

COMPETITIVE MANY-TO-ONE MATCHING: SORTING VS. EQUALITY

ANTON KOLOTILIN

School of Economics, UNSW Business School

ALEXANDER WOLITZKY

Department of Economics, MIT

We study many-to-one matching with transfers and peer effects, such as matching workers to firms, students to schools, residents to neighborhoods, or consumers to status goods. With flexible prices (as in the labor market), competitive equilibrium exists and is efficient under general conditions. We characterize when workforces are segregated by skill and matched to firms in a positively assortative manner. In general, equilibrium features alternating intervals of workforce segregation and compression (mixing). Comparative statics characterize when workforces are more segregated or more compressed, and when profits and wages are more or less unequal. With uniform prices (as in school or neighborhood choice), the value generated by peer effects accrues to schools rather than students, and equilibrium can be excessively segregated. Our model generalizes both assignment models (optimal transport) and Bayesian persuasion.

KEYWORDS: matching, competitive equilibrium, segregation, compression, inequality, assignment model, optimal transport, Bayesian persuasion.

JEL CLASSIFICATION. C78, D51, D82, J00.

Anton Kolotilin: akolotilin@gmail.com

Alexander Wolitzky: wolitzky@mit.edu

Date: May 28th, 2026

We thank Roberto Corrao for helpful early stage discussions on this research. We are grateful for feedback from seminar participants at AETW, APIOC, Bristol, BC, Brown, CMU, Columbia, ESWC, Harvard, LSE, MIT, Monash, NTU, NYU, Oxford, Penn, Penn State, and UCL. Kolotilin gratefully acknowledges financial support from the ARC DP240103257 and FT240100731.

1. INTRODUCTION

Many economic settings involve many-to-one matching among heterogeneous individuals. Firms of varying productivity hire workers of varying skill. Schools of varying quality admit students of varying ability. Neighborhoods with varying amenities house heterogeneous residents. Status goods are purchased by heterogeneous consumers.

What matching patterns arise in these markets? Do workers segregate by skill, so the best firms hire only the best workers? Or, do firms hire a mix of workers with different skill levels, leading to a more equal distribution of workforces? When are workforces (or student bodies, or neighborhoods) more sorted, and when are they more equal?

These questions can be seen at the heart of current social and economic debates. Influential research has linked rising wage inequality to rising assortativity between firms and workers (Card et al., 2013, Song et al., 2019, Kline, 2024). Proposed explanations include skill-biased technological change (Acemoglu and Autor, 2011, Håkanson et al., 2021, Cortes et al., 2023), outsourcing (Goldschmidt and Schmieder, 2017, Handwerker, 2023), and factor supply and demand shocks (Haanwinckel, 2025). Yet, standard assignment models focus on one-to-one matching (Sattinger, 1993, Terviö, 2008)—where firms match with individual workers, not *workforces*—and are thus silent as to what changes in production technologies lead firms to hire more or less homogeneous workforces. Another longstanding debate concerns whether neighborhoods, schools, and other social institutions are inefficiently segregated by income or ability: i.e., whether society is overly sorted or meritocratic. See, e.g., Benabou (1996), Becker and Murphy (2000), Durlauf (2004), and Bishop and Cushing (2009) on residential segregation, Arnott and Rowse (1987) and Epple and Romano (1998) on schools, and Markovits (2019) and Sandel (2020) on society more broadly. Again, existing models do not provide a tractable framework for studying equilibrium matching patterns, utilities, or comparative statics in these settings, especially with rich heterogeneity on both sides of the market.

The current paper proposes such a framework. We study competitive equilibrium in continuum-agent, many-to-one matching markets with quasilinear utility and both *match effects*—output depends on the pattern of cross-side matching, such as workers to firms—and *peer effects*—output also depends on the pattern of within-side matching, such as workers to coworkers. Competitive equilibrium exists and is efficient. The key tradeoff in the

model is between sorting and equality. Suppose a firm of productivity x that hires a workforce with mean skill z generates value $v(x, z)$. (This is the baseline, *mean-measurable* case of our model, where the skill distribution in a workforce affects output only through its mean z .) Assume that there are increasing differences between firm productivity and workforce skill ($v_{xz} > 0$), but that there are also decreasing returns to workforce skill ($v_{zz} < 0$). Increasing differences implies that, whatever distribution of workforce means z forms in equilibrium, they will be matched with firms in a positively assortative manner. Thus, when the distribution of workforces is more spread out, more value is created by matching better workforces to better firms—but, at the same time, more value is wasted due to decreasing returns to skill. There is thus a tradeoff between sorting workers across workforces—to capitalize on increasing differences—and mixing workers to create more equal workforces—to economize on decreasing returns.

We show how this tradeoff resolves. In the mean-measurable case, equilibrium is described by the solution to a concave maximization problem, subject to majorization constraint, which says that the equilibrium distribution of workforce skills is compressed relative to the population distribution of worker skills. If increasing differences are stronger than decreasing returns (or if there are increasing returns, $v_{zz} \geq 0$), the equilibrium is *positive assortative segregation (PAS)*: workforces are segregated by skill and matched to firms in a positively assortative manner. PAS plays an important role in our model, as the many-to-one analogue of the positive assortative matching pattern characterized by [Becker \(1973\)](#). If instead increasing differences are weaker than decreasing returns, the equilibrium is *full compression*: essentially, all workforces are mixed, but more productive firms still hire strictly more skilled workforces. We also provide conditions for other simple matching patterns, such as *upper compression*, where less productive firms hire segregated workforces, while more productive firms hire mixed ones. In general, equilibrium is characterized by alternating intervals of segregation and compression, where at each point either the majorization constraint binds or the marginal value of workforce skill is increasing.

We also obtain comparative statics on the equilibrium assignment and utilities. In particular, we show what changes in the production function $v(x, z)$ make the equilibrium assignment more segregated by skill. Essentially, equilibrium becomes more segregated if and only if increasing differences become stronger relative to decreasing returns. We also show that comparative statics for wages have both similarities and differences from

the one-to-one matching case (e.g., [Terviö, 2008](#)). One notable difference is that, in our model, a worker’s wage may fall when higher-skilled workers become even more productive, whereas her wage is unaffected by this change under one-to-one matching. The reason is that workers do not directly compete with higher-skilled workers under one-to-one matching, while in our model all workers are in competition, as their skills substitute for each other in improving workforces.

Our baseline, mean-measurable model admits a unique equilibrium matching between firm productivities x and mean workforce skills z . However, many workforce compositions imply the same distribution of z , so the equilibrium matching between firms and individual workers is not uniquely determined. We show how this indeterminacy resolves when firms of different productivities x value workforces according to slightly different moments, all of which are close to the mean z . In particular, we show when the equilibrium assignment is *strictly single-dipped*—meaning that more productive firms employ more heterogeneous workforces—or, conversely, *strictly single-peaked*. These results refine the relationship between firm productivity and workforce composition predicted in the mean-measurable case. In addition, we also characterize when the equilibrium is PAS in the general, non-mean-measurable case.

The above results pertain to competitive equilibrium, which is always efficient. However, competitive equilibrium implicitly allows personalized prices, which are not reasonable in all of our applications: firms pay different wages to different workers, but schools might or might not charge different tuition to different students, and the price of a house in a given neighborhood is generally not personalized to different prospective buyers. We therefore contrast competitive equilibrium with *uniform price equilibrium*, where there is a single price to match with each agent on the “one” side of the market (e.g., a single price for each school or neighborhood). In markets without personalized prices, uniform price equilibrium is the appropriate descriptive equilibrium concept, and competitive equilibrium serves as an efficiency benchmark. Generalizing earlier results of [Benabou \(1996\)](#) and [Becker and Murphy \(2000\)](#) (among others), we characterize when uniform price equilibria are *inefficiently segregated*, meaning that the uniform price equilibrium is PAS, but the competitive equilibrium is not. We also obtain the apparently novel result that moving from competitive equilibrium to uniform price equilibrium impacts utilities precisely by shifting the value

generated by peer effects from students to schools. This result suggests that agreements among colleges to forego merit scholarships may warrant regulatory scrutiny.¹

Related Literature. Our model is a minimal mutual generalization of optimal transport and Bayesian persuasion. It also generalizes prior work on many-to-one matching with peer effects. These connections are explored in more detail in the body of the paper—here, we give a brief overview.

The special case of our model with no peer effects is one-to-one matching with transfers and a continuum of agents, aka the continuum assignment model, aka optimal transport. The discrete version of this model traces to [Koopmans and Beckmann \(1957\)](#), [Shapley and Shubik \(1971\)](#), and [Becker \(1973\)](#). [Sattinger \(1975, 1993\)](#) derived the wage equation, [Terviö \(2008\)](#) derived its comparative statics, and [Gabaix and Landier \(2008\)](#) applied it to study CEO compensation. The continuum assignment model is due to [Gretsky et al. \(1992\)](#). These and other applications of optimal transport are surveyed by [Galichon \(2018\)](#).

The special case of our model with no match effects is mathematically equivalent to the Bayesian persuasion problem of [Kamenica and Gentzkow \(2011\)](#), where a “sender” splits a distribution of states into a distribution of posteriors, to maximize the expectation of a function of the posterior. As far as we know, the mean-measurable version of this problem first appeared in [Saint-Paul \(2001\)](#), in the labor market context. In the persuasion literature, it was studied by [Gentzkow and Kamenica \(2016\)](#), [Kolotilin et al. \(2017\)](#), [Kolotilin \(2018\)](#), [Dworczak and Martini \(2019\)](#), and [Kleiner et al. \(2021\)](#), among others.

A few prior papers study many-to-one matching with linear peer effects. [Kremer \(1993\)](#) studies the production function $v(x, z) = xe^z$, with $x > 0$ for all firms. This function has increasing returns to skill ($v_{zz} \geq 0$), which directly implies that equilibrium is PAS: see Remark 10. [Boerma et al. \(2025\)](#) study the production function $v(x, z) = xe^z$, with $x < 0$ for all firms. This function has decreasing returns to skill ($v_{zz} < 0$) and is thus covered

¹Such an agreement among elite US colleges was the subject of the “Overlap case” ([U.S. Court of Appeals for the Third Circuit, 1993](#)) and [Henry v. Brown \(U.S. District Court for the Northern District of Illinois, 2022\)](#). A counterpoint is that the money colleges saved by forgoing merit aid supported need-based aid and other pro-social goals ([Carlton et al., 1995](#)).

by our main results.² Other papers that study matching with linear peer effects in particular contexts include [Arnott and Rowse \(1987\)](#) and [Epple and Romano \(1998\)](#) on schools, [De Bartolome \(1990\)](#), [Benabou \(1996\)](#), and [Becker and Murphy \(2000\)](#) on neighborhoods, and [Rayo \(2013\)](#) on status goods.³

The literature on many-to-one matching in discrete economies dates to [Kelso and Crawford \(1982\)](#). Other than requiring substitutability between workers, their model is very general; consequently, equilibrium lacks a tight characterization. Continuum-agent generalizations, including [Che et al. \(2019\)](#), [Azevedo and Hatfield \(2018\)](#), and [Leshno \(2022\)](#), likewise mostly focus on existence and other basic properties, rather than tightly characterizing matching patterns and comparative statics. The same comment applies to the literature on general equilibrium with “clubs” (e.g., [Ellickson et al., 1999](#), [Rahman, 2026](#)).

We also relate to a mathematics literature on *weak optimal transport*, initiated by [Gozlan et al. \(2017\)](#) (see Remark 5). The closest papers in this literature are [Alibert et al. \(2019\)](#) and [Backhoff-Veraguas et al. \(2020\)](#), whose results anticipate parts of our Theorem 1 and Corollary 1, respectively.

The remainder of the paper is organized as follows. Section 2 present our model and competitive equilibrium notion. Section 3 establishes equilibrium existence and efficiency. Section 4 characterizes when equilibrium is PAS. Section 5 specializes the model to the mean-measurable case. This section forms the core of our analysis: it characterizes the equilibrium matching and utilities and provides comparative statics. Section 6 shows how perturbing the model away from mean-measurability determines individual workers’ assignments. Section 7 defines uniform price equilibrium and contrasts it with competitive equilibrium. Section 8 discusses possible extensions.

²[Kremer \(1993\)](#) and [Boerma et al. \(2025\)](#) assume that production depends on the product of worker abilities, so their production functions are mean-measurable after taking log’s. [Boerma et al.](#)’s production function is then $v(\tilde{x}, z) = \tilde{x}(1 - e^{-z})$ with $\tilde{x} > 0$, which is equivalent to $v(x, z) = xe^z$, with $x = -\tilde{x} < 0$, after dropping the constant \tilde{x} . A difference from our model is that these authors assume that each firm hires a finite number of workers, rather than a continuum.

³Farther afield, [Lindenlaub \(2017\)](#), [Edmond and Mongey \(2022\)](#), and [Chone et al. \(2025\)](#) study sorting and bundling workers with multidimensional skills.

2. MODEL

We study many-to-one matching between unit measures of heterogeneous firms and workers. Firm types (productivity) $x \in X$ are distributed $\mu \in \Delta(X)$, and worker types (skill) $y \in Y$ are distributed $\nu \in \Delta(Y)$, where X and Y are compact metric spaces.⁴ Each firm x must hire a unit mass of workers $\eta \in \Delta(Y)$, which we call a *workforce*, and each worker must find employment. Thus, an *assignment* is a distribution $\gamma \in \Delta(X \times \Delta(Y))$ over (*firm, workforce*) pairs (x, η) such that its marginal over X equals μ (each firm hires a unit mass of workers) and the average workforce η equals ν (each worker is employed): that is, $\gamma_X = \mu$ and $\mathbb{E}_\gamma[\eta] = \nu$.

All agents (firms and workers) have quasi-linear utility. A matched firm-workforce pair (x, η) generates *value* $V(x, \eta)$. Our main technical assumption is that V is Lipschitz continuous: $V \in \text{Lip}(X \times \Delta(Y))$.

We define a competitive equilibrium as follows.

DEFINITION 1: A *competitive equilibrium* is an assignment γ together with utilities (or *prices*) $(p, q) \in \text{Lip}(X) \times \text{Lip}(Y)$ satisfying

$$\begin{aligned} p(x) + \mathbb{E}_\eta[q(y)] &= V(x, \eta), & \text{for all } (x, \eta) \in \text{supp}(\gamma), \\ p(x) + \mathbb{E}_\eta[q(y)] &\geq V(x, \eta), & \text{for all } (x, \eta) \in X \times \Delta(Y). \end{aligned}$$

The first condition says that each firm x makes profit $p(x)$ from hiring workforce η at wages q ; the second condition says that firm x cannot make more profit by hiring a different workforce. Equivalently, $(\gamma; p, q)$ is a core or group-stable outcome (see Remark 4).

Note that if $(\gamma; p, q)$ is an equilibrium, then so is $(\gamma; p + c, q - c)$, for any constant $c \in \mathbb{R}$. That is, utilities are only determined up to a constant that divides the surplus between the

⁴For any compact metric space X , ρ_X denotes its metric, $\text{Lip}(X)$ denotes the set of Lipschitz functions on X , and $\Delta(X)$ denotes the set of Borel probability measures on X , endowed with the Kantorovich-Rubinstein distance, induced by the norm (16). A subset of X will always mean a Borel subset of X . For any $\mu \in \Delta(X)$, its support $\text{supp}(\mu)$ is the smallest compact subset of X of measure one.

two sides of the market. This feature is typical of assignment models (Sattinger, 1993, Boerma et al., 2025).⁵

REMARK 1—Relation to Assignment Models and Bayesian Persuasion: The model is a minimal generalization of assignment models (one-to-one matching) and Bayesian persuasion. If V is linear in η , it reduces to one-to-one matching: i.e., the continuum assignment model as in Gretsky et al. (1992, 1999), or the optimal transport problem as in Santambrogio (2015) or Galichon (2018).⁶ If instead X is a singleton, it reduces to Bayesian persuasion as in Kamenica and Gentzkow (2011).⁷

REMARK 2—Matching with Peer Effects and Personalized Prices: Our model captures matching with peer effects, because if a worker with skill y working at a firm of productivity x with co-workers with skill distribution η generates value $W(x, y, \eta)$, we recover our model with $V(x, \eta) = \mathbb{E}_\eta[W(x, y, \eta)]$. With this interpretation, the special case of our model with no peer effects—where W does not depend on η —is the continuum assignment model/optimal transport problem. Conversely, the special case with no match effects—where W does not depend on x —is the Bayesian persuasion problem.

In the worker-firm interpretation of the model, the value $W(x, y, \eta)$ generated by firm-worker-workforce triple (x, y, η) accrues to the firm: the firm’s profit is $p(x) = \mathbb{E}_\eta[W(x, y, \eta) - q(y)]$, and worker y ’s wage is $q(y)$. In other applications, $W(x, y, \eta)$ instead accrues to the “worker” (the agent on the “many” side of the market). For instance, in the student-school interpretation, $W(x, y, \eta)$ is the willingness-to-pay of a type- y student to attend a type- x school with student body η ; the tuition charged to this student is

⁵Some assignment models pin down the constant by introducing outside options (Kremer, 1993, Terviö, 2008) or a size imbalance between the sides of the market. These features could likewise be introduced in our model.

⁶A minor difference with Gretsky et al. (1992, 1999) is that they assume free disposal, so an assignment γ is a positive measure on $X \times \Delta(Y)$ such that $\gamma_X \leq \mu$ and $\int \eta \gamma(dx, d\eta) \leq \nu$. If $V \geq 0$ and $\mu(X) = \nu(Y)$, free disposal is immaterial, as $\gamma_X = \mu$ and $\int \eta \gamma(dx, d\eta) = \nu$ in any competitive equilibrium.

⁷Explicitly, our model with singleton X is described by a distribution ν of workers, which must be split into a distribution $\tau \in \Delta(\Delta(Y))$ of workforces $\eta \in \Delta(Y)$, with $\mathbb{E}_\tau[\eta] = \nu$, to maximize $\mathbb{E}_\tau[V(\eta)]$. This is precisely the Bayesian persuasion problem of Kamenica and Gentzkow (2011), with workers replacing states of the world, the distribution τ replacing an experiment, workforces replacing posteriors, and the production function $V(\eta)$ replacing the sender’s indirect utility from inducing posterior η .

$W(x, y, \eta) - q(y)$ (leaving her utility $W(x, y, \eta) - (W(x, y, \eta) - q(y)) = q(y)$); and the total tuition received by the school is $p(x) = \mathbb{E}_\eta[W(x, y, \eta) - q(y)]$. The neighborhood effects and status goods applications are similar.⁸

In applications where $W(x, y, \eta)$ accrues to the “worker,” competitive equilibrium implicitly involves personalized prices. For example, if two students y and y' attend the same school x with student body η , their tuitions $W(x, y, \eta) - q(y)$ and $W(x, y', \eta) - q(y')$ can differ. Personalized prices may not be very realistic in the student-school and status goods applications (merit scholarships and giveaways to influencers notwithstanding), and they are likely unrealistic in the neighborhood application (as emphasized by [Becker and Murphy, 2000](#)). Thus, in these applications, our competitive equilibrium notion is relevant as an efficiency benchmark rather than a positive prediction, and the corresponding positive theory is instead given by the uniform-price equilibrium notion developed in Section 7.

3. EXISTENCE AND WELFARE THEOREMS

Our first result is that (i) competitive equilibria exist, (ii) equilibrium assignments maximize total value, and (iii) equilibrium prices minimize total utility subject to a no-entry condition. Formally, define the *planner’s problem*

$$\begin{aligned} & \max_{\gamma \in \Delta(X \times \Delta(Y))} \mathbb{E}_\gamma[V(x, \eta)] \\ \text{s.t. } & \gamma_X = \mu \quad \text{and} \quad \mathbb{E}_\gamma[\eta] = \nu, \end{aligned} \tag{P}$$

and define the *dual problem*

$$\begin{aligned} & \min_{(p, q) \in \text{Lip}(X) \times \text{Lip}(Y)} \mathbb{E}_\mu[p(x)] + \mathbb{E}_\nu[q(y)] \\ \text{s.t. } & p(x) + \mathbb{E}_\eta[q(y)] \geq V(x, \eta), \quad \text{for all } (x, \eta) \in X \times \Delta(Y). \end{aligned} \tag{D}$$

The planner’s problem is to choose an assignment to maximize total value. This problem can be broken into a Bayesian persuasion problem followed by an optimal transport problem: first, choose a distribution $\tau \in \Delta(\Delta(Y))$ over workforces η satisfying $\mathbb{E}_\tau[\eta] = \nu$;

⁸For example, in the neighborhood interpretation, the willingness-to-pay of a type- y household to live in a type- x neighborhood with composition η is $W(x, y, \eta)$; the rent charged to this household is $W(x, y, \eta) - q(y)$; and the average rent in the neighborhood is $p(x) = \mathbb{E}_\eta[W(x, y, \eta) - q(y)]$.

then, choose a transport plan $\gamma \in \Delta(X \times \Delta(Y))$ between the (exogenous) distribution μ of firms and the (endogenous) distribution τ of workforces satisfying $\gamma_X = \mu$ and $\gamma_{\Delta(Y)} = \tau$. The dual problem is to minimize total utility subject to the constraint that no entering (*firm, workforce*) pair can make a strictly positive profit.⁹

THEOREM 1: *A competitive equilibrium exists. Moreover, $(\gamma; p, q)$ is a competitive equilibrium if and only if γ solves the planner's problem (P) and (p, q) solves the dual problem (D).*

This strong duality result forms the basis of our analysis, as it lets us study competitive equilibria via the optimization problems (P) and (D). The bulk of the proof consists in showing that the value of (P) as a function of the distributions μ and ν , denoted $U(\mu, \nu)$, inherits Lipschitz continuity from $V(x, \eta)$, at which point the theorem follows from an infinite-dimensional separating hyperplane theorem (namely, the Duality Theorem in [Gale, 1967](#)). In turn, Lipschitz preservation follows the usual line of argument for why the value function in parameterized optimization problems inherits Lipschitz continuity from the objective function ([Milgrom and Segal, 2002](#)), once we show that the set of feasible assignments varies with the population distributions in a well-behaved manner. Specifically, the key technical step is showing that if γ is a feasible assignment for population measures (μ, ν) , then for any population measures $(\hat{\mu}, \hat{\nu})$, there exists a feasible assignment $\hat{\gamma}$ such that $\|\gamma - \hat{\gamma}\|_{KR} \leq \|\mu - \hat{\mu}\|_{KR} + \|\nu - \hat{\nu}\|_{KR}$, where $\|\cdot\|_{KR}$ is the Kantorovich-Rubinstein norm. This is proved by using optimal transport plans between μ and $\hat{\mu}$, and between ν and $\hat{\nu}$, to construct $\hat{\gamma}$.¹⁰

REMARK 3—Price Functions: The equilibrium utilities satisfy

$$p(x) = \max_{\eta \in \Delta(Y)} \{V(x, \eta) - \mathbb{E}_\eta[q(y)]\}, \quad \text{for all } x \in \text{supp}(\mu), \quad \text{and}$$

⁹Mathematically, the dual formulation is justified by the Kantorovich-Rubinstein theorem that a continuous linear function on $M(X) \times M(Y)$ (where $M(X)$ is the space of signed measures on X endowed with the Kantorovich-Rubinstein norm) is represented by $\int p d\mu + \int q d\nu$ for some Lipschitz functions p and q (e.g., Theorem 0 in [Hanin, 1992](#)).

¹⁰In the special case of Bayesian persuasion (where X is a singleton), this argument replicates the proof of Theorem 4 in [Dworczak and Kolotilin \(2024\)](#).

$$q \in \arg \min_{\hat{q} \in \text{Lip}(Y)} \{ \mathbb{E}_\nu[\hat{q}(y)] : \mathbb{E}_\eta[\hat{q}(y)] \geq \max_{x \in X} \{ V(x, \eta) - p(x) \}, \text{ for all } \eta \in \Delta(Y) \}.$$

The first equation says that the profit $p(x)$ of firm x equals its maximum profit from hiring any workforce η at wages q . This is analogous to the profit equation in optimal transport (e.g., [Santambrogio, 2015](#), Proposition 1.11; [Galichon, 2018](#), Proposition 2.3). The second equation says that the wage schedule $q(y)$ is the lowest one such that the total wage $\mathbb{E}_\eta[\hat{q}(y)]$ of any workforce η is at least the value it generates at any firm x , net of the firm's profit $p(x)$. This is analogous to the dual problem in Bayesian persuasion ([Dworczak and Kolotilin, 2024](#)).

REMARK 4—Relation to General Equilibrium: Our economy is convex as we have a continuum of firms and workers, as in [Aumann \(1964\)](#): mathematically, $U(\mu, \nu)$ is concave as the value of a linear program. As in [Aumann \(1964\)](#) and [Gretsky et al. \(1992, 1999\)](#), say that $(p, q) \in \text{Lip}(X) \times \text{Lip}(Y)$ is the *core* if $U(\hat{\mu}, \hat{\nu}) \leq \mathbb{E}_{\hat{\mu}}[p(x)] + \mathbb{E}_{\hat{\nu}}[q(y)]$ for all $(\hat{\mu}, \hat{\nu}) \in \Delta(X) \times \Delta(Y)$, with equality at (μ, ν) , meaning that there is no blocking coalition of firms and workers. Mathematically, this says that (p, q) is a *supergradient* of U at (μ, ν) . By conjugate duality (e.g., Theorem 16 in [Rockafellar, 1974](#)), [\(D\)](#) has a solution and the values of [\(P\)](#) and [\(D\)](#) coincide at (μ, ν) if and only if U has a supergradient at (μ, ν) .¹¹ Core and competitive equilibrium are thus equivalent in our model (as in [Aumann, 1964](#) and [Gretsky et al., 1992, 1999](#)), and the contribution of Theorem 1 is showing that if V is Lipschitz continuous then U has a supergradient. To do so, we build on [Gale \(1967\)](#), whose duality theorem implies that U has a supergradient at (μ, ν) if and only if it has *bounded steepness* at (μ, ν) , meaning that there exists a constant L such that $U(\hat{\mu}, \hat{\nu}) - U(\mu, \nu) \leq L \|(\hat{\mu}, \hat{\nu}) - (\mu, \nu)\|_{KR}$, for all $(\hat{\mu}, \hat{\nu}) \in \Delta(X) \times \Delta(Y)$. So, bounded steepness is equivalent to equilibrium existence,¹² and is implied by Lipschitz continuity. Finally, as described above, we show that U inherits Lipschitz continuity from V by combining an envelope

¹¹[\(P\)](#) has a solution by the Weierstrass Theorem.

¹²This observation applies to any general equilibrium setting with a normed vector space of commodities and quasi-linear utility (e.g., [Ostroy, 1984](#), [Gretsky et al., 2002](#)). For comparison, the main technical result in the infinite-dimensional general equilibrium literature is that *properness* of preferences is equivalent to quasi-equilibrium existence in a one-consumer economy ([Mas-Colell, 1986](#), [Mas-Colell and Zame, 1991](#)). Since any equilibrium is a quasi-equilibrium, bounded steepness implies properness under quasi-linear utility. Under mild

argument as in [Milgrom and Segal \(2002\)](#) and a generalization of the optimal transport argument in [Dworczak and Kolotilin \(2024\)](#).

