_{K-j+1}. The function G is also convex for $x \in [0, K-\underline{t}]$ but concave for $x \in (K-\underline{t}, K]$. Therefore, H coincides with G for $x \in [0, K-\underline{t}]$, but is linear for $x \in (K-\underline{t}, K]$.

By (B.5), $\sum_{j\in[K]} \hat{w}_j X_j$ is Schur convex since $\hat{w}_1 \ge \ldots \ge \hat{w}_{\underline{t}} \ge \ldots \ge \hat{w}_K$. Therefore, by Proposition 2, the assortative partition $\widetilde{\mathbf{X}}$ is optimal. However, we will show that there exists a $\underline{t}$-upper-uniform partition $\mathbf{X}^*$ which is also optimal for $\sum_{j\in[K]} \hat{w}_j X_j$.

Indeed, consider the $\underline{t}$-upper-uniform partition $\mathbf{X}^*$ with $X_j^* = \frac{1}{\underline{t}} \sum_{j\le\underline{t}} \widetilde{X}_j$ and $X_j^* = \widetilde{X}_j$ for $j > \underline{t}$. (Clearly, $\mathbf{X}^* \prec \widetilde{\mathbf{X}}$ thus $\mathbf{X}^* \in P(\widetilde{\mathbf{X}})$.) Since $\hat{w}_j = \frac{1}{\underline{t}} \sum_{1\le k\le\underline{t}} w_k$, for all $j \le \underline{t}$ and by the construction of $\mathbf{X}^*$, it follows that

$$\sum_{j\in[K]} \hat{w}_j \widetilde{X}_j = \sum_{j\in[K]} \hat{w}_j X_j^*,$$

therefore $\mathbf{X}^*$ is also optimal.

For the last step, observe that Claim 2 guarantees that $\mathbf{X}^*$ also maximizes $\sum_{j\in[K]} w_j X_j$, since (B.4) holds with equality for $\mathbf{X}^*$. Consequently, the $\underline{t}$-upper-uniform partition $\mathbf{X}^*$ maximizes the total weighted effort. □

COROLLARY 3. *Let $w_j = 1$ for all $j \in [K]$. Then any feasible partition attains the optimal total effort under the LIM model.*

Proof. By Lemma B.5, $e(\pi_j)$ is proportional to $X_j^\pi = \sum_{i\in\pi_j} x_i$. Therefore,

$$E(\pi) = \sum_{j\in[K]} e(\pi_j) \propto \sum_{j\in[K]} X_j^\pi = \sum_{i\in[N]} x_i,$$

which is the same under any partition $\pi \in \Pi_{\mathcal{I}}$ and any $M \in \Pi_{\mathcal{F}}$. □

LEMMA 1. *Under the LIM model and for any partition $\pi \in \Pi_{\mathcal{I}}$, $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\widetilde{\mathbf{y}}]$. Furthermore, there is no partition $\pi \in \Pi_{\mathcal{I}}$ with $\mathbf{X}^\pi \neq \underline{\mathbf{X}}$ such that $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\underline{\mathbf{y}}]$.*

Proof. We break the proof into steps.

Step 1: We derive $\mathbb{E}[\widetilde{y}_{[i]}], \mathbb{E}[\underline{y}_{[i]}]$, *and* $\mathbb{E}[y_{[i]}]$. By (4.7), for any effort vector y^π, $\pi \in \Pi_{\mathcal{I}}$, where $[i] \in \pi_j$ for some $j \in [K]$, we have:

$$\mathbb{E}[y_{[i]}] = \frac{\gamma(n-1)}{n-1+\theta_1} x_{[i]} + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} X^\pi_j.$$

For the assortative partition, $x_{[i]} = x_i$ and $j = \lceil i/n \rceil$, we have that

$$\mathbb{E}[\widetilde{y}_{[i]}] = \mathbb{E}[\widetilde{y}_i] = \frac{\gamma(n-1)}{n-1+\theta_1} x_i + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \widetilde{X}_{\lceil i/n \rceil}.$$

Similarly, for the uniform partition's $\underline{y}$, we find that

$$\begin{aligned}
\mathbb{E}[\underline{y}_{[i]}] &= \frac{\gamma(n-1)}{n-1+\theta_1} x_{[i]} + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \left(\sum_{j\in[K]} \underline{\mu}_{ij} \sum_{k\in[N]} \underline{\mu}_{kj} x_k \right) \\
&= \frac{\gamma(n-1)}{n-1+\theta_1} x_{[i]} + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \left(\sum_{j\in[K]} \underline{\mu}_{ij} \underline{X}_j \right) \\
&= \frac{\gamma(n-1)}{n-1+\theta_1} x_{[i]} + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \frac{\sum_{i\in[N]} x_i}{K},
\end{aligned}$$

where the last step follows from the fact that for the uniform partition, $\underline{X}_j = \frac{1}{K} \sum_{i\in[N]} x_i$ and $\underline{\mu}_{ij} = 1/K$ for all $i \in [N]$ and $j \in [K]$. Since $x_1 \geq \ldots \geq x_N$,

$$\mathbb{E}[\underline{y}_{[i]}] = \mathbb{E}[\underline{y}_i] = \frac{\gamma(n-1)}{n-1+\theta_1} x_i + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \frac{\sum_{i\in[N]} x_i}{K} = \frac{\gamma(n-1)}{n-1+\theta_1} x_i + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \underline{X}_{\lceil i/n \rceil}. \tag{B.7}$$

Step 2: We show that $\mathbb{E}[\widetilde{\mathbf{y}}] \succ \mathbb{E}[\mathbf{y}^\pi]$ *for any* $\pi \in \Pi_{\mathcal{I}}$. By Corollary 3, $\sum_{i\in[N]} \mathbb{E}[y^\pi_{[i]}] = \sum_{i\in[N]} \mathbb{E}[\widetilde{y}_{[i]}]$.

It remains to show that, for any $k \in [N]$,

$$\sum_{j\in[k]} \mathbb{E}[\widetilde{y}_j] \geq \sum_{j\in[k]} \mathbb{E}[y^\pi_{[j]}].$$

We have that

$$\sum_{j\in[k]} \mathbb{E}[\widetilde{y}_j] = \sum_{j\in[k]} \left(\frac{\gamma(n-1)}{n-1+\theta_1} x_j + \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)} \widetilde{X}_{\lceil j/n \rceil} \right). \tag{B.8}$$

Consider the index sequence of $y^\pi_{[1]} \geq \ldots \geq y^\pi_{[N]}$, namely $l_1, \ldots, l_N$. Then, for any rearrangement of the decreasing vector $\mathbf{x}$, including $x_{l_1}, \ldots, x_{l_k}$, we have:

$$\sum_{j\in[k]} x_j \geq \sum_{j\in[k]} x_{l_j}.$$

For every $j \in [k]$, let $m_j \in [K]$ be such that $l_j \in \pi_{m_j}$. Since $\widetilde{\mathbf{X}} \succ \mathbf{X}^\pi$, the majorization property implies that for any $k \in [N]$,

$$\sum_{j\in[k]} \widetilde{X}_{\lceil j/n \rceil} \geq \sum_{j\in[k]} X^\pi_{m_j}.$$

Summing together the last inequalities with the appropriate weights, we get (B.8).

Step 3: We show that, for $\theta_1 > 0$, *there is no* $\pi \in \Pi_{\mathcal{I}}$ *with* $\mathbf{X}^\pi \neq \underline{\mathbf{X}}$ *such that* $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\underline{\mathbf{y}}]$. Let $A = \frac{\gamma(n-1)}{n-1+\theta_1}$,

$B = \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)}$. By Step 1, $\mathbb{E}[y_i^\pi] = Ax_i + BX_{s(i)}^\pi$, where $s(i)$ denotes the section of student i under π, and $\mathbb{E}[\underline{y}_i] = Ax_i + B\underline{X}$. Grouping students by section gives

$$\sum_{i\in[N]} X_{s(i)}^\pi = n\sum_{j\in[K]} X_j^\pi = N\underline{X}, \qquad \sum_{i\in[N]} x_i X_{s(i)}^\pi = \sum_{j\in[K]} X_j^\pi \sum_{i\in\pi_j} x_i = \sum_{j\in[K]} (X_j^\pi)^2.$$

Consequently,

$$\sum_{i\in[N]} \mathbb{E}[y_i^\pi]^2 - \sum_{i\in[N]} \mathbb{E}[\underline{y}_i]^2 = \left(2AB + nB^2\right) \sum_{j\in[K]} \left(X_j^\pi - \underline{X}\right)^2 > 0 \tag{B.9}$$

for $\theta_1 > 0$ and $\mathbf{X}^\pi \neq \underline{\mathbf{X}}$. Suppose, for the sake of contradiction, that such a π satisfies $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\underline{\mathbf{y}}]$. Applying the Karamata inequality (Proposition A.12) with the convex function $f(x) = x^2$ yields $\sum_i \mathbb{E}[y_i^\pi]^2 \leq \sum_i \mathbb{E}[\underline{y}_i]^2$, contradicting (B.9). □

PROPOSITION 7 **(LIM, Inequality)**. *Under the LIM model, the following hold:*

(i) There exists threshold $\underline{\theta} \in (0,1)$ such that if peer effects are strong ($\theta_1 > \underline{\theta}$), then every partition minimizing the Gini index of effort over $\Pi_{\mathcal{I}}$ is balanced.

(ii) The assortative partition achieves the maximum Gini index of effort over $\Pi_{\mathcal{I}}$. In particular, if $x_1 > \ldots > x_N$, the assortative partition is the only partition with this property.

(iii) There exist cohorts where no (fractional) partition in $\Pi_{\mathcal{F}}$ can achieve perfect equality in effort.

Proof. We prove each part separately:

Proof of Part (i). As we showed in the proof of Lemma B.6, minimizing the Gini index of effort over $\Pi_{\mathcal{I}}$ is equivalent to minimizing its numerator $\Delta(\pi)$ in (B.13). Furthermore, the expected efforts satisfy $\mathbb{E}[y_i^\pi] = Ax_i + BX_{s(i)}^\pi$ with $A, B > 0$ independent of π, and B/A is increasing in θ_1 (note that these properties do not rely on the assumption $x_1 > \ldots > x_N$).

Set $\Gamma(\pi) = \sum_{j,k\in[K]} \left|X_j^\pi - X_k^\pi\right|$. We get that

$$\begin{aligned}
\left|\Delta(\pi) - Bn^2\Gamma(\pi)\right| &= \left| \sum_{i,i'\in[N]} \left(\left|\mathbb{E}[y_i^\pi] - \mathbb{E}[y_{i'}^\pi]\right| - B\left|X_{s(i)}^\pi - X_{s(i')}^\pi\right| \right) \right| \\
&\leq \sum_{i,i'\in[N]} \left| \left|\mathbb{E}[y_i^\pi] - \mathbb{E}[y_{i'}^\pi]\right| - B\left|X_{s(i)}^\pi - X_{s(i')}^\pi\right| \right| \\
&= \sum_{i,i'\in[N]} \left| \left|A(x_i - x_{i'}) + B(X_{s(i)}^\pi - X_{s(i')}^\pi)\right| - B\left|X_{s(i)}^\pi - X_{s(i')}^\pi\right| \right| \\
&\leq A \sum_{i,i'\in[N]} \left|x_i - x_{i'}\right|,
\end{aligned} \tag{B.10}$$

where the last step follows from the triangle inequality.

We first show that every partition attaining $\Gamma^* = \min_{\pi\in\Pi_{\mathcal{I}}} \Gamma(\pi)$ is balanced. As in (B.14),

$$\Gamma(\pi) = 2\sum_{j\in[K]} (K + 1 - 2j)\, X_{[j]}^\pi,$$

thus $-\gamma(\mathbf{X}^\pi) = -\Gamma(\pi)$ is strictly Schur concave. Thus, Proposition 3 applies, implying that every partition attaining Γ^* is balanced.

It remains to show that, for θ_1 large enough, every minimizer of Δ attains Γ^*. If Γ is constant on $\Pi_{\mathcal{I}}$, then every partition attains Γ^* and the claim holds for any $\underline{\theta} \in (0,1)$. Otherwise, the attributes are not all equal, so $\sum_{i,i' \in [N]} |x_i - x_{i'}| > 0$ and, since $\Pi_{\mathcal{I}}$ is finite, $\varepsilon = \min\{\Gamma(\pi) - \Gamma^* : \pi \in \Pi_{\mathcal{I}},\ \Gamma(\pi) > \Gamma^*\} > 0$.