REMARK 5—Relation to Weak Optimal Transport: If $V(x, \eta)$ is concave in η , then, by Jensen’s inequality, there exists an optimal assignment γ such that $(x, \eta), (x, \hat{\eta}) \in \text{supp}(\gamma) \implies \eta = \hat{\eta}$, so **(P)** can be reformulated as

$$\begin{aligned} & \max_{\pi \in \Delta(X \times Y)} \mathbb{E}_{\pi_X} [V(x, \pi_x)] \\ & \text{s.t. } \pi_X = \mu \quad \text{and} \quad \pi_Y = \nu, \end{aligned} \tag{P'}$$

where π_x is the conditional distribution of y given x under π . This formulation is the premise of the weak optimal transport literature (e.g., [Gozlan et al., 2017](#), [Alibert et al., 2019](#), [Backhoff-Veraguas et al., 2019](#)). This literature establishes primal attainment and no duality gap (i.e., $\max(\mathbf{P}') = \inf(\mathbf{D})$), assuming that V is concave in η but allowing V to be upper-semicontinuous and X and Y to be non-compact.¹³ However, none of these papers provides tractable sufficient conditions on V for dual attainment and hence for equilibrium existence.¹⁴

4. POSITIVE ASSORTATIVE SEGREGATION

This short section characterizes when workforces are segregated by skill and matched to firms in a positively assortative manner. This is the many-to-one analogue of positive assortative matching. This section also serves as a benchmark for [Section 7](#), where we compare conditions for (and payoffs under) segregation with competitive and uniform prices.

For the remainder of the paper, we assume that firm and worker types are 1-dimensional: $X = [\underline{x}, \bar{x}]$ and $Y = [\underline{y}, \bar{y}]$, with $\underline{x} < \bar{x}$ and $\underline{y} < \bar{y}$. We represent measures μ and ν by their cdfs $F \in \Delta(X)$ and $G \in \Delta(Y)$, which are assumed to have continuous and strictly positive

additional conditions (e.g., positivity of V and full support of (μ, ν)), any quasi-equilibrium is an equilibrium, so properness and bounded steepness are equivalent.

¹³Without concavity, we can have $\sup(\mathbf{P}') < \sup(\mathbf{P}) = \inf(\mathbf{D})$, which clarifies the role of the concavity assumption in the weak optimal transport literature.

¹⁴The most related of these papers is [Alibert et al. \(2019\)](#), who establish dual attainment assuming Lipschitz continuity of U (Theorem 5.1) and in some special cases (Theorems 5.6 and 6.8).

densities f and g .¹⁵ We also assume that $V(x, \delta_y)$ is twice continuously differentiable on $X \times Y$, where δ_y is the Dirac measure at y , and we denote partial derivatives of $(x, y) \mapsto V(x, \delta_y)$ with subscripts.

In one-to-one matching, strictly increasing differences in (x, y) imply positive assortative matching between firms and workers: each firm x hires worker $y = S(x)$, where $S = G^{-1} \circ F$ is the quantile matching function from F to G (Becker, 1973). In our many-to-one model, the same matching arises (where each firm x hires workforce $\delta_{S(x)}$) if all workforces are segregated and matched to firms positively assortatively. We thus refer to $S = G^{-1} \circ F$ as the *positive assortative segregation (PAS) matching*, where the corresponding PAS assignment $\phi \in \Delta(X \times \Delta(Y))$ is given by

$$\phi(A, D) = \int_A \mathbf{1}\{\delta_{S(x)} \in D\} \mu(dx), \quad \text{for all } A \subset X \text{ and } D \subset \Delta(Y).$$

The following result characterizes when the equilibrium is PAS.

THEOREM 2: *The following hold:*

1. *PAS is a competitive equilibrium if and only if*

$$\int_{\underline{y}}^{\bar{y}} \int_{S(x)}^y V_y(S^{-1}(\tilde{y}), \delta_{\tilde{y}}) d\tilde{y} \eta(dy) \geq V(x, \eta) - V(x, \delta_{S(x)}), \quad (1)$$

for all $(x, \eta) \in X \times \Delta(Y)$.

In addition, if (1) holds strictly for all $(x, \eta) \in X \times \Delta(Y)$ with $\eta \neq \delta_{S(x)}$, then PAS is the unique competitive equilibrium.

2. *PAS is a competitive equilibrium for all (F, G) if and only if $V(x, \delta_y)$ is supermodular in (x, y) ,*

$$V_{xy}(x, \delta_y) \geq 0, \quad \text{for all } (x, y) \in X \times Y, \quad (2)$$

and $V(x, \eta)$ is outer convex in η ,

$$\int_Y V(x, \delta_y) \eta(dy) \geq V(x, \eta), \quad \text{for all } (x, \eta) \in X \times \Delta(Y). \quad (3)$$

¹⁵Most results extend straightforwardly to general F and G , which need not have full support or densities.

In addition, if (2) holds strictly for all (x, y) , then PAS is the unique competitive equilibrium for all (F, G) .

3. When PAS is a competitive equilibrium, functions (p, q) are corresponding equilibrium prices if and only if there exists $c \in \mathbb{R}$ such that

$$p(x) = \int_{\underline{x}}^x V_x(\tilde{x}, \delta_{S(\tilde{x})}) d\tilde{x} + c, \quad \text{for all } x \in X, \quad (4)$$

$$q(y) = V(\underline{x}, \delta_{\underline{y}}) + \int_{\underline{y}}^y V_y(S^{-1}(\tilde{y}), \delta_{\tilde{y}}) d\tilde{y} - c, \quad \text{for all } y \in Y, \quad (5)$$

The third part shows that, when PAS is a competitive equilibrium, the corresponding utilities are given by the envelope theorem (i.e., $p'(x) = V_x(x, \delta_{S(x)})$ and $q'(y) = V_y(S^{-1}(y), \delta_y)$). The first part then follows from substituting these utilities in Definition 1.

The second part says that the equilibrium is PAS for any distribution of firm and worker types if and only if $V(x, \delta_y)$ is supermodular in (x, y) and $V(x, \eta)$ is outer convex in η .¹⁶ Intuitively, segregation by skill is optimal if and only if $V(x, \eta)$ is outer convex; and, if workforces are segregated, positive assortative matching between firms and workforces is optimal if and only if $V(x, \delta_y)$ is supermodular.

REMARK 6—Sufficiency of Checking 2-Type Workforces: Suppose there exist functions $v : X \times Y \times \mathbb{R} \rightarrow \mathbb{R}$ and $m : X \times Y \rightarrow \mathbb{R}$ such that $V(x, \eta) = \mathbb{E}_\eta[v(x, y, \mathbb{E}_\eta[m(x, \tilde{y})])]$ for all $(x, \eta) \in X \times \Delta(Y)$. This assumption holds throughout the remainder of the paper: in Section 6, v does not depend on y ; in the linear peer effects model in Section 7, $m(x, y) = y$; and in Section 5, both of these additional assumptions hold. Then, to check the conditions of Theorem 2, it suffices to consider workforces with at most two worker types: (1) and (3) hold if and only if they hold for all $(x, \eta) \in X \times \Delta(Y)$ with $|\text{supp}(\eta)| \leq 2$.¹⁷

¹⁶This condition is a mutual generalization of the conditions that characterize positive assortative matching in optimal transport (supermodularity of V ; Becker, 1973) and full disclosure in Bayesian persuasion (outer convexity of V ; Dworzak and Martini, 2019).

¹⁷In addition, if (1) holds strictly for all (x, η) with $\eta \neq \delta_{S(x)}$ and $|\text{supp}(\eta)| \leq 2$, then it holds strictly for all (x, η) with $\eta \neq \delta_{S(x)}$.

5. MEAN-MEASURABLE PRODUCTION

We now focus on the tractable, baseline model specification where the productivity of a workforce is determined by the mean skill of the workers it contains: there exists a function $v \in \text{Lip}(X \times Y)$ such that

$$V(x, \eta) = v(x, \mathbb{E}_\eta[y]), \quad \text{for all } (x, \eta) \in X \times \Delta(Y).$$

We refer to this specification as the *mean-measurable case*.

The mean-measurable case generalizes [Saint-Paul \(2001\)](#) (as well as mean-measurable Bayesian persuasion following [Gentzkow and Kamenica, 2016](#), [Kolotilin, 2018](#), and [Dworczak and Martini, 2019](#)) by allowing multiple firm types. Other special cases include [Kremer \(1993\)](#), who assume that $v(x, z) = xe^z$ with $x > 0$ for all firms, and [Boerma et al. \(2025\)](#), who assume that $v(x, z) = xe^z$ with $x < 0$ for all firms. The mean-measurable case also generalizes models with binary worker types, such as [Becker and Murphy \(2000\)](#) and [Benabou \(1996\)](#), as well as models where workers contribute different quantities of efficiency units of labor.

This section characterizes competitive equilibrium in the mean-measurable case. Section 5.1 shows how the primal and dual problems defining a competitive equilibrium simplify. Section 5.2 characterizes the equilibrium, showing how the sorting and pooling forces determine the pattern of workforce segregation and compression. Section 5.3 characterizes when simple matching patterns arise, such as PAS or full compression. Section 5.4 gives comparative statics on the equilibrium matching and prices.

5.1. Competitive Equilibrium

In the mean-measurable case, a workforce $\eta \in \Delta(Y)$ can be identified with its mean, $z = \mathbb{E}_\eta[y] \in Y$. We can thus abstract from the details of how individual workers and firms match (returning to this question only in Section 6), and instead simply ask what distribution H of workforce mean skills z forms, and how workforces with different mean skills z match with firms with different productivities x . We therefore define an *assignment* as a cdf $J \in \Delta(X \times Y)$ over (*firm, workforce mean*) pairs (x, z) such that its X -marginal is $J_X = F$ and its Y -marginal $J_Y = H$ can be obtained by splitting the population worker distribution G into a distribution of workforces η that induces distribution H of workforce

means. By [Blackwell \(1953\)](#), such a distribution H is feasible if and only if it *majorizes* G , denoted $H \succeq G$, meaning that $\int_{\underline{y}}^y H(\tilde{y})d\tilde{y} \leq \int_{\underline{y}}^y G(\tilde{y})d\tilde{y}$ for all $y \in Y$, with equality at $y = \bar{y}$.¹⁸ More generally, the same integral inequality defines the majorization order for any pair of increasing functions on Y .

The definition of competitive equilibrium in the mean-measurable case can be simplified by simultaneously viewing $q(z)$ as both the wage of a type z worker and the total wage bill associated with hiring a workforce with mean skill z . For such a wage schedule q to be consistent with a competitive assignment J , it must be convex—as otherwise a workforce z could be obtained more cheaply by hiring a distribution of workers η such that $\mathbb{E}_\eta[q(y)] < q(\mathbb{E}_\eta[z])$ —and it must satisfy $\mathbb{E}_{J_Y}[q(z)] = \mathbb{E}_G[q(y)]$ —so that the total wages paid by firms, $\mathbb{E}_{J_Y}[q(z)]$, equals the total wages received by workers, $\mathbb{E}_G[q(y)]$. We can thus define a competitive equilibrium as follows (where $\text{Conv}(Y)$ is the set of convex functions on Y).

DEFINITION 2: A *competitive equilibrium* is an assignment J together with prices $(p, q) \in \text{Lip}(X) \times \text{Lip}(Y)$ satisfying

$$\begin{aligned} p(x) + q(z) &= v(x, z), & \text{for all } (x, z) \in \text{supp}(J), \\ p(x) + q(z) &\geq v(x, z), & \text{for all } (x, z) \in X \times Y, \\ q &\in \text{Conv}(Y), & \text{and } \mathbb{E}_{J_Y}[q(z)] = \mathbb{E}_G[q(z)]. \end{aligned}$$

This definition coincides with the general one.

LEMMA 1: *Suppose that $\gamma \in \Delta(X \times \Delta(Y))$ induces $J \in \Delta(X \times Y)$.¹⁹ Then, $(J; p, q)$ is a competitive equilibrium in the sense of [Definition 2](#) if and only if $(\gamma; p, q)$ is a competitive equilibrium in the sense of [Definition 1](#).*

The primal and dual problems simplify analogously: the planner's problem is

$$\begin{aligned} &\max_{J \in \Delta(X \times Y)} \mathbb{E}_J[v(x, z)], \\ \text{s.t. } &J_X = F \quad \text{and} \quad J_Y \succeq G, \end{aligned} \tag{P_L}$$

¹⁸Equivalently, $H \succeq G$ if and only if G is a mean-preserving spread of H (or, if and only if G dominates H in the convex order).

¹⁹Formally, $J(\tilde{x}, \tilde{z}) = \mathbb{E}_\gamma[\mathbf{1}\{(x, \mathbb{E}_\eta[y]) \in [\underline{x}, \tilde{x}] \times [\underline{y}, \tilde{z}]\}]$ for all $\tilde{x} \in X$ and $\tilde{z} \in Y$.

and the dual problem is

$$\begin{aligned} & \min_{(p,q) \in \text{Lip}(X) \times \text{Lip}(Y)} \mathbb{E}_F[p(x)] + \mathbb{E}_G[q(z)] \\ \text{s.t. } & p(x) + q(z) \geq v(x, z), \quad \text{for all } (x, z) \in X \times Y, \\ & q \in \text{Conv}(Y). \end{aligned} \tag{D_L}$$

Theorem 1 specializes to the mean-measurable case as follows.²⁰

THEOREM 3: *A competitive equilibrium exists. Moreover, $(J; p, q)$ is a competitive equilibrium if and only if J solves the planner's problem (P_L) and (p, q) solves the dual problem (D_L) .*

Theorem 3 lets us characterize the competitive equilibrium assignment and utilities via the primal and dual problems (P_L) and (D_L) .

5.2. Equilibrium Characterization

For the remainder of this section, we assume that v is twice continuously differentiable, with strictly increasing differences between firm productivity and workforce skill ($v_{xz}(x, z) > 0$ for all (x, z)) and strictly decreasing returns to workforce skill ($v_{zz}(x, z) < 0$ for all (x, z)).²¹

By the same logic as in one-to-one matching, strictly increasing differences implies positive assortative matching between firms and workforces: each firm x hires workforce $z = H^{-1}(F(x)) = T(x)$, where $H = J_Y$ is the endogenous distribution of workforce means, and T is the quantile matching function from F to H . Thus, (P_L) simplifies to

$$\max_{H \succeq G} \int_X v(x, H^{-1}(F(x))) F(dx),$$

²⁰Theorem 3 simplifies Theorem 1 analogously to how Dworzak and Martini (2019) simplifies Kamenica and Gentzkow (2011) in the mean-measurable case in Bayesian persuasion.

²¹The increasing returns to skill case, $v_{zz}(x, z) > 0$ for all (x, z) , is also economically relevant—e.g., Kremer (1993) assumes increasing returns—and we characterize the equilibrium in this case in Remark 10.

or, equivalently, by the equivalence of $H \succeq G$ and $H^{-1} \preceq G^{-1}$ (Shaked and Shanthikumar, 1994, Theorem 3.A.5),

$$\max_{T: X \rightarrow Y} \int_X v(x, T(x)) F(dx), \quad (\text{M})$$

s.t. T is increasing, and

$$\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) \geq \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}), \quad \text{for all } x \in X, \quad \text{with equality at } x = \bar{x}, \quad (\text{F})$$

where, again, $S = G^{-1} \circ F$ is the segregation matching that arises when $H = G$.²²

Finding a competitive equilibrium thus simplifies to finding the increasing matching function T that maximizes $\int v(x, T(x)) F(dx)$, such that the resulting distribution of workforce means $F \circ T^{-1}$ majorizes G . In particular, if F is uniform—as is without loss up to a change of variables—then a matching T is feasible if and only if $T \preceq S$. As we have assumed that $v_{zz} < 0$, the objective is concave, so we have a concave maximization problem with a majorization constraint.

Given an increasing matching function T satisfying the feasibility constraint (F), we can partition the set of firm types X into the *segregation* and *compression* sets

$$X^S = \left\{ x \in X : \int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) = \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) \right\}, \quad \text{and}$$

$$X^C = \left\{ x \in X : \int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) > \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) \right\} = X \setminus X^S.$$

As we show in Lemma 7 in Appendix A, if $x \in X^S$, then all workers of types $\tilde{y} < S(x)$ are employed by firms of types $\tilde{x} < x$, and all workers of types $\tilde{y} > S(x)$ are employed by firms of types $\tilde{x} > x$. Consequently, if T is continuous at $x \in (\underline{x}, \bar{x})$ (which will hold for all x in equilibrium), then $T(x) = S(x)$. If instead $x \in X^C$, then some workers of types $\tilde{y} < S(x)$ are employed by firms of types $\tilde{x} > x$, and some workers of types $\tilde{y} > S(x)$ are

²²In addition, H^{-1} denotes the generalized inverse, although we will see that in equilibrium T is continuous and strictly increasing, so it coincides with the ordinary inverse.

employed by firms of types $\tilde{x} < x$. Note that X^C is open and thus is a union of at most countably many disjoint open intervals, $X^C = \bigcup_i (\underline{x}_i, \bar{x}_i)$.²³

The following theorem characterizes the equilibrium matching T and prices (p, q) .

THEOREM 4: *The following hold:*

1. *There is a unique equilibrium matching T , which is continuous and strictly increasing. In particular, more productive firms hire strictly more skilled workforces.*
2. *The equilibrium matching is the unique increasing matching T such that the marginal value of skill, $v_z(x, T(x))$, is increasing and is constant on each compression interval $(\underline{x}_i, \bar{x}_i)$.*
3. *Functions (p, q) are corresponding equilibrium prices if and only if there exists $c \in \mathbb{R}$ such that*

$$p(x) = \int_{\underline{x}}^x v_x(\tilde{x}, T(\tilde{x})) d\tilde{x} + c, \quad \text{for all } x \in X, \quad (6)$$

$$q(y) = \begin{cases} v(x, T(\underline{x})) + \int_{T(\underline{x})}^y v_z(T^{-1}(\tilde{y}), \tilde{y}) d\tilde{y} - c, & \text{if } y \in [T(\underline{x}), T(\bar{x})], \\ q(T(\underline{x})) - (T(\underline{x}) - y)v_z(\underline{x}, T(\underline{x})), & \text{if } y \in [\underline{y}, T(\underline{x})], \\ q(T(\bar{x})) + (y - T(\bar{x}))v_z(\bar{x}, T(\bar{x})), & \text{if } y \in (T(\bar{x}), \bar{y}]. \end{cases} \quad (7)$$

The intuition is as follows. For part 1, the equilibrium matching is unique, because it solves (M), which has a strictly concave objective and a convex constraint set. It is increasing, because $v_{xz} > 0$ implies positive assortativity between firms and workforces. It is strictly increasing, because if $x < x'$ but $T(x) = T(x')$, then $v_z(x, T(x)) < v_z(x', T(x'))$ by $v_{xz} > 0$, so the assignment is *under-sorted*: marginally segregating workforces $T(x)$ and $T(x')$ (e.g., swapping a few high-skill firm x employees for a few low-skill firm x' employees) would increase total value. It is continuous, because if T jumps up at x then $v_z(x, T(x))$ jumps down at x by $v_{zz} < 0$, but then marginally compressing workforces $T(x - \varepsilon)$ and $T(x + \varepsilon)$ would increase total value.

For part 2, $v_z(x, T(x))$ is increasing, because if $x < x'$ but $v_z(x, T(x)) > v_z(x', T(x'))$ then the assignment is *over-sorted*: marginally compressing workforces $T(x)$ and $T(x')$

²³In the optimal transport literature, the intervals $(\underline{x}_i, \bar{x}_i)$ are called *irreducible* with respect to $(F \circ T^{-1}, G)$ (Beiglböck and Juillet, 2016).

would increase total value. It is constant on compression intervals, because if x and x' are in the same compression interval but $v_z(x, T(x)) < v_z(x', T(x'))$, then marginally segregating workforces $T(x)$ and $T(x')$ would increase total value. Conversely, if $v_z(x, T(x))$ is increasing and is constant on compression intervals, then it is optimal, because the matching cannot be improved on either compression intervals (where $v_z(x, T(x))$ is constant, so further compression or segregation are both suboptimal) or segregation intervals (where $v_z(x, T(x))$ is increasing and workforces are already segregated, so compression is suboptimal and further segregation is infeasible).

For part 3, equation (6) and the middle equation in (7) follow from the envelope theorem, which implies that $p'(x) = v_x(x, T(x))$ and $q'(y) = v_z(T^{-1}(y), y)$.²⁴ Note that the wage schedule q is linear for workers employed by firms in compression intervals (i.e., workers y where $T^{-1}(y) \in X^C$) and is convex for workers employed by firms in segregation intervals (workers y where $T^{-1}(y) \in X^S$). Given the equilibrium matching function, equations (6)–(7) are the same as in one-to-one matching (Sattinger, 1993).²⁵ However, in one-to-one matching, the matching function is always S (under $v_{xz} > 0$), while in our model the matching function—and hence profits and wages—also depends on the type distributions F and G . This feature will have important implications for our comparative statics results in Section 5.4.

REMARK 7—The Convex-Loss Special Case: An especially tractable special case of the model is the *convex-loss case*, where $v(x, z) = a(z - x) + b(x) + cz$ with a strictly concave function a (e.g., $v(x, z) = xz - z^2/2$).²⁶ In this case, $v_z(x, T(x))$ is increasing in x —so the matching function T avoids over-sorting—if and only if $T'(x) \leq 1$: that is, iff T is 1-Lipschitz. The equilibrium matching is the same for any production function

²⁴The second and third equations in (7) follow from linearly extending the first equation to $[y, T(\underline{x})]$ and $(T(\bar{x}), \bar{y}]$.

²⁵A subtlety is that, even though in our model type y workers may be employed by a range of firms in equilibrium—not just firms with productivity $T^{-1}(y)$ —their marginal return to skill is still uniquely determined as $v_z(T^{-1}(y), y)$.

²⁶This case has received attention in the weak optimal transport literature. In particular, parts 1 and 3 of Corollary 1 follow from Backhoff-Veraguas et al. (2020), while part 2 is novel. It also covers Boerma et al. (2025) (as $|x|e^z = e^{z-\tilde{x}}$, with $\tilde{x} = -\log|x|$).

v in the convex-loss case. It can be characterized in three equivalent ways. First, T is the unique increasing 1-Lipschitz matching with slope 1 on each compression interval. Second, T results from ironing the function $x - S(x)$ (cf. Myerson, 1981, Section 6; Kleiner et al., 2021, Section 3.2.1): more precisely, defining $E : [0, 1] \rightarrow \mathbb{R}$ by $E(t) = \int_0^t (F^{-1}(\tilde{t}) - G^{-1}(\tilde{t})) d\tilde{t}$ for all $t \in [0, 1]$, letting $\bar{E} : [0, 1] \rightarrow \mathbb{R}$ be the convex envelope of E (i.e., the largest convex function that lies below E), and letting $\bar{e} = \bar{E}'$ be its continuous derivative, we have $T(x) = x - \bar{e}(F(x))$ for all $x \in X$. Third, T is the most segregated 1-Lipschitz matching. In sum, letting \mathcal{T} be the set of increasing, 1-Lipschitz maps $T : X \rightarrow Y$ satisfying (F), we have:

COROLLARY 1: *In the convex-loss case, the unique equilibrium matching T has the following three equivalent characterizations:*

1. $T \in \mathcal{T}$ such that $T'(x) = 1$ for all $x \in X^C$.
2. $T(x) = x - \bar{e}(F(x))$ for all $x \in X$.
3. $T \in \mathcal{T}$ such that $T \circ F^{-1} \succeq \tilde{T} \circ F^{-1}$ for all $\tilde{T} \in \mathcal{T}$.

REMARK 8—Concave Optimization with Majorization Constraints: Much recent interest in mechanism and information design centers on linear optimization problems over distributions satisfying majorization constraint (Bergemann et al., 2026, Kleiner et al., 2026). Solutions to such problems lie at the extreme points of the constraint set, which were characterized by Kleiner et al. (2021). In contrast, our problem (M) is a concave optimization problem over distributions satisfying majorization constraints. Theorem 4 shows that the solution everywhere either satisfies the majorization constraint with equality ($x \in X^S$) or satisfies a first-order condition ($\frac{d}{dx} v_z(x, T(x)) = 0$). Investigating the generality of this property is a possible avenue for future research.

5.3. Conditions for Simple Matching Patterns

Theorem 4 implies transparent sufficient conditions for simple matching patterns. We say that the matching function T is *PAS* if $X^S = X$ (or, equivalently, $T = S$), *full compression* if $X^C = (\underline{x}, \bar{x})$, *upper compression* if there exists x^* such that $X^C = (x^*, \bar{x})$, and *lower compression* if there exists x^* such that $X^C = (\underline{x}, x^*)$. Roughly speaking, under PAS, all workforces are segregated; under full compression, all workforces are mixed; under upper

compression, low-skill workforces are segregated and high-skill workforces are mixed; and under lower compression, low-skill workforces are mixed and high-skill workforces are segregated.²⁷

THEOREM 5: *The equilibrium matching is:*

1. *PAS if and only if $v_z(x, S(x))$ is increasing on X .*
2. *Full compression if $v_z(x, S(x))$ is strictly decreasing on X .*
3. *Upper compression if $v_z(x, S(x))$ is strictly quasiconcave on X .*
4. *Lower compression if $v_z(x, S(x))$ is strictly quasiconvex on X .*

To interpret Theorem 5, note that

$$\frac{d}{dx}v_z(x, S(x)) = v_{xz}(x, S(x)) + v_{zz}(x, S(x))S'(x). \quad (8)$$

Thus, $v_z(x, S(x))$ is increasing if and only if the sorting force v_{xz} is stronger than the pooling force v_{zz} , in that the magnitude of v_{xz} is larger than that of v_{zz} , scaled by the change-of-measure term $S'(x) = f(x)/g(S(x))$.

To see the intuition, consider matching firm types $x < x'$ with worker types $S(x) < S(x')$. Under segregation, these matches generate total value

$$v(x, S(x)) + v(x', S(x')).$$

If instead the workforces are slightly mixed, they generate total value

$$v(x, S(x) + \varepsilon) + v(x', S(x') - \varepsilon).$$

Taking $\varepsilon \downarrow 0$, slight mixing is suboptimal if and only if

$$v_z(x, S(x)) \leq v_z(x', S(x')).$$

Moreover, by $v_{zz} < 0$, if slight mixing is suboptimal, then so is any larger mixing. Thus, PAS is optimal (and hence arises in equilibrium) if and only if $v_z(x, S(x)) \leq v_z(x', S(x'))$

²⁷This sentence is imprecise because a firm in a compression interval need not actually hire a mixed workforce in the equilibrium assignment (see, e.g., Figure 5(a)). However, such a firm is always willing to hire a mixed workforce at the equilibrium wages.

for all $x < x'$: that is, if and only if $v_z(x, S(x))$ is increasing. Conversely, if $v_z(x, S(x))$ is decreasing, then it is suboptimal to segregate any interval of workforces, so the equilibrium is full compression. Similarly, if $v_z(x, S(x))$ is first increasing and then decreasing (i.e., quasiconcave), then it is optimal to segregate low-skill workforces and mix high-skill workforces, so the equilibrium is upper compression. Likewise, if $v_z(x, S(x))$ is quasiconvex, then the equilibrium is lower compression.²⁸

We illustrate Theorem 5 and Corollary 1 with some examples, where the production function is assumed to lie in the convex-loss case (e.g., $v(x, z) = xz - z^2/2$). Recall that, by Corollary 1, the equilibrium matching is the same for any such production function, and is the unique increasing 1-Lipschitz matching with slope 1 on each compression interval.

Figure 1 illustrates the segregation mapping $S = G^{-1} \circ F$ and the equilibrium matching T for three different type distributions. In panel (a), $F = U[0, 1]$ and $G = U[1/4, 3/4]$ —so, in the population, worker types are concentrated relative to firm types. The segregation mapping S has slope $1/2 < 1$ and thus coincides with the equilibrium matching T . Hence, equilibrium is PAS.

In panel (b), $F = U[1/4, 3/4]$ and $G = U[0, 1]$. Now, worker types are dispersed relative to firm types. The segregation mapping S has slope $2 > 1$, so the equilibrium matching T is everywhere compressed, with slope 1. Equilibrium is fully compressed.

In panel (c), $F = U[0, 1]$ and $G(y) = \sqrt{y}$. Here, low-skill workers are plentiful relative to low-productivity firms, but high-skill workers are scarce relative to high-productivity firms. The segregation mapping is $S(x) = x^2$, so $v_z(x, S(x)) = a'(x^2 - x) + c$. As this function is strictly quasiconcave (as a is strictly concave), Theorem 5 implies that the equilibrium matching is upper compression.²⁹

REMARK 9—Comparison to Bayesian Persuasion: Theorem 5 parallels results on the optimality of full disclosure, no disclosure, and upper and lower censorship (i.e., disclosing states below a cutoff while concealing states above it, or vice versa) in the linear Bayesian persuasion problem (Kolotilin et al., 2017, Kolotilin, 2018, Dworzak and Martini, 2019, Kolotilin et al., 2022). In persuasion, if the sender’s marginal utility from increasing the

²⁸Part 1 of Theorem 5, as well as Remark 10, can alternatively be obtained using Theorem 2 and Remark 6.

²⁹Specifically, the equilibrium matching is given by $T(x) = x^2$ if $x \leq 1/4$, and $T(x) = x - 3/16$ if $x > 1/4$.

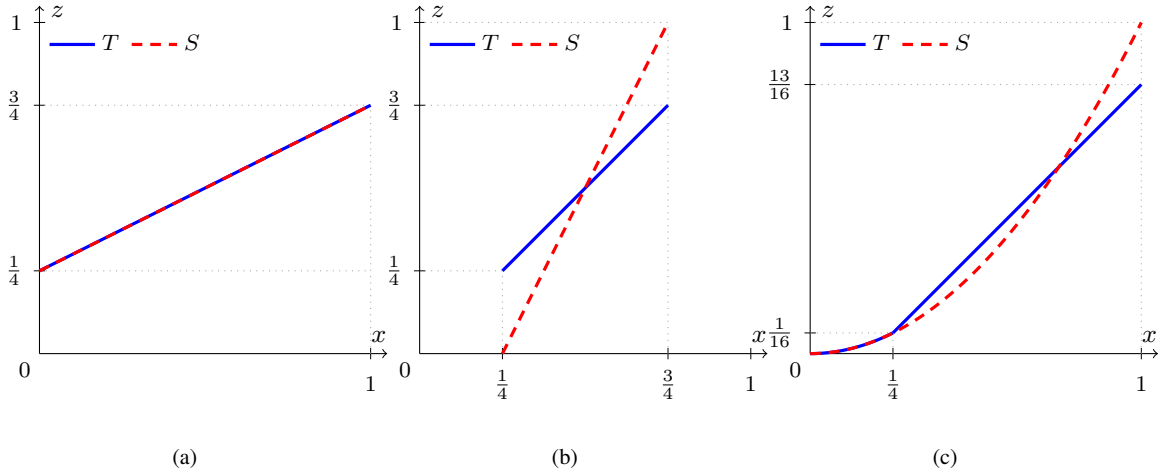


FIGURE 1.—Equilibrium Matchings. (a) When $F = U[0, 1]$, $G = U[1/4, 3/4]$, equilibrium is PAS. (b) When $F = U[1/4, 3/4]$, $G = U[0, 1]$, equilibrium is full compression. (c) When $F = U[0, 1]$, $G(y) = \sqrt{y}$, equilibrium is upper compression.

receiver’s posterior mean belief is v_z , then full disclosure is optimal if and only if v_z is increasing, no disclosure is optimal if v_z is decreasing, upper censorship is optimal if v_z is quasiconcave, and lower censorship is optimal if v_z is quasiconvex. However, there does not appear to be a tight formal connection between these two sets of results. For example, no disclosure and upper censorship are deterministic policies, whereas full compression and upper compression involve mixing.

REMARK 10—Increasing Returns to Skill: Under increasing differences and decreasing returns to skill ($v_{zz} < 0 < v_{xz}$), the sign of (8)—and hence the optimality of segregation or compression—depends on distributions F and G . In contrast, under increasing differences and *increasing* returns to skill ($v_{xz} > 0$ and $v_{zz} \geq 0$), PAS is optimal for any distributions F and G . Mathematically, this follows from the Fan-Lorentz theorem (e.g., [Kleiner et al., 2021](#), Theorem 4), which says that $\int_X v(x, T(x)) dx \leq \int_X v(x, S(x)) dx$ for all $T \preceq S$ if and only if $v(x, z)$ is supermodular in (x, z) and convex in z .³⁰ Intuitively, if $v_{xz} > 0$ and $v_{zz} \geq 0$ then there is no conflict between forces favoring sorting and pooling, but rather a

³⁰To apply the theorem, first re-scale x to make F uniform. This re-scaling does not affect supermodularity in (x, z) or convexity in z .

confluence of forces favoring sorting and separation. This observation generalizes [Kremer \(1993\)](#) (see also [Lazear, 2001](#)), where $v(x, z) = xe^z$, with $x > 0$ for all firms.

5.4. Comparative Statics

We now consider how changes in the production function and the distributions of firm and worker types affect equilibrium matching patterns and utilities.

We first give conditions for workforces to become more segregated, in that the distribution H of workforce means spreads out (equivalently, the matching $T \circ F^{-1}$ increases in the majorization order). This occurs if the sorting force v_{xz} becomes stronger relative to the pooling force v_{zz} , as captured by the following single-crossing condition.

DEFINITION 3: Given two production functions v and \hat{v} , the marginal return to skill under \hat{v} has a higher *marginal rate of substitution (MRS)* between firm productivity and workforce skill than that under v , denoted $\hat{v}_z \triangleright v_z$ if, for all $x' \geq x$ and $z' \geq z$,

$$v_z(x', z') \geq v_z(x, z) \implies \hat{v}_z(x', z') \geq \hat{v}_z(x, z).$$

In words, \hat{v} has a higher MRS than v if, whenever the marginal return to worker skill under v is higher at a high-productivity firm x' with a high-skill workforce z' than at a low-productivity firm x with a low-skill workforce z , then the same is true under \hat{v} .

REMARK 11: [MRS and Magnitude of Sorting vs. Pooling Forces] The condition that \hat{v} has a higher MRS than v is closely related to the ratio of the magnitudes of the sorting and pooling forces, $v_{xz}/(-v_{zz})$, being higher under \hat{v} . Specifically, if $\hat{v}_z \triangleright v_z$ then

$$\frac{\hat{v}_{xz}(x, z)}{-\hat{v}_{zz}(x, z)} \geq \frac{v_{xz}(x, z)}{-v_{zz}(x, z)} \quad \text{for all } (x, z);$$

and, conversely, if

$$\frac{\hat{v}_{xz}(x, z)}{-\hat{v}_{zz}(x', z')} \geq \frac{v_{xz}(x, z)}{-v_{zz}(x', z')} \quad \text{for all } (x, z) \text{ and } (x', z'),$$

then $\hat{v}_z \triangleright v_z$. In turn, the latter condition is equivalent to the existence of a constant $c > 0$ such that $\hat{v}_{xz}(x, z) \geq cv_{xz}(x, z)$ and $-\hat{v}_{zz}(x, z) \leq -cv_{zz}(x, z)$ for all (x, z) . These conditions are the ‘‘Spence–Mirrlees’’ analogues of the single-crossing condition given in Definition 3 ([Milgrom and Shannon, 1994](#)).

Our next result shows that matching is more segregated if the sorting force is stronger relative to the pooling force, or if workers are more heterogeneous.³¹

THEOREM 6: *Let $H_{v,F,G}$ be the equilibrium workforce distribution with production function v , firm type distribution F , and worker type distribution G . Then:*

1. $H_{\hat{v},F,G} \preceq H_{v,F,G}$ for all (F, G) if and only if $\hat{v}_z \succeq v_z$. (Workforces are more heterogeneous if and only if the sorting force is stronger relative to pooling force.)
2. $H_{v,F,\hat{G}} \preceq H_{v,F,G}$ for all (v, F) if and only if $\hat{G} \preceq G$. (Workforces are more heterogeneous if and only if workers are more heterogeneous.)

To see the logic, suppose by contradiction that $\hat{v}_z \succeq v_z$ but there exists an interval of firm types (x_0, x_1) on which \hat{T} is more compressed than T , in that $\int_{\underline{x}}^x \hat{T}(\tilde{x}) F(d\tilde{x}) > \int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x})$ for all $x \in (x_0, x_1)$, with equality at x_0 and x_1 . Then, there exist $\tilde{x}_0 < \tilde{x}_1$ in (x_0, x_1) such that firm \tilde{x}_0 's workforce is better under \hat{T} ($T(\tilde{x}_0) < \hat{T}(\tilde{x}_0)$), firm \tilde{x}_1 's workforce is better under T ($T(\tilde{x}_1) > \hat{T}(\tilde{x}_1)$), and these firms face the same marginal return to skill under (\hat{v}, \hat{T}) ($\hat{v}_z(\tilde{x}_0, \hat{T}(\tilde{x}_0)) = \hat{v}_z(\tilde{x}_1, \hat{T}(\tilde{x}_1))$). This implies that

$$\hat{v}_z(\tilde{x}_0, T(\tilde{x}_0)) > \hat{v}_z(\tilde{x}_0, \hat{T}(\tilde{x}_0)) = \hat{v}_z(\tilde{x}_1, \hat{T}(\tilde{x}_1)) > \hat{v}_z(\tilde{x}_1, T(\tilde{x}_1)),$$

where the inequalities hold because $T(\tilde{x}_0) < \hat{T}(\tilde{x}_0)$, $\hat{T}(\tilde{x}_1) < T(\tilde{x}_1)$, and $v_{zz} < 0$. But then $\hat{v}_z \succeq v_z$ implies that $v_z(\tilde{x}_0, T(\tilde{x}_0)) > v_z(\tilde{x}_1, T(\tilde{x}_1))$, contradicting that $v_z(x, T(x))$ is increasing.

Recent trends in wage inequality have been linked to rising assortativity between firms and workers (Card et al., 2013, Song et al., 2019, Kline, 2024). Theorem 6 contributes to this literature by characterizing what changes in production functions and the worker type distribution lead to rising workforce homogeneity in a simple competitive model. In particular, we identify the relative strength of complementarity between firm productivity and workforce skill and of decreasing returns to workforce skill, $v_{xz}/(-v_{zz})$, as the key de-

³¹One might guess that matching is also more segregated when firms are more heterogeneous, but this is only true under very restrictive conditions. To see this, suppose that $\hat{T} \circ \hat{F}^{-1} \succeq T \circ F^{-1}$ for all (v, G) . Taking $v(x, z) = k(x)z - z^2/2$ and a sufficiently spread out, zero-mean distribution G , the equilibrium matchings are $T(x) = k(x) - \mathbb{E}_F[k(x)]$ and $\hat{T}(x) = k(x) - \mathbb{E}_{\hat{F}}[k(x)]$. Then $k \circ \hat{F}^{-1} - \mathbb{E}_{\hat{F}}[k(x)] \succeq k \circ F^{-1} - \mathbb{E}_F[k(x)]$ for all increasing k , which implies that $\hat{F} = F$, so $\hat{T} \circ \hat{F}^{-1} \succeq T \circ F^{-1}$ for all (v, G) only in the trivial case $\hat{F} = F$.

terminant of the extent of workforce segregation by skill. Specifically, greater firm-worker complementarity and less sharply decreasing returns to workforce skill (or a switch from decreasing to increasing returns) increase workforce segregation by skill.

Part 2 of Theorem 6 says that, when the population distribution of worker skill becomes more heterogeneous, so does the equilibrium distribution of mean workforce skill. This comparative static is illustrated in panels (a) and (b) of Figure 1, after re-scaling x and y in panel (b) so that F and v are the same in both panels (i.e., $\hat{x} = 2x - 1/2$ and $\hat{y} = 2y - 1/2$): as the distribution of worker skills spreads out by a factor of 4 (from $U[1/4, 3/4]$ to $U[-1/2, 3/2]$), the distribution of workforce skills spreads out by a factor of 2 (from $U[1/4, 3/4]$ to $U[0, 1]$). The same panels also illustrate Part 1 of Theorem 6, after re-scaling x and y in panel (b) so that F and G are the same in both panels (i.e., $\hat{x} = 2x - 1/2$ and $\hat{y} = y/2 + 1/4$): as the sorting force weakens by a factor of 4 relative to the pooling force (e.g., v_z changes from $x - z$ to $x - 4z$), the distribution of workforce skills contracts by a factor of 2 (from $U[1/4, 3/4]$ to $[3/8, 5/8]$).

Kremer and Maskin (1996) and Acemoglu (1999) study different mechanisms that give a similar comparative static as Part 2 of Theorem 6.³² Moreover, these authors emphasize that changes in equilibrium matching patterns can exacerbate rising skill dispersion. For example, a headline result of Kremer and Maskin is that moving mass toward the tails of the worker skill distribution can trigger a shift from mixed workforces to segregated ones.

Figure 2 shows that, in our model, the equilibrium matching response can either exacerbate or mitigate rising skill dispersion. The production function is again convex-loss, and firm types are uniformly distributed: $F = U[0, 1]$. In panel (a), $g(y) = 2 - |2 - 4y|$. Here, extreme worker types are scarce relative to extreme firm types, so extreme firms employ mixed workforces. In panel (b), the skill distribution is more dispersed— $g(y) = 1$ —and equal to the firm productivity distribution, so the equilibrium is PAS. Thus, in moving from panel (a) to panel (b), the change in equilibrium matching (from partial compression to PAS) further spreads out the distribution of mean workforce skills, similarly to the effect

³²Kremer and Maskin consider a model of one-to-one matching between workers, and show that increasing the supply of high-skill workers can eliminate cross-skill matches and reduce low-skill workers' wages. Acemoglu studies a search model with capital investment that complements skill, and shows that increasing the supply of high-skill workers leads firms to invest and then decline to hire low-skill workers.

emphasized by [Kremer and Maskin \(1996\)](#) and [Acemoglu \(1999\)](#). However, in panel (c), the skill distribution is yet more dispersed— $g(y) = |2 - 4y|$ —so now moderate worker types are scarce relative to moderate firm types, which leads all firms to employ mixed workforces. Therefore, in moving from panel (b) to panel (c), the change in equilibrium matching (from PAS to full compression) counteracts the rise in skill heterogeneity.

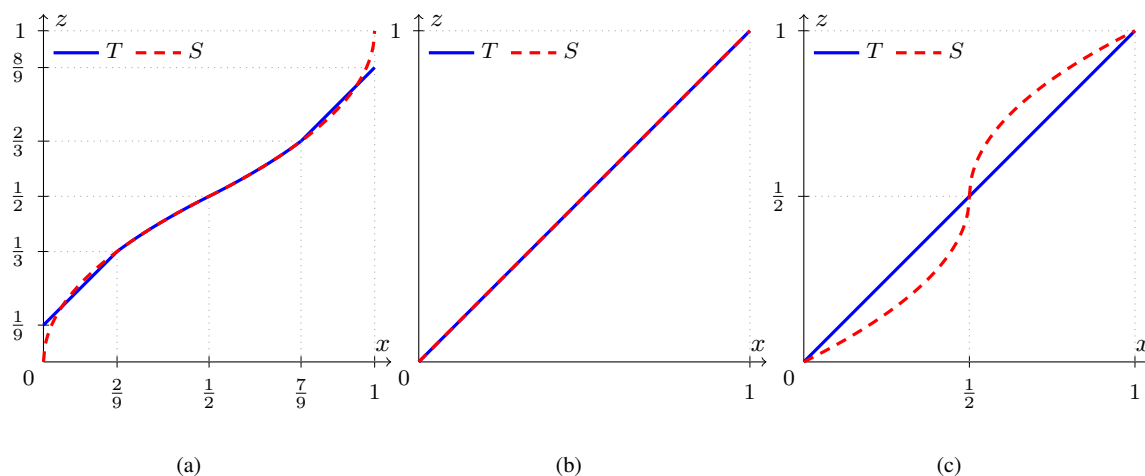


FIGURE 2.—Impact of Rising Skill Polarization. In all panels, $F = U[0, 1]$. In panel (a), $g(y) = 2 - |2 - 4y|$, and equilibrium is partially compressed. In panel (b), $g(y) = 1$, and equilibrium is PAS. In panel (c), $g(y) = |2 - 4y|$, and equilibrium is full compression.

We next consider comparative statics on profits and wages. The most pertinent result in the one-to-one matching context is due to [Terviö \(2008\)](#), who shows that, under increasing differences, firms' marginal return to productivity is higher, and workers' marginal return to skill is lower, when the distribution of firm productivity is lower or the distribution of worker skill is higher in first-order stochastic dominance (FOSD). Thus, inequality in payoffs on either side of the market increases when the type distribution on the opposite side of the market improves.³³ Under segregation, this result follows immediately from the price equations (6)–(7), because (for example) decreasing F or increasing G in the FOSD order increases $S(x)$ for all x , which increases $v_x(x, S(x))$ for all x , and hence increases

³³This force has been argued to play a role in trends in wage inequality, e.g., among CEOs ([Gabaix and Landier, 2008](#)). Moreover, if the utility of the lowest type on each side of the market is determined by an outside option, as in [Terviö \(2008\)](#), comparative statics on payoff differences imply comparative statics on payoff levels.

$p(x') - p(x)$ for all $x < x'$. In many-to-one matching, the analogous result is not immediate, because it is not obvious that decreasing F or increasing G in the FOSD order increases $T(x)$, once changes in the equilibrium matching pattern are accounted for—that is, that improving the distribution of worker types relative to firm types implies that each firm type hires a more skilled workforce. However, the following result shows that this does hold. An intuition is that increasing G clearly increases $T(x)$ for all x under PAS, and it also increases $T(x)$ for all x under full compression, because $v_z(x, T(x))$ must decrease for all x ; and the theorem shows that the same conclusion holds in the general case where T alternates between intervals of segregation and compression.

THEOREM 7: *Let $T_{v,F,G}$ be the equilibrium matching and let $(p_{v,F,G}, q_{v,F,G})$ be any equilibrium prices with production function v , firm type distribution F and worker type distribution G . Then, writing \geq for pointwise inequality, we have:*

1. $T_{v,\hat{F},G} \geq T_{v,F,G}$ for all (v, G) if and only if $F \leq \hat{F}$. (Each firm type hires a better workforce if and only if the firm type distribution worsens.)
2. $T_{v,F,\hat{G}} \geq T_{v,F,G}$ for all (v, F) if and only if $\hat{G} \leq G$. (Each firm type hires a better workforce if and only if the worker type distribution improves.)
3. For any v , $T_{v,\hat{F},\hat{G}} \geq T_{v,F,G}$ if and only if $p'_{v,\hat{F},\hat{G}} \geq p'_{v,F,G}$ if and only if $q'_{v,\hat{F},\hat{G}} \leq q'_{v,F,G}$. (Firm returns to productivity increase, and worker returns to skill decrease, if and only if each firm type hires a better workforce.)

Theorem 7 shows that the main price comparative statics in one-to-one assignment models extend to our many-to-one model. Our model also features some new possibilities, due to the dependence of the equilibrium matching pattern on the type distributions. In one-to-one matching with increasing differences, a worker's marginal return to skill equals $q'(y) = v_z(F^{-1}(G(y)), y)$, which depends on the distribution of worker types G only insofar as this changes the quantile of y . For example, holding fixed the wage of the lowest-skill workers in the market, an increase in the supply of high-skill workers does not affect the wages of low-skill workers in one-to-one matching.³⁴ This result no longer holds in many-to-one matching, because increasing the supply of high-skill workers reduces the productivity $T^{-1}(y)$ of firms matched to low-skill workers, and hence reduces these workers' returns

³⁴Fixing the lowest wage is appropriate if it is determined by a fixed outside option, as in [Terviö \(2008\)](#).

to skill $q'(y) = v_z(T^{-1}(y), y)$. Intuitively, in one-to-one matching, a worker faces competitive pressure only from lower-skill workers, whereas in many-to-one matching, all workers are potentially in competition. For instance, a medium-skill worker can be harmed by an increase in the supply of high-skill workers, because she is now more easily replaced by a mix of low-skill and high-skill workers. Thus, whereas in one-to-one assignment models a worker's return to skill is only reduced by an increase in the skill of less-skilled workers, in our model it is also reduced by an increase in the skill of more-skilled workers.

Figure 3 illustrates this possibility. Panel (a) reproduces panel (b) of Figure 1, where F and G are both uniform, and the equilibrium is full compression. Panel (b) then stretches the right tail of G : high-skill workers get even more skilled. This leads to an upward shift of the entire equilibrium matching function, which increases all firms' returns to productivity, and decreases all workers' returns to skill.

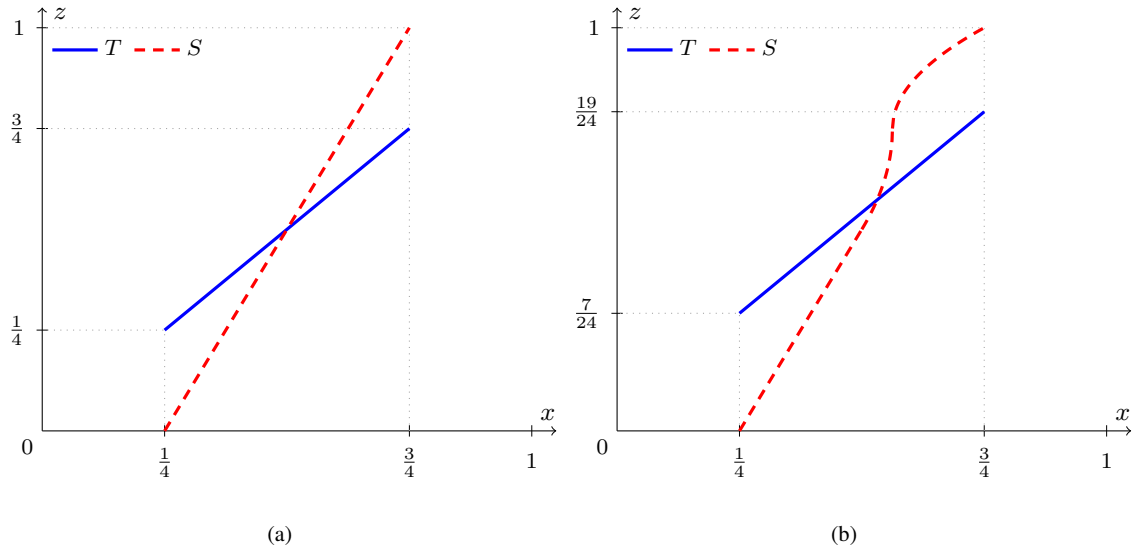


FIGURE 3.—Impact of a Fatter Right-Tail of Worker Skills. In both panels, $F = U[1/4, 3/4]$. In panel (a), $g(y) = 1$ for all y ; in panel (b), $g(y) = 1$, for $y \in [0, 1/2]$, $g(y) = 3 - 4y$ for $y \in (1/2, 3/4]$, and $g(y) = 12y - 9$ for $y \in (3/4, 1]$. All firms hire more skilled workforces in panel (b).

6. MOMENT-MEASURABLE PRODUCTION

Theorem 4 shows that our baseline, mean-measurable model admits a unique equilibrium matching between firm productivities x and mean workforce skills z . However, the

equilibrium matching between firms and individual workers is not uniquely determined, because many workforce compositions imply the same distribution of z . For example, in Figure 1(b) or 1(c), the equilibrium distribution of z can be attained when each firm in the compression interval hires a (different) mixture of all workers in the corresponding skill interval, or, alternatively, when each firm hires a mixture of only two worker types, with workers and co-workers matched in a negatively assortative manner. In this section, we show how this indeterminacy resolves for a class of production functions that approximate the mean-measurable case.

Specifically, we generalize the mean-measurable case by assuming that there exist two twice continuously differentiable functions $v : X \times \mathbb{R} \rightarrow \mathbb{R}$ and $m : X \times Y \rightarrow \mathbb{R}$, satisfying $v_z > 0$ and $m_y > 0$, such that

$$V(x, \eta) = v(x, \mathbb{E}_\eta[m(x, y)]), \quad \text{for all } (x, \eta) \in X \times \Delta(Y).$$

In this specification, the function v determines output as a function of firm productivity x and an aggregate labor input z , and the function m determines the labor input of a worker of skill y at a firm of productivity x . We refer to this specification as the *moment-measurable case*. We also continue to assume that $v_{zz}(x, z) < 0 < v_{xz}(x, z)$ for all (x, z) .

The mean-measurable case is the special case of the moment-measurable case where $m(x, y) = m^*(x, y) := y$, for all x . Thus, the moment-measurable case retains the assumption that a workforce η impacts production at firm x only through a single moment, but it allows this moment to differ across firms.³⁵

For example, consider a CES aggregator of labor inputs, where the elasticity of substitution between skill levels depends on firm productivity. Specifically, there exists $\alpha : X \rightarrow \mathbb{R}$ such that $V(x, \eta)$ depends on η only through the CES aggregator $(\mathbb{E}_\eta[y^{\alpha(x)}])^{1/\alpha(x)}$, so $m(x, y) = (y^{\alpha(x)} - 1)/\alpha(x)$. This aggregator is more dependent on the contribution of higher-skill workers at firms with higher $\alpha(x)$. For instance, note that

$$\lim_{\alpha(x) \rightarrow -\infty} (\mathbb{E}_\eta[y^{\alpha(x)}])^{\frac{1}{\alpha(x)}} = \inf_{y \in \text{supp}(\eta)} y,$$

³⁵As far as we know, the moment-measurable case is mathematically novel. In the Bayesian persuasion case of our model where the firm type distribution is degenerate, the moment-measurable case reduces to the linear case. In contrast, Dworzak and Kolotilin (2024) study Bayesian persuasion with multiple payoff-relevant moments.

$$\lim_{\alpha(x) \rightarrow 0} (\mathbb{E}_\eta[y^{\alpha(x)}])^{\frac{1}{\alpha(x)}} = \exp(\mathbb{E}_\eta[\log(y)]), \quad \text{and}$$

$$\lim_{\alpha(x) \rightarrow +\infty} (\mathbb{E}_\eta[y^{\alpha(x)}])^{\frac{1}{\alpha(x)}} = \sup_{y \in \text{supp}(\eta)} y,$$

so the aggregator interpolates between *weakest-link* (similar to [Kremer, 1993](#)) and *best-shot* (similar to [Boerma et al., 2025](#)) as $\alpha(x)$ increases from $-\infty$ to $+\infty$.