Set

$$\underline{\theta} = \frac{2(n-1)\sum_{i,i' \in [N]} |x_i - x_{i'}|}{n^2\varepsilon + 2(n-1)\sum_{i,i' \in [N]} |x_i - x_{i'}|} \in (0,1).$$

Since $B/A = \theta_1/\big((1-\theta_1)(n-1)\big)$ is strictly increasing in θ_1, the condition $\theta_1 > \underline{\theta}$ is equivalent to

$$Bn^2\varepsilon > 2A \sum_{i,i' \in [N]} |x_i - x_{i'}| \tag{B.11}$$

Let π^* minimize $\Delta(\pi)$ while $\Gamma(\pi^*) > \Gamma^*$. Consider π' that attains Γ^*. Then

$$\begin{aligned} Bn^2(\Gamma^* + \varepsilon) - A \sum_{i,i' \in [N]} |x_i - x_{i'}| &\leq Bn^2\Gamma(\pi^*) - A \sum_{i,i' \in [N]} |x_i - x_{i'}| \\ &\leq \Delta(\pi^*) \\ &\leq \Delta(\pi') \\ &\leq Bn^2\Gamma(\pi') + A \sum_{i,i' \in [N]} |x_i - x_{i'}| \\ &= Bn^2\Gamma^* + A \sum_{i,i' \in [N]} |x_i - x_{i'}|, \end{aligned}$$

where the first inequality holds because $\Gamma(\pi^*) > \Gamma^*$ implies $\Gamma(\pi^*) \geq \Gamma^* + \varepsilon$ by the definition of ε, the second and fourth inequalities follow from (B.10), the third inequality holds due to the optimality of π^*, and the equality follows from the definition of π'. Consequently, $Bn^2\varepsilon \leq 2A\sum_{i,i' \in [N]} |x_i - x_{i'}|$, which contradicts that $\theta_1 > \underline{\theta}$. Therefore, $\Gamma(\pi^*) = \Gamma^*$, and by the previous argument, it follows that π^* is balanced. □

Proof of Part (ii). By Lemma 1, $\mathbb{E}[\mathbf{y}^\pi] \prec \mathbb{E}[\widetilde{\mathbf{y}}]$ for every $\pi \in \Pi_{\mathcal{I}}$, and the Gini index is strictly Schur convex (Dalton 1920), therefore $\widetilde{\pi}$ attains the maximum.

Suppose $x_1 > \cdots > x_N$ and let π also be optimal. Strict Schur convexity forces $\mathbb{E}[\mathbf{y}^\pi]$ to be a permutation of $\mathbb{E}[\widetilde{\mathbf{y}}]$. As in Step 3 of the proof of Lemma 1,

$$\sum_{i \in [N]} \mathbb{E}[y_i^\pi]^2 = A^2 \sum_{i \in [N]} x_i^2 + \big(2AB + nB^2\big) \sum_{j \in [K]} \big(X_j^\pi\big)^2,$$

and $2AB + nB^2 > 0$, so $\sum_j (X_j^\pi)^2 = \sum_j (\widetilde{X}_j)^2$. Since $\mathbf{X}^\pi \prec \widetilde{\mathbf{X}}$ by Proposition 1 and $\sum_j (X_j)^2$ is strictly Schur convex, $\mathbf{X}^\pi$ must be a permutation of $\widetilde{\mathbf{X}}$. Finally, $x_1 + \cdots + x_n$ is the largest attainable section sum and for $x_1 > \ldots > x_N$, it is attained only by the section $\{1, \ldots, n\}$. Repeating the argument on the remaining students gives $\pi = \widetilde{\pi}$ up to a relabeling of the sections.

Proof of Part (iii). We show that, in the general case, no (fractional) partition can achieve perfect equality in effort. For this we construct a counterexample. Consider a cohort of $N = 4$ students, $x_1 = x_{1'} < x_2 = x_{2'}$, to be partitioned to two sections ($K = 2$). Let $M \in \Pi_{\mathcal{F}}$ be a fractional partition that achieves equality in effort. Then, the expected effort of student i is equal to the expected effort of student i', i.e.,

$$\begin{aligned} &\gamma x_i + \theta_1\gamma\left(\mu_{i1}\mu_{i'1}x_{i'} + \mu_{i1}(\mu_{j1}x_j + \mu_{j'1}x_{j'}) + (1-\mu_{i1})(1-\mu_{i'1})x_{i'} + (1-\mu_{i1})((1-\mu_{j1})x_j + (1-\mu_{j'1})x_{j'})\right) \\ =&\gamma x_{i'} + \theta_1\gamma\left(\mu_{i'1}\mu_{i1}x_i + \mu_{i'1}(\mu_{j1}x_j + \mu_{j'1}x_{j'}) + (1-\mu_{i'1})(1-\mu_{i1})x_i + (1-\mu_{i'1})((1-\mu_{j1})x_j + (1-\mu_{j'1})x_{j'})\right), \end{aligned}$$

which simplifies to

$$(\mu_{i1} - \mu_{i'1})(\mu_{j1} + \mu_{j'1} - 1) = 0.$$

Similarly, we find that

$$(\mu_{j1} - \mu_{j'1})(\mu_{i1} + \mu_{i'1} - 1) = 0.$$

Since M is feasible, $\mu_{i1} + \mu_{i'1} + \mu_{j1} + \mu_{j'1} = n = 2$, so $\mu_{i1} + \mu_{i'1} = 1$ if and only if $\mu_{j1} + \mu_{j'1} = 1$. We have two cases.

Case 1: $\mu_{i1} + \mu_{i'1} \neq 1$. The two equations above give $\mu_{i1} = \mu_{i'1}$ and $\mu_{j1} = \mu_{j'1}$. Writing $a = \mu_{i1} = \mu_{i'1}$, feasibility gives $\mu_{j1} = \mu_{j'1} = 1 - a$, and simple algebra yields

$$\mathbb{E}[y_i] - \mathbb{E}[y_j] = \gamma(x_i - x_j)\left(1 + \theta_1(6a^2 - 6a + 1)\right).$$

Since $6a^2 - 6a + 1 \geq -1/2$ for $a \in [0,1]$, the second term is at least $1 - \theta_1/2 > 0$, and $x_i \neq x_j$. Thus, equality in effort fails.

Case 2: $\mu_{i1} + \mu_{i'1} = \mu_{j1} + \mu_{j'1} = 1$. The expected effort of student i is equal to the expected effort of student j, i.e.,

$$\begin{aligned}&\gamma x_i + \theta_1\gamma\left(\mu_{i1}\mu_{i'1}x_{i'} + \mu_{i1}(\mu_{j1}x_j + \mu_{j'1}x_{j'}) + (1-\mu_{i1})(1-\mu_{i'1})x_{i'} + (1-\mu_{i1})((1-\mu_{j1})x_j + (1-\mu_{j'1})x_{j'})\right)\\ =&\gamma x_j + \theta_1\gamma\left(\mu_{j1}\mu_{j'1}x_{j'} + \mu_{j1}(\mu_{i1}x_i + \mu_{i'1}x_{i'}) + (1-\mu_{j1})(1-\mu_{j'1})x_{j'} + (1-\mu_{j1})((1-\mu_{i1})x_i + (1-\mu_{i'1})x_{i'})\right),\end{aligned}$$

which given the above equations simplifies to

$$\gamma x_i + \theta_1\gamma(2\mu_{i1}(1-\mu_{i1})x_i + x_j) = \gamma x_j + \theta_1\gamma(2\mu_{j1}(1-\mu_{j1})x_j + x_i),$$

or

$$\frac{x_i}{x_j} = \frac{1 - \theta_1 + 2\theta_1\mu_{j1}(1-\mu_{j1})}{1 - \theta_1 + 2\theta_1\mu_{i1}(1-\mu_{i1})}.$$

Note that the RHS is minimized for $\mu_{j1} \in \{0,1\}$, $\mu_{i1} = 1/2$ thus is lower-bounded by $\frac{1-\theta_1}{1-0.5\theta_1}$, where $\theta_1 \in (0,1)$ is fixed. However the LHS can grow arbitrary small. Therefore, whenever $x_i/x_j < \frac{1-\theta_1}{1-0.5\theta_1}$, the above system of equations is infeasible and no such $M \in \Pi_{\mathcal{F}}$ exists. □

LEMMA B.6. *For a partition $\pi \in \Pi_{\mathcal{I}}$, define the total attribute spread between peers, weighted by their distance in rank, as*

$$W(\pi) = \sum_{j \in [K]} \sum_{i,i' \in \pi_j, i < i'} (i' - i)(x_i - x_{i'}).$$

Under the LIM model, fix $x_1 > x_2 > \ldots > x_N$. There exists $\bar{\theta} \in (0,1)$, such that for all $\theta_1 \in (0, \bar{\theta})$ and all $\pi, \pi' \in \Pi_{\mathcal{I}}$, $G_E(\mathbf{y}^{\pi}) \leq G_E(\mathbf{y}^{\pi'})$ if and only if $W(\pi) \geq W(\pi')$.

Proof. The expected equilibrium effort of student $i \in \pi_j$ under any partition $\pi \in \Pi_{\mathcal{I}}$ is

$$\mathbb{E}[y_i^{\pi}] = Ax_i + BX_j^{\pi}, \qquad \text{where} \quad A = \frac{\gamma(n-1)}{n - 1 + \theta_1}, \qquad B = \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)}, \tag{B.12}$$

and $A, B > 0$ for $\theta_1 \in (0,1)$. By Corollary 3, the total effort is the same under every partition in $\Pi_{\mathcal{I}}$, so we can focus only on the numerator of the Gini index:

$$\Delta(\pi) = \sum_{i \in [N]} \sum_{j \in [N]} \left|\mathbb{E}[y_i^{\pi}] - \mathbb{E}[y_j^{\pi}]\right|. \tag{B.13}$$

We show that $\Delta(\pi)$ is a decreasing linear function of $W(\pi)$. Let $\delta = \min_{i\in[N-1]}(x_i - x_{i+1}) > 0$ and $R = \widetilde{X}_{[1]} - \widetilde{X}_{[K]}$. Note that $|X^\pi_j - X^\pi_k| \le R$ for all $\pi \in \Pi_{\mathcal{I}}$ and $j,k \in [K]$. Set

$$\bar{\theta} = \frac{\delta(n-1)}{R + \delta(n-1)} \in (0,1).$$

Then, for any $\pi \in \Pi_{\mathcal{I}}$ and $i < i'$, (B.12) gives

$$\mathbb{E}[y^\pi_i] - \mathbb{E}[y^\pi_{i'}] = A(x_i - x_{i'}) + B\big(X^\pi_{s(i)} - X^\pi_{s(i')}\big) \ge A\delta - BR > 0,$$

where $s(i)$ denotes the section of student i and the inequality follows from the fact that the condition $\theta_1 < \bar{\theta}$ is equivalent to $BR < A\delta$ since $B/A = \theta_1/\big((1-\theta_1)(n-1)\big)$ is strictly increasing in θ_1. Therefore, given $\theta_1 \in (0,\bar{\theta})$, $\mathbb{E}[y^\pi_i]$ are strictly decreasing in i under any partition $\pi \in \Pi_{\mathcal{I}}$.

Next, observe that for any vector $z_1 > z_2 > \ldots > z_N$,

$$\sum_{i\in[N]}\sum_{j\in[N]} |z_i - z_j| = 2\sum_{i<j}(z_i - z_j) = 2\sum_{i\in[N]} c_i\, z_i, \tag{B.14}$$

where $c_i = N + 1 - 2i$. Thus, applying (B.14) with $z_i = \mathbb{E}[y^\pi_i]$ and substituting (B.12), we get that

$$\begin{aligned}
\Delta(\pi) &= 2A\sum_{i\in[N]} c_i x_i + 2B\sum_{i\in[N]} c_i X^\pi_{s(i)} \\
&= 2A\sum_{i\in[N]} c_i x_i + 2B\sum_{j\in[K]}\sum_{i\in\pi_j}\sum_{i'\in\pi_j} c_i\, x_{i'} \\
&= 2A\sum_{i\in[N]} c_i x_i + 2B\left(n\sum_{i\in[N]} c_i x_i - 2\,W(\pi)\right) \\
&= 2(A + nB)\sum_{i\in[N]} c_i x_i - 4B\,W(\pi),
\end{aligned}$$

where the first term does not depend on π. Therefore, $\Delta(\pi) \le \Delta(\pi')$ if and only if $W(\pi) \ge W(\pi')$, which proves the lemma. □

Definition B.10 **(Snake partition)**. *The* snake *(or* serpentine*) partition $\breve{\pi} \in \Pi_{\mathcal{I}}$ assigns students to sections in n rounds of K consecutive students each, reversing the order of sections in every other round. Formally, for $j \in [K]$,*

$$\breve{\pi}_j = \big\{(r-1)K + j : r \in [n],\ r \text{ odd}\big\} \cup \big\{rK + 1 - j : r \in [n],\ r \text{ even}\big\}.$$

Proposition B.14. *Let $x_1 > x_2 > \ldots > x_N$ and $\theta_1 \in (0,\bar{\theta})$.*

(i) If $n = 2$, the snake partition $\breve{\pi}$ is the unique minimizer of the Gini index of effort over $\Pi_{\mathcal{I}}$.

(ii) If n is even and the attributes x_i are a linear function of the rank, i.e., $x_i = c - \beta i$ for all $i \in [N]$ with $\beta > 0$, then $\breve{\pi}$ minimizes the Gini index of effort over $\Pi_{\mathcal{I}}$.

Proof. By Lemma B.6, for $\theta_1 \in (0, \bar{\theta})$ it suffices to show that $\breve{\pi}$ maximizes the within-section dispersion W over $\Pi_{\mathcal{I}}$, and specifically uniquely for part (i).

We prove each part separately:

Proof of Part (i). For $n = 2$, every section is a pair, so we can write $W(\pi) = \sum_{\{i,i'\} \in \pi,\, i<i'} w(i,i')$ with $w(i,i') = (i'-i)(x_i - x_{i'}) > 0$.

For any fixed ranks $i < j < k < l$, first observe that

$$w(i,l) + w(j,k) > \max\big\{w(i,k) + w(j,l), w(i,j) + w(k,l)\big\}. \tag{B.15}$$

Indeed, simple algebra gives

$$w(i,l) + w(j,k) - w(i,k) - w(j,l) = (j-i)(x_k - x_l) + (l-k)(x_i - x_j) > 0, \tag{B.16}$$

$$w(i,l) + w(j,k) - w(i,j) - w(k,l) = (l-j)(x_i - x_k) + (k-i)(x_j - x_l) > 0, \tag{B.17}$$

since $i < j < k < l$ and $x_1 > \ldots > x_N$.

Now suppose for the sake of contradiction that π maximizes W and assume $\{1, N\} \notin \pi$; instead, $\{1, b\}, \{c, N\} \in \pi$ for some $1 < b, c < N$ and $b \neq c$. However, replacing these two pairs by $\{1, N\}$ and $\{b, c\}$ strictly increases W, due to (B.15) where $(i, j, k, l) = (1, \min\{b, c\}, \max\{b, c\}, N)$. This is a contradiction, so it must hold that $\{1, N\} \in \pi$. Since $W(\pi) = w(1, N) + W(\pi \setminus \{\{1, N\}\})$, induction on K gives $\pi = \breve{\pi}$. Thus $\breve{\pi}$ is the unique maximizer of W and Lemma B.6 applies.

Proof of Part (ii). For $\pi \in \Pi_{\mathcal{I}}$ and $j \in [K]$, let $C_j^{\pi} = \sum_{i \in \pi_j} (N + 1 - 2i) = n(N+1) - 2\sum_{i \in \pi_j} i$, and note that $\sum_{j \in [K]} C_j^{\pi} = 0$. For the given form of x_i and $c_i = N + 1 - 2i$, we have that

$$\begin{aligned}
\sum_{j \in [K]} C_j^{\pi} X_j^{\pi} &= \sum_{j \in [K]} \Big(\sum_{i \in \pi_j} c_i\Big)\Big(\sum_{i' \in \pi_j} x_{i'}\Big) = \sum_{j \in [K]} \sum_{i \in \pi_j} \sum_{i' \in \pi_j} c_i\, x_{i'} \\
&= \sum_{j \in [K]} \sum_{i \in \pi_j} c_i x_i + \sum_{j \in [K]} \sum_{\substack{i, i' \in \pi_j \\ i < i'}} \big(c_i x_{i'} + c_{i'} x_i\big) \\
&= \sum_{i \in [N]} c_i x_i + \sum_{j \in [K]} \sum_{\substack{i, i' \in \pi_j \\ i < i'}} \Big(c_i x_i + c_{i'} x_{i'} - 2(i'-i)(x_i - x_{i'})\Big) \\
&= \sum_{i \in [N]} c_i x_i + (n-1) \sum_{i \in [N]} c_i x_i - 2W(\pi) \\
&= n \sum_{i \in [N]} c_i x_i - 2W(\pi).
\end{aligned}$$

Therefore, maximizing W over $\Pi_{\mathcal{I}}$ is equivalent to minimizing $\sum_j C_j^{\pi} X_j^{\pi}$. The minimum value zero is attained exactly when $C_j^{\pi} = 0$ for all j. Specifically, the snake partition attains it: for any pair of

consecutive rounds $2m-1, 2m$ with $m \in [n/2]$, section j receives the students of ranks $(2m-2)K+j$ and $2mK+1-j$, therefore $c_{(2m-2)K+j} + c_{2mK+1-j} = 2N-(8m-4)K$, which does not depend on j. Summing over the $n/2$ round pairs, $C_j^{\breve{\pi}}$ becomes the same for all j, and $\sum_j C_j^{\breve{\pi}} = 0$ forces $C_j^{\breve{\pi}} = 0$. Thus, $\breve{\pi}$ maximizes W, and Lemma B.6 applies. □

Example B.2. The uniform partition does not minimize the Gini index of effort over $\Pi_{\mathcal{F}}$ in general. For $N=6$, $K=2$, $n=3$, $\theta_1 = 1/5$ and $\mathbf{x} = (12,5,4,3,2,1)$, the integral balanced partition $\pi = \big\{\{12,2,1\},\{5,4,3\}\big\}$ attains $G_E(\mathbf{y}^{\pi}) = 511/1782 < 520/1782 = G_E(\underline{\mathbf{y}})$. Since $\Pi_{\mathcal{I}} \subset \Pi_{\mathcal{F}}$, the uniform partition is therefore not optimal over $\Pi_{\mathcal{F}}$ either.

Proposition B.15 **(LIM, Welfare)**. *Under the LIM model, the assortative partition maximizes total utility.*

Proof. By Lemma B.5, the total utility under a partition M with summation vector $\mathbf{X}^M$ can be written as

$$S(M) = \frac{\gamma^2}{2}\frac{(n-1)^2}{(n-1+\theta_1)^2}\sum_{i\in[N]} x_i^2 + \frac{\gamma^2}{2}\cdot\frac{\theta_1\big(2(n-1)(1-\theta_1)+n\theta_1\big)}{(1-\theta_1)^2(n-1+\theta_1)^2}\sum_{j\in[K]}\big(X_j^M\big)^2 + C\sigma^2,$$

where $C>0$ is a constant. Since the first and third terms are constant across partitions, maximizing $S(M)$ reduces to maximizing

$$f(\mathbf{X}^M) = \frac{\gamma^2}{2}\cdot\frac{\theta_1\big(2(n-1)(1-\theta_1)+n\theta_1\big)}{(1-\theta_1)^2(n-1+\theta_1)^2}\sum_{j\in[K]}\big(X_j^M\big)^2.$$

For $0 \le \theta_1 < 1$, the coefficient above is nonnegative, thus by Proposition C.1 in Marshall et al. (2011), f is Schur convex as the sum of the convex functions $g(X_j^M) = (X_j^M)^2$. Consequently, Proposition 2 implies that the assortative partition maximizes total utility in the LIM model. □

Proposition B.16 **(Diversity)**. *The uniform partition (respectively, a balanced partition) attains the maximum total variance over $\Pi_{\mathcal{F}}$ (respectively, $\Pi_{\mathcal{I}}$).*

Proof. We rewrite $V(\pi) = \sum_{j\in[K]}\sum_{i\in\pi_j}(x_i - \overline{x}_j)^2$ as

$$V(\pi) = \sum_{i\in[N]} x_i^2 - \frac{1}{n}\sum_{j\in[K]}(X_j^{\pi})^2,$$

and analogously $V(M) = \sum_{i\in[N]} x_i^2 - \frac{1}{n}\sum_{j\in[K]}(X_j^M)^2$ for $M \in \Pi_{\mathcal{F}}$. Since $\sum_{i\in[N]} x_i^2$ is constant across partitions, the problem reduces to maximizing the Schur concave term $-\sum_{j\in[K]}(X_j^M)^2$. By Proposition C.1 in Marshall et al. (2011), $\sum_{j\in[K]}(X_j^M)^2$ is Schur convex as a sum of convex functions, hence $-\sum_{j\in[K]}(X_j^M)^2$ is Schur concave. Proposition 3 thus implies that the uniform partition (respectively, a balanced partition) is optimal over $\Pi_{\mathcal{F}}$ (respectively, $\Pi_{\mathcal{I}}$). □

B.2.2. Proofs from Subsection 4.3.1

PROPOSITION B.17 (**Bergstrom and Varian (1985)**). *Let $n \geq 3$. Fix a peer effects model $\mathcal{Q} = (\mathbf{U}, n, K, \mathbf{x})$. Consider any partition $\pi \in \Pi_{\mathcal{I}}$ with Nash equilibrium $\mathbf{y} \in \mathbb{R}^N$ and assume that $\mathbf{y}$ can be expressed in the form $y_i = f(x_i, \sum_{l \in \pi_j} y_l)$, $i \in \pi_j$, with f being a continuous function. Then, the equilibrium value of $e(\pi_j) = \sum_{i \in \pi_j} y_i$, $j \in [K]$, depends on π_j only through $X_j^{\pi} = \sum_{i \in \pi_j} x_i$ if and only if f takes the form:*

$$f\left(x_i, \sum_{i \in \pi_j} y_i\right) = r\left(\sum_{i \in \pi_j} y_i\right) + v\left(\sum_{i \in \pi_j} y_i\right) x_i, \tag{B.18}$$

where $r, v : \mathbb{R} \to \mathbb{R}$.

Note that if $n < 3$ in the above proposition, then condition (B.18) is sufficient but not necessary.

PROPOSITION 8. *Let $n \geq 3$. Consider a peer effects model $\mathcal{Q} = (\mathbf{U}, h, n, K)$, with h and each U_i being continuous, and assume that, for every $\mathbf{x} \in \mathbb{R}^{nK}$, each partition $\pi \in \Pi_{\mathcal{I}}$ induces a unique Nash equilibrium $\mathbf{y} \in \mathbb{R}^{nK}$ that satisfies $y_i = f\left(x_i, \sum_{k \in \pi_j} y_k\right)$, $i \in \pi_j$, $j \in [K]$, for some continuous function f. Then, the section effort $e(\pi_j)$ depends on π_j only through $\sum_{k \in \pi_j} x_k$ and $\mathcal{Q}$ is sorting-neutral, if and only if $\mathbf{y}$ satisfies:*

$$y_i = r\Big(c \sum_{k \in \pi_j} x_k + a\Big) + v\Big(c \sum_{k \in \pi_j} x_k + a\Big) x_i, \qquad i \in \pi_j,\ j \in [K], \tag{4.9}$$

for some constants $c, a \in \mathbb{R}$, and functions $v, r : \mathbb{R} \to \mathbb{R}$ such that, when $c \neq 0$,

$$z = n\, r(z) + v(z)\, \frac{z - a}{c}, \qquad z \in \mathbb{R}. \tag{4.10}$$

Proof. To prove sufficiency, suppose $\mathbf{y}$ satisfies (4.9), and write $z_j = cX_j^{\pi} + a$. Summing (4.9) over $i \in \pi_j$ gives

$$e(\pi_j) \;=\; n\, r(z_j) \;+\; v(z_j)\, X_j^{\pi}.$$

If $c \neq 0$, then $X_j^{\pi} = (z_j - a)/c$ and (4.10) gives $e(\pi_j) = z_j = cX_j^{\pi} + a$. If $c = 0$, then $z_j = a$ for every j and $e(\pi_j) = n\, r(a) + v(a)\, X_j^{\pi}$. In either case $e(\pi_j)$ is linear in X_j^{π} and depends on π_j only through $X_j^{\pi} = \sum_{k \in \pi_j} x_k$, thus

$$E(\pi) = \sum_{j \in [K]} e(\pi_j) = nK r(a) + v(a) \sum_{i \in [N]} x_i,$$

which is independent of π. Therefore, $\mathcal{Q}$ is sorting-neutral. Note that this direction holds for any $n \geq 2$.

For the other direction, let $n \geq 3$, and suppose $\mathcal{Q}$ is sorting-neutral and $e(\pi_j)$ depends on π_j only through $\sum_{k \in \pi_j} x_k$. We proceed in three steps.

Step 1: For all $j \in [K]$, we show that the section effort must equal $e(\pi_j) = c\sum_{i\in\pi_j} x_i + a$. First we write the section effort $e(\pi_j)$ as a function of $\{x_i, i \in \pi_j\}$. For any enumeration $\{x_1^{(j)}, \dots, x_n^{(j)}\}$ of $\{x_i, i \in \pi_j\}$, let $e(\pi_j) = \hat{e}(\mathbf{x}^{(j)}) = \hat{e}(x_1^{(j)}, \dots, x_n^{(j)})$, where the function $\hat{e} : \mathbb{R}^n \to \mathbb{R}$ is continuous and symmetric.

Fix $\alpha, \beta \in \mathbb{R}$, $\mathbf{u}, \mathbf{w} \in \mathbb{R}^{n-1}$, and consider the cohort with attributes α, β, the entries of $\mathbf{u}$ and of $\mathbf{w}$, together with any other $(K-2)n$ attributes. The partition placing $(\alpha, \mathbf{u})$ and $(\beta, \mathbf{w})$ in the first two sections and the partition placing $(\beta, \mathbf{u})$ and $(\alpha, \mathbf{w})$ in the first two sections agree on all remaining sections, so sorting-neutrality gives

$$\hat{e}(\alpha, \mathbf{u}) + \hat{e}(\beta, \mathbf{w}) = \hat{e}(\beta, \mathbf{u}) + \hat{e}(\alpha, \mathbf{w}), \qquad \alpha, \beta \in \mathbb{R},\ \mathbf{u}, \mathbf{w} \in \mathbb{R}^{n-1}. \tag{B.19}$$

Thus $\hat{e}(\alpha, \mathbf{u}) - \hat{e}(\beta, \mathbf{u})$ does not depend on $\mathbf{u}$. Taking $\beta = 0$ and $\mathbf{w} = \mathbf{0}$ yields $\hat{e}(\alpha, \mathbf{u}) = \varphi(\alpha) + \chi(\mathbf{u})$, where $\varphi(\alpha) = \hat{e}(\alpha, \mathbf{0}) - \hat{e}(\mathbf{0})$ and $\chi(\mathbf{u}) = \hat{e}(0, \mathbf{u})$. Since $\hat{e}$ is symmetric, χ is a symmetric function on $\mathbb{R}^{n-1}$ satisfying (B.19) in its own coordinates. By induction on n, we get that

$$\hat{e}(\mathbf{x}) = \sum_{l\in[n]} u(x_l) + a_0, \qquad \mathbf{x} \in \mathbb{R}^n, \tag{B.20}$$

for a continuous $u : \mathbb{R} \to \mathbb{R}$ and a constant $a_0 \in \mathbb{R}$.