For another example (in the spirit of [Garicano, 2000](#) and [Garicano and Rossi-Hansberg, 2006](#)), suppose that each worker tries to solve an independent problem whose difficulty depends on the firm's type, and the aggregate labor input is the total measure of problems solved. That is, there exists an increasing function $L : [y - \bar{x}, \bar{y} - \underline{x}] \rightarrow [0, 1]$ such that a skill- y worker at a type- x firm solves her problem with probability $L(y - x)$, and $V(x, \eta)$ depends on η only through $\mathbb{E}_\eta[L(y - x)]$, so $m(x, y) = L(y - x)$.

We now consider the relationship between firm productivity and workforce heterogeneity in the moment-measurable case. We focus on the question of when more productive firms employ more or less heterogeneous workforces. We formalize this using the notion of single-dipped or single-peaked assignments.³⁶

DEFINITION 4: An assignment γ is *strictly single-dipped* (resp., *strictly single-peaked*) if there exists a μ -measure-1 set X^* and functions $\rho : X^* \rightarrow [0, 1]$ and $T_d, T_u : X^* \rightarrow Y$, with $T_d(x) \leq T_u(x)$ for all $x \in X^*$, such that

1. $\eta = (1 - \rho(x))\delta_{T_d(x)} + \rho(x)\delta_{T_u(x)}$ for all $(x, \eta) \in \text{supp}(\gamma)$ with $x \in X^*$; and
2. $T_d(x'), T_u(x') \notin (T_d(x), T_u(x))$ for all $x < x'$ in X^* .

That is, an assignment is strictly single-dipped if whenever firm 0 employs both low- and high-skill workers and firm 1 employs medium-skill workers, firm 0 is more productive than firm 1. In other words, the mapping from worker skill y to the productivity x of the firm where this worker is employed is single-dipped over nested sets of workforces. Conversely, an assignment is strictly single-peaked if firms that employ a mix of low- and high-skill workers are less productive than firms that employ medium-skill workers.

³⁶[Kolotilin et al. \(2025\)](#) introduced a related notion of single-dipped matching in Bayesian persuasion. The notion in [Kolotilin et al. \(2025\)](#) was previously introduced by [Beiglböck and Juillet \(2016\)](#) in optimal transport, under the name “left-curtain coupling.”

Our next result says that the equilibrium assignment is strictly single-dipped if m_y is strictly log-supermodular: $m_{xy}(x, y)/m_y(x, y)$ is strictly increasing in y , for all x . Intuitively, $m_y(x, y)$ is log-supermodular if labor input m is more convex in skill y at more productive firms.

THEOREM 8: *If m_y is strictly log-supermodular (resp., log-submodular), then the equilibrium assignment is strictly single-dipped (resp., single-peaked).*

For example, with a CES labor aggregator where $m(x, y) = (y^{\alpha(x)} - 1)/\alpha(x)$, Theorem 8 implies that equilibrium is strictly single-dipped (resp., single-peaked) if α is increasing (resp., decreasing). Intuitively, since labor input interpolates between weakest-link and best-shot as $\alpha(x)$ increases, low- $\alpha(x)$ firms hire homogeneous workforces, while high- $\alpha(x)$ firms hire heterogeneous ones. Similarly, with a Garicano-style aggregator where $m(x, y) = L(y - x)$, equilibrium is strictly single-dipped (resp., single-peaked) if L' is strictly log-concave (resp., log-convex).

To see the logic of Theorem 8, suppose for contradiction that m_y is strictly log-supermodular but that in equilibrium firm x_0 hires workers y_0 and y'_0 and firm x_1 hires worker y_1 , where $x_0 < x_1$ and $y_0 < y_1 < y'_0$. Consider swapping a mix of y_0 and y'_0 workers from firm x_0 for an equal mass of y_1 workers from firm x_1 , such that total labor input at firm x_1 , $\mathbb{E}_{\eta_1}[m(x_1, y)]$, remains constant. By strict log-supermodularity of m_y , the marginal rate of substitution in the production of labor input between “improving a worker’s skill from y_0 to y_1 ” and “improving a worker’s skill from y_1 to y'_0 ” is higher at firm x_0 than firm x_1 , which implies that this swap strictly increases total labor input at firm x_0 , $\mathbb{E}_{\eta_0}[m(x_0, y)]$. The swap therefore leaves output at firm x_1 unchanged while increasing output at firm x_0 , contradicting the optimality of the pre-swap assignment.³⁷ Finally, x deterministically maps to z because $v_{zz} < 0$, which, together with this swapping argument, implies that the equilibrium assignment is unique and is described as in Definition 4.

³⁷This logic is similar to that of Theorem 4 of Kolotilin et al. (2025), with the difference that the current result involves matching worker types and (exogenous) firm types, rather than matching states and (endogenous) actions.

We now show that if a mean-measurable production function is perturbed in the direction of log-supermodularity, the equilibrium matching is perturbed in the direction of strict single-dippedness.³⁸ This resolves the indeterminacy in the matching γ in Theorem 4.

THEOREM 9: *Suppose that m^n converges uniformly to m^* , and each m_y^n is strictly log-supermodular (resp., log-submodular). Let T be the equilibrium matching function under m^* , given by Theorem 4, and let γ be the unique strictly single-dipped (resp., single-peaked) assignment that induces T . Then the equilibrium assignment γ^n under m^n converges weakly to γ .*

We illustrate Theorem 9 in the context of the examples in Figure 1(b) and 1(c).

First, consider the convex-loss case with $F = U[1/4, 3/4]$ and $G = U[0, 1]$, as in Figure 1(b). The equilibrium matching is $T(x) = x$. The corresponding strictly single-dipped assignment, illustrated in Figure 4(a), is given by

$$\rho(x) = \frac{3}{4}, \quad T_d(x) = \frac{3}{8} - \frac{x}{2}, \quad T_u(x) = \frac{3x}{2} - \frac{1}{8}, \quad \text{for all } x \in X.$$

The corresponding strictly single-peaked assignment, illustrated in Figure 4(b), is given by

$$\rho(x) = \frac{1}{4}, \quad T_d(x) = \frac{3x}{2} - \frac{3}{8}, \quad T_u(x) = \frac{9}{8} - \frac{x}{2}, \quad \text{for all } x \in X.$$

Theorem 9 implies that, if the production function is moment-measurable and approximately convex-loss but m_y is strictly log-supermodular, then the equilibrium assignment approximates the one in Figure 4(a); and if the production function is moment-measurable and approximately convex-loss but m_y is strictly log-submodular, then the equilibrium assignment approximates the one in Figure 4(b).

Next, consider the convex-loss case with $F = U[0, 1]$ and $G(y) = \sqrt{y}$, as in Figure 1(c). Recall that $S(x) = x^2$ and

$$T(x) = \begin{cases} x^2, & x \in [0, \frac{1}{4}], \\ x - \frac{3}{16}, & x \in (\frac{1}{4}, 1]. \end{cases}$$

³⁸For example, up to re-scaling, the CES aggregator with $m(x, y) = (y^{\alpha(x)} - 1)/\alpha(x)$ converges to m^* if α converges to 1, and the Garicano-style aggregator with $m(x, y) = L(y - x)$ converges to m^* if L' converges to a constant.

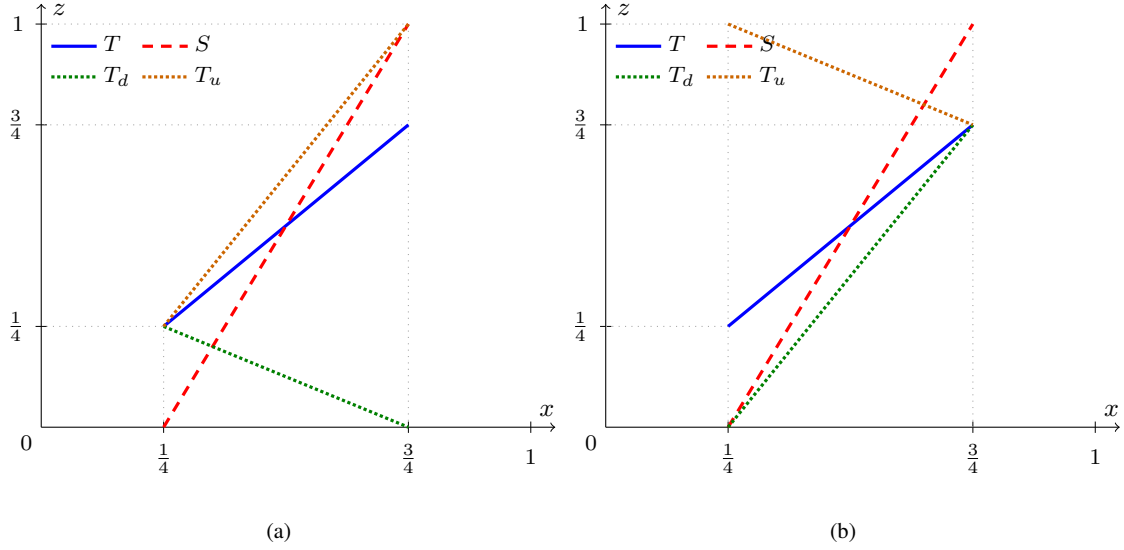


FIGURE 4.—Strictly Single-Dipped and Single-Peaked Full Compression. In both panels, $F = U[1/4, 3/4]$, $G = U[0, 1]$, and equilibrium is full compression. In panel (a), the assignment is strictly single-dipped; in panel (b), it is strictly single-peaked.

The corresponding strictly single-dipped assignment, illustrated in Figure 5(a), is given by

$$T_d(x) = T_u(x) = T(x), \quad \text{for } x \in [0, \frac{7}{16}], \quad \text{and}$$

$$\rho(x) = \frac{1}{d(x)}, \quad T_d(x) = \left(\frac{3}{2} - d(x)\right)^2, \quad T_u(x) = \left(2d(x) - \frac{3}{2}\right)^2, \quad \text{for } x \in \left(\frac{7}{16}, 1\right],$$

with $d(x) = \sqrt{x + 9/16}$.³⁹ Notably, firm types $x \in (1/4, 7/16)$ lie in the compression interval $(1/4, 1)$, yet they must hire homogeneous workforces. This shows that firms in compression intervals need not hire mixed workforces in the equilibrium assignment, even though they are always willing to do so at the equilibrium wages. The corresponding strictly single-peaked assignment, illustrated in Figure 5(b), is given by

$$T_d(x) = T_u(x) = T(x) = S(x), \quad \text{for } x \in [0, \frac{1}{4}], \quad \text{and}$$

$$\rho(x) = \frac{1}{2} - \frac{4-x}{6d(x)}, \quad T_d(x) = \frac{1}{4}(d(x) + x - 1)^2, \quad T_u(x) = \frac{1}{4}(d(x) - x + 1)^2, \quad \text{for } x \in \left(\frac{1}{4}, 1\right],$$

³⁹Formally, this can be proved using Theorem 4.5 in Henry-Labordère and Touzi (2016). Specifically, the bifurcation point $m = 7/16$ and the functions ρ , T_d , and T_u on $[m, 1]$ are determined by the boundary condition, $T_d(m) = T_u(m) = T(m)$, the martingale condition, $T(x) = (1 - \rho(x))T_d(x) + \rho(x)T_u(x)$, and the mass conservation conditions, $dG(T_u(x)) = \rho(x) dF(x)$ and $dG(T_d(x)) - dF(T^{-1}(T_d(x))) = -(1 - \rho(x)) dF(x)$.

with $d(x) = \sqrt{(75 - 4(x - 4)^2)/12}$. Theorem 9 implies that, if the production function is moment-measurable and approximately convex-loss but m_y is strictly log-supermodular, then the equilibrium assignment approximates the one in Figure 5(a); and if the production function is moment-measurable and approximately convex-loss but m_y is strictly log-submodular, then the equilibrium assignment approximates the one in Figure 5(b).

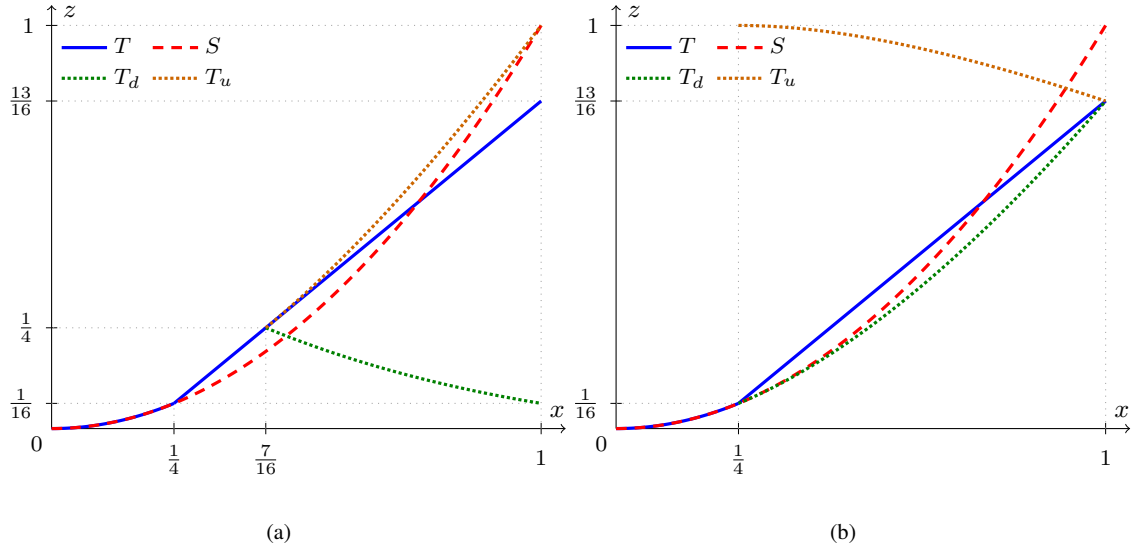


FIGURE 5.—Strictly Single-Dipped and Single-Peaked Upper Compression. In both panels, $F = U[0, 1]$, $G(y) = \sqrt{y}$, and equilibrium is upper compression. In panel (a), the assignment is strictly single-dipped; in panel (b), it is strictly single-peaked.

7. UNIFORM PRICE EQUILIBRIUM

We now return to the social economics applications described in Remark 2, where match value accrues to agents on the “many” side of the market, such as students attending a school or households moving to a neighborhood. In these settings, competitive equilibrium implicitly involves personalized prices, which may be impractical. In this section, we introduce an alternative equilibrium notion, where there is a single price to match with each agent on the “one” side of the market (e.g., a single price for each school or neighborhood). This alternative notion is more realistic in some applications, and we will compare it to competitive equilibrium to assess the impact of uniform pricing on equilibrium assignments and utilities.

The model and the definition of an assignment are unchanged from Section 2 and Remark 2, except that in this section we will refer to “students” and “schools” rather than “workers” and “firms.” Our alternative equilibrium notion is as follows.

DEFINITION 5: A *uniform price equilibrium (UPE)* is an assignment γ together with prices $p : \text{supp}(\gamma) \rightarrow \mathbb{R}$ for schools and utilities $q : Y \rightarrow \mathbb{R}$ for students satisfying

$$\begin{aligned} p(x, \eta) + q(y) &= W(x, y, \eta), & \text{for all } (x, \eta) \in \text{supp}(\gamma) \text{ and } y \in \text{supp}(\eta), \\ p(x, \eta) + q(y) &\geq W(x, y, \eta), & \text{for all } (x, \eta) \in \text{supp}(\gamma) \text{ and } y \in Y. \end{aligned}$$

The first condition says that each student y obtains utility $q(y) = W(x, y, \eta) - p(x, \eta)$ from attending school x with student body η and paying tuition $p(x, \eta)$; and the second condition says that student y cannot get a higher utility by attending a different school. Equivalently, γ describes a Nash equilibrium assignment (and q describes the corresponding payoffs) in the game where students simultaneously decide what schools to attend, given prices p .

UPE requires that two students who attend the same school must pay the same tuition. This could result from an agreement among schools, as in the “Overlap case” ([U.S. Court of Appeals for the Third Circuit, 1993](#)). It would also result if schools cannot observe students’ types. In the residential choice context, UPE requires that two individuals who want to buy the same house face the same price. This could result from anti-discrimination laws, from the possibility of resale, or simply from the infeasibility of residents paying off to their departing neighbors to sell to more desirable buyers.⁴⁰

In the one-to-one matching case where W is independent of η , competitive equilibrium (henceforth, CE) and UPE both reduce to the usual equilibrium notion for assignment games, which can equivalently be described as a competitive equilibrium, an optimal assignment, a group stable (core) assignment, or a pairwise stable assignment ([Gretsky et al., 1992](#)). However, they differ in many-to-one matching. For instance, in a CE, each school

⁴⁰[Becker and Murphy](#)—hardly skeptics as to the prevalence of flexible prices—emphasized the latter rationale, writing, “[W]ith large neighborhoods, such as those in a reasonably sized town or city district, residents generally do not collectively bid for other residents. Instead, they compete for houses, taking account of the composition of the residents in a neighborhood,” ([Becker and Murphy, 2000](#), p. 61).

gets the same payoff from each attending student, inclusive of her peer effect on others, which implies that different students pay different tuitions. Instead, in a UPE, each school gets the same payoff from each student, exclusive of her peer effect, which implies that all students pay the same tuition.⁴¹

CE and UPE are non-nested. In a CE, student y pays tuition $W(x, y, \eta) - q(y)$ to attend school x with student body η . Since this tuition may differ across students, a CE may not be a UPE. In other words, in a CE, $p(x) = \mathbb{E}_\eta[W(x, y, \eta) - q(y)]$ for all $(x, \eta) \in \text{supp}(\gamma)$, but it is not necessarily true that $p(x) = W(x, y, \eta) - q(y)$ for all $y \in \text{supp}(\eta)$, so the UPE conditions may not hold. Conversely, in a UPE, a school cannot increase its (uniform) tuition and continue to attract students, but an entrant that can charge personalized tuitions to attract a different student body could make a profit. That is, in a UPE, $p(x, \eta) \geq W(x, y, \eta) - q(y)$ for all $(x, \eta) \in \text{supp}(\gamma)$ and $y \in Y$, but it is not necessarily true that $p(x, \eta) \geq \mathbb{E}_{\tilde{\eta}}[W(x, y, \tilde{\eta}) - q(y)]$ for all $\tilde{\eta} \in \Delta(Y)$, so the second condition in the definition of a CE may not hold.⁴²

For the remainder of this section, we assume that peer effects are *moment-measurable*, meaning that there exist two twice continuously differentiable functions $w : X \times Y \times \mathbb{R} \rightarrow \mathbb{R}$ and $m : X \times Y \rightarrow \mathbb{R}$ such that

$$W(x, y, \eta) = w(x, y, \mathbb{E}_\eta[m(x, \tilde{y})]), \quad \text{for all } (x, y, \eta) \in X \times Y \times \Delta(Y).$$

For example, the usual case of *linear peer effects* (which generalizes [Benabou, 1996](#) and [Becker and Murphy, 2000](#)) is the special case where $m(x, y) = y$.

We ask when UPE are positively assortatively segregated, and how this compares to CE (i.e., when PAS is efficient). Recall that PAS is described by the matching function $S(x) = G^{-1}(F(x))$, where student bodies are segregated by ability and matched with schools in a positively assortative manner. We denote partial derivatives of $(x, y, z) \mapsto W(x, y, \delta_z)$ with subscripts, and similarly for w .

⁴¹UPE generalizes the equilibrium notion in [Benabou \(1996\)](#) and [Becker and Murphy \(2000\)](#), discussed below. Another related concept is the UPE notion of [Mailath et al. \(2013\)](#). They consider one-to-one matching, but otherwise their notion is more complex, as their model involves pre-match investment and match utilities on both sides of the market.

⁴²Another difference is that in CE each type x school gets the same payoff $p(x)$, while in UPE two type x schools with different student bodies $\eta \neq \eta'$ can get different payoffs $p(x, \eta) \neq p(x, \eta')$.

THEOREM 10: *The following hold:*

1. *PAS is a UPE if and only if*

$$\int_{S(x)}^y \int_x^{S^{-1}(\tilde{y})} \frac{d^2 W(\tilde{x}, \tilde{y}, \delta_{S(\tilde{x})})}{d\tilde{x} d\tilde{y}} d\tilde{x} d\tilde{y} \geq 0, \quad \text{for all } (x, y) \in X \times Y. \quad (9)$$

2. *PAS is a UPE for all (F, G) if and only if*

$$W_{xy}(x, y, \delta_y) \geq 0, \quad \text{for all } (x, y) \in X \times Y, \quad \text{and} \quad (10)$$

$$\int_z^y \int_z^{\tilde{y}} W_{yz}(x, \tilde{y}, \delta_{\tilde{z}}) d\tilde{z} d\tilde{y} \geq 0, \quad \text{for all } (x, y, z) \in X \times Y \times Y. \quad (11)$$

3. *PAS is the unique UPE for all (F, G) if*

$$W_{xy}(x, y, \delta_y) > 0, \quad \text{for all } (x, y) \in X \times Y, \quad \text{and} \quad (12)$$

$$W_{yz}(x, y, \delta_z) = 0, \quad \text{for all } (x, y, z) \in X \times Y \times Y. \quad (13)$$

4. *When PAS is a UPE, functions (p, q) are corresponding equilibrium prices if and only if there exists $c \in \mathbb{R}$ such that*

$$p(x) = \int_{\underline{x}}^x (W_x(\tilde{x}, S(\tilde{x}), \delta_{S(\tilde{x})}) + W_z(\tilde{x}, S(\tilde{x}), \delta_{S(\tilde{x})}) S'(\tilde{x})) d\tilde{x} + c, \quad \text{for all } x \in X, \quad (14)$$

$$q(y) = W(\underline{x}, \underline{y}, \delta_{\underline{y}}) + \int_{\underline{y}}^y W_y(S^{-1}(\tilde{y}), \tilde{y}, \delta_{\tilde{y}}) d\tilde{y} - c, \quad \text{for all } y \in Y. \quad (15)$$

Parts 1 and 2 imply that PAS is a UPE if $W(x, y, \delta_{S(x)})$ has increasing differences in (x, y) , and PAS is a UPE for all (F, G) if $W_{xy}(x, y, \delta_y) \geq 0$ and $W_{yz}(x, y, \delta_z) \geq 0$. Intuitively, PAS is a UPE if better students are willing to pay more to attend better schools with better peers. Conversely, PAS is not a UPE if $W(x, y, \delta_{S(x)})$ ever has strictly decreasing differences locally at $y = S(x)$, and PAS is not a UPE for some (F, G) if $W_{xy}(x, y, \delta_y) < 0$ or $W_{yz}(x, y, \delta_y) < 0$ for some (x, y) . Intuitively, if $W_{xy}(x, y, \delta_y) < 0$ then PAS is not a UPE if f is much more dispersed than g around (x, y) (so match effects dominate), and if $W_{yz}(x, y, \delta_y) < 0$ then PAS is not a UPE if g is much more dispersed than f around (x, y) (so peer effects dominate).

Part 3 says that PAS is the unique UPE if $W_{xy}(x, y, \delta_y) > 0$ and $W_{yz}(x, y, \delta_z) = 0$. These conditions ensure that *negative assortative segregation*—where student bodies are segregated but are matched with schools in a *negatively* assortative manner—is not a UPE.

To understand part 4, consider for simplicity the case where $f = g$, so $S(x) = x$. Here, the theorem says that the UPE prices $(p_{\text{UPE}}, q_{\text{UPE}})$ satisfy

$$p'_{\text{UPE}}(x) = W_x(x, x, \delta_x) + W_z(x, x, \delta_x) \quad \text{and} \quad q'_{\text{UPE}}(x) = W_y(x, x, \delta_x), \quad \text{for all } x \in X.$$

In contrast, when PAS is a competitive equilibrium, we have

$$p'_{\text{CE}}(x) = W_x(x, x, \delta_x) \quad \text{and} \quad q'_{\text{CE}}(x) = W_y(x, x, \delta_x) + W_z(x, x, \delta_x), \quad \text{for all } x \in X.$$

Thus, in either case, schools receive their marginal contribution to social surplus, W_x , and students receive their marginal contribution exclusive of peer effects, W_y , but the two cases differ in which side of the market receives the marginal contribution of peer effects, W_z . In a CE, tuition reflects students' contribution to schools inclusive of peer effects, so this value accrues to students. In a UPE, tuition reflects schools' value to students inclusive of peer effects, so this value accrues to schools.

This comparison has obvious implications for the distributional consequences of regulating private school tuition. Assume that (i) the utility of the lowest type of students and schools is fixed by outside options, (ii) the CE and UPE are both PAS, and (iii) peer effects are positive: $W_z > 0$. Then an agreement among schools that each school must charge the same tuition to all its attendees (i.e., no merit scholarships) increases the profit of all schools and reduces the welfare of all students (except the lowest types on each side). Similarly, an agreement among luxury goods retailers not to give discounts to high-status consumers (e.g., “influencers”) benefits all retailers and harms all consumers.

Parts 1–3 of Theorem 10 simplify in the mean-measurable case where $m(x, y) = y$ and w is affine in y . In this case, $V(x, \eta) = \mathbb{E}_\eta[w(x, y, \mathbb{E}_\eta[\tilde{y}])] = w(x, \mathbb{E}_\eta[y], \mathbb{E}_\eta[y])$, so the production function is mean-measurable, as in Section 5.

COROLLARY 2: *In the mean-measurable case, the following hold.*⁴³

⁴³The proof of the “only if” direction of part 3 uses densities that are strictly positive on subintervals of X and Y , while elsewhere in the paper we assume strictly positive densities on X and Y .

1. PAS is a UPE if and only if $w_y(x, S(x), S(x))$ is increasing in x .
2. PAS is a UPE for all (F, G) if and only if $w_y(x, z, z)$ is increasing in both x and z .
3. PAS is the unique UPE for all (F, G) if and only if $w_y(x, z, z)$ is strictly increasing in x and is independent of z .

We can now compare the conditions in Theorem 10 and Corollary 2 for PAS to be a UPE with conditions for PAS to be efficient. The following result is an immediate implication of Theorems 2 and 5 and Remarks 6 and 10.

COROLLARY 3: *PAS is efficient for all (F, G) if and only if*

$$\begin{aligned} &W_{xy}(x, y, \delta_y) + W_{xz}(x, y, \delta_y) \geq 0, \quad \text{and} \\ &(1 - \rho)W(x, y, \delta_y) + \rho W(x, y', \delta_{y'}) \\ &\geq (1 - \rho)W(x, y, (1 - \rho)\delta_y + \rho\delta_{y'}) + \rho W(x, y', (1 - \rho)\delta_y + \rho\delta_{y'}), \end{aligned}$$

for all $(x, y, y') \in X \times Y \times Y$ and all $\rho \in (0, 1)$.