By hypothesis, $\hat{e}(\mathbf{x}) = \phi\big(\sum_{l\in[n]} x_l\big)$ for some $\phi : \mathbb{R} \to \mathbb{R}$. Substituting into (B.20),

$$\phi(x_1 + \cdots + x_n) = u(x_1) + \cdots + u(x_n) + a_0, \qquad \mathbf{x} \in \mathbb{R}^n. \tag{B.21}$$

Equation (B.21) is a special form of a Pexider functional equation and by Theorem 2, p. 348, in Aczél (1966), its general continuous solution is

$$u(t) = c\,t + a_1, \qquad \phi(s) = c\,s + a, \qquad a = n a_1 + a_0.$$

It follows that

$$e(\pi_j) = c\sum_{i\in\pi_j} x_i + a, \qquad j \in [K].$$

Step 2: We show that $y_i = r\big(\sum_{k\in\pi_j} y_k\big) + v\big(\sum_{k\in\pi_j} y_k\big)x_i$ for some $r, v : \mathbb{R} \to \mathbb{R}$ and all $i \in [N]$. By hypothesis, $e(\pi_j)$ depends on π_j only through $\sum_{k\in\pi_j} x_k$, so this is exactly Proposition B.17.

Step 3: We complete the proof by showing that $z = n\,r(z) + v(z)\frac{z-a}{c}$. By Steps 1 and 2, $\sum_{k\in\pi_j} y_k = e(\pi_j) = cX_j^\pi + a$ with $X_j^\pi = \sum_{k\in\pi_j} x_k$, and substituting into Step 2 gives (4.9). Summing (4.9) over $i \in \pi_j$ and using $e(\pi_j) = cX_j^\pi + a$ once more,

$$cX_j^\pi + a \;=\; n\,r\big(cX_j^\pi + a\big) + v\big(cX_j^\pi + a\big)\,X_j^\pi, \qquad j \in [K].$$

If $c \neq 0$, writing $z = cX_j^\pi + a$ gives $X_j^\pi = (z-a)/c$ and the above equation becomes (4.10). □

B.2.3. Proof from Subsection 4.3.2

PROPOSITION 9. *In the quadratic-in-means model, the assortative partition maximizes total effort if* $\gamma_{22}+\gamma_{12}>0$*, whereas the uniform partition maximizes total effort over* $\Pi_{\mathcal{F}}$ *and a balanced partition maximizes total effort over* $\Pi_{\mathcal{I}}$ *if* $\gamma_{22}+\gamma_{12}<0$*.*

Proof. In (4.12), the first three terms are constant across all partitions. Hence, maximizing $E(\pi)$ is equivalent to maximizing $\sum_{j=1}^{K}\left(\sum_{i\in\pi_j}x_i\right)^2$ when $\gamma_{22}+\gamma_{12}>0$, and minimizing it when $\gamma_{22}+\gamma_{12}<0$. Since $f(\mathbf{z})=\sum_{j=1}^{K}z_j^2$ is Schur-convex, Proposition 2 applies in the former case, yielding the assortative partition as optimal. In the latter case, Proposition 3 applies, yielding the uniform partition as optimal over $\Pi_{\mathcal{F}}$ and a balanced partition as optimal over $\Pi_{\mathcal{I}}$. □

B.2.4. Proofs from Subsection 4.3.3

LEMMA B.7 **(Proposition 1 in Boucher et al. (2024))**. *In the social norm model, assume that* $0<\frac{\theta_1}{1+\theta_2}<1$ *and* $0<\frac{\theta_1+\theta_2}{1+\theta_2}<1$*. Then, there exists a unique Nash equilibrium.*

Throughout the following proofs, we set

$$a_i=\frac{q_i}{1+\theta_2},\quad b=\frac{\theta_1+\theta_2}{(1+\theta_2)(n-1)^2},$$

and for any $\pi\in\Pi_{\mathcal{I}}$, we define

$$S_j:=\sum_{i\in\pi_j}y_i,\qquad \widehat{S}_j:=\sum_{i\in\pi_j}\sqrt{y_i}.$$

By the assumptions in Proposition B.7, it follows that $0<b<\frac{1}{(n-1)^2}$, thus

$$1+(n-1)b-n\sqrt{b}=(1-\sqrt{b})(1-(n-1)\sqrt{b})>0.$$

LEMMA B.8. *Consider the social norm model for* $\beta=1/2$*. For* $\pi\in\Pi_{\mathcal{I}}$*,* $\widehat{S}_j=\sum_{i\in\pi_j}\sqrt{y_i}$ *is the unique positive root of*

$$(1+(n-1)b)\widehat{S}_j=\sum_{i\in\pi_j}\sqrt{b\widehat{S}_j^2+(1-b)a_i}. \tag{B.22}$$

Then the section j *effort satisfies*

$$e(\pi_j)=S_j=\frac{\sum_{i\in\pi_j}a_i+b(n-2)\widehat{S}_j^2}{1-b}. \tag{B.23}$$

Proof. The first-order condition of (4.14) is

$$y_i=a_i+b\left(\sum_{l\in\pi_j,l\neq i}\sqrt{y_l}\right)^2=a_i+b\left(\hat{S}_j-\sqrt{y_i}\right)^2. \tag{B.24}$$

Then, each y_i and $\sqrt{y_i}$ can be written as follows:

$$y_i = \frac{a_i + b\widehat{S}_j^2 - 2b\widehat{S}_j\sqrt{y_i}}{1-b}, \tag{B.25}$$

$$\sqrt{y_i} = \frac{-b\widehat{S}_j + \sqrt{b\widehat{S}_j^2 + (1-b)a_i}}{1-b}, \tag{B.26}$$

where note that (B.26) is the unique nonnegative root. Summing (B.26) over $i \in \pi_j$, we get (B.22). Let $\Phi(x) = \sum_{i\in\pi_j} \sqrt{bx^2 + (1-b)a_i} - (1 + (n-1)b)x$. Since

$$\Phi'(x) = \sum_{i\in\pi_j} \frac{bx}{\sqrt{bx^2 + (1-b)a_i}} - (1 + (n-1)b) < n\sqrt{b} - (1 + (n-1)b) < 0, \tag{B.27}$$

Φ is strictly decreasing on $[0,\infty)$, with $\Phi(0) = \sum_{i\in\pi_j} \sqrt{(1-b)a_i} > 0$ and $\lim_{x\to+\infty} \Phi(x) = -\infty$, therefore (B.22) has a unique positive root $\widehat{S}_j$ by the Intermediate Value Theorem.

By definition, $e(\pi_j) = S = \sum_{i\in\pi_j} y_i = \sum_{i\in\pi_j} (\sqrt{y_i})^2$. Summing (B.25) over i, we get

$$\hat{S} = \sum_{i\pi_j} a_i + b(n-2)\widehat{S}_j^2 + bS,$$

which in turn implies (B.23). □

By Lemma B.8, for $n \geq 3$, maximizing total effort is equivalent to maximizing $\sum_{j\in[K]} Z_j^\pi$ where $Z_j^\pi = (\widehat{S}_j(\pi_j))^2$. For $n = 2$, total effort is the same under every partition. For any set $A \subseteq [N]$, $|A| = n$, let $\widehat{S}(A)$ denote the root of (B.22) for that set A and define $Z(A) = (\widehat{S}(A))^2$.

LEMMA B.9. *Let $\pi_j = \{i_1, i_2, \ldots, i_n\}$, $\pi_{j'} = \{i'_1, i'_2, \ldots, i'_n\}$, with $i_1 \geq i_2 \geq \ldots \geq i_n$ and $i'_1 \geq i'_2 \geq \ldots \geq i'_n$. Let also $\widehat{S}(\pi_j) = \widehat{S}_j$ denote the unique root of (B.22) for any n-element set $\pi_j \subseteq [N]$. If $a_{i'_l} \geq a_{i_l}$ for all $l \in [n]$, then $\widehat{S}(\pi'_j) \geq \widehat{S}(\pi_j)$, with strict inequality if $a_{i'_l} > a_{i_l}$ for some $l \in [n]$.*

Proof. For an n-element set $A \subseteq [N]$ and $x > 0$, write $\Phi_S(x) = \sum_{i\in S} \sqrt{bx^2 + (1-b)a_i} - (1 + (n-1)b)x$, so that $\widehat{S}(A)$ is the root of Φ_A. By (B.27), Φ_A is strictly decreasing in x; hence its root is unique and $\Phi_A(x) \geq 0$ if and only if $x \leq \widehat{S}(A)$.

Fix $x > 0$. Then $\Phi_S(x)$ is nondecreasing in each attribute a_i, since

$$\frac{\partial}{\partial a_i}\sqrt{bx^2 + (1-b)a_i} = \frac{1-b}{2\sqrt{bx^2 + (1-b)a_i}} > 0.$$

$a_{i'_l} \geq a_{i_l}$ for every $l \in [n]$. Raising the attributes one coordinate at a time from $(a_{i_1}, \ldots, a_{i_n})$ to $(a_{i'_1}, \ldots, a_{i'_n})$ thus only increases Φ at every point, so

$$\Phi_{\pi_{j'}}(x) \;\geq\; \Phi_{\pi_j}(x) \qquad \text{for all } x > 0.$$

Evaluating at $x = \widehat{S}(\pi_j)$ gives $\Phi_{\pi_{j'}}\big(\widehat{S}(\pi_j)\big) \geq \Phi_{\pi_j}\big(\widehat{S}(\pi_j)\big) = 0$. However, $\Phi_{\pi_{j'}}$ is strictly decreasing with root $\widehat{S}(\pi_{j'})$, so $\Phi_{\pi_{j'}}(x) \geq 0$ only on $[0, \widehat{S}(\pi_{j'})]$. This implies that $\widehat{S}(\pi_j) \in [0, \widehat{S}(\pi_{j'})]$, i.e., $\widehat{S}(\pi_j) \leq \widehat{S}(\pi_{j'})$. If moreover $a_{i'_l} > a_{i_l}$ for some l, the inequality above is strict at $x = \widehat{S}(\pi_j)$, so $\widehat{S}(\pi_{j'}) > \widehat{S}(\pi_j)$. □

LEMMA B.10. *Let $a_1 > a_2 > \ldots > a_{2n} > 0$, $n \geq 3$. Among all 2-way partitions of $\{a_1, a_2, \ldots, a_{2n}\}$, the assortative partition $\widetilde{\pi}_1 = \{1, \ldots, n\}$, $\widetilde{\pi}_2 = \{n+1, \ldots, 2n\}$ uniquely maximizes $Z(\pi_1) + Z(\pi_2)$.*

Proof. We break the proof into several steps.

Step 1: We consider a continuous relaxation. We first divide (B.22) by $\widehat{S}$ and set $Z = \widehat{S}^2$. Next we replace the integral π_j with the fraction of individual i placed in j, namely $\lambda_i \in [0,1]$. For $\boldsymbol{\lambda} \in [0,1]^{2n}$ with $\sum_i \lambda_i = n$, let $Z(\boldsymbol{\lambda})$ be the unique positive root[9] of

$$\sum_{i=1}^{2n} \lambda_i \sqrt{b + (1-b)a_i/Z} = 1 + (n-1)b. \tag{B.28}$$

Our goal is to maximize $\widehat{S}(\pi_1)^2 + \widehat{S}(\pi_2)^2$ over all feasible 2-way integral partitions $\pi in \Pi_{\mathcal{I}}$. Since $Z(\mathbf{1}_{\pi_j}) = \widehat{S}(\pi_j)^2$ and $\mathbf{1} - \mathbf{1}_{\pi_1} = \mathbf{1}_{\pi_2}$,

$$F(\boldsymbol{\lambda}) = Z(\boldsymbol{\lambda}) + Z(\mathbf{1} - \boldsymbol{\lambda})$$

over $\Lambda = \{\boldsymbol{\lambda} \in [0,1]^{2n} : \sum_i \lambda_i = n\}$. The vertices of Λ correspond exactly to the integral 2-way partitions. We will show that every maximizer of F corresponds to the assortative partition, that is, either of the vertices $\mathbf{1}_{\widetilde{\pi}_1}$, $\mathbf{1}_{\widetilde{\pi}_2}$.

Step 2: For a maximizer $\lambda^* \in \Lambda$, we derive the marginal value $\eta(a)$ of reassigning an individual with attribute a from section 2 to section 1, and show that optimality enforces a threshold rule, that is, $\eta \geq \tau$ for section 1 and $\eta \leq \tau$ for section 2, for some $\tau \in \mathbb{R}$.