Moreover, in the mean-measurable case,

1. PAS is efficient if and only if $w_y(x, S(x), S(x)) + w_z(x, S(x), S(x))$ is increasing in x .
2. PAS is efficient for all (F, G) if and only if $w_y(x, z, z) + w_z(x, z, z)$ is increasing in x and z .

Comparing Corollaries 2 and 3 shows that PAS can be a UPE (and even the unique one) even if segregation is inefficient. In particular, by Corollaries 2 and 3, PAS is a UPE but is not efficient if and only if $w_y(x, S(x), S(x))$ is increasing in x but $w_y(x, S(x), S(x)) + w_z(x, S(x), S(x))$ is not increasing in x , which tends to occur if $w_z(x, S(x), S(x))$ is decreasing in x . In turn, $w_z(x, S(x), S(x))$ can be decreasing in x if $w_{xz} < 0$ (“good schools” and “good peers” are substitutes), if $w_{yz} < 0$ (good students value good peers less), or if $w_{zz} < 0$ (there are decreasing returns to peer quality). Intuitively, in these cases, a high-ability student’s private return to attending a high-quality school tends to exceed the social return, resulting in inefficient segregation.

Benabou (1996) and Becker and Murphy (2000) consider related models of neighborhood choice. Benabou assumes that neighborhoods are identical ex ante (corresponding

to degenerate x in our model) and that there are two household types (corresponding to binary y). He shows that in this case the efficiency of PAS depends on the sign of $d(w_y(x, z, z) + w_z(x, z, z))/dz$. Part 2 of the mean-measurable case of Corollary 3 generalizes this result to heterogeneous neighborhoods and continuous household types. [Becker and Murphy](#) focus on the case where there are two neighborhood types and two household types with equal mass, and $w(x, y, z) = a(z) + yb(x)$. Besides generality, the cases studied by these authors omit some potentially important effects. For instance, [Benabou and Becker and Murphy](#) both assume that $w_{xz} = 0$, while we saw that $w_{xz} < 0$ creates a force for inefficient segregation. This case seems realistic, for example if learning from teachers and learning from peers are substitutes.⁴⁴

8. DISCUSSION

We conclude by mentioning three possible extensions of our model.

Investment. Several authors have considered productivity-enhancing investment prior to matching (e.g., [Nöldeke and Samuelson, 2015](#)). Our model gives a simple and stark result in this context: if workers are ex ante identical and the cost for a worker to acquire skill y is $C(y)$, where

$$V(x, \eta) - \mathbb{E}_\eta[C(y)] < V(x, \delta_{\mathbb{E}_\eta[y]}) - C(\mathbb{E}_\eta[y]), \quad \text{for all } (x, \eta) \text{ s.t. } |\text{supp}(\eta)| \geq 2,$$

then the competitive equilibrium must be segregated, because it is inefficient to form any nondegenerate workforce η , rather than having the same workers each acquire skill $\mathbb{E}_\eta[y]$. For example, this condition holds in the mean-measurable case if C is strictly convex. From this perspective, equilibrium workforce compression results from either ex ante worker heterogeneity (prior to investment) or from ex post inefficient investment decisions.

Scale. [Eeckhout and Kircher \(2018\)](#) study a model of one-to-one matching with endogenous *scale*, where each firm x hires a single type y of worker, but can hire any measure of them (e.g., because the output of worker y at firm x depends only on x, y , and the total measure of workers in the firm, so there is no reason for a firm to hire heterogeneous workers). Their main result characterizes positive assortative matching in this model. In contrast, our

⁴⁴[Benabou](#) and [Becker and Murphy](#) also do not compare utilities between competitive and uniform price equilibria.

model allows flexible workforce composition, but a fixed scale. A natural mutual generalization would allow flexibility in both composition and scale. For instance, characterizing PAS in such a model is a clear open direction.

Many-to-Many Matching. Our results extend straightforwardly to many-to-many matching, such as forming schools consisting of students and faculty. The many-to-many case is actually slightly simpler, because the possibility of mixing on both sides symmetrizes the analysis. For example, if the production function is mean-measurable on both sides, equilibrium features mean-preserving contractions of F and G , with compression regions corresponding to affine regions of p and q , which are both convex.

REFERENCES

- ACEMOGLU, DARON (1999): “Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence,” *American Economic Review*, 89 (5), 1259–1278. [26, 27]
- ACEMOGLU, DARON AND DAVID AUTOR (2011): “Skills, Tasks and Technologies: Implications for Employment and Earnings,” *Handbook of Labor Economics*, 4, 1043–1171. [1]
- ALIBERT, J-J, GUY BOUCHITTÉ, AND THIERRY CHAMPION (2019): “A New Class of Costs for Optimal Transport Planning,” *European Journal of Applied Mathematics*, 30 (6), 1229–1263. [5, 11]
- ALIPRANTIS, CHARALAMBOS D. AND KIM BORDER (2006): *Infinite Dimensional Analysis: A Hitchhiker’s Guide*, Springer-Verlag, Berlin. [70, 76]
- ARNOTT, RICHARD AND JOHN ROWSE (1987): “Peer Group Effects and Educational Attainment,” *Journal of Public Economics*, 32 (3), 287–305. [1, 5]
- AUMANN, ROBERT J (1964): “Markets with a Continuum of Traders,” *Econometrica*, 32, 39–50. [10]
- AZEVEDO, EDUARDO M AND JOHN WILLIAM HATFIELD (2018): “Existence of Equilibrium in Large Matching Markets with Complementarities,” *SSRN Working Papers*. [5]
- BACKHOFF-VERAGUAS, JULIO, MATHIAS BEIGLBÖCK, AND GUDMUN PAMMER (2019): “Existence, Duality, and Cyclical Monotonicity for Weak Transport Costs,” *Calculus of Variations and Partial Differential Equations*, 58 (6), 203. [11]
- BACKHOFF-VERAGUAS, JULIO, MATHIAS BEIGLBÖCK, AND GUDMUND PAMMER (2020): “Weak Monotone Rearrangement on the Line,” *Electronic Communications in Probability*, 25, 1–16. [5, 19]
- BECKER, GARY S (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81 (4), 813–846. [2, 4, 12, 13]
- BECKER, GARY S AND KEVIN M MURPHY (2000): *Social Economics: Market Behavior in a Social Environment*, Harvard University Press. [1, 3, 5, 8, 14, 36, 37, 40, 41]
- BEIGLBÖCK, MATHIAS AND NICOLAS JUILLET (2016): “On a Problem of Optimal Transport under Marginal Martingale Constraints,” *Annals of Probability*, 44 (1), 42–106. [18, 31, 76]

- BENABOU, ROLAND (1996): “Equity and Efficiency in Human Capital Investment: The Local Connection,” *Review of Economic Studies*, 63 (2), 237–264. [1, 3, 5, 14, 37, 40, 41]
- BERGEMANN, DIRK, TIBOR HEUMANN, AND STEPHEN MORRIS (2026): “Information Design and Mechanism Design: An Integrated Framework,” *CEPR Discussion Papers*. [20]
- BISHOP, BILL AND ROBERT G CUSHING (2009): *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart*, Houghton Mifflin Harcourt. [1]
- BLACKWELL, DAVID (1953): “Equivalent Comparisons of Experiments,” *Annals of Mathematical Statistics*, 24 (2), 265–272. [15, 54]
- BOERMA, JOB, ALEH TSYVINSKI, AND ALEXANDER P ZIMIN (2025): “Sorting with Teams,” *Journal of Political Economy*, 133 (2), 421–454. [4, 5, 7, 14, 19, 31]
- BOGACHEV, VLADIMIR (2007): *Measure Theory*, Springer-Verlag, Berlin. [47, 48, 50]
- CARD, DAVID, JÖRG HEINING, AND PATRICK KLINE (2013): “Workplace Heterogeneity and the Rise of West German Wage Inequality,” *Quarterly Journal of Economics*, 128 (3), 967–1015. [1, 25]
- CARLTON, DENNIS W, GUSTAVO BAMBERGER, AND ROY J EPSTEIN (1995): “Antitrust and Higher Education: Was There a Conspiracy to Restrict Financial Aid?” *NBER Working Papers*. [4]
- CHE, YEON-KOO, JINWOO KIM, AND FUHITO KOJIMA (2019): “Stable Matching in Large Economies,” *Econometrica*, 87 (1), 65–110. [5]
- CHONE, PHILIPPE, FRANCIS KRAMARZ, AND OSKAR NORDSTROM SKANS (2025): “Wage Markdowns and the Sorting of Workers to Firms when Skills are Bundled,” *CEPR Working Paper*. [5]
- CORTES, GUIDO MATIAS, ADRIAN LERCHE, UTA SCHÖNBERG, AND JEANNE TSCHOPP (2023): “Technological Change, Firm Heterogeneity and Wage Inequality,” *IZA Discussion Papers*. [1]
- DE BARTOLOME, CHARLES AM (1990): “Equilibrium and Inefficiency in a Community Model with Peer Group Effects,” *Journal of Political Economy*, 98 (1), 110–133. [5]
- DURLAUF, STEVEN N (2004): “Neighborhood Effects,” *Handbook of Regional and Urban Economics*, 4, 2173–2242. [1]
- DWORCZAK, PIOTR AND ANTON KOLOTILIN (2024): “The Persuasion Duality,” *Theoretical Economics*, 19 (4), 1701–1755. [9, 10, 11, 30, 49]
- DWORCZAK, PIOTR AND GIORGIO MARTINI (2019): “The Simple Economics of Optimal Persuasion,” *Journal of Political Economy*, 127 (5), 1993–2048. [4, 13, 14, 16, 22]
- EDMOND, CHRIS AND SIMON MONGEY (2022): “Unbundling Labor,” *Working Paper*. [5]
- ECKHOUT, JAN AND PHILIPP KIRCHER (2018): “Assortative Matching with Large Firms,” *Econometrica*, 86 (1), 85–132. [41]
- ELLICKSON, BRYAN, BIRGIT GRODAL, SUZANNE SCOTCHMER, AND WILLIAM R ZAME (1999): “Clubs and the Market,” *Econometrica*, 67 (5), 1185–1217. [5]
- EPPLE, DENNIS AND RICHARD E ROMANO (1998): “Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects,” *American Economic Review*, 88 (1), 33–62. [1, 5]
- GABAIX, XAVIER AND AUGUSTIN LANDIER (2008): “Why Has CEO Pay Increased So Much?” *Quarterly Journal of Economics*, 123 (1), 49–100. [4, 27]

- GALE, DAVID (1967): “A Geometric Duality Theorem with Economic Applications,” *Review of Economic Studies*, 34 (1), 19–24. [9, 10, 49]
- GALICHON, ALFRED (2018): *Optimal Transport Methods in Economics*, Princeton University Press. [4, 7, 10]
- GARICANO, LUIS (2000): “Hierarchies and the Organization of Knowledge in Production,” *Journal of Political Economy*, 108 (5), 874–904. [31]
- GARICANO, LUIS AND ESTEBAN ROSSI-HANSBERG (2006): “Organization and Inequality in a Knowledge Economy,” *Quarterly Journal of Economics*, 121 (4), 1383–1435. [31]
- GENTZKOW, MATTHEW AND EMIR KAMENICA (2016): “A Rothschild-Stiglitz Approach to Bayesian Persuasion,” *American Economic Review*, 106 (5), 597–601. [4, 14]
- GOLDSCHMIDT, DEBORAH AND JOHANNES F SCHMIEDER (2017): “The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure,” *Quarterly Journal of Economics*, 132 (3), 1165–1217. [1]
- GOZLAN, NATHAEL, CYRIL ROBERTO, PAUL-MARIE SAMSON, AND PRASAD TETALI (2017): “Kantorovich Duality for General Transport Costs and Applications,” *Journal of Functional Analysis*, 273 (11), 3327–3405. [5, 11]
- GRETSKY, NEIL E, JOSEPH M OSTROY, AND WILLIAM R ZAME (1992): “The Nonatomic Assignment Model,” *Economic Theory*, 2 (1), 103–127. [4, 7, 10, 36]
- (1999): “Perfect Competition in the Continuous Assignment Model,” *Journal of Economic Theory*, 88 (1), 60–118. [7, 10]
- (2002): “Subdifferentiability and the Duality Gap,” *Positivity*, 6 (3), 261–274. [10]
- HAANWINCKEL, DANIEL (2025): “Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution,” *American Economic Review*, 115 (12), 4137–4182. [1]
- HÅKANSON, CHRISTINA, ERIK LINDQVIST, AND JONAS VLACHOS (2021): “Firms and Skills: The Evolution of Worker Sorting,” *Journal of Human Resources*, 56 (2), 512–538. [1]
- HANDWERKER, ELIZABETH WEBER (2023): “Outsourcing, Occupationally Homogeneous Employers, and Wage Inequality in the United States,” *Journal of Labor Economics*, 41 (S1), S173–S203. [1]
- HANIN, LEONID G. (1992): “Kantorovich-Rubinstein Norm and its Application in the Theory of Lipschitz Spaces,” *Proceedings of the American Mathematical Society*, 115, 345–352. [9, 50]
- HENRY-LABORDÈRE, PIERRE AND NIZAR TOUZI (2016): “An Explicit Martingale Version of the One-Dimensional Brenier Theorem,” *Finance and Stochastics*, 20 (3), 635–668. [34]
- KAMENICA, EMIR AND MATTHEW GENTZKOW (2011): “Bayesian Persuasion,” *American Economic Review*, 101, 2590–2615. [4, 7, 16]
- KELSO, ALEXANDER S AND VINCENT P CRAWFORD (1982): “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica*, 50 (6), 1483–1504. [5]
- KLEINER, ANDREAS, BENNY MOLDOVANU, AND PHILIPP STRACK (2021): “Extreme Points and Majorization: Economic Applications,” *Econometrica*, 89 (4), 1557–1593. [4, 20, 23]
- (2026): “Extreme Points and Majorization,” *Cowles Foundation Discussion Papers*. [20]
- KLINER, PATRICK (2024): “Firm Wage Effects,” *Handbook of Labor Economics*, 5, 115–181. [1, 25]

- KOLOTILIN, ANTON (2018): “Optimal Information Disclosure: A Linear Programming Approach,” *Theoretical Economics*, 13, 607–636. [4, 14, 22]
- KOLOTILIN, ANTON, ROBERTO CORRAO, AND ALEXANDER WOLITZKY (2025): “Persuasion and Matching: Optimal Productive Transport,” *Journal of Political Economy*, 133 (4), 1334–1381. [31, 32]
- KOLOTILIN, ANTON, TIMOFIY MYLOVANOV, AND ANDRIY ZAPECHELNYUK (2022): “Censorship as Optimal Persuasion,” *Theoretical Economics*, 17 (2), 561–585. [22]
- KOLOTILIN, ANTON, TYMOFIY MYLOVANOV, ANDRIY ZAPECHELNYUK, AND MING LI (2017): “Persuasion of a Privately Informed Receiver,” *Econometrica*, 85, 1949–1964. [4, 22]
- KOOPMANS, TJALLING C AND MARTIN BECKMANN (1957): “Assignment Problems and the Location of Economic Activities,” *Econometrica*, 25 (1), 53–76. [4]
- KREMER, MICHAEL (1993): “The O-Ring Theory of Economic Development,” *Quarterly Journal of Economics*, 108 (3), 551–575. [4, 5, 7, 14, 16, 24, 31]
- KREMER, MICHAEL AND ERIC MASKIN (1996): “Wage Inequality and Segregation by Skill,” *NBER Working Papers*. [26, 27]
- LAZEAR, EDWARD P (2001): “Educational Production,” *Quarterly Journal of Economics*, 116 (3), 777–803. [24]
- LESHNO, JACOB (2022): “Stable Matching with Peer-Dependent Preferences in Large Markets: Existence and Cutoff Characterization,” *SSRN Working Papers*. [5]
- LINDENLAUB, ILSE (2017): “Sorting Multidimensional Types: Theory and Application,” *Review of Economic Studies*, 84 (2), 718–789. [5]
- MAILATH, GEORGE J, ANDREW POSTLEWAITE, AND LARRY SAMUELSON (2013): “Pricing and Investments in Matching Markets,” *Theoretical Economics*, 8 (2), 535–590. [37]
- MARKOVITS, DANIEL (2019): *The Meritocracy Trap*, Penguin UK. [1]
- MAS-COLELL, ANDREU (1986): “The Price Equilibrium Existence Problem in Topological Vector Lattices,” *Econometrica*, 54 (5), 1039–1053. [10]
- MAS-COLELL, ANDREU AND WILLIAM R ZAME (1991): “Equilibrium Theory in Infinite Dimensional Spaces,” *Handbook of Mathematical Economics*, 4, 1835–1898. [10]
- MILGROM, PAUL AND ILYA SEGAL (2002): “Envelope Theorems for Arbitrary Choice Sets,” *Econometrica*, 70 (2), 583–601. [9, 11, 52, 60, 76]
- MILGROM, PAUL AND CHRIS SHANNON (1994): “Monotone Comparative Statics,” *Econometrica*, 62 (1), 157–180. [24]
- MYERSON, ROGER B (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6 (1), 58–73. [20]
- NÖLDEKE, GEORG AND LARRY SAMUELSON (2015): “Investment and Competitive Matching,” *Econometrica*, 83 (3), 835–896. [41]
- OSTROY, JOSEPH M (1984): “A Reformulation of the Marginal Productivity Theory of Distribution,” *Econometrica*, 599–630. [10]
- RAHMAN, DAVID M (2026): “Organized Competition with Translucent Teams,” *SSRN Working Papers*. [5]
- RAYO, LUIS (2013): “Monopolistic Signal Provision,” *B.E. Journal of Theoretical Economics*, 13 (1), 27–58. [5]

- ROCKAFELLAR, R TYRRELL (1974): *Conjugate Duality and Optimization*, SIAM. [10]
- SAINT-PAUL, GILLES (2001): “On the Distribution of Income and Worker Assignment under Intrafirm Spillovers, with an Application to Ideas and Networks,” *Journal of Political Economy*, 109 (1), 1–37. [4, 14]
- SANDEL, MICHAEL (2020): *The Tyranny of Merit: What’s Become of the Common Good?*, Penguin UK. [1]
- SANTAMBROGIO, FILIPPO (2015): *Optimal Transport for Applied Mathematicians*, Birkäuser, New York, NY. [7, 10, 71]
- SATTINGER, MICHAEL (1975): “Comparative Advantage and the Distributions of Earnings and Abilities,” *Econometrica*, 43 (3), 455–468. [4]
- (1993): “Assignment Models of the Distribution of Earnings,” *Journal of Economic Literature*, 31 (2), 831–880. [1, 4, 7, 19]
- SHAKED, MOSHE AND J GEORGE SHANTHIKUMAR (1994): *Stochastic Orders*, Springer, New York. [17]
- SHAPLEY, LLOYD S AND MARTIN SHUBIK (1971): “The Assignment Game I: The Core,” *International Journal of Game Theory*, 1 (1), 111–130. [4]
- SONG, JAE, DAVID J PRICE, FATIH GUVENEN, NICHOLAS BLOOM, AND TILL VON WACHTER (2019): “Firming Up Inequality,” *Quarterly Journal of Economics*, 134 (1), 1–50. [1, 25]
- TERVIÖ, MARKO (2008): “The Difference that CEOs Make: An Assignment Model Approach,” *American Economic Review*, 98 (3), 642–668. [1, 3, 4, 7, 27, 28]
- U.S. COURT OF APPEALS FOR THE THIRD CIRCUIT (1993): “United States v. Brown University,” 5 F.3d 658 (3d Cir. 1993). [4, 36]
- U.S. DISTRICT COURT FOR THE NORTHERN DISTRICT OF ILLINOIS (2022): “Henry et al. v. Brown University et al.” No. 1:22-cv-00125 (N.D. Ill. filed Jan. 9, 2022). [4]
- VILLANI, CÉDRIC (2009): *Optimal Transport, Old and New*, Springer-Verlag, Berlin. [47]
- WINKLER, GERHARD (1988): “Extreme Points of Moment Sets,” *Mathematics of Operations Research*, 13 (4), 581–587. [53]

APPENDIX A: PROOFS

A.1. Duality in the General Case

PROOF OF THEOREM 1: We start with mathematical preliminaries. A function $q : Y \rightarrow \mathbb{R}$ is L -Lipschitz, denoted by $q \in \text{Lip}_L(Y)$, if there exists a constant L such that $|q(y) - q(\hat{y})| \leq L\rho_Y(y, \hat{y})$, for all $y, \hat{y} \in Y$, where ρ_Y is a metric on Y . As in Section 8.10(viii) in [Bogachev \(2007\)](#), we define the Kantorovich-Rubinstein norm on $M(Y)$ by

$$\|\eta\|_{KR} := |\eta(Y)| + \sup \left\{ \int_Y q(y)\eta(dy) : q \in \text{Lip}_1(Y), q(y_0) = 0 \right\}, \quad (16)$$

for a fixed $y_0 \in Y$. By Theorem 6.9 and Remark 6.19 in [Villani \(2009\)](#), $\|\cdot\|_{KR}$ metrizes the weak* topology on the compact metric space $\Delta(Y)$. So $\Delta(X \times \Delta(Y))$ is also a compact metric space.

Primal attainment follows from standard compactness arguments. Let $\mathcal{P}(\mu, \nu)$ denote the feasible set in **P** (i.e., the set of all assignments). Because the functions $\gamma \rightarrow \gamma_X$ and $\gamma \rightarrow \mathbb{E}_\gamma[\eta]$ are continuous, $\mathcal{P}(\mu, \nu)$ is compact, as a closed subset of the compact set $\Delta(X \times \Delta(Y))$. Moreover, $\mathcal{P}(\mu, \nu)$ is non-empty, as it contains $\gamma = \mu \otimes \delta_\nu$ (i.e., $\gamma(A, D) = \mu(A)\mathbf{1}\{\nu \in D\}$ for all $A \subset X$ and $D \subset \Delta(Y)$). Since V is Lipschitz continuous, the function $\gamma \rightarrow \mathbb{E}_\gamma[V(x, \eta)]$ is also continuous and thus attains its maximum on the compact set $\mathcal{P}(\mu, \nu)$, by the Weierstrass Theorem. Thus, **(P)** has a solution γ .

Next, the value of **(P)** is concave and Lipschitz in (μ, ν) . Define $U : \Delta(X) \times \Delta(Y) \rightarrow \mathbb{R}$ by $U(\mu, \nu) = \max_{\gamma \in \mathcal{P}(\mu, \nu)} \mathbb{E}_\gamma[V(x, \eta)]$ for all $(\mu, \nu) \in \Delta(X) \times \Delta(Y)$. Fix $\mu, \hat{\mu} \in \Delta(X)$, $\nu, \hat{\nu} \in \Delta(Y)$, and $\lambda \in [0, 1]$. By primal attainment, there exist $\gamma \in \mathcal{P}(\mu, \nu)$ and $\hat{\gamma} \in \mathcal{P}(\hat{\mu}, \hat{\nu})$ such that $U(\mu, \nu) = \mathbb{E}_\gamma[V(x, \eta)]$ and $U(\hat{\mu}, \hat{\nu}) = \mathbb{E}_{\hat{\gamma}}[V(x, \eta)]$. Note that $\gamma_\lambda \in \mathcal{P}(\mu_\lambda, \nu_\lambda)$, where $\mu_\lambda = (1 - \lambda)\mu + \lambda\hat{\mu}$, $\nu_\lambda = (1 - \lambda)\nu + \lambda\hat{\nu}$, and $\gamma_\lambda = (1 - \lambda)\gamma + \lambda\hat{\gamma}$. Thus, $U(\mu_\lambda, \nu_\lambda) \geq \mathbb{E}_{\gamma_\lambda}[V(x, \eta)] = (1 - \lambda)U(\mu, \nu) + \lambda U(\hat{\mu}, \hat{\nu})$, showing that U is concave.

To show that U is Lipschitz, we rely on the following lemma.

LEMMA 2: *Let $\mu, \hat{\mu} \in \Delta(X)$, $\nu, \hat{\nu} \in \Delta(Y)$, and $\gamma \in \mathcal{P}(\mu, \nu)$. There exists $\hat{\gamma} \in \mathcal{P}(\hat{\mu}, \hat{\nu})$ such that $\|\gamma - \hat{\gamma}\|_{KR} \leq \|\mu - \hat{\mu}\|_{KR} + \|\nu - \hat{\nu}\|_{KR}$.*

Letting $\Gamma(\gamma)$ denote the assignment in Lemma 2, and letting L denote the Lipschitz constant for V , we have

$$\begin{aligned} U(\mu, \nu) - U(\hat{\mu}, \hat{\nu}) &= \max_{\gamma \in \mathcal{P}(\mu, \nu)} \mathbb{E}_\gamma[V(x, \eta)] - \max_{\hat{\gamma} \in \mathcal{P}(\hat{\mu}, \hat{\nu})} \mathbb{E}_{\hat{\gamma}}[V(x, \eta)] \\ &\leq \max_{\gamma \in \mathcal{P}(\mu, \nu)} (\mathbb{E}_\gamma[V(x, \eta)] - \mathbb{E}_{\Gamma(\gamma)}[V(x, \eta)]) \\ &\leq \max_{\gamma \in \mathcal{P}(\mu, \nu)} L \|\gamma - \Gamma(\gamma)\|_{KR} \leq L(\|\mu - \hat{\mu}\|_{KR} + \|\nu - \hat{\nu}\|_{KR}), \end{aligned}$$

establishing Lipschitz continuity (after interchanging (μ, ν) and $(\hat{\mu}, \hat{\nu})$).

PROOF OF LEMMA 2: For notational convenience, denote $\hat{X} = X$ and $\hat{Y} = Y$. By the Kantorovich-Rubinstein theorem (Theorem 8.10.45 in Bogachev, 2007), there exist optimal transport plans between μ and $\hat{\mu}$ and between ν and $\hat{\nu}$: $\alpha \in \Delta(X \times \hat{X})$ with $\alpha_X = \mu$ and $\alpha_{\hat{X}} = \hat{\mu}$ and $\beta \in \Delta(Y \times \hat{Y})$ with $\beta_Y = \nu$ and $\beta_{\hat{Y}} = \hat{\nu}$ such that

$$\|\mu - \hat{\mu}\|_{KR} = \int_{X \times \hat{X}} \rho_X(x, \hat{x}) \alpha(dx, d\hat{x}) \quad \text{and} \quad \|\nu - \hat{\nu}\|_{KR} = \int_{Y \times \hat{Y}} \rho_Y(y, \hat{y}) \beta(dy, d\hat{y}).$$

By Theorem 10.4.5 in Bogachev (2007), α and β have conditional distributions $\alpha(\hat{x}|x)$ and $\beta(\hat{y}|y)$. Define the pushforward mapping $N: \Delta(Y) \rightarrow \Delta(Y)$ by $N(\eta)(\hat{A}) = \int_Y \beta(\hat{A}|y) \eta(dy)$ for all $\eta \in \Delta(Y)$ and $\hat{A} \subset \hat{Y}$. Define the joint distribution $\zeta \in \Delta(X \times \hat{X} \times \Delta(Y) \times \Delta(\hat{Y}))$ by

$$\zeta(A, \hat{A}, D, \hat{D}) = \int_{A \times D} \mathbf{1}\{N(\eta) \in \hat{D}\} \alpha(\hat{A}|x) \gamma(dx, d\eta),$$

for all $A, \hat{A} \subset X$ and $D, \hat{D} \subset \Delta(Y)$. Note that $\zeta_{X \times \Delta(Y)} = \gamma$ and $\zeta_{X \times X} = \alpha$. Define the projection $\hat{\gamma} = \zeta_{\hat{X} \times \Delta(\hat{Y})}$, and note that $\hat{\gamma} \in \mathcal{P}(\hat{\mu}, \hat{\nu})$, as $\hat{\gamma}_{\hat{X}} = \hat{\mu}$ and $\mathbb{E}_{\hat{\gamma}}[\hat{\eta}] = \hat{\nu}$. Finally, define $\sigma = \zeta_{\Delta(Y) \times \Delta(\hat{Y})}$, and note that $\mathbb{E}_\zeta[\eta] = \nu$ and $\mathbb{E}_\zeta[\hat{\eta}] = \hat{\nu}$.

Then, $\|\gamma - \hat{\gamma}\|_{KR} \leq \|\mu - \hat{\mu}\|_{KR} + \|\nu - \hat{\nu}\|_{KR}$, as, for any $\hat{V} \in \text{Lip}_1(X \times \Delta(Y))$, we have

$$\begin{aligned} &\int_{X \times \Delta(Y)} \hat{V}(x, \eta) (\gamma - \hat{\gamma})(dx, d\eta) \\ &= \int_{X \times \hat{X} \times \Delta(Y) \times \Delta(\hat{Y})} (\hat{V}(x, \eta) - \hat{V}(\hat{x}, \hat{\eta})) \zeta(dx, d\hat{x}, d\eta, d\hat{\eta}) \end{aligned}$$

$$\begin{aligned}
&= \int_{X \times \hat{X} \times \Delta(Y) \times \Delta(\hat{Y})} (\hat{V}(x, \eta) - \hat{V}(\hat{x}, \eta) + \hat{V}(\hat{x}, \eta) - \hat{V}(\hat{x}, \hat{\eta})) \zeta(dx, d\hat{x}, d\eta, d\hat{\eta}) \\
&\leq \int_{X \times \hat{X}} \rho_X(x, \hat{x}) \alpha(dx, d\hat{x}) + \int_{\Delta(Y) \times \Delta(\hat{Y})} \|\eta - \hat{\eta}\|_{KR} \sigma(d\eta, d\hat{\eta}) \\
&= \|\mu - \hat{\mu}\|_{KR} + \|\nu - \hat{\nu}\|_{KR},
\end{aligned}$$

where the last line is by Lemma 3 in [Dworczak and Kolotilin \(2024\)](#), restated as Claim 1.

CLAIM 1: $\int_{\Delta(Y) \times \Delta(\hat{Y})} \|\eta - \hat{\eta}\|_{KR} \sigma(d\eta, d\hat{\eta}) = \|\nu - \hat{\nu}\|_{KR}$.

PROOF: Define $\tau = \sigma_{\Delta(Y)}$, $\chi \in \Delta(Y \times \Delta(Y))$ by $\chi(B, D) = \int_D \eta(B) \tau(d\eta)$ for all $B \subset Y$ and $D \subset \Delta(Y)$, and $\xi \in \Delta(Y \times \hat{Y} \times \Delta(Y))$ by $\xi(B, \hat{B}, D) = \int_B \beta(\hat{B}|y) \chi(D|y) \nu(dy)$ for all $B, \hat{B} \subset Y$ and $D \subset \Delta(Y)$. Note that $\xi(B, \hat{Y}|\eta) = \eta(B)$ and $\xi(Y, \hat{B}|\eta) = N(\eta)(\hat{B})$ for all $B, \hat{B} \subset Y$, so $\|\eta - N(\eta)\|_{KR} \leq \int_{Y \times \hat{Y}} \rho_Y(y, \hat{y}) \xi(dy, d\hat{y}|\eta)$. Then

$$\begin{aligned}
\int_{\Delta(Y) \times \Delta(\hat{Y})} \|\eta - \hat{\eta}\|_{KR} \sigma(d\eta, d\hat{\eta}) &= \int_{\Delta(Y)} \|\eta - N(\eta)\|_{KR} \tau(d\eta) \\
&\leq \int_{\Delta(Y)} \int_{Y \times \hat{Y}} \rho_Y(y, \hat{y}) \xi(dy, d\hat{y}|\eta) \tau(d\eta) \\
&= \int_{Y \times \hat{Y}} \rho_Y(y, \hat{y}) \beta(dy, d\hat{y}) \\
&= \|\nu - \hat{\nu}\|_{KR} \\
&= \left\| \int_{\Delta(Y)} \eta \tau(d\eta) - \int_{\Delta(Y)} N(\eta) \tau(d\eta) \right\|_{KR} \\
&\leq \int_{\Delta(Y)} \|\eta - N(\eta)\|_{KR} \tau(d\eta),
\end{aligned}$$

implying that both inequalities hold with equality. *Q.E.D.*

Q.E.D.

Next, U is superdifferentiable, because it is concave and has bounded steepness at (μ, ν) (which is implied by Lipschitz continuity). Indeed, the Duality Theorem in [Gale \(1967\)](#) implies that there exists a continuous linear function H on $M(X) \times M(Y)$ such that $H(\hat{\mu} - \mu, \hat{\nu} - \nu) \geq U(\hat{\mu}, \hat{\nu}) - U(\mu, \nu)$ for all $\hat{\mu} \in \Delta(X)$ and $\hat{\nu} \in \Delta(Y)$. By the

Kantorovich-Rubinstein theorem (Exercise 8.10.143 in [Bogachev, 2007](#) or Theorem 0 in [Hanin, 1992](#)), there exist $p \in \text{Lip}(X)$ and $q \in \text{Lip}(Y)$ such that $H(\hat{\mu} - \mu, \hat{\nu} - \nu) = \int_X p(x)(\hat{\mu} - \mu)(dx) + \int_Y q(y)(\hat{\nu} - \nu)(dy)$ for all $\hat{\mu} \in \Delta(X)$ and $\hat{\nu} \in \Delta(Y)$. Adding a constant to p if necessary, we obtain

$$U(\mu, \nu) = \int_X p(x)\mu(dx) + \int_Y q(y)\nu(dy), \quad (17)$$

$$U(\hat{\mu}, \hat{\nu}) \leq \int_X p(x)\hat{\mu}(dx) + \int_Y q(y)\hat{\nu}(dy), \quad \text{for all } \hat{\mu} \in \Delta(X) \text{ and } \hat{\nu} \in \Delta(Y). \quad (18)$$

Given that U has a supergradient (p, q) at (μ, ν) , standard duality arguments imply equilibrium existence and its equivalence to solutions to **(P)** and **(D)**. Indeed, note that $(p, q) \in \mathcal{D}(V)$, where $\mathcal{D}(V)$ denotes the set of feasible solutions to **(D)**, as, for each $x \in X$ and $\eta \in \Delta(Y)$, we have

$$V(x, \eta) \leq U(\delta_x, \eta) \leq p(x) + \int_Y q(y)\eta(dy), \quad (19)$$

where the first inequality is by $\delta_{(x, \eta)} \in \mathcal{P}(\delta_x, \eta)$, and the second inequality is by (18) evaluated at $(\hat{\mu}, \hat{\nu}) = (\delta_x, \eta)$. Letting γ be a solution to **(P)**, we have

$$\begin{aligned} & \int_{X \times \Delta(Y)} \left(V(x, \eta) - p(x) - \int_Y q(y)\eta(dy) \right) \gamma(dx, d\eta) \\ &= U(\mu, \nu) - \int_X p(x)\mu(dx) - \int_Y q(y)\nu(dy) = 0, \end{aligned}$$

where the first equality is by the definition of $\gamma \in \mathcal{P}(\mu, \nu)$, and the second equality is by (17). Hence, (19) holds with equality for γ -almost all (x, η) and thus for all $(x, \eta) \in \text{supp}(\gamma)$, by continuity of $V(x, \eta) - p(x) - \int_Y q(y)\eta(dy)$ in (x, η) . So, $(\gamma; p, q)$ is a competitive equilibrium, which proves equilibrium existence. Next, if $(\gamma; p, q)$ is an equilibrium, then γ solves **(P)** and (p, q) solves **(D)**, because, for each $\tilde{\gamma} \in \mathcal{P}(\mu, \nu)$ and $(\tilde{p}, \tilde{q}) \in \mathcal{D}(V)$, we have

$$\begin{aligned} \int_X \tilde{p}(x)\mu(dx) + \int_Y \tilde{q}(y)\nu(dy) &= \int_{X \times \Delta(Y)} \left(\tilde{p}(x) + \int_Y \tilde{q}(y)\eta(dy) \right) \gamma(dx, d\eta) \\ &\geq \int_{X \times \Delta(Y)} V(x, \eta)\gamma(dx, d\eta) \end{aligned}$$

$$\begin{aligned}
&= \int_{X \times \Delta(Y)} \left(p(x) + \int_Y q(y) \eta(dy) \right) \gamma(dx, d\eta) \\
&= \int_X p(x) \mu(dx) + \int_Y q(y) \nu(dy) \\
&= \int_{X \times \Delta(Y)} \left(p(x) + \int_Y q(y) \eta(dy) \right) \tilde{\gamma}(dx, d\eta) \\
&\geq \int_{X \times \Delta(Y)} V(x, \eta) \tilde{\gamma}(dx, d\eta).
\end{aligned}$$

Conversely, if γ solves **(P)** and (p, q) solves **(D)**, then $(\gamma; p, q)$ is an equilibrium. Indeed, since there exists an equilibrium $(\tilde{\gamma}; \tilde{p}, \tilde{q})$, we have

$$\begin{aligned}
0 &= \int_{X \times \Delta(Y)} \left(V(x, \eta) - \tilde{p}(x) - \int_Y \tilde{q}(y) \eta(dy) \right) \tilde{\gamma}(dx, d\eta) \\
&= \int_{X \times \Delta(Y)} V(x, \eta) \tilde{\gamma}(dx, d\eta) - \int_X \tilde{p}(x) \mu(dx) - \int_Y \tilde{q}(y) \nu(dy) \\
&\leq \int_{X \times \Delta(Y)} V(x, \eta) \gamma(dx, d\eta) - \int_X p(x) \mu(dx) - \int_Y q(y) \nu(dy) \\
&= \int_{X \times \Delta(Y)} \left(V(x, \eta) - p(x) - \int_Y q(y) \eta(dy) \right) \gamma(dx, d\eta) \leq 0,
\end{aligned}$$

showing that the inequalities hold with equality, so $V(x, \eta) - p(x) - \int_Y q(y) \eta(dy) = 0$ for all $(x, \eta) \in \text{supp}(\gamma)$. *Q.E.D.*

PROOF OF REMARK 3: The remark follows from standard optimal transport arguments. Fix a solution (p, q) to **(D)**. Let $p^* : X \rightarrow \mathbb{R}$ and $V^* : \Delta(Y) \rightarrow \mathbb{R}$ be given by, for all $x \in X$ and $\eta \in \Delta(Y)$,

$$p^*(x) = \max_{\eta \in \Delta(Y)} \left\{ V(x, \eta) - \int_Y q(y) \eta(dy) \right\} \quad \text{and} \quad V^*(\eta) = \max_{x \in X} \{ V(x, \eta) - p(x) \},$$

where the maximums are attained by continuity and compactness.

By a standard envelope-theorem argument, p^* and V^* are L -Lipschitz, given that V is L -Lipschitz. Indeed, for all $x, \hat{x} \in X$ and $\eta, \hat{\eta} \in \Delta(Y)$, we have

$$p^*(x) \geq V(x, \eta) - \int_Y q(y) \eta(dy) \geq V(\hat{x}, \eta) - \int_Y q(y) \eta(dy) - L\rho_X(x, \hat{x}),$$

$$V^*(\eta) \geq V(x, \eta) - p(x) \geq V(x, \hat{\eta}) - p(x) - L \|\eta - \hat{\eta}\|_{KR},$$

so taking the maximum over $\eta \in \Delta(Y)$ and $x \in X$ on the right-hand sides yields $p^*(x) \geq p^*(\hat{x}) - L\rho_X(x, \hat{x})$ and $V^*(\eta) \geq V^*(\hat{\eta}) - L \|\eta - \hat{\eta}\|_{KR}$. Then, since (D) has a solution when $\mu = \delta_x$ and $V(x, \eta) = V^*(\eta)$ for all $x \in X$ and $\eta \in \Delta(Y)$, there exists q^* that solves

$$\min_{\hat{q} \in \text{Lip}(Y)} \int_Y \hat{q}(y) \nu(dy) \quad \text{s.t.} \quad \int_Y \hat{q}(y) \eta(dy) \geq V^*(\eta), \quad \text{for all } \eta \in \Delta(Y). \quad (20)$$

Note that $(p^*, q) \in \mathcal{D}(V)$ and $p^*(x) \leq p(x)$ for all $x \in X$, so $p(x) = p^*(x)$ for all $x \in \text{supp}(\mu)$, since (p, q) solves (D). Similarly, $(p, q^*) \in \mathcal{D}(V)$ and $\int_Y q^*(y) \nu(dy) \leq \int_Y q(y) \nu(dy)$, so q also solves (20), since (p, q) solves (D). Q.E.D.

A.2. Positive Assortative Segregation

PROOF OF THEOREM 2: Part 3. By Definition 1 (take $\eta = \delta_y$), if $(\phi; p, q)$ is a competitive equilibrium, then

$$p(x) = V(x, \delta_{S(x)}) - q(S(x)) = \max_{y \in Y} \{V(x, \delta_y) - q(y)\},$$

so, by the envelope theorem (Milgrom and Segal, 2002, Theorem 3) and continuity of V_x and S , p has a derivative $p'(x) = V_x(x, \delta_{S(x)})$ at each $x \in (\underline{x}, \bar{x})$, yielding (4); and (5) follows from $p(S^{-1}(y)) + q(y) = V(S^{-1}(y), \delta_y)$.

Part 1. By part 3, Definition 1, and some algebra, PAS is a competitive equilibrium if and only if (1) holds. In addition, if (1) holds strictly for all (x, η) with $\eta \neq \delta_{S(x)}$, then, by Theorem 1 and continuity of V , p , and q , for any equilibrium assignment γ , we have

$$\text{supp}(\gamma) \subset \left\{ (x, \eta) \in X \times \Delta(Y) : p(x) + \int_Y q(y) \eta(dy) = V(x, \eta) \right\} = \{(x, \delta_{S(x)}) : x \in X\},$$

showing that $\gamma = \phi$.

Part 2. Suppose (2) and (3) hold. Let p and q be given by (4) and (5). Then

$$\begin{aligned} p(x) + \int_Y q(y) \eta(dy) &= \int_Y \left(V(x, \delta_y) + \int_{S(x)}^y \int_x^{S^{-1}(\tilde{y})} V_{xy}(\tilde{x}, \delta_{\tilde{y}}) d\tilde{x} d\tilde{y} \right) \eta(dy) \\ &\geq \int_Y V(x, \delta_y) \eta(dy) \geq V(x, \eta), \end{aligned}$$

for all $(x, \eta) \in X \times \Delta(Y)$, so ϕ is an equilibrium assignment. In addition, if $V_{xy}(x, \delta_y) > 0$ for all $(x, y) \in X \times Y$, then the first inequality holds strictly for all (x, η) with $\eta \neq \delta_{S(x)}$, so ϕ is the unique equilibrium assignment.

Conversely, suppose that $V_{xy}(x, \delta_y) < 0$ for some $(x, y) \in X \times Y$. By continuity of V_{xy} , there exist $x' > x$ in X and $y' > y$ in Y such that $V_{xy}(\tilde{x}, \delta_{\tilde{y}}) < 0$ for all $(\tilde{x}, \tilde{y}) \in [x, x'] \times [y, y']$. Let μ and ν be such that $S(x) = y$ and $S(x') = y'$. Then, for p and q given by (4) and (5),

$$p(x) + q(y') - V(x, \delta_{y'}) = \int_y^{y'} \int_x^{S^{-1}(\tilde{y})} V_{xy}(\tilde{x}, \delta_{\tilde{y}}) d\tilde{x} d\tilde{y} < 0,$$

showing that (1) fails at $(x, \delta_{y'})$, so ϕ is not an equilibrium assignment.

Finally, suppose that $\int_Y V(x, \delta_y) \eta(dy) + \varepsilon < V(x, \eta)$ for some $(x, \eta) \in X \times \Delta(Y)$ and $\varepsilon > 0$. By continuity of V_{xy} , there exist $\mu \in \Delta(X)$ (with small enough $\|\mu - \delta_x\|_{KR}$) and $\nu \in \Delta(Y)$ such that $\int_{S(x)}^y \int_x^{S^{-1}(\tilde{y})} V_{xy}(\tilde{x}, \delta_{\tilde{y}}) d\tilde{x} d\tilde{y} < \varepsilon$ for all $y \in Y$. Then, for p and q given by (4) and (5),

$$\begin{aligned} p(x) + \int_Y q(y) \eta(dy) &= \int_Y \left(V(x, \delta_y) + \int_{S(x)}^y \int_x^{S^{-1}(\tilde{y})} V_{xy}(\tilde{x}, \delta_{\tilde{y}}) d\tilde{x} d\tilde{y} \right) \eta(dy) \\ &< \int_Y V(x, \delta_y) \eta(dy) + \varepsilon < V(x, \eta), \end{aligned}$$

showing that (1) fails at (x, η) , so ϕ is not an equilibrium assignment. *Q.E.D.*

PROOF OF REMARK 6: Fix $x \in X$ and $\eta \in \Delta(Y)$. By the Choquet Theorem (Winkler, 1988, Theorem 3.1) and the Richter-Rogosinsky Theorem (Winkler, 1988, Theorem 2.1), there exists $\tau \in \Delta(\Delta(Y))$ such that $\mathbb{E}_\tau[\tilde{\eta}] = \eta$, and, for each $\tilde{\eta} \in \text{supp}(\tau)$, we have $|\text{supp}(\tilde{\eta})| \leq 2$ and $\mathbb{E}_{\tilde{\eta}}[m(x, y)] = \mathbb{E}_\eta[m(x, y)]$. Thus, if (1) holds at all $\tilde{\eta} \in \text{supp}(\tau)$, then

$$\begin{aligned} p(x) + \mathbb{E}_\eta[q(\tilde{y})] &= p(x) + \mathbb{E}_\tau[\mathbb{E}_{\tilde{\eta}}[q(\tilde{y})]] \\ &\geq \mathbb{E}_\tau[\mathbb{E}_{\tilde{\eta}}[v(x, y, \mathbb{E}_{\tilde{\eta}}[m(x, \tilde{y})])]] = \mathbb{E}_\eta[v(x, y, \mathbb{E}_\eta[m(x, \tilde{y})])], \end{aligned}$$

showing that (1) holds at η . In addition, if $\eta \neq \delta_{S(x)}$ and (1) holds strictly for all $\tilde{\eta} \in \text{supp}(\tau)$ with $\tilde{\eta} \neq \delta_{S(x)}$, then $\text{supp}(\tau)$ contains some $\tilde{\eta} \neq \delta_{S(x)}$ (by $\mathbb{E}_\tau[\tilde{\eta}] = \eta \neq \delta_{S(x)}$), at which the

inequality is strict, so (1) holds strictly at η . Similarly, if (3) holds at all $\tilde{\eta} \in \text{supp}(\tau)$, then

$$\int_Y V(x, \delta_y) \eta(dy) = \int_{\Delta(Y)} \int_Y V(x, \delta_y) \tilde{\eta}(dy) \tau(d\tilde{\eta}) \geq \int_{\Delta(Y)} V(x, \tilde{\eta}) \tau(d\tilde{\eta}) = V(x, \eta),$$

showing that (3) holds at η .

Q.E.D.

A.3. Duality in the Mean-Measurable Case

This appendix allows for general $v \in \text{Lip}(X \times Y)$, which may not be continuously differentiable, and general $F \in \Delta(X)$ and $G \in \Delta(Y)$, which may not have full support or densities. Theorem 3 continues to hold verbatim, and Lemma 1 is generalized to Lemma 6.

We start with three lemmas.

LEMMA 3: *If $v \in \text{Lip}_L(X \times Y)$, then $V \in \text{Lip}_L(X \times \Delta(Y))$.*

PROOF: If v is L -Lipschitz, then so is V , as, for all $x, \hat{x} \in X$ and $\eta, \hat{\eta} \in \Delta(Y)$, we have

$$\begin{aligned} V(x, \eta) - V(\hat{x}, \hat{\eta}) &= v(x, \mathbb{E}_\eta[y]) - v(\hat{x}, \mathbb{E}_\eta[y]) + v(\hat{x}, \mathbb{E}_\eta[y]) - v(\hat{x}, \mathbb{E}_{\hat{\eta}}[y]) \\ &\leq L \left(|x - \hat{x}| + \left| \int_Y y \eta(dy) - \int_Y y \hat{\eta}(dy) \right| \right) \\ &\leq L (|x - \hat{x}| + \|\eta - \hat{\eta}\|_{KR}). \end{aligned}$$

Q.E.D.

LEMMA 4: *There exists $\gamma \in \mathcal{P}(\mu, \nu)$ that induces $J \in \Delta(X \times Y)$ if and only if $J_X = F$ and $J_Y \succeq G$.*

PROOF: Let $\gamma \in \mathcal{P}(\mu, \nu)$ and let $J^\gamma \in \Delta(X \times Y)$ denote the joint distribution of $(x, \mathbb{E}_\eta[y])$ induced by γ . Then $J_X^\gamma = F$, by $\gamma_X = \mu$, and $J_Y^\gamma \succeq G$, because, for each $h \in \text{Conv}(Y)$, we have, by Jensen's inequality and $\mathbb{E}_\gamma[\eta] = \nu$,

$$\mathbb{E}_{J_Y^\gamma}[h(y)] = \mathbb{E}_\gamma[h(\mathbb{E}_\eta[y])] \leq \mathbb{E}_\gamma[\mathbb{E}_\eta[h(y)]] = \mathbb{E}_G[h(y)].$$

Conversely, let $J \in \Delta(X \times Y)$ be such that $J_X = F$ and $J_Y \succeq G$, and let μ and ν denote the measures induced by the cdfs F and G . By Blackwell (1953), there exists $\mathcal{N} : Y \rightarrow \Delta(Y)$ such that $\mathbb{E}_{\mathcal{N}(z)}[y] = z$, for all $z \in Y$, and $\mathbb{E}_{J_Y}[\mathcal{N}(z)] = \nu$. Define $\gamma \in \Delta(X \times \Delta(Y))$ by

$$\gamma(A, D) = \int_{A \times Y} \mathbf{1}\{\mathcal{N}(z) \in D\} J(dx, dz), \quad \text{for all } A \subset X \text{ and } D \subset \Delta(Y).$$

Then $J_X^\gamma = J_X$ and $J_Y^\gamma = J_Y$. Moreover, $\gamma_X = \mu$, by $J_X = F$, and $\mathbb{E}_\gamma[\eta] = \mathbb{E}_{J_Y}[\mathcal{N}(z)] = \nu$, by the construction of \mathcal{N} . *Q.E.D.*

Let $\mathcal{D}_L(v)$ denote the set of feasible solutions to **(D_L)**. For any $q \in \text{Lip}(Y)$, define $\check{q} : Y \rightarrow \mathbb{R}$ by

$$\check{q}(z) = \min_{\eta \in \Delta(Y)} \mathbb{E}_\eta[q(y)] \quad \text{s.t.} \quad \mathbb{E}_\eta[y] = z, \quad \text{for all } z \in Y. \quad (21)$$

LEMMA 5: *If $(p, q) \in \mathcal{D}_L(v)$, then $(p, q) \in \mathcal{D}(V)$. If $(p, q) \in \mathcal{D}(V)$, then $(p, \check{q}) \in \mathcal{D}_L(v)$.*

PROOF: If $(p, q) \in \mathcal{D}_L(v)$, then, for all $x \in X$ and $\eta \in \Delta(Y)$, we have

$$p(x) + \mathbb{E}_\eta[q(y)] \geq p(x) + q(\mathbb{E}_\eta[y]) \geq v(x, \mathbb{E}_\eta[y]) = V(x, \eta),$$

so $(p, q) \in \mathcal{D}(V)$.

Let $(p, q) \in \mathcal{D}(V)$. Up to notation and sign change, **(21)** features the same optimization problem as **(P)** when $|\text{supp}(\mu)| = 1$ and $|\text{supp}(\nu)| = 2$. Thus, as follows from Section 3, \check{q} is Lipschitz continuous and convex. Moreover, letting η_z denote a minimizer in **(21)**, $(p, q) \in \mathcal{D}(V)$ yields

$$p(x) + \check{q}(z) = p(x) + \mathbb{E}_{\eta_z}[q(y)] \geq V(x, \eta_z) = v(x, z),$$

showing that $(p, \check{q}) \in \mathcal{D}_L(v)$. *Q.E.D.*

LEMMA 6: *Suppose that $\gamma \in \Delta(X \times \Delta(Y))$ induces $J \in \Delta(X \times Y)$. If $(J; p, q)$ is a competitive equilibrium in the sense of Definition 2, then $(\gamma; p, q)$ is a competitive equilibrium in the sense of Definition 1. Conversely, if $(\gamma; p, q)$ is a competitive equilibrium in the sense of Definition 1, then $(J; p, \check{q})$ is a competitive equilibrium in the sense of Definition 2, and $\check{q}(y) = q(y)$ for all $y \in \text{supp}(G)$.*

PROOF: If $(J; p, q)$ is a competitive equilibrium in the sense of Definition 2, then $(\gamma; p, q)$ is a competitive equilibrium in the sense of Definition 1, because

$$\begin{aligned} \int_{X \times \Delta(Y)} V(x, \eta) \gamma(\mathrm{d}x, \mathrm{d}\eta) &\leq \int_{X \times \Delta(Y)} \left(p(x) + \int_Y q(y) \eta(\mathrm{d}y) \right) \gamma(\mathrm{d}x, \mathrm{d}\eta) \\ &= \int_X p(x) \mu(\mathrm{d}x) + \int_Y q(y) \nu(\mathrm{d}y) \end{aligned}$$

$$\begin{aligned}
&= \int_X p(x) F(dx) + \int_Y q(z) J_Y(dz) \\
&= \int_{X \times Y} v(x, z) J(dx, dz) \\
&= \int_{X \times \Delta(Y)} V(x, \eta) \gamma(dx, d\eta),
\end{aligned}$$

where the inequality holds by $(p, q) \in \mathcal{D}(V)$ (see Lemma 5). Conversely, if $(\gamma; p, q)$ is a competitive equilibrium in the sense of Definition 1, then $(J; p, \check{q})$ is a competitive equilibrium in the sense of Definition 2, and $\check{q}(y) = q(y)$ for all $y \in \text{supp}(G)$, because

$$\begin{aligned}
\int_{X \times Y} v(x, z) J(dx, dz) &\leq \int_{X \times Y} (p(x) + \check{q}(z)) J(dx, dz) \\
&= \int_X p(x) F(dx) + \int_Y \check{q}(z) J_Y(dz) \\
&\leq \int_X p(x) \mu(dx) + \int_Y \check{q}(y) \nu(dy) \\
&\leq \int_X p(x) \mu(dx) + \int_Y q(y) \nu(dy) \\
&= \int_{X \times \Delta(Y)} V(x, \eta) \gamma(dx, d\eta) \\
&= \int_{X \times Y} v(x, z) J(dx, dz),
\end{aligned}$$

where the first inequality holds by $(p, \check{q}) \in \mathcal{D}_L(v)$, the second inequality holds by $\check{q} \in \text{Conv}(Y)$ and $J_Y \succeq G$, and the third inequality holds by $\check{q}(y) \leq q(y)$ for all $y \in Y$. *Q.E.D.*

PROOF OF THEOREM 3: By Lemma 3, Theorem 1 implies that there exists a competitive equilibrium $(\gamma; p, q)$ in the sense of Definition 1. Let J denote the joint distribution of $(x, \mathbb{E}_\eta[y])$ induced by γ . By Lemma 6, $(J; p, \check{q})$ is a competitive equilibrium in the sense of Definition 2. Moreover, by Theorem 1 and Lemmas 4–6, $(J; p, q)$ is a competitive equilibrium if and only if J solves (\mathbf{P}_L) and (p, q) solves (\mathbf{D}_L) . *Q.E.D.*

A.4. Equilibrium Characterization in the Mean-Measurable Case

We first formalize the intuition for the segregation set X^S .