For any $\boldsymbol{\lambda} \in \Lambda$, consider $Z = Z(\boldsymbol{\lambda})$ that solves (B.28). The LHS of (B.28) has derivative

$$\begin{aligned} L'(Z) &= -\frac{1}{2Z} \sum_{i=1}^{2n} \lambda_i \left(\sqrt{b + \frac{(1-b)a_i}{Z}} - \frac{b}{\sqrt{b + \frac{(1-b)a_i}{Z}}} \right) \\ &= -\frac{1}{2Z} \left(1 + (n-1)b - \sum_{l=1}^{2n} \lambda_l \frac{b}{\sqrt{b + \frac{(1-b)a_l}{Z}}} \right), \end{aligned}$$

where the second equality follows by substituting (B.28).

Since $Z(\boldsymbol{\lambda})$ is defined by $L(Z) = 1 + (n-1)b$ in (B.28), implicit differentiation gives

$$\frac{\partial Z(\boldsymbol{\lambda})}{\partial \lambda_i} = -\frac{\partial L / \partial \lambda_i}{\partial L / \partial Z} = \frac{2\sqrt{bZ^2 + (1-b)a_i Z}}{1 + (n-1)b - \sum_{l=1}^{2n} \lambda_l \frac{b}{\sqrt{b + \frac{(1-b)a_l}{Z}}}}.$$

[9] The left side is strictly decreasing in Z, tends to $+\infty$ as $Z \downarrow 0$ and to $\sqrt{b}\sum_i \lambda_i = n\sqrt{b} < 1 + (n-1)b$ as $Z \to \infty$, so the root exists and is unique.

For a maximizer $\lambda^* \in \Lambda$, $\eta(a)$ is the corresponding marginal change in F when an infinitesimal amount of attribute a is moved from section $j=2$ to section $j=1$. Note that $\eta(a)$ is given by

$$\eta(a) = 2\left(\frac{\sqrt{bZ(\boldsymbol{\lambda}^*)^2 + (1-b)aZ(\boldsymbol{\lambda}^*)}}{1+(n-1)b - \sum_{l=1}^{2n} \lambda_l^* \frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\boldsymbol{\lambda}^*)}}}} - \frac{\sqrt{bZ(\mathbf{1}-\boldsymbol{\lambda}^*)^2 + (1-b)aZ(\mathbf{1}-\boldsymbol{\lambda}^*)}}{1+(n-1)b - \sum_{l=1}^{2n} (1-\lambda_l^*) \frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\mathbf{1}-\boldsymbol{\lambda}^*)}}}} \right). \quad \text{(B.29)}$$

At a maximizer $\boldsymbol{\lambda}^*$ over the convex set Λ, since F is continuously differentiable on a neighborhood of Λ with $\partial F/\partial \lambda_i = \eta(a_i)$, the maximizer $\boldsymbol{\lambda}^*$ satisfies

$$\nabla F(\boldsymbol{\lambda}^*)^\top (\boldsymbol{\lambda} - \boldsymbol{\lambda}^*) = \sum_{i=1}^{2n} \eta(a_i)\,(\lambda_i - \lambda_i^*) \le 0 \qquad \text{for all } \boldsymbol{\lambda} \in \Lambda. \quad \text{(B.30)}$$

Equivalently, $\boldsymbol{\lambda}^*$ maximizes the linear functional $\sum_{i=1}^{2n} \eta(a_i)\,\lambda_i$ over $\boldsymbol{\lambda} \in \Lambda$. By the Bathtub Principle, there exists a threshold $\tau \in \mathbb{R}$ such that if $\lambda_i^* > 0$ then $\eta(a_i) \ge \tau$ and if $\lambda_i^* < 1$ then $\eta(a_i) \le \tau$, while $\eta(a_i) = \tau$ for any fractional coordinate. Note that $\boldsymbol{\lambda}^*$ cannot have exactly one fractional coordinate since $\sum_i \lambda_i^* = n$.

Step 3: We show that η' changes sign at most once, and characterize the feasible shapes of η. We conclude that $\boldsymbol{\lambda}^* \in \{\mathbf{1}_{\widetilde{\pi}_1}, \mathbf{1}_{\widetilde{\pi}_2}\}$. We have that

$$\eta'(a) = \frac{(1-b)Z(\boldsymbol{\lambda}^*)}{\left(1+(n-1)b - \sum_{l=1}^{2n} \lambda_l^* \frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\boldsymbol{\lambda}^*)}}}\right)\sqrt{bZ(\boldsymbol{\lambda}^*)^2 + (1-b)\,a\,Z(\boldsymbol{\lambda}^*)}}$$
$$- \frac{(1-b)Z(\mathbf{1}-\boldsymbol{\lambda}^*)}{\left(1+(n-1)b - \sum_{l=1}^{2n} (1-\lambda_l^*) \frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\mathbf{1}-\boldsymbol{\lambda}^*)}}}\right)\sqrt{bZ(\mathbf{1}-\boldsymbol{\lambda}^*)^2 + (1-b)\,a\,Z(\mathbf{1}-\boldsymbol{\lambda}^*)}}.$$

Note that $\eta'(a) > 0$ if and only if

$$bZ(\boldsymbol{\lambda}^*)Z(\mathbf{1}-\boldsymbol{\lambda}^*)\left[\left(1+(n-1)b - \sum_{l=1}^{2n}(1-\lambda_l^*)\frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\mathbf{1}-\boldsymbol{\lambda}^*)}}}\right)^2 - \left(1+(n-1)b - \sum_{l=1}^{2n}\lambda_l^*\frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\boldsymbol{\lambda}^*)}}}\right)^2\right]$$
$$+(1-b)a\left[Z(\boldsymbol{\lambda}^*)\left(1+(n-1)b - \sum_{l=1}^{2n}(1-\lambda_l^*)\frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\mathbf{1}-\boldsymbol{\lambda}^*)}}}\right)^2 - Z(\mathbf{1}-\boldsymbol{\lambda}^*)\left(1+(n-1)b - \sum_{l=1}^{2n}\lambda_l^*\frac{b}{\sqrt{b+\frac{(1-b)a_l}{Z(\boldsymbol{\lambda}^*)}}}\right)^2\right] > 0,$$

where the LHS is affine in a and therefore changes sign at most once, so η is (i) (strictly) monotone, (ii) decreasing-then-increasing, (iii) increasing-then-decreasing, or (iv) constant. Consider the superlevel set $U := \{a > 0 : \eta(a) > \tau\}$. Every individual with attribute in U is fully assigned to section 1 ($\lambda_i^* = 1$), every individual with $\eta(a_i) < \tau$ is fully assigned to section 2 ($\lambda_i^* = 0$), and only individuals with $\eta(a_i) = \tau$ may be split across the two sections. We consider the following cases.

Case (i): η strictly monotone. The level set $\{a>0:\eta(a)=\tau\}$ contains at most one attribute, so at most one coordinate of $\boldsymbol{\lambda}^*$ is fractional, which is impossible since $\sum_i \lambda_i = n$. Hence $\lambda_i^* \in \{0,1\}$ for all i, with $\lambda_i^* = 1$ exactly for the n individuals whose attributes give the n largest values of η, i.e., the n largest attributes if η is increasing, or the n smallest if η is decreasing.

Case (ii): η increasing-then-decreasing. We have four subcases based on U. If $U=\emptyset$, then all mass n of section 1 must concentrate on the level set $\{a>0:\eta(a)=\tau\}$. However, since η is single-peaked, the level set $\{a>0:\eta(a)=\tau\}$ contains at most two attributes, thus $n\le 2$, which is impossible. If $U=(0,\infty)$, $\lambda_i^*=1$ for all $i\in[2n]$, contradicting $\sum_i \lambda_i^* = n$. If either $U=(0,p)$ or $U=(q,\infty)$, the level set $\{a>0:\eta(a)=\tau\}$ consists of the single finite endpoint, therefore Case (i) applies. Otherwise $U=(q,p)$ with $0<q<p<\infty$, which we rule out next.

If $U=(q,p)$, then we must have $Z_1 \neq Z_2$, where we use $Z_1 = Z(\boldsymbol{\lambda}^*)$ and $Z_2 = Z(\mathbf{1}-\boldsymbol{\lambda}^*)$. This is because if $Z_1 = Z_2$, the sign of (B.2.4) does not depend on a, so $\eta(\alpha)$ would be monotone or constant.

First, suppose that $Z_1 > Z_2$, We write κ_1, κ_2 for the first and second denominator in (B.29); note that κ_1, κ_2 do not depend on a. For $x>y>0$, consider the ratio

$$R(x,y)=\frac{\sqrt{bZ_1^2+(1-b)xZ_1}-\sqrt{bZ_1^2+(1-b)yZ_1}}{\sqrt{bZ_2^2+(1-b)xZ_2}-\sqrt{bZ_2^2+(1-b)yZ_2}}=\frac{W_1(x)-W_1(y)}{W_2(x)-W_2(y)}=\frac{Z_1}{Z_2}\cdot\frac{W_2(x)+W_2(y)}{W_1(x)+W_1(y)},$$

where the denominator is positive. Consequently, for $x>y$,

$$\eta(x)\gtrless\eta(y) \quad\Longleftrightarrow\quad R(x,y)\gtrless\frac{\kappa_1}{\kappa_2}. \tag{B.31}$$

We claim that R is strictly increasing in each of x and y; since $R(x,y)$ is symmetric in x and y, it suffices to prove strict monotonicity only in x. Since $W_t'(x)=(1-b)Z_t/2W_t(x)$, the sign of $\partial R/\partial x$ is the sign of

$$\Big[Z_2W_1(x)^2-Z_1W_2(x)^2\Big]+\Big[Z_2W_1(x)W_1(y)-Z_1W_2(x)W_2(y)\Big].$$

The first bracket equals $Z_1Z_2\big(bZ_1+(1-b)x\big)-Z_1Z_2\big(bZ_2+(1-b)x\big)=bZ_1Z_2(Z_1-Z_2)>0$. The second bracket is a difference of two positive terms whose squares are $Z_1^2Z_2^2\big(bZ_1+(1-b)x\big)\big(bZ_1+(1-b)y\big)$ and $Z_1^2Z_2^2\big(bZ_2+(1-b)x\big)\big(bZ_2+(1-b)y\big)$; since $bZ_1+(1-b)t>bZ_2+(1-b)t$ for every t, the bracket is positive. Therefore, R is strictly increasing in each argument.

We now rule out that $U=(q,p)$. Since $U=(q,p)$ is the strict superlevel set of the continuous function η at level τ, its endpoints satisfy

$$\eta(q)=\eta(p)=\tau. \tag{B.32}$$

By the threshold rule in Step 2, every individual in section 1 has $\eta(a_i)\ge\tau$, therefore $a_i\in[q,p]$, since $\eta<\tau$ outside $[q,p]$. Since the attributes are distinct, at most two individuals have $a_i\in\{q,p\}$,

and they contribute at most $2 < n$ mass to section 1. Consequently, there must be some remaining individual with attribute $a' \in (q,p)$ that contributes additional mass to section 1, and

$$\eta(a') > \tau.$$

Now apply (B.31) to the points $q < a' < p$. The equality (B.32), with $x = p > y = q$, gives $R(p,q) = \kappa_1/\kappa_2$; the strict inequality $\eta(p) = \tau < \eta(a')$, with $x = p > y = a'$, gives $R(p,a') < \kappa_1/\kappa_2$. Since $a' > q$ and R is strictly increasing in its second argument,

$$\frac{\kappa_1}{\kappa_2} = R(p,q) < R(p,a') < \frac{\kappa_1}{\kappa_2},$$

a contradiction. Hence no window with $Z_1 > Z_2$ exists.

Suppose now that $Z_1 < Z_2$. For $r \in \{q,p\}$ write

$$\zeta_r := \sqrt{b + (1-b)r/Z_1}, \qquad \xi_r := \sqrt{b + (1-b)r/Z_2}.$$

Since $Z_1 < Z_2$, we have $\zeta_r > \xi_r > \sqrt{b}$. Note that $W_1(r) = Z_1\zeta_r$ and $W_2(r) = Z_2\xi_r$, so $R(p,q) = (\xi_q + \xi_p)/(\zeta_q + \zeta_p)$, which combined with $\eta(q) = \eta(p) = \tau$ and (B.31), yields

$$\kappa_1(\zeta_q + \zeta_p) = \kappa_2(\xi_q + \xi_p). \tag{B.33}$$

We derive a contradiction by showing that the LHS strictly exceeds the RHS.

Equation (B.28), applied to $\boldsymbol{\lambda}^*$ and to $\mathbf{1} - \boldsymbol{\lambda}^*$, gives

$$\sum_{l=1}^{2n} \lambda_l^* \sqrt{b + \frac{(1-b)a_l}{Z_1}} = \sum_{l=1}^{2n} (1-\lambda_l^*) \sqrt{b + \frac{(1-b)a_l}{Z_2}} = 1 + (n-1)b, \qquad \sum_{l=1}^{2n} \lambda_l^* = \sum_{l=1}^{2n} (1-\lambda_l^*) = n.$$

By the threshold rule, $\lambda_l^* > 0$ implies $a_l \in [q,p]$, and ζ_r is increasing in r, so $\sqrt{b + (1-b)a_l/Z_1} \in [\zeta_q, \zeta_p]$, while $\lambda_l^* < 1$ implies $a_l \notin (q,p)$, so $\sqrt{b + (1-b)a_l/Z_2} \notin (\xi_q, \xi_p)$.

Furthermore, for $x > 0$ and $\sqrt{b} < u < v$, we have that $1/x \le (u+v-x)/uv$ for $x \in [u,v]$, with the reverse inequality for $x \notin (u,v)$. Applying this term by term, with $(u,v) = (\zeta_q, \zeta_p)$ in the sum for κ_1, and $(u,v) = (\xi_q, \xi_p)$ in the sum for κ_2, we get that

$$\sum_{l=1}^{2n} \frac{\lambda_l^*}{\sqrt{b + \frac{(1-b)a_l}{Z_1}}} \le \frac{n(\zeta_q + \zeta_p) - \big(1 + (n-1)b\big)}{\zeta_q \zeta_p}, \qquad \sum_{l=1}^{2n} \frac{1-\lambda_l^*}{\sqrt{b + \frac{(1-b)a_l}{Z_2}}} \ge \frac{n(\xi_q + \xi_p) - \big(1 + (n-1)b\big)}{\xi_q \xi_p},$$

and therefore

$$\kappa_1 \ge 1 + (n-1)b - b\,\frac{n(\zeta_q + \zeta_p) - \big(1 + (n-1)b\big)}{\zeta_q \zeta_p}, \qquad \kappa_2 \le 1 + (n-1)b - b\,\frac{n(\xi_q + \xi_p) - \big(1 + (n-1)b\big)}{\xi_q \xi_p}.$$

For $u, v > \sqrt{b}$, define

$$\Psi(u,v) = (u+v)\,\frac{\big(1 + (n-1)b\big)(uv + b) - bn(u+v)}{uv},$$

so the previous inequalities can be rewritten as

$$\kappa_1(\zeta_q+\zeta_p) \;\geq\; \Psi(\zeta_q,\zeta_p), \qquad \kappa_2(\xi_q+\xi_p) \;\leq\; \Psi(\xi_q,\xi_p).$$

Since Ψ is strictly increasing in each coordinate on $(\sqrt{b},\infty)^2$, for $\zeta_q>\xi_q$ and $\zeta_p>\xi_p$,

$$\kappa_1(\zeta_q+\zeta_p) \;\geq\; \Psi(\zeta_q,\zeta_p) \;>\; \Psi(\xi_q,\xi_p) \;\geq\; \kappa_2(\xi_q+\xi_p),$$

contradicting (B.33). Thus U cannot take the form $U=(q,p)$, completing Case (ii).

Case (iii): η decreasing-then-increasing. We replace $\boldsymbol{\lambda}^*$ by $\mathbf{1}-\boldsymbol{\lambda}^*$, which preserves F and replaces η by $-\eta$, which is increasing-then-decreasing, thus Case (ii) applies.

Case (iv): η constant. If η is constant, then $\kappa_1=\kappa_2=\kappa>0$ and then $Z_1=Z_2=Z$ in the expression of $\eta'(a)=0$. This implies that $\eta\equiv 0$. We will show that $\boldsymbol{\lambda}^*\in\{\mathbf{1}_{\widetilde{\pi}_1},\mathbf{1}_{\widetilde{\pi}_2}\}$.

The function $h(x)=x-\frac{b}{x}$ is strictly increasing and strictly concave on $(\sqrt{b},\infty)$. A direction $\mathbf{d}\in\mathbb{R}^{2n}$ is feasible if $\sum_i d_i=0$, $d_i\geq 0$ whenever $\lambda_i^*=0$, and $d_i\leq 0$ whenever $\lambda_i^*=1$. Twice differentiating (B.28) shows that, for any feasible direction d satisfying $\sum_i d_i\, h\big(\sqrt{b+(1-b)a_i/\bar{Z}}\big)=0$,

$$F''(\boldsymbol{\lambda}^*;\mathbf{d})=\frac{2Z}{\kappa^3}\left[\sum_i \frac{\frac{(1-b)a_i}{Z}\left(\frac{3(1-b)a_i}{Z}+4b\right)}{\left(b+\frac{(1-b)a_i}{Z}\right)^{3/2}}\right]\left(\sum_i d_i\sqrt{b+\frac{(1-b)a_i}{Z}}\right)^2, \tag{B.34}$$

where $F''(\boldsymbol{\lambda}^*;d)>0$ whenever $\sum_i d_i\sqrt{b+(1-b)a_i/Z}\neq 0$.

Suppose that $\boldsymbol{\lambda}^*\notin\{\mathbf{1}_{\widetilde{\pi}_1},\mathbf{1}_{\widetilde{\pi}_2}\}$. Then there exist feasible exchanges $\mathbf{u}=\mathbf{e}_k-\mathbf{e}_i$ and $\mathbf{w}=\mathbf{e}_{k'}-\mathbf{e}_{i'}$ with $a_k>a_i$ and $a_{k'}<a_{i'}$. The exchanges $\mathbf{u}$ and $\mathbf{w}$ may be chosen so that

$$\left(b+\tfrac{(1-b)a_k}{\bar{Z}}\right)\left(b+\tfrac{(1-b)a_i}{\bar{Z}}\right) \;\neq\; \left(b+\tfrac{(1-b)a_{k'}}{\bar{Z}}\right)\left(b+\tfrac{(1-b)a_{i'}}{\bar{Z}}\right). \tag{B.35}$$

To see why, suppose every pair of exchanges makes (B.35) an equality, thus all feasible exchanges have the same product. Since $\sum_k(1-\lambda_k^*)=n\geq 3$ and $0\leq\lambda_k^*\leq 1$, at least three coordinates satisfy $\lambda_k^*<1$; fix one coordinate i with $\lambda_i^*>0$ and two coordinates $k\neq k'$ with $\lambda_k^*,\lambda_{k'}^*<1$, distinct from i. The feasible exchanges $\mathbf{e}_k-\mathbf{e}_i$ and $\mathbf{e}_{k'}-\mathbf{e}_i$ then satisfy

$$\left(b+\tfrac{(1-b)a_i}{Z}\right)\left(b+\tfrac{(1-b)a_k}{Z}\right)=\left(b+\tfrac{(1-b)a_i}{Z}\right)\left(b+\tfrac{(1-b)a_{k'}}{Z}\right),$$

thus $a_k=a_{k'}$, which is a contradiction

For

$$\alpha=h\left(\sqrt{b+\tfrac{(1-b)a_k}{\bar{Z}}}\right)-h\left(\sqrt{b+\tfrac{(1-b)a_i}{\bar{Z}}}\right)>0, \qquad \gamma=h\left(\sqrt{b+\tfrac{(1-b)a_{i'}}{\bar{Z}}}\right)-h\left(\sqrt{b+\tfrac{(1-b)a_{k'}}{\bar{Z}}}\right)>0,$$

set

$$d=\gamma u+\alpha w,$$

where $\sum_i d_i\, h\big(\sqrt{b+(1-b)a_i/Z}\big) = \gamma\alpha - \alpha\gamma = 0$. Substituting and simplifying, we get

$$\sum_i d_i \sqrt{b+\frac{(1-b)a_i}{\bar{Z}}} - \sqrt{b+\tfrac{(1-b)a_i}{\bar{Z}}}\Big) - \sqrt{b+\tfrac{(1-b)a_{k'}}{\bar{Z}}}\Big) \;\neq\; 0,$$

Consequently, (B.34) gives $F''(\boldsymbol{\lambda}^*; d) > 0$, contradicting maximality.

To sum up, Cases (i)–(iv) show that $\boldsymbol{\lambda}^* \in \{\mathbf{1}_{\widetilde{\pi}_1}, \mathbf{1}_{\widetilde{\pi}_2}\}$. Since a maximizer exists and $F(\mathbf{1}_{\widetilde{\pi}_1}) = F(\mathbf{1}_{\widetilde{\pi}_2})$, the assortative partition uniquely maximizes $Z(\pi_1) + Z(\pi_2)$ among all 2-way partitions of $\{a_1, \ldots, a_{2n}\}$. □

LEMMA B.11. *For every integral partition $\pi \in \Pi_{\mathcal{I}}$, $\mathbf{Z}^{\pi} \prec_w \mathbf{Z}^{\widetilde{\pi}}$.*

Proof. Fix $k \in [K]$, and let T be the union of the k sections of π with the largest section values $Z(\pi_j)$. Among the finitely many partitions of T into k sets of size n, pick one, say $(T_1, \ldots, T_k)$, that maximizes the sum $\sum_{\ell \le k} Z(T_\ell)$. Since the k chosen sections of π themselves form such a partition of T,

$$\sum_{j \le k} Z^{\pi}_{[j]} \le \sum_{\ell \le k} Z(T_\ell).$$

At the maximizing partition, every pair $(T_\ell, T_{\ell'})$ maximizes $Z(\cdot) + Z(\cdot)$ over all splits of $T_\ell \cup T_{\ell'}$ into two sets of size n, the other sets being fixed; by Lemma B.10, every pair is therefore split assortatively. Hence the sets $T_1, \ldots, T_k$ are pairwise non-interleaving and, after relabeling in decreasing attribute order, are the consecutive n-blocks of T. Since $T \subseteq [N]$, the r-th largest attribute of T is at most the r-th largest attribute of $[N]$; as T_ℓ is the ℓ-th consecutive block of T, its r-th largest attribute is the $\big((\ell-1)n + r\big)$-th largest of T, hence at most the $\big((\ell-1)n + r\big)$-th largest of $[N]$, which is the r-th largest attribute of $\widetilde{\pi}_\ell$. By Lemma B.9, $Z(T_\ell) \le Z(\widetilde{\pi}_\ell)$ for every $\ell \le k$. Summing,

$$\sum_{j \le k} Z^{\pi}_{[j]} \le \sum_{\ell \le k} Z(\widetilde{\pi}_\ell) = \sum_{\ell \le k} Z^{\widetilde{\pi}}_{\ell}.$$

Finally, section $\widetilde{\pi}_\ell$ elementwise dominates $\widetilde{\pi}_{\ell+1}$, so $Z^{\widetilde{\pi}}_1 \ge \cdots \ge Z^{\widetilde{\pi}}_K$ by Lemma B.9, whence $\sum_{\ell \le k} Z^{\widetilde{\pi}}_{\ell} = \sum_{j \le k} Z^{\widetilde{\pi}}_{[j]}$. As $k \in [K]$ was arbitrary, this is the definition of $Z^{\pi} \prec_w Z^{\widetilde{\pi}}$. □

PROPOSITION 10. *Consider the social norm model with $\beta = 1/2$. For $n \ge 3$, the assortative partition $\widetilde{\pi}$ maximizes total effort over $\Pi_{\mathcal{I}}$. For $n = 2$, the model is sorting-neutral.*

Proof. By Lemma B.8, for $n \ge 3$ maximizing total effort over $\Pi_{\mathcal{I}}$ is equivalent to maximizing $\sum_{j \in [K]} Z(\pi_j)$.

Assume first that the attributes are distinct. Taking $k = K$ in Lemma B.11 gives $\sum_{j \in [K]} Z(\pi_j) \le \sum_{j \in [K]} Z(\widetilde{\pi}_j)$ for every $\pi \in \Pi_{\mathcal{I}}$, so $\widetilde{\pi}$ is a maximizer.

With ties among the attributes, perturb them to distinct values preserving an assortative ordering, apply the distinct case, and pass to the limit: the unique positive root of (B.22) is continuous in the attributes, so the inequality $\sum_j Z(\pi_j) \le \sum_j Z(\widetilde{\pi}_j)$ is preserved for every $\pi \in \Pi_{\mathcal{I}}$.

Finally, when $n = 2$, the coefficient $b(n-2)$ in (B.23) vanishes, so by Lemma B.8 total effort equals $\sum_i a_i/(1-b)$ under every partition. □

B.2.5. Case study: The USAFA experiment (Carrell et al. 2013) In the USAFA experiment (Carrell et al. 2013), the goal of the intervention was to improve the mean performance of the bottom 1/3 of the students without harming the remaining 2/3 of the cohort. There are many mathematical interpretations of this statement. We examine various mathematical formulations and their implications on performance under the LIM model:

—*Satisfying a majorization constraint*: The new partition M should satisfy

$$\sum_{j=2K/3+1}^{K} X^M_{[j]} > \sum_{j=2K/3+1}^{K} \underline{X}_{[j]}, \quad \sum_{j=K/3+1}^{2K/3} X^M_{[j]} \ge \sum_{j=K/3+1}^{2K/3} \underline{X}_{[j]}, \quad \sum_{j=1}^{K/3} X^M_{[j]} \ge \sum_{j=1}^{K/3} \underline{X}_{[j]},$$

where $\underline{\mathbf{X}}$ represents the summation vector of the initial uniform policy. However, no such M exists. The three blocks partition $[K]$, and the sum of attribute is invariant across partitions, since

$$\sum_{j\in[K]} X^M_{[j]} = \sum_{j\in[K]} \sum_{i\in[N]} \mu_{ij} x_i = \sum_{i\in[N]} x_i = \sum_{j\in[K]} \underline{X}_{[j]}.$$

Summing the three constraints above would therefore give $\sum_{i\in[N]} x_i > \sum_{i\in[N]} x_i$, a contradiction. Under this formulation, the planner's goal is thus infeasible, since no reassignment can weakly increase every block of section sums and strictly increase one.