LEMMA 7: Let $\gamma \in \mathcal{P}(\mu, \nu)$ induce T satisfying (F), and let $I \in \Delta(X \times Y)$ be given by

$$I(x, y) = \int_{X \times \Delta(Y)} \eta([\underline{y}, y]) \mathbf{1}\{\tilde{x} \leq x\} \gamma(d\tilde{x}, d\eta).$$

Then $x \in X^S$ if and only if $I(x, S(x)) = F(x)$. Moreover, if $x \in X^S$, with $x \neq \underline{x}, \bar{x}$, and T is continuous, then $T(x) = S(x)$.

PROOF: Note that $I_X = F$ and $I_Y = G$. Thus, $I(x, y) \leq \min(F(x), G(y))$ for all $(x, y) \in X \times Y$. Moreover,

$$\begin{aligned} \int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) &= \int_{X \times \Delta(Y)} \int y \eta(dy) \mathbf{1}\{\tilde{x} \leq x\} \gamma(d\tilde{x}, d\eta) \\ &= \int_{X \times \Delta(Y)} \left(\underline{y} + \int_{\underline{y}}^{\bar{y}} (1 - \eta([\underline{y}, y])) dy \right) \mathbf{1}\{\tilde{x} \leq x\} \gamma(d\tilde{x}, d\eta) \\ &= \underline{y}F(x) + \int_{\underline{y}}^{\bar{y}} \int_{X \times \Delta(Y)} (1 - \eta([\underline{y}, y])) \mathbf{1}\{\tilde{x} \leq x\} \gamma(d\tilde{x}, d\eta) dy \\ &= \underline{y}F(x) + \int_{\underline{y}}^{\bar{y}} (F(x) - I(x, y)) dy, \end{aligned}$$

and, similarly,

$$\int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) = \underline{y}F(x) + \int_{\underline{y}}^{\bar{y}} (F(x) - \min(F(x), G(y))) dy,$$

so

$$\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) - \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) = \int_{\underline{y}}^{\bar{y}} (\min(F(x), G(y)) - I(x, y)) dy.$$

Since the integrand is nonnegative, $x \in X^S$ if and only if $I(x, y) = \min(F(x), G(y))$ for all $y \in Y$, which is equivalent to $I(x, S(x)) = F(x)$, since $G(S(x)) = F(x)$.

Finally, if $x \in X^S \cap (\underline{x}, \bar{x})$ and T is continuous, then $\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) - \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x})$ attains a minimum of 0 at an interior point x , so $T(x) = S(x)$ by the first-order condition.

Q.E.D.

PROOF OF THEOREM 4: By Theorem 3, J solves (P_L) if and only if there exists $(p, q) \in \mathcal{D}_L(v)$ satisfying

$$p(x) + q(z) = v(x, z), \quad \text{for all } (x, z) \in \text{supp}(J),$$

$$\int_Y q(z) J_Y(dz) = \int_Y q(z) G(dz).$$

Define the contact set

$$\Gamma := \{(x, z) \in X \times Y : p(x) + q(z) = v(x, z)\},$$

and notice that Γ is a compact set containing $\text{supp}(J)$. We first prove several lemmas.

LEMMA 8: *There exists a continuous, increasing $T : X \rightarrow Y$ such that*

$$\Gamma = \{(x, T(x)) : x \in X\}.$$

PROOF: Since Γ is closed and $J_X = F$ has full support on X , the set $\Gamma_x := \{z \in Y : (x, z) \in \Gamma\}$ is nonempty for all $x \in X$. Suppose for contradiction that there exist $x_0 \leq x_1$ in X and $z_0 > z_1$ in Y such that (x_0, z_0) and (x_1, z_1) are in Γ . Then, denoting $z = (z_0 + z_1)/2$, we have

$$\begin{aligned} p(x_0) + q(z_0) + p(x_1) + q(z_1) &\geq p(x_0) + p(x_1) + 2q(z) \geq v(x_0, z) + v(x_1, z) \\ &> \frac{1}{2}(v(x_0, z_0) + v(x_0, z_1)) + \frac{1}{2}(v(x_1, z_0) + v(x_1, z_1)) \\ &\geq v(x_0, z_0) + v(x_1, z_1) = p(x_0) + q(z_0) + p(x_1) + q(z_1), \end{aligned}$$

where the first inequality is by $q \in \text{Conv}(Y)$, the second inequality is by $(p, q) \in \mathcal{D}_L(v)$, the third inequality is by strict concavity of v in z , the fourth inequality is by supermodularity of v , and the equality is by the definition of Γ . *Q.E.D.*

Fix a continuous, increasing map T satisfying **(F)**. Let $J \in \Delta(X \times Y)$ be given by $J(x, z) = \min(F(x), H(z))$ for all $(x, z) \in X \times Y$, with $H \in \Delta(Y)$ given by $H(z) = F(T^{-1}(z))$ for all $z \in Y$, where T^{-1} is a generalized inverse of T , defined by $T^{-1}(z) = \max\{x \in X : T(x) \leq z\}$, for all $z \geq T(\underline{x})$, and $T^{-1}(z) = \underline{x}$, for all $z < T(\underline{x})$. Then J

solves (P_L) if and only if there exists $(p, q) \in \mathcal{D}_L(v)$ such that

$$p(x) + q(T(x)) = v(x, T(x)), \quad \text{for all } x \in X, \quad (22)$$

$$\int_X q(S(x)) F(dx) = \int_X q(T(x)) F(dx). \quad (23)$$

Define the sets

$$Y^S = \left\{ y \in Y : \int_{\underline{y}}^y G(\tilde{y}) d\tilde{y} = \int_{\underline{y}}^y H(\tilde{y}) d\tilde{y} \right\}, \quad \text{and}$$

$$Y^C = \left\{ y \in Y : \int_{\underline{y}}^y G(\tilde{y}) d\tilde{y} > \int_{\underline{y}}^y H(\tilde{y}) d\tilde{y} \right\} = Y \setminus Y^S.$$

Notice that Y^C is open and thus is a union of at most countably many disjoint open intervals, $Y^C = \bigcup_k (\underline{y}_k, \bar{y}_k)$. Moreover, recall that any $q \in \text{Lip}(Y) \cap \text{Conv}(Y)$ admits a right-continuous, bounded, and increasing subderivative q' .

LEMMA 9: *For $q \in \text{Lip}(Y) \cap \text{Conv}(Y)$, (23) holds if and only if q is affine on each $(\underline{y}_k, \bar{y}_k)$.*

PROOF: We have

$$\begin{aligned} \int_X (q(S(x)) - q(T(x))) F(dx) &= \int_Y q(y) dG(y) - \int_Y q(y) dH(y) \\ &= \int_Y (H(y) - G(y)) q(dy) = \int_Y \left(\int_{\underline{y}}^y G(\tilde{y}) d\tilde{y} - \int_{\underline{y}}^y H(\tilde{y}) d\tilde{y} \right) q'(dy), \end{aligned}$$

where the first equality is by a change of variables, the second equality is by integration by parts and $G(\bar{y}) = H(\bar{y}) = 1$, and the third equality is by integration by parts and $\int_{\underline{y}}^{\bar{y}} G(y) dy = \int_{\underline{y}}^{\bar{y}} H(y) dy$. Since the integrand is (weakly) positive and q' is increasing, (23) holds if and only if q' is constant on each $(\underline{y}_k, \bar{y}_k)$, where the integrand is strictly positive. *Q.E.D.*

LEMMA 10: *Fix $x \in (\underline{x}, \bar{x})$. Then $x \in X^C$ if and only if $T(x) \in Y^C$.*

PROOF: Integration by parts gives

$$\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) + \int_{\underline{y}}^{T(x)} H(y) dy = F(x)T(x), \quad (24)$$

$$\int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) + \int_{\underline{y}}^{S(x)} G(y) dy = F(x)S(x). \quad (25)$$

If $x \in X^S$, then $T(x) = S(x)$ by Lemma 7. Thus, $T(x) \in Y^S$ by (24) and (25). Conversely, if $T(x) \in Y^S$, then $G(T(x)) = H(T(x))$ (equivalently, $T(x) = S(x)$), since $\int_{\underline{y}}^y G(\tilde{y}) d\tilde{y} - \int_{\underline{y}}^y H(\tilde{y}) d\tilde{y}$ attains a minimum of 0 at $T(x)$. Thus, $x \in X^S$ by (24) and (25). *Q.E.D.*

LEMMA 11: *If J solves (P_L) and (p, q) solves (D_L) , then, for all $x_0, x_1 \in (\underline{x}_i, \bar{x}_i)$, we have*

$$q'(T(x_0)) = v_z(x_0, T(x_0)) = v_z(x_1, T(x_1)) = q'(T(x_1)).$$

PROOF: By Lemma 10, there exists k such that $T(x) \in (\underline{y}_k, \bar{y}_k)$ for all $x \in (\underline{x}_i, \bar{x}_i)$. By Lemma 9, q has the same derivative at $T(x_0)$ and $T(x_1)$, i.e., $q'(T(x_0)) = q'(T(x_1))$. For $j = 0, 1$,

$$p(x_j) = v(x_j, T(x_j)) - q(T(x_j)) = \max_{y \in Y} \{v(x_j, y) - q(y)\},$$

where the first equality is by $(x_j, T(x_j)) \in \Gamma$ and the second equality is by $(p, q) \in \mathcal{D}_L(v)$. Since v and q are continuously differentiable in y on $(\underline{y}_k, \bar{y}_k)$, we get $q'(T(x_j)) = v_z(x_j, T(x_j))$. *Q.E.D.*

LEMMA 12: *If J solves (P_L) and (p, q) solves (D_L) , then (6) and (7) hold for some $c \in \mathbb{R}$. Moreover, $v_z(x, T(x))$ is increasing on X .*

PROOF: Fix $x_0 \in (\underline{x}, \bar{x})$. By $(x_0, T(x_0)) \in \Gamma$ and $(p, q) \in \mathcal{D}_L(v)$, we have

$$p(x_0) = v(x_0, T(x_0)) - q(T(x_0)) = \max_{y \in Y} \{v(x_0, y) - q(y)\}.$$

By continuity of v_x and T , the envelope theorem (Milgrom and Segal, 2002, Theorem 3) implies that p has a derivative $p'(x_0) = v_x(x_0, T(x_0))$ at x_0 , so (6) holds for some $c \in \mathbb{R}$.

Moreover, for all $y \in [T(\underline{x}), T(\bar{x})]$, by (6), (22), and a change of variables, we have

$$q(y) = v(T^{-1}(y), y) - p(T^{-1}(y)) = v(\underline{x}, T(\underline{x})) + \int_{T(\underline{x})}^y v_z(T^{-1}(\tilde{y}), \tilde{y}) d\tilde{y} - c,$$

so (7) holds for all $y \in [T(\underline{x}), T(\bar{x})]$. Next, if $T(\underline{x}) > \underline{y}$, then, since G has positive density on Y , there exists $y_0 \in (T(\underline{x}), \bar{y}]$ such that $\int_{\underline{y}}^y G(\tilde{y}) d\tilde{y} > \int_{\underline{y}}^y H(\tilde{y}) d\tilde{y}$ for all $y \in (\underline{y}, y_0)$. Thus, by Lemmas 9–11, (7) holds for all $y \in [\underline{y}, T(\underline{x})]$. Similarly, if $T(\bar{x}) < \bar{y}$, then, since G has positive density on Y , we have that (7) holds for all $y \in (T(\bar{x}), \bar{y}]$.

Since q is given by (7), and $q \in \text{Conv}(Y)$, it follows that $v_z(T^{-1}(y), y)$ is increasing on $[T(\underline{x}), T(\bar{x})]$. Since v_z is increasing in x and T^{-1} is increasing, it follows that $v_z(x, T(x))$ is increasing on X . Q.E.D.

LEMMA 13: *If $v_z(x, T(x))$ is increasing on X and is constant on each $(\underline{x}_i, \bar{x}_i)$, then (p, q) given by (6) and (7), for any $c \in \mathbb{R}$, is in $\mathcal{D}_L(v)$ and satisfies (22) and (23).*

PROOF: Substituting (p, q) from (6) and (7), and using that v_x is increasing in z and v_z is increasing in x and decreasing in z , we get

$$p(x) + q(T(\hat{x})) - v(x, T(\hat{x})) = \int_{\hat{x}}^x (v_x(\tilde{x}, T(\tilde{x})) - v_x(\tilde{x}, T(\hat{x}))) d\tilde{x} \geq 0, \quad (26)$$

for all $x, \hat{x} \in X$, with equality at $\hat{x} = x$,

$$\begin{aligned} p(x) + q(y) - v(x, y) &= \int_{\underline{x}}^x (v_x(\tilde{x}, T(\tilde{x})) - v_x(\tilde{x}, T(\underline{x}))) d\tilde{x} \\ &+ \int_{\underline{y}}^{T(\underline{x})} (v_z(x, \tilde{y}) - v_z(\underline{x}, T(\underline{x}))) d\tilde{y} \geq 0, \quad \text{for all } x \in X \text{ and } y \in [\underline{y}, T(\underline{x})], \end{aligned} \quad (27)$$

$$\begin{aligned} p(x) + q(y) - v(x, y) &= \int_x^{\bar{x}} (v_x(\tilde{x}, T(\bar{x})) - v_x(\tilde{x}, T(\tilde{x}))) d\tilde{x} \\ &+ \int_{T(\bar{x})}^y (v_z(\bar{x}, T(\bar{x})) - v_z(x, \tilde{y})) d\tilde{y} \geq 0, \quad \text{for all } x \in X \text{ and } y \in (T(\bar{x}), \bar{y}]. \end{aligned} \quad (28)$$

Since $v_z(x, T(x))$ is increasing in x , we have $q \in \text{Conv}(Y)$. Thus, by (26)–(28), (p, q) is in $\mathcal{D}_L(v)$ and satisfies (22). Finally, (23) holds by Lemmas 9 and 10, given that $v_z(x, T(x))$ is constant on each $(\underline{x}_i, \bar{x}_i)$, $q'(y) = q'(T(\underline{x}))$ for all $y \in [\underline{y}, T(\underline{x})]$, and $q'(y) = q'(T(\bar{x}))$ for all $y \in (T(\bar{x}), \bar{y}]$. Q.E.D.

We are now ready to prove Theorem 4.

Part 1. By Theorem 3, any equilibrium assignment J satisfies $\text{supp}(J) \subset \Gamma$. Then, by Lemma 8, there is a unique equilibrium matching T , which is continuous and increasing. We now show that T is strictly increasing. Suppose for contradiction that there exist $x_0 < x_1$ in X and $z \in Y$ such that $z = T(x_0) = T(x_1)$. Note that $v_z(x, T(x))$ is increasing on X by Lemma 12 and is constant on each $(\underline{x}_i, \bar{x}_i)$ by Lemma 11. Since v_z is strictly increasing in x and strictly decreasing in z , it follows that T is strictly increasing on each $(\underline{x}_i, \bar{x}_i)$, so $[x_0, x_1] \subset X^S$, and $\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) = \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x})$, for all $x \in [x_0, x_1]$. Then, by Lemma 7, $S(x) = T(x) = z$ for all $x \in (x_0, x_1)$, yielding $G(z) - G(z_-) \geq F(x_1) - F(x_0) > 0$, contradicting that G has a density on Y .

Part 2. By part 1, there is a unique increasing equilibrium matching T . By Lemmas 11 and 12, $v_z(x, T(x))$ is increasing on X and is constant on each $(\underline{x}_i, \bar{x}_i)$. Conversely, consider an increasing matching T such that $v_z(x, T(x))$ is increasing on X and is constant on each $(\underline{x}_i, \bar{x}_i)$. Since T and $v_z(x, T(x))$ are increasing, and v_z is continuous and strictly decreasing in z , it follows that T cannot have upward jumps and thus is continuous. Then by Theorem 3 and Lemma 13, T is the equilibrium matching.

Part 3. By Theorem 3, the only if part follows from Lemma 12, and the if part follows from Lemma 13. *Q.E.D.*

PROOF OF COROLLARY 1: *Part 1.* Follows from Theorem 4, since, for an increasing T , $v_z(x, T(x)) = a'(T(x) - x) + c$ is increasing in x with $dv_z(x, T(x))/dx = 0$ for all $x \in X^C$ if and only if T is 1-Lipschitz with $T'(x) = 1$ for all $x \in X^C$.

Part 2. It suffices to show that T given by part 2 satisfies part 1. Since $T(x) = x - E'(F(x)) = S(x)$ if $\bar{E}(F(x)) = E(F(x))$, and $T'(x) = 1 - \bar{e}'(F(x))F'(x) = 1$ if $\bar{E}(F(x)) < E(F(x))$, it follows that T is increasing. Then, by $\bar{E} \in \text{Conv}([0, 1])$ and $F \in \Delta(Y)$, $\bar{e} \circ F$ is increasing, so T is 1-Lipschitz. Finally, T satisfies (F) and $x \in X^C$ if and only if $E(F(x)) > \bar{E}(F(x))$ (and hence $T'(x) = 1$), because

$$\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) - \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) = E(F(x)) - \bar{E}(F(x)) \geq 0,$$

for all $x \in X$, with equality at $x = \bar{x}$ by $E(1) = \bar{E}(1)$, and with strict inequality if and only if $E(F(x)) > \bar{E}(F(x))$.

Part 3. Take any $\tilde{T} \in \mathcal{T}$, and define $\tilde{E} : [0, 1] \rightarrow \mathbb{R}$ by $\tilde{E}(t) = \int_0^t (F^{-1}(\tilde{t}) - \tilde{T}(F^{-1}(\tilde{t}))) d\tilde{t}$ for all $t \in [0, 1]$. Then \tilde{E} is convex, because \tilde{T} is 1-Lipschitz. Moreover

$$E(t) - \tilde{E}(t) = \int_{\underline{x}}^{F^{-1}(t)} \tilde{T}(x) F(dx) - \int_{\underline{x}}^{F^{-1}(t)} S(x) F(dx) \geq 0,$$

for all $t \in [0, 1]$, because \tilde{T} satisfies (F). Hence $\tilde{E}(t) \leq \bar{E}(t)$ for all $t \in [0, 1]$, because \bar{E} is the largest convex function that lies below E . Consequently, $T \circ F^{-1} \succeq \tilde{T} \circ F^{-1}$, because

$$\int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) - \int_{\underline{x}}^x \tilde{T}(\tilde{x}) F(d\tilde{x}) = \tilde{E}(F(x)) - \bar{E}(F(x)) \leq 0,$$

for all $x \in X$.

Q.E.D.

PROOF OF THEOREM 5: *Part 1.* Follows from Theorem 4.

Part 2. Follows from Lemma 14 with $\hat{x} = \underline{x}$.

Part 3. By strict quasiconcavity, there exists $\hat{x} \in X$ such that $v_z(x, S(x))$ is strictly increasing on $[\underline{x}, \hat{x})$ and strictly decreasing on $(\hat{x}, \bar{x}]$. Assume that $\hat{x} < \bar{x}$, as otherwise part 1 applies. Then part 3 follows from the next lemma.

LEMMA 14: *There exists $x^* \in [\underline{x}, \hat{x}]$ and an increasing $T : X \rightarrow Y$ such that*

$$\int_{x^*}^{\bar{x}} T(x) F(dx) = \int_{x^*}^{\bar{x}} S(x) F(dx), \quad (29)$$

$$v_z(x^*, T(x^*)) = v_z(x, T(x)), \quad \text{for all } x \in [x^*, \bar{x}], \quad (30)$$

$$S(x^*) \leq T(x^*), \quad \text{with equality if } x^* \neq \underline{x}. \quad (31)$$

Moreover, the equilibrium matching is upper compression with $X^C = (x^*, \bar{x})$.

PROOF: Define $O(x; z) = \int_{\underline{x}}^{\bar{x}} (T(\tilde{x}) - S(\tilde{x})) F(d\tilde{x})$, where $T(\tilde{x})$ is given by $v_z(x, z) = v_z(\tilde{x}, T(\tilde{x}))$. First, consider the case $O(\underline{x}; S(\underline{x})) \leq 0$. In this case, $v_z(\underline{x}, S(\underline{x})) > v_z(\bar{x}, S(\bar{x}))$, as otherwise, for all $\tilde{x} \in (\underline{x}, \bar{x})$, $v_z(\tilde{x}, T(\tilde{x})) < v_z(\tilde{x}, S(\tilde{x}))$ because $v_z(x, S(x))$ is strictly quasiconcave, so $T(\tilde{x}) > S(\tilde{x})$ because v_z is strictly decreasing in z , and thus $O(\underline{x}; S(\underline{x})) > 0$, yielding a contradiction. Let \hat{z} be given by $v_z(\underline{x}, \hat{z}) = v_z(\bar{x}, S(\bar{x}))$. Then, for all $\tilde{x} \in (\underline{x}, \bar{x})$, $v_z(\tilde{x}, T(\tilde{x})) < v_z(\tilde{x}, S(\tilde{x}))$, so $T(\tilde{x}) > S(\tilde{x})$, and thus $O(\underline{x}; \hat{z}) > 0$. By the intermediate value theorem, there exists $z^* \in [S(\underline{x}), \hat{z})$ such that $O(\underline{x}; z^*) = 0$, so (29)–(31) hold. Since $v_z(x, S(x))$ is strictly quasiconcave and v_z is strictly decreasing in z ,

there exists $x_0 \in (\underline{x}, \bar{x})$ such that $S(\tilde{x}) < T(\tilde{x})$ for all $\tilde{x} \in (\underline{x}, x_0)$ and $S(\tilde{x}) > T(\tilde{x})$ for all $\tilde{x} \in (x_0, \bar{x}]$, implying that (F) holds strictly for all $x \in (\underline{x}, \bar{x})$. Thus, the equilibrium matching is full compression by Theorem 4.

Next, consider the case $O(\underline{x}; S(\underline{x})) > 0$. Note that $O(\hat{x}; S(\hat{x})) < 0$, as, for all $\tilde{x} \in (\hat{x}, \bar{x}]$, $v_z(\tilde{x}, T(\tilde{x})) = v_z(\hat{x}, S(\hat{x})) > v_z(\tilde{x}, S(\tilde{x}))$, so $T(\tilde{x}) < S(\tilde{x})$. By the intermediate value theorem, there exists $x^* \in (\underline{x}, \hat{x})$ such that $O(x^*; S(x^*)) = 0$, so (29)–(31) hold. Then there exists $x_0 \in (x^*, \bar{x})$ such that $S(\tilde{x}) < T(\tilde{x})$ for all $\tilde{x} \in (x^*, x_0)$ and $S(\tilde{x}) > T(\tilde{x})$ for all $\tilde{x} \in (x_0, \bar{x}]$. Thus the equilibrium matching is upper compression with $X^C = (x^*, \bar{x})$ by Theorem 4. Q.E.D.

Part 4. Follows from part 3 by symmetry. Q.E.D.

A.5. Comparative Statics in the Mean-Measurable Case

PROOF OF REMARK 11: Suppose by contradiction that $\hat{v}_z \geq v_z$ but there exist $x \in (\underline{x}, \bar{x})$, $z \in (\underline{y}, \bar{y})$, and $k > 0$ such that

$$\frac{\hat{v}_{xz}(x, z)}{-\hat{v}_{zz}(x, z)} < k < \frac{v_{xz}(x, z)}{-v_{zz}(x, z)}.$$

Then, by a Taylor expansion, there exists $\varepsilon > 0$ such that $v_z(x, z) < v_z(x + \varepsilon, z + k\varepsilon)$ and $\hat{v}_z(x, z) > \hat{v}_z(x + \varepsilon, z + k\varepsilon)$, contradicting that $\hat{v}_z \geq v_z$.

Next, by rearrangement,

$$\frac{\hat{v}_{xz}(x, z)}{-\hat{v}_{zz}(x', z')} \geq \frac{v_{xz}(x, z)}{-v_{zz}(x', z')},$$

for all (x, z) and (x', z') in $X \times Y$ is equivalent to

$$\inf_{(x, z) \in X \times Y} \frac{\hat{v}_{xz}(x, z)}{v_{xz}(x, z)} \geq \sup_{(x', z') \in X \times Y} \frac{-\hat{v}_{zz}(x', z')}{-v_{zz}(x', z')},$$

which is equivalent to the existence of $c > 0$ such that

$$\frac{\hat{v}_{xz}(x, z)}{v_{xz}(x, z)} \geq c \geq \frac{-\hat{v}_{zz}(x', z')}{-v_{zz}(x', z')},$$

for all (x, z) and (x', z') in $X \times Y$, meaning that $\hat{v}_{xz}(x, z) \geq cv_{xz}(x, z)$ and $-\hat{v}_{zz}(x, z) \leq -cv_{zz}(x, z)$ for all (x, z) .

Now, suppose that such $c > 0$ exists. If $x' \geq x$, $z' \geq z$, and $v_z(x', z') \geq v_z(x, z)$, then

$$\begin{aligned} \hat{v}_z(x', z') - \hat{v}_z(x, z) &= \int_x^{x'} \hat{v}_{xz}(\tilde{x}, z) d\tilde{x} + \int_z^{z'} \hat{v}_{zz}(x', \tilde{z}) d\tilde{z} \\ &\geq c \int_x^{x'} v_{xz}(\tilde{x}, z) d\tilde{x} + c \int_z^{z'} v_{zz}(x', \tilde{z}) d\tilde{z} \\ &= c(v_z(x', z') - v_z(x, z)) \geq 0, \end{aligned}$$

implying that $\hat{v}_z \supseteq v_z$.

Q.E.D.