—*Maximizing $\sum_{j\in[K]} w_j e(\pi_{[j]})$ with increasing section weights $\mathbf{w} \in \mathbb{R}^K$:* Another choice would be to maximize the total weighted effort with increasing weights so that the planner prioritizes the performance of the bottom 1/3 of sections. In this case, the uniform partition over $\Pi_{\mathcal{F}}$ (or a balanced partition over $\Pi_{\mathcal{I}}$) would be optimal (Proposition 6).

—*Maximizing $\sum_{i\in[N]} \beta_i \mathbb{E}[y_i]$ with individual weights $\beta \in \mathbb{R}^N_+$ decreasing in ability:* A third alternative is to assign individual weights β_i to students based on their attributes x_i and maximize the resulting weighted total effort. Unlike Corollary 3, this objective is not sorting-neutral. Setting $B^\pi_j = \sum_{i\in\pi_j} \beta_i$, $A = \frac{\gamma(n-1)}{n-1+\theta_1}$ and $B = \frac{\gamma\theta_1}{(1-\theta_1)(n-1+\theta_1)}$, we get that

$$\sum_{i\in[N]} \beta_i \mathbb{E}[y_i] = A \sum_{i\in[N]} \beta_i x_i + B \sum_{j\in[K]} B^\pi_j X^\pi_j, \tag{B.36}$$

where the first term is independent of π while the second is not. Note that the objective depends on section π_j through two summation vectors, B^π_j and X^π_j.

We consider the following special case. Let $\beta_i = p + r x_i$, then $B_j^\pi = np + r X_j^\pi$, and (B.36) becomes

$$\sum_{i\in[N]} \beta_i \mathbb{E}[y_i] = A \sum_{i\in[N]} \beta_i x_i + Bnp \sum_{i\in[N]} x_i \; + \; Br \sum_{j\in[K]} \left(X_j^\pi\right)^2,$$

so that only the last term varies with π. Since $B > 0$ for $\theta_1 \in (0,1)$ and $\sum_{j\in[K]}(X_j^\pi)^2$ is strictly Schur convex, Propositions 2 and 3 imply that the assortative partition is optimal when $r > 0$, whereas the uniform partition (resp., a balanced partition) is optimal over Π_F (resp., Π_I) when $r < 0$.

Appendix C: Applications beyond education

In this section, we consider two partition problems in non-educational applications: the assignment of experts to teams (Section C.1) and the design of Tullock contests with weakly heterogeneous agents (Section C.2).

C.1. Optimal assignment of experts to teams

Chade and Eeckhout (2018) study the optimal allocation of N experts to K teams of equal size n, when experts differ in the precision of their information, namely their "level of expertise" $x_i \in [\underline{x}, \overline{x}]$, $0 < \underline{x} < \overline{x} < \infty$. Agents solve a joint decision problem indexed by an unknown state s. The prior belief $\widetilde{s}$ is distributed according to $\mathcal{N}(\mu_0, \frac{1}{\tau})$. When assigned to a team, an agent with expertise x_i draws a signal $\widetilde{\xi}_i \sim \mathcal{N}(s, \frac{1}{x_i})$. The pairwise covariance between the signals of two agents with expertise levels x_i and x_j equals $\frac{\rho}{\sqrt{x_i x_j}}$, where $\rho \in (-\frac{1}{n-1}, 1)$. The agents in a team jointly take an action $a \in \mathbb{R}$ to maximize the expected value of $p - (a-s)^2$, where $p > \frac{1}{\tau}$ is an exogenous profit level. Proposition 1 in Chade and Eeckhout (2018) finds that the *team value* equals

$$V(\pi_j) = p - \frac{1}{\tau + C_\rho(\pi_j)},$$

where

$$C_\rho(\pi_j) = \frac{(1+(n-2)\rho)\sum_{i\in\pi_j} x_i - \rho \sum_{i\in\pi_j}\sum_{k\in\pi_j, k\neq i}\sqrt{x_i x_k}}{(1-\rho)(1+(n-1)\rho)}. \tag{C.1}$$

The objective of the planner is to find a partition that maximizes the total value $\sum_{j\in[K]} V(\pi_j)$. Due to the complexity of $C_\rho(\pi_j)$, Chade and Eeckhout (2018) focus on finding the optimal partition in the case of conditional independence, i.e., when $\rho = 0$. Then $V(\pi_j)$ simplifies to

$$V(\pi_j) = p - \frac{1}{\tau + \sum_{i\in\pi_j} x_i}.$$

At $\rho = 0$ the team value is a strictly concave, increasing function of the section sum $\sum_{i\in\pi_j} x_i$, so the total value $\sum_{j\in[K]} V(\pi_j)$ is Schur concave. Proposition 3 therefore applies.

PROPOSITION C.18 **(Propositions 4 and 5 in Chade and Eeckhout (2018)).** *Assume conditional independence, i.e., $\rho = 0$. Then the uniform partition attains the optimal total team value over $\Pi_{\mathcal{F}}$, and a balanced partition attains it over $\Pi_{\mathcal{I}}$.*

Our framework allows us to recover their result at any negative correlation level $\rho \in \left(-\frac{1}{n-1}, 0\right]$. We formally state this result in the following proposition.

PROPOSITION C.19 **(Proposition 8 in Chade and Eeckhout (2018)).** *Fix $\rho \in \left(-\frac{1}{n-1}, 0\right]$. The uniform partition attains the optimal total team value over $\Pi_{\mathcal{F}}$.*

Proof. Let $M=(\mu_{ij})\in\Pi_{\mathcal{F}}$, where μ_{ij} is the fraction of expert i assigned to team j. Under the fractional signal technology of Chade and Eeckhout (2018), expert i contributes a signal of precision $\mu_{ij}x_i$ to team j, so team j has value $\phi_\rho(\mathbf{Z}_j^M)$, where $\mathbf{Z}_j^M=(\mu_{1j}x_1,\dots,\mu_{Nj}x_N)$ and, for $\mathbf{z}\in\mathbb{R}_+^N$,

$$\phi_\rho(\mathbf{z})=p-\frac{1}{\tau+C_\rho(\mathbf{z})},\qquad C_\rho(\mathbf{z})=\frac{(1+(n-2)\rho)\sum_{i\in[N]}z_i-2\rho\sum_{1\le i<k\le N}\sqrt{z_iz_k}}{(1-\rho)(1+(n-1)\rho)}.$$

Fix $\rho\in\left(-\frac{1}{n-1},0\right]$. Note that C_ρ is concave and nonnegative on $\mathbb{R}_+^N$, so ϕ_ρ is also concave.

Since $\sum_{j\in[K]}\mu_{ij}=1$, the uniform partition satisfies $\mathbf{Z}_j^{\underline{M}}=\frac{1}{K}\sum_{l\in[K]}\mathbf{Z}_l^M$, so $\mathbf{Z}^{\underline{M}}\prec\mathbf{Z}^M$ by Lemma A.3; concavity of ϕ_ρ in turn implies $\sum_{j\in[K]}\phi_\rho(\mathbf{Z}_j^{\underline{M}})\ge\sum_{j\in[K]}\phi_\rho(\mathbf{Z}_j^M)$. Thus $\underline{M}$ is optimal over $\Pi_{\mathcal{F}}$. □

C.2. Group contests

Ryvkin (2011) studies how the sorting of heterogeneous players into groups affects the total effort expended in a group contest, using a quadratic approximation of the seminal competition model of Tullock (1980).[10] There are N players partitioned into K groups, and the groups compete in a Tullock contest. A player i's attribute x_i serves as a multiplicative bias on their effort cost; equivalently, x_i can be interpreted as heterogeneity in prize valuations. Let $E_j=\sum_{i\in\pi_j}y_i$ denote the aggregate effort of group j. Each player $i\in\pi_j$ chooses effort $y_i\ge0$ to maximize her payoff

$$U_i(\mathbf{y})=\frac{E_j}{\sum_{\ell\in[K]}E_\ell}-x_i\,g(y_i),\tag{C.2}$$

where g satisfies the following assumption.

Assumption 1. *Let $\underline{c}=\min_{i\in[N]}\{x_i\}$. We assume that: (i) $g(0)=0$; (ii) $g'(0)=0$; (iii) $g'(y)>0$ for $y\in(0,\underline{c}]$; (iv) $g''(y)>0$ for $y\in(0,\underline{c}]$; (v) $g'''(y)$ exists and is finite for $y\in(0,\underline{c}]$.*

The above assumption states that players have a costless drop-out option (part (i)), the marginal cost at zero effort is zero (part (ii)), and the cost function is strictly increasing (part (iii)). Part (iv) ensures the existence and uniqueness of an equilibrium in which all players exert positive effort. Part (v) is a smoothness condition used for the quadratic approximation. Ryvkin (2011) also assumes weak heterogeneity among players' attributes:

Assumption 2. *For all $i\in[N]$, $x_i=\overline{x}(1-r_i)$ for some $r_i\ll1$.*

By construction, Assumption 2 implies that $\sum_{i=1}^N r_i=0$. This technical assumption allows for a quadratic approximation.

[10] Rosenzweig and Xu (2023) recently adopted Tullock's model to model peer effects across two types of students, high- and low-performing.

PROPOSITION C.20 **(Propositions 1 and 2 in Ryvkin (2011))**. *Under Assumption 1, the group contest has a unique Nash equilibrium* $\mathbf{y}$ *where each player exerts effort* $y_i > 0$*, such that*

$$\frac{\sum_{\ell \neq j} E_\ell}{\left(\sum_{\ell \in [K]} E_\ell\right)^2} = x_i\, g'(y_i), \qquad i \in \pi_j,\ j \in [K]. \tag{C.3}$$

Let $\overline{y}_s$ *be the symmetric equilibrium effort level, and set* $\zeta_2 = \frac{g''(\overline{y}_s)\,\overline{y}_s}{g'(\overline{y}_s)}, \zeta_3 = \frac{g'''(\overline{y}_s)\,\overline{y}_s}{g''(\overline{y}_s)}$*. Then, in the quadratic approximation, the deviation of total effort from the symmetric level is*

$$\hat{Y}^{(2)} = \frac{1}{2\zeta_2(1+\zeta_2)} \left[(2\zeta_2 - \zeta_3) \sum_{i \in [N]} r_i^2 - \frac{2\zeta_2 - \zeta_3 + 2\zeta_2(K-1)(\zeta_2 - \zeta_3)}{n\left[(K-1)\zeta_2 + 1\right]^2} \sum_{j \in [K]} \left(\sum_{i \in \pi_j} r_i \right)^2 \right]. \tag{C.4}$$

Therefore, the effort-maximizing partition of players into groups minimizes (respectively, maximizes) the variance in aggregate cost across groups when $\sigma > 0$ *(respectively,* $\sigma < 0$*), where* $\sigma = 2\zeta_2 - \zeta_3 + 2\zeta_2(K-1)(\zeta_2 - \zeta_3)$.

Proposition C.20 implies that the controllable component of aggregate effort is the term $\sum_{j \in [K]} \left(\sum_{i \in \pi_j} r_i\right)^2$. Proposition C.21 characterizes the optimal partition policy based on the value of σ.

PROPOSITION C.21. *In the contest between groups, in the quadratic approximation, the uniform partition (resp., a balanced partition) attains the optimal total effort over* Π_F *(resp.,* Π_I*) when* $\sigma > 0$*, and the assortative partition attains the optimal total effort over both* Π_F *and* Π_I *when* $\sigma < 0$*, where* σ *is defined as in Proposition C.20.*

Proof. By Proposition C.20, the partition that maximizes effort must minimize $\sum_{j \in [K]} \left(\sum_{i \in \pi_j} r_i\right)^2$ when $\sigma > 0$ and maximize it when $\sigma < 0$.

Since by construction $x_i = \overline{x}(1 - r_i)$ for all $i \in [N]$ and $\sum_{i \in [N]} r_i = 0$, it follows that $r_1 \leq r_2 \leq \ldots \leq r_N$, and in particular $\mathbf{r}$ takes negative values. As Propositions 1 and 2 are stated for nonnegative attributes, set $c = \max_{i \in [N]} |r_i|$ and consider the vector $\mathbf{v} \in \mathbb{R}^N_+$ where $v_i = r_i + c \geq 0$. Since $\sum_{i \in [N]} r_i = 0$,

$$\sum_{j \in [K]} \left(\sum_{i \in \pi_j} v_i \right)^2 = \sum_{j \in [K]} \left(\sum_{i \in \pi_j} r_i \right)^2 + K n^2 c^2,$$

so the two objectives differ by a constant and thus share the same maximizers and minimizers. By Proposition C.1 in Marshall et al. (2011), the function $f(\mathbf{V}) = \sum_{j \in [K]} V_j^2$ is Schur convex as a sum of the convex functions $g(V_j) = V_j^2$, so $-f$ is Schur concave.

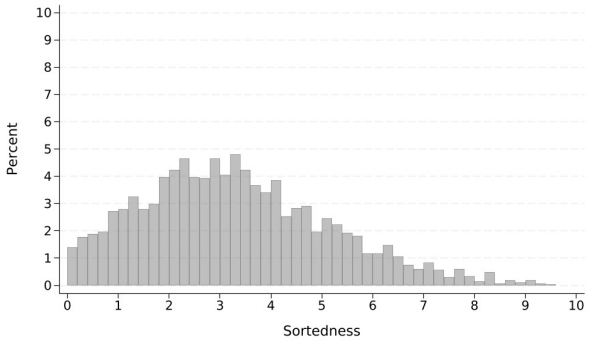
When $\sigma < 0$, Proposition 2 applies to the attribute vector $\mathbf{v}$, implying that the assortative partition with respect to $\mathbf{v}$ is optimal. Since $v_i = 1 + c - x_i/\overline{x}$ is strictly decreasing in x_i and section labels are interchangeable in our setting, this partition is the assortative partition $\widetilde{\pi}$.

When $\sigma > 0$, Proposition 3 applies to $-f$, implying that the uniform partition is optimal over Π_F and that there exists a balanced partition that is optimal over Π_I. Note that v_i is linear and decreasing function of x_i, and we have $|V_j - V_k| = |X_j - X_k|/\overline{x}$ for all $j, k \in [K]$. Therefore, the pair (π_j, π_k) minimizes $|V_j - V_k|$ if and only if it minimizes $|X_j - X_k|$. Thus, a partition is balanced with respect to $\mathbf{v}$ if and only if it is balanced with respect to $\mathbf{x}$. □

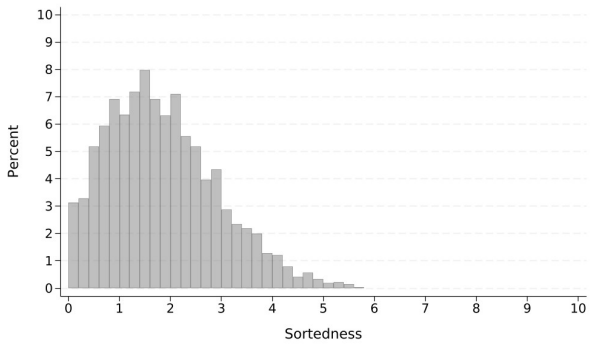
Appendix D: Supplementary material for the empirical results

Figure D.7 Histograms of Cohort Partition Sortedness

Panel A: Mathematics



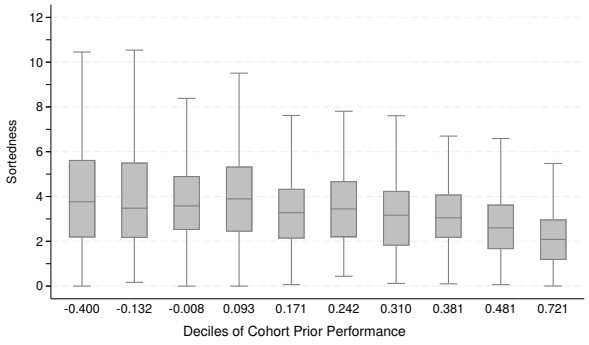
Panel B: Modern Greek Language



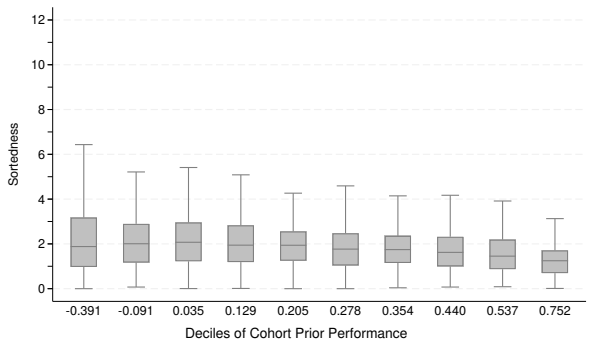
Notes: This figure presents the histogram of our measure of classroom sortedness across all school cohorts. Sortedness is defined as the percentage difference between the Lorenz curve area of the empirical partition generated by quasi-random alphabetical classroom assignment and a theoretically uniform partition. To limit the influence of extreme values, we exclude observations falling outside the 1^{st}–99^{th} percentile range of the distribution, leaving 2,642 observations in Mathematics and 2,645 in Modern Greek.

Figure D.8 Variation in Cohort Partition Sortedness with Cohort Prior Performance

Panel A: Mathematics



Panel B: Modern Greek Language



Notes: This graph shows the relationship between each cohort's partition sortedness in grade 10 and its prior-year performance from grade 9. Sortedness is defined as the percentage difference in Lorenz curve areas between the empirical partition and a theoretically uniform partition. For each decile of cohort mean prior performance, the box plot shows the 25^{th} percentile, median, 75^{th} percentile, minimum, and maximum of the sortedness distribution. The mean value for each decile is also reported as labels on the horizontal axis. To limit the influence of extreme values, we exclude observations falling outside the 1^{st}–99^{th} percentile range of the distribution, leaving 2,642 observations in Mathematics and 2,645 in Modern Greek.

Table D.3 Balancing Tests

	Independent Variable		
Dependent Variable	Cohort Partition Sortedness		
Prior Student Performance, Mathematics	0.00039	0.00093	0.00522
	(0.00063)	(0.00073)	(0.00498)
Prior Student Performance, Modern Greek	0.00122	0.00218	0.01217
	(0.00176)	(0.00276)	(0.01533)
Female (1=Yes)	-0.00056		
	(0.00062)		
School FE	✓	✓	✓
Year FE	✓	✓	✓
Cohort Mean & SD Prior Performance	✓	✓	✓
Student Gender		✓	✓
Teacher FE			✓

Notes: This table reports balance checks of the cohort partition sortedness with respect to student characteristics. Sortedness is defined as the percentage difference between the areas under a theoretically uniform Lorenz curve and the empirical Lorenz curve. Each row corresponds to a regression where the listed student characteristic from grade 9 is the dependent variable and the student's cohort partition sortedness in the grade 10 is the independent variable. All regressions control for the mean and standard deviation of prior cohort performance, school fixed effects, and year fixed effects. Clustered standard errors at school level are shown in parentheses.

Table D.4 Linear Effects of Cohort Partition Sortedness on Academic Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	Mathematics			Modern Greek Language		
All Students	-0.012	-0.012	-0.016	0.023	0.024	0.029
	(0.011)	(0.011)	(0.011)	(0.022)	(0.022)	(0.022)
By Tertile of Prior Student Performance						
Bottom Tertile	-0.002	-0.002	-0.004	0.038	0.037	0.043*
	(0.011)	(0.011)	(0.011)	(0.023)	(0.023)	(0.023)
Middle Tertile	-0.015	-0.015	-0.017	0.018	0.019	0.028
	(0.011)	(0.011)	(0.011)	(0.023)	(0.023)	(0.022)
Top Tertile	-0.018*	-0.018*	-0.021*	0.017	0.018	0.029
	(0.011)	(0.011)	(0.011)	(0.022)	(0.022)	(0.022)
Observations	201,354	201,354	201,354	201,127	201,126	201,126
Y Mean	0.096	0.096	0.096	0.023	0.023	0.023
Y SD	1.017	1.017	1.017	0.982	0.982	0.982
School FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Cohort Mean Prior Perf.	✓	✓	✓	✓	✓	✓
Cohort SD Prior Perf.	✓	✓	✓	✓	✓	✓
Student Prior Perf.	✓	✓	✓	✓	✓	✓
Student Gender		✓	✓		✓	✓
Teacher FE			✓			✓

Notes: Academic performance is measured by standardized grade 10 final exam scores, with prior performance captured by standardized grade 9 final exam scores. Sortedness is defined as the percentage difference in Lorenz curve areas between the empirical partition and a theoretically uniform partition. Clustered standard errors at school level are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table D.5 Nonlinear Effects of Cohort Partition Sortedness on Academic Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Sortedness Quintile	Mathematics			Modern Greek Language		
Second Quintile	0.064***	0.064***	0.033	0.013	0.014	-0.011
	(0.022)	(0.022)	(0.024)	(0.021)	(0.021)	(0.021)
Third Quintile	0.056**	0.056**	0.044*	0.019	0.019	0.006
	(0.022)	(0.022)	(0.024)	(0.030)	(0.029)	(0.029)
Fourth Quintile	0.077***	0.077***	0.046*	0.046	0.045	0.033
	(0.025)	(0.025)	(0.027)	(0.040)	(0.040)	(0.040)
Top Quintile	0.117***	0.117***	0.076**	0.066	0.066	0.036
	(0.032)	(0.032)	(0.035)	(0.062)	(0.062)	(0.059)
School FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Cohort Mean Prior Perf.	✓	✓	✓	✓	✓	✓
Cohort SD Prior Perf.	✓	✓	✓	✓	✓	✓
Student Prior Perf.	✓	✓	✓	✓	✓	✓
Student Gender		✓	✓		✓	✓
Teacher FE			✓			✓

Notes: Academic performance is measured by standardized grade 10 final examination scores. Prior performance is measured by standardized grade 9 final examination scores. Sortedness is defined as the percentage difference between the areas under a theoretically uniform Lorenz curve and the empirical Lorenz curve. Higher values indicate greater sorting of students across classroom partitions. Coefficients represent differences relative to the bottom quintile of cohort partition sortedness. Standard errors clustered at school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table D.6 Nonlinear Heterogeneous Effects of Cohort Partition Sortedness by Prior Student Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	Mathematics			Modern Greek Language		
Bottom Tertile: Q2	0.054**	0.054**	0.020	0.047*	0.044*	0.018
	(0.025)	(0.025)	(0.026)	(0.026)	(0.026)	(0.024)
Bottom Tertile: Q3	0.050**	0.050**	0.038	0.066*	0.062*	0.041
	(0.024)	(0.024)	(0.026)	(0.036)	(0.035)	(0.034)
Bottom Tertile: Q4	0.069**	0.069**	0.035	0.118**	0.112**	0.093*
	(0.027)	(0.027)	(0.029)	(0.050)	(0.050)	(0.048)
Bottom Tertile: Q5	0.128***	0.128***	0.083**	0.201***	0.196***	0.141**
	(0.034)	(0.034)	(0.037)	(0.076)	(0.075)	(0.071)
Middle Tertile: Q2	0.079***	0.079***	0.045	0.009	0.012	-0.009
	(0.027)	(0.027)	(0.028)	(0.023)	(0.023)	(0.023)
Middle Tertile: Q3	0.071***	0.071***	0.056**	0.011	0.015	0.014
	(0.027)	(0.027)	(0.028)	(0.032)	(0.032)	(0.031)
Middle Tertile: Q4	0.085***	0.085***	0.049	0.031	0.031	0.038
	(0.029)	(0.029)	(0.031)	(0.044)	(0.043)	(0.043)
Middle Tertile: Q5	0.133***	0.133***	0.089**	0.048	0.051	0.050
	(0.038)	(0.038)	(0.040)	(0.070)	(0.070)	(0.065)
Top Tertile: Q2	0.055**	0.055**	0.022	-0.032	-0.028	-0.049**
	(0.024)	(0.024)	(0.026)	(0.021)	(0.021)	(0.022)
Top Tertile: Q3	0.042	0.042	0.029	-0.042	-0.040	-0.045
	(0.026)	(0.026)	(0.028)	(0.030)	(0.030)	(0.029)
Top Tertile: Q4	0.080***	0.080***	0.043	-0.041	-0.038	-0.044
	(0.029)	(0.029)	(0.031)	(0.041)	(0.040)	(0.042)
Top Tertile: Q5	0.089**	0.090**	0.045	-0.104	-0.098	-0.107
	(0.036)	(0.036)	(0.041)	(0.064)	(0.064)	(0.065)
Observations	201,354	201,354	201,354	201,127	201,126	201,126
Y Mean	0.096	0.096	0.096	0.023	0.023	0.023
Y SD	1.017	1.017	1.017	0.982	0.982	0.982
School FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Cohort Mean Prior Perf.	✓	✓	✓	✓	✓	✓
Cohort SD Prior Perf.	✓	✓	✓	✓	✓	✓
Student Prior Perf.	✓	✓	✓	✓	✓	✓
Student Gender		✓	✓		✓	✓
Teacher FE			✓			✓

Notes: Academic performance is measured by standardized grade 10 final examination scores. Prior performance is measured by standardized grade 9 final examination scores. Coefficients represent differences relative to the bottom quintile of cohort partition sortedness. Standard errors clustered at school level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table D.7 Effect of Cohort Partition Sortedness on Academic Performance Without Controlling for the Distribution of Cohort Prior Achievement

	(1)	(2)	(3)	(4)	(5)	(6)
	Mathematics			Modern Greek Language		
All Students	0.006*	0.006*	0.005	0.016***	0.015**	0.029***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)
By Tertile of Prior Student Performance						
Bottom Tertile	0.011***	0.011***	0.011***	0.018**	0.017**	0.030***
	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)
Middle Tertile	-0.002	-0.002	-0.002	0.002	0.002	0.017***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)
Top Tertile	-0.004	-0.004	-0.005	0.003	0.003	0.018***
	(0.003)	(0.003)	(0.004)	(0.006)	(0.006)	(0.006)
Observations	201,354	201,354	201,354	201,127	201,126	201,126
Y Mean	0.096	0.096	0.096	0.023	0.023	0.023
Y SD	1.017	1.017	1.017	0.982	0.982	0.982
School FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Student Prior Perf.	✓	✓	✓	✓	✓	✓
Student Gender		✓	✓		✓	✓
Teacher FE			✓			✓

Notes: Academic performance is measured by standardized grade 10 final exam scores, with prior performance captured by standardized grade 9 final exam scores. Sortedness is defined as the percentage difference in Lorenz curve areas between the empirical partition and a theoretically uniform partition. Higher values indicate greater sorting of students across classroom partitions. Clustered standard errors at school level are shown in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.