PROOF OF THEOREM 6: *If parts of both parts.* Suppose that $\hat{v}_z \supseteq v_z$ and $\hat{G} \preceq G$. Let T and \hat{T} denote the equilibrium matchings under (v, F, G) and (\hat{v}, F, \hat{G}) , and let $S = G^{-1} \circ F$ and $\hat{S} = \hat{G}^{-1} \circ F$. Suppose by contradiction that $\hat{T} \circ F^{-1} \not\subseteq T \circ F^{-1}$. Then there exist $x_0 < x_1$ in X such that $\int_{\underline{x}}^{x_i} \hat{T}(\tilde{x}) F(d\tilde{x}) = \int_{\underline{x}}^{x_i} T(\tilde{x}) F(d\tilde{x})$ for $i = 0, 1$ and $\int_{\underline{x}}^x \hat{T}(\tilde{x}) F(d\tilde{x}) > \int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x})$ for all $x \in (x_0, x_1)$. In turn, there exist $\tilde{x}_0 < \tilde{x}_1$ in (x_0, x_1) such that $\hat{T}(\tilde{x}_0) > T(\tilde{x}_0)$ and $\hat{T}(\tilde{x}_1) < T(\tilde{x}_1)$. Next, $\hat{v}_z(\tilde{x}_0, \hat{T}(\tilde{x}_0)) = \hat{v}_z(\tilde{x}_1, \hat{T}(\tilde{x}_1))$ by Theorem 4, because

$$\int_{\underline{x}}^x \hat{T}(\tilde{x}) F(d\tilde{x}) > \int_{\underline{x}}^x T(\tilde{x}) F(d\tilde{x}) \geq \int_{\underline{x}}^x S(\tilde{x}) F(d\tilde{x}) \geq \int_{\underline{x}}^x \hat{S}(\tilde{x}) F(d\tilde{x}),$$

for all $x \in (x_0, x_1)$. But then

$$\hat{v}_z(\tilde{x}_0, T(\tilde{x}_0)) > \hat{v}_z(\tilde{x}_0, \hat{T}(\tilde{x}_0)) = \hat{v}_z(\tilde{x}_1, \hat{T}(\tilde{x}_1)) > \hat{v}_z(\tilde{x}_1, T(\tilde{x}_1)),$$

which implies, by $\hat{v}_z \supseteq v_z$, that $v_z(\tilde{x}_0, T(\tilde{x}_0)) > v_z(\tilde{x}_1, T(\tilde{x}_1))$, contradicting Theorem 4, which shows that $v_z(x, T(x))$ is increasing in x .

Only if part of part 1. Suppose that $\hat{v}_z \not\supseteq v_z$, so there exist $x_1 \geq x_0$ in X and $y_1 \geq y_0$ in Y such that $v_z(x_1, y_1) \geq v_z(x_0, y_0)$ but $\hat{v}_z(x_1, y_1) < \hat{v}_z(x_0, y_0)$. Consider discrete distributions $F^d = \delta_{x_0}/2 + \delta_{x_1}/2$ and $G^d = \delta_{y_0}/2 + \delta_{y_1}/2$. Let $J^d = \delta_{(x_0, y_0)}/2 + \delta_{(x_1, y_1)}/2$, $r \in [v_z(x_0, y_0), v_z(x_1, y_1)]$, $p(x_i) = v(x_i, y_i) - ry_i$, and $q(z) = rz$ for all $z \in [y_0, y_1]$ and $i = 0, 1$. Then J^d is the unique equilibrium assignment under (v, F^d, G^d) by Theorem 3, as

$$\begin{aligned} v(x_i, y_i) - ry_i + rz - v(x_i, z) &= r(z - y_i) - \int_{y_i}^z v_z(x_i, \tilde{z}) d\tilde{z} \\ &\geq \int_{y_i}^z (v_z(x_i, y_i) - v_z(x_i, \tilde{z})) d\tilde{z} \geq 0, \end{aligned}$$

for all $z \in [y_0, y_1]$ and $i = 0, 1$, with both inequalities holding as equality if and only if $z = y_i$. Next, there exist unique $\hat{z}_0 < \hat{z}_1$ in (y_0, y_1) such that $\hat{z}_0 + \hat{z}_1 = y_0 + y_1$ and $\hat{v}_z(x_0, \hat{z}_0) = \hat{v}_z(x_1, \hat{z}_1)$. Let $\hat{J}^d = \delta_{(x_0, \hat{z}_0)}/2 + \delta_{(x_1, \hat{z}_1)}/2$, $\hat{r} = \hat{v}_z(x_0, \hat{z}_0) = \hat{v}_z(x_1, \hat{z}_1)$, $\hat{p}(x_i) = \hat{v}(x_i, \hat{z}_i) - \hat{r}\hat{z}_i$, and $\hat{q}(z) = \hat{r}z$ for all $z \in [y_0, y_1]$ and $i = 0, 1$. Then \hat{J}^d is the unique equilibrium assignment under (\hat{v}, F^d, G^d) by Theorem 3, as

$$\hat{v}(x_i, \hat{z}_i) - \hat{r}\hat{z}_i + \hat{r}z - \hat{v}(x_i, z) = \int_{\hat{z}_i}^z (\hat{v}_z(x_i, \hat{z}_i) - \hat{v}_z(x_i, \tilde{z})) d\tilde{z} \geq 0,$$

for all $z \in [y_0, y_1]$ and $i = 0, 1$, with the inequality holding as equality if and only if $z = \hat{z}_i$. Clearly, $\hat{J}_Y^d \not\leq J_Y^d$. By continuity and approximation, there exist distributions F and G with continuous positive densities such that the equilibrium assignments under (\hat{v}, F, G) and (v, F, G) satisfy $\hat{J}_Y \not\leq J_Y$, and thus $\hat{T} \circ F^{-1} \not\leq T \circ F^{-1}$.

Only if part of part 2. Suppose that $\hat{G} \not\leq G$. Consider the convex-loss case with $F = G$, so that $T(x) = S(x) = x$ by Corollary 1. Then $\hat{T} \circ F^{-1} \not\leq T \circ F^{-1}$, as otherwise $\hat{G}^{-1} = \hat{S} \circ F^{-1} \succeq \hat{T} \circ F^{-1} \succeq T \circ F^{-1} = S \circ F^{-1} = G^{-1}$, contradicting that $\hat{G} \not\leq G$. *Q.E.D.*

PROOF OF THEOREM 7: *If parts of parts 1 and 2.* Suppose that $F \leq \hat{F}$ and $\hat{G} \leq G$. Let T and \hat{T} denote the equilibrium matchings under (v, F, G) and (v, \hat{F}, \hat{G}) , and let $S = G^{-1} \circ F$ and $\hat{S} = \hat{G}^{-1} \circ \hat{F}$. Suppose by contradiction that $\hat{T} \not\leq T$. Then the set $A := \{x \in X : \hat{T}(x) < T(x)\}$ is non-empty. By continuity of \hat{T} and T , the set A is open and is thus a union of at most countably many disjoint open intervals. Take one such interval (\underline{a}, \bar{a}) , with $\underline{x} \leq \underline{a} < \bar{a} \leq \bar{x}$. Note that $\int_{\underline{x}}^{\underline{a}} \hat{T}(x) \hat{F}(dx) = \int_{\underline{x}}^{\underline{a}} \hat{S}(x) \hat{F}(dx)$ and thus $\hat{T}(\underline{a}) \geq \hat{S}(\underline{a}) = \hat{G}^{-1}(\hat{F}(\underline{a}))$ (with equality if $\underline{a} > \underline{x}$), as otherwise, $\int_{\underline{x}}^{\underline{a}} \hat{T}(x) \hat{F}(dx) > \int_{\underline{x}}^{\underline{a}} \hat{S}(x) \hat{F}(dx)$, so $\underline{a} > \underline{x}$ and, for some $x \in (\underline{a}, \bar{a})$,

$$v_z(\underline{a}, T(\underline{a})) = v_z(\underline{a}, \hat{T}(\underline{a})) = v_z(x, \hat{T}(x)) > v_z(x, T(x)) \geq v_z(\underline{a}, T(\underline{a})).$$

Similarly, $\int_{\bar{x}}^{\bar{a}} T(x) F(dx) = \int_{\bar{x}}^{\bar{a}} S(x) F(dx)$ and thus $T(\bar{a}) \leq S(\bar{a})$ (with equality if $\bar{a} < \bar{x}$), as otherwise $\int_{\bar{x}}^{\bar{a}} T(x) F(dx) > \int_{\bar{x}}^{\bar{a}} S(x) F(dx)$, so $\bar{a} < \bar{x}$ and, for some $x \in (\underline{a}, \bar{a})$,

$$v_z(\bar{a}, \hat{T}(\bar{a})) \geq v_z(x, \hat{T}(x)) > v_z(x, T(x)) = v_z(\bar{a}, T(\bar{a})) = v_z(\bar{a}, \hat{T}(\bar{a})).$$

Next,

$$\begin{aligned}
0 &\stackrel{(a)}{\leq} \int_{\underline{x}}^{\bar{a}} (\hat{T}(x) - \hat{S}(x)) \hat{F}(dx) \stackrel{(b)}{=} \int_{\underline{a}}^{\bar{a}} (\hat{T}(x) - \hat{S}(x)) \hat{F}(dx) \\
&\stackrel{(c)}{=} \int_{\hat{F}(\underline{a})}^{\hat{F}(\bar{a})} (\hat{T}(\hat{F}^{-1}(t)) - \hat{G}^{-1}(t)) dt \stackrel{(d)}{\leq} \int_{\hat{F}(\underline{a})}^{\hat{F}(\bar{a})} (T(\hat{F}^{-1}(t)) - \hat{G}^{-1}(t)) dt,
\end{aligned} \tag{32}$$

where (a) holds because $\hat{T} \circ \hat{F}^{-1} \preceq \hat{G}^{-1}$, (b) holds because $\int_{\underline{x}}^{\bar{a}} \hat{T}(x) \hat{F}(dx) = \int_{\underline{a}}^{\bar{a}} \hat{S}(x) \hat{F}(dx)$, (c) holds by a change of variable, and (d) holds because $\hat{T}(x) < T(x)$ for $x \in (\underline{a}, \bar{a})$.

Note that

$$T(\hat{F}^{-1}(t)) \leq T(\bar{a}) \leq G^{-1}(F(\bar{a})) \leq G^{-1}(t) \leq \hat{G}^{-1}(t), \quad \text{for all } t \in [F(\bar{a}), \hat{F}(\bar{a})], \tag{33}$$

where the first inequality holds because T is increasing and $t \leq \hat{F}(\bar{a})$, the second inequality holds because $T(\bar{a}) \leq S(\bar{a})$, the third inequality holds because G^{-1} is increasing and $t \geq F(\bar{a})$, and the fourth inequality holds because $G^{-1} \leq \hat{G}^{-1}$. Similarly, we have

$$\hat{T}(F^{-1}(t)) \geq \hat{T}(\underline{a}) \geq \hat{G}^{-1}(\hat{F}(\underline{a})) \geq \hat{G}^{-1}(t) \geq G^{-1}(t), \quad \text{for all } t \in [F(\underline{a}), \hat{F}(\underline{a})], \tag{34}$$

where the first inequality holds because \hat{T} is increasing and $t \geq F(\underline{a})$, the second inequality holds because $\hat{T}(\underline{a}) \geq \hat{S}(\underline{a})$, the third inequality holds because \hat{G}^{-1} is increasing and $t \leq \hat{F}(\underline{a})$, and the fourth inequality holds because $\hat{G}^{-1} \geq G^{-1}$.

Finally, if $F(\bar{a}) \leq \hat{F}(\underline{a})$, then $[\hat{F}(\underline{a}), \hat{F}(\bar{a})] \subset [F(\bar{a}), \hat{F}(\bar{a})]$, so, by (33), the integrand in (32) is negative, yielding a contradiction. Otherwise (if $F(\bar{a}) > \hat{F}(\underline{a})$), we again get a contradiction,

$$\begin{aligned}
0 &\stackrel{(e)}{\leq} \int_{\hat{F}(\underline{a})}^{\hat{F}(\bar{a})} (T(\hat{F}^{-1}(t)) - \hat{G}^{-1}(t)) dt \stackrel{(f)}{\leq} \int_{\hat{F}(\underline{a})}^{F(\bar{a})} (T(\hat{F}^{-1}(t)) - \hat{G}^{-1}(t)) dt \\
&\stackrel{(g)}{\leq} \int_{\hat{F}(\underline{a})}^{F(\bar{a})} (T(F^{-1}(t)) - G^{-1}(t)) dt \stackrel{(h)}{\leq} \int_{F(\underline{a})}^{F(\bar{a})} (T(F^{-1}(t)) - G^{-1}(t)) dt \\
&\stackrel{(i)}{=} \int_{\underline{a}}^{\bar{a}} (T(x) - S(x)) F(dx) \stackrel{(j)}{=} - \int_{\underline{x}}^{\bar{a}} (T(x) - S(x)) F(dx) \stackrel{(k)}{\leq} 0,
\end{aligned}$$

where (e) holds by (32), (f) holds by (33), (g) holds because T is increasing, $\hat{F}^{-1} \leq F^{-1}$, and $\hat{G}^{-1} \geq G^{-1}$, (h) holds because $T(F^{-1}(t)) \geq G^{-1}(t)$ for all $t \in [F(\underline{a}), \hat{F}(\underline{a})]$ by (34)

and $T(F^{-1}(t)) \geq \hat{T}(F^{-1}(t))$ for all $t \in [F(\underline{a}), F(\bar{a})] \supset [F(\underline{a}), \hat{F}(\underline{a})]$, (i) holds by a change of variable, (j) holds because $\int_{\underline{x}}^{\bar{a}} T(x) F(dx) = \int_{\underline{x}}^{\bar{a}} S(x) F(dx)$, and (k) holds because $T \circ F^{-1} \preceq G^{-1}$.

Only if part of part 1. Suppose that $F \not\preceq \hat{F}$, so that there exists $x_0 \in (\underline{x}, \bar{x})$ such that $F(x_0) > \hat{F}(x_0)$. Consider the convex-loss case with G uniform on $[0, \varepsilon]$ such that, for all $x \in X$, we have $\varepsilon \leq 1/F'(x)$ and $\varepsilon \leq 1/\hat{F}'(x)$, so $S'(x) = F'(x)\varepsilon \leq 1$ and $\hat{S}'(x) = \hat{F}'(x)\varepsilon \leq 1$. By Corollary 1, the equilibrium matchings are $T = S$ and $\hat{T} = \hat{S}$, so $\hat{T}(x_0) = G^{-1}(\hat{F}(x_0)) < G^{-1}(F(x_0)) = T(x_0)$, showing that $\hat{T} \not\preceq T$.

Only if part of part 2. Suppose that $\hat{G} \not\preceq G$, so that there exists $y_0 \in (\underline{y}, \bar{y})$ such that $\hat{G}(y_0) > G(y_0)$. Consider the convex-loss case with F uniform on $[0, 1/\varepsilon]$ such that, for all $y \in Y$, we have $\varepsilon \leq G'(y)$ and $\varepsilon \leq \hat{G}'(y)$, so $S'(x) = \varepsilon/G'(S(x)) \leq 1$ and $\hat{S}'(x) = \varepsilon/\hat{G}'(\hat{S}(x)) \leq 1$. By Corollary 1, the equilibrium matchings are $T = S$ and $\hat{T} = \hat{S}$, so, letting $x_0 = F^{-1}(G(y_0))$, we have $\hat{T}(x_0) = \hat{G}^{-1}(F(x_0)) < G^{-1}(F(x_0)) = T(x_0)$, showing that $\hat{T} \not\preceq T$.

Part 3. Follows from part 3 of Theorem 4.

Q.E.D.

A.6. Duality in the Moment-Measurable Case

We first show that Theorem 1 applies to the moment-measurable case.

LEMMA 15: *If $v \in \text{Lip}(X \times \mathbb{R})$ and $m \in \text{Lip}(X \times Y)$, then $V(x, \eta) = v(x, \mathbb{E}_\eta[m(x, y)]) \in \text{Lip}(X \times \Delta(Y))$.*

PROOF: Letting L_v and L_m denote Lipschitz constants of v and m , for all $x, \hat{x} \in X$ and $\eta, \hat{\eta} \in \Delta(Y)$, we have

$$\begin{aligned} V(x, \eta) - V(\hat{x}, \hat{\eta}) &= \left(v \left(x, \int_Y m(x, y) \eta(dy) \right) - v \left(\hat{x}, \int_Y m(x, y) \eta(dy) \right) \right) \\ &\quad + \left(v \left(\hat{x}, \int_Y m(x, y) \eta(dy) \right) - v \left(\hat{x}, \int_Y m(\hat{x}, y) \eta(dy) \right) \right) \\ &\quad + \left(v \left(\hat{x}, \int_Y m(\hat{x}, y) \eta(dy) \right) - v \left(\hat{x}, \int_Y m(\hat{x}, y) \hat{\eta}(dy) \right) \right) \\ &\leq L_v \rho_X(x, \hat{x}) + L_v L_m \rho_X(x, \hat{x}) + L_v L_m \|\eta - \hat{\eta}\|_{KR}, \end{aligned}$$

so $V \in \text{Lip}(X \times \Delta(Y))$

Q.E.D.

Next, we simplify the primal and dual problems. Let $Z = [\underline{z}, \bar{z}]$, where $\underline{z} = \min_{x \in X} m(x, \underline{y})$ and $\bar{z} = \max_{x \in X} m(x, \bar{y})$. Moreover, for $\pi \in \Delta(X \times Y \times Z)$, let $\pi_{x,z}$ denote the conditional distribution of y given (x, z) . The planner's problem is

$$\begin{aligned} & \max_{\pi \in \Delta(X \times Y \times Z)} \mathbb{E}_{\pi}[v(x, z)] \\ \text{s.t. } & \pi_X = \mu, \quad \pi_Y = \nu, \quad \text{and} \quad \mathbb{E}_{\pi_{x,z}}[m(x, y)] = z, \quad \text{for all } (x, z) \in X \times Z. \end{aligned} \tag{P_M}$$

Let $B(X \times Z)$ denote the set of bounded, measurable functions on $X \times Z$. The dual problem is

$$\begin{aligned} & \min_{(p,q,r) \in \text{Lip}(X) \times \text{Lip}(Y) \times B(X \times Z)} \mathbb{E}_{\mu}[p(x)] + \mathbb{E}_{\nu}[q(y)] \\ \text{s.t. } & p(x) + q(y) \geq v(x, z) + r(x, z)(m(x, y) - z), \quad \text{for all } (x, y, z) \in X \times Y \times Z. \end{aligned} \tag{D_M}$$

Let $\mathcal{D}_M(v, m)$ denote the set of feasible solutions to (D_M).

LEMMA 16: *Suppose that v is continuously differentiable and m is Lipschitz. Then $(p, q) \in \mathcal{D}(V)$ if and only if there exists $r \in B(X \times Z)$ such that $(p, q, r) \in \mathcal{D}_M(v, m)$.*

PROOF: Suppose that $(p, q, r) \in \mathcal{D}_M(v, m)$. Then, for all $\eta \in \Delta(Y)$ and $z = \mathbb{E}_{\eta}[m(x, y)]$, we have

$$\int_Y (p(x) + q(y))\eta(dy) \geq \int_Y (v(x, z) + r(x, z)(m(x, y) - z))\eta(dy) = v(x, z) = V(x, \eta),$$

so $(p, q) \in \mathcal{D}(V)$.

Suppose now that $(p, q) \in \mathcal{D}(V)$. Considering all $\eta \in \Delta(Y)$ with $|\text{supp}(\eta)| = 1$ gives

$$p(x) + q(y) \geq v(x, m(x, y)), \quad \text{for all } (x, y) \in X \times Y. \tag{35}$$

Considering all $\eta \in \Delta(Y)$ with $|\text{supp}(\eta)| = 2$ gives

$$\frac{p(x) + q(y_1) - v(x, z)}{m(x, y_1) - z} \leq \frac{p(x) + q(y_2) - v(x, z)}{m(x, y_2) - z}, \tag{36}$$

for all $y_1, y_2 \in Y$ and $z \in Z$ such that $m(x, y_1) < z < m(x, y_2)$.

Define \tilde{r} on $X \times Y \times Z$ by

$$\tilde{r}(x, y, z) = \begin{cases} v_z(x, z), & z = m(x, y), \\ \frac{v(x, m(x, y)) - v(x, z)}{m(x, y) - z}, & z \neq m(x, y). \end{cases}$$

Note that \tilde{r} is continuous on $X \times Y \times Z$, because, by the mean value theorem, for each $(x, y, z) \in X \times Y \times Z$, there exists \hat{z} between z and $m(x, y)$ such that $\tilde{r}(x, y, z) = v_z(x, \hat{z})$. Thus, there exists C such that $|\tilde{r}(x, y, z)| \leq C$ for all $(x, y, z) \in X \times Y \times Z$.

Define \underline{r} and \bar{r} on $X \times Z$ by

$$\underline{r}(x, z) = \begin{cases} \sup_{y_1 \in Y: m(x, y_1) < z} \frac{p(x) + q(y_1) - v(x, z)}{m(x, y_1) - z}, & z > m(x, \underline{y}), \\ -\infty, & z \leq m(x, \underline{y}), \end{cases}$$

$$\bar{r}(x, z) = \begin{cases} \inf_{y_2 \in Y: m(x, y_2) > z} \frac{p(x) + q(y_2) - v(x, z)}{m(x, y_2) - z}, & z < m(x, \bar{y}), \\ +\infty, & z \geq m(x, \bar{y}). \end{cases}$$

Note that the correspondence R defined by $R(x, z) = [\underline{r}(x, z), \bar{r}(x, z)]$, for all $(x, z) \in X \times Z$, is nonempty valued, by (36). Moreover, for all $(x, z) \in X \times Z$, the set $R(x, z) \cap [-C, C]$ is nonempty, because $\bar{r}(x, z) \geq -C$ and $\underline{r}(x, z) \leq C$ for all $(x, z) \in X \times Z$, as, by (35), we have

$$\frac{p(x) + q(y_2) - v(x, z)}{m(x, y_2) - z} \geq \frac{v(x, m(x, y_2)) - v(x, z)}{m(x, y_2) - z} = \tilde{r}(x, y_2, z), \quad z < m(x, y_2),$$

$$\frac{p(x) + q(y_1) - v(x, z)}{m(x, y_1) - z} \leq \frac{v(x, m(x, y_1)) - v(x, z)}{m(x, y_1) - z} = \tilde{r}(x, y_1, z), \quad z > m(x, y_1).$$

Then r given by $r(x, z) = \arg \min_{t \in R(x, z)} |t|$, for all $(x, z) \in X \times Z$, is a bounded, measurable function on $X \times Z$, by the measurable maximum theorem (Aliprantis and Border, 2006, Theorem 18.19). Finally, it is easy to verify that $(p, q, r) \in \mathcal{D}_M(v, m)$. *Q.E.D.*

By Theorem 1 and Lemmas 15 and 16, there exists a solution (p, q, r) to (\mathbf{D}_M) such that (p, q) is a solution to (\mathbf{D}) . Define the contact set

$$\Lambda = \{(x, \eta) \in X \times \Delta(Y) : p(x) + \mathbb{E}_\eta[q(y)] = v(x, \mathbb{E}_\eta[m(x, y)])\},$$

which is compact by continuity in (x, η) . By Theorem 1, $\gamma \in \mathcal{P}(\mu, \nu)$ solves (P) if and only if $\text{supp}(\gamma) \subset \Lambda$. Moreover, for every $(x, \eta) \in \Lambda$ and $z = \mathbb{E}_\eta[m(x, y)]$, we have

$$\int_Y (p(x) + q(y) - v(x, z) - r(x, z)(m(x, y) - z))\eta(dy) = p(x) + \mathbb{E}_\eta[q(y)] - v(x, z) = 0.$$

Since the integrand is nonnegative, by $(p, q, r) \in \mathcal{D}_M(v, m)$, and continuous in y , we have, for all $y \in \text{supp}(\eta)$,

$$\begin{aligned} p(x) + q(y) &= v(x, z) + r(x, z)(m(x, y) - z), \\ \text{for all } (x, \eta) \in \Lambda \text{ and } z &= \mathbb{E}_\eta[m(x, y)]. \end{aligned} \tag{37}$$

LEMMA 17: *Suppose that v and m are continuously differentiable. Let (p, q, r) be a solution to (\mathbf{D}_M) , let $(x, \eta) \in \Lambda$ with nondegenerate η , and let $z = \mathbb{E}_\eta[m(x, y)]$. Then $r(x, z) = v_z(x, z)$. Moreover, for Lebesgue almost every such x , r has a partial derivative $r_x(x, z)$ with respect to x at (x, z) satisfying, for all $y \in \text{supp}(\eta)$,*

$$v_z(x, z)\mathbb{E}_\eta[m_x(x, y)] = v_z(x, z)m_x(x, y) + r_x(x, z)(m(x, y) - z). \tag{38}$$

PROOF: Since η is nondegenerate and $m_y(x, y) > 0$, for all $(x, y) \in X \times Y$, there exist $y_0, y_1 \in \text{supp}(\eta)$ such that $m(x, y_0) < z < m(x, y_1)$. By (37) and $(p, q, r) \in \mathcal{D}_M(v, m)$, we have, for all $\hat{z} \in (m(x, y_0), m(x, y_1))$ and $i = 0, 1$,

$$p(x) + q(y_i) = v(x, z) + r(x, z)(m(x, y_i) - z) \geq v(x, \hat{z}) + r(x, \hat{z})(m(x, y_i) - \hat{z}),$$

yielding

$$\begin{aligned} (p(x) + q(y_0))\frac{m(x, y_1) - \hat{z}}{m(x, y_1) - m(x, y_0)} + (p(x) + q(y_1))\frac{\hat{z} - m(x, y_0)}{m(x, y_1) - m(x, y_0)} \\ = v(x, z) + r(x, z)(\hat{z} - z) \geq v(x, \hat{z}). \end{aligned}$$

Thus, $r(x, z) = v_z(x, z)$, as follows from

$$v_z(x, z) = \lim_{\hat{z} \downarrow z} \frac{v(x, \hat{z}) - v(x, z)}{\hat{z} - z} \leq r(x, z) \leq \lim_{\hat{z} \uparrow z} \frac{v(x, \hat{z}) - v(x, z)}{\hat{z} - z} = v_z(x, z).$$

Let \hat{X} be the set of interior points x of X where p has a derivative, which we denote by $p'(x)$. By Rademacher's theorem (e.g., Santambrogio, 2015, Box 1.9), $X \setminus \hat{X}$ has zero

Lebesgue measure. Fix $(x, \eta) \in \Lambda$ with $x \in \hat{X}$ and $|\text{supp}(\eta)| > 1$. Recall that there exist $y_0, y_1 \in \text{supp}(\eta)$ such that $m(x, y_0) < z < m(x, y_1)$. By (37) and $(p, q, r) \in \mathcal{D}_M(v, m)$, we have, for all $\hat{x} \in X$ with $m(\hat{x}, y_0) < z < m(\hat{x}, y_1)$, and $y \in \text{supp}(\eta)$,

$$q(y) = v(x, z) + r(x, z)(m(x, y) - z) - p(x) \geq v(\hat{x}, z) + r(\hat{x}, z)(m(\hat{x}, y) - z) - p(\hat{x}). \quad (39)$$

Denote, for $i = 0, 1$,

$$C_i = \frac{v_x(x, z) + r(x, z)m_x(x, y_i) - p'(x)}{z - m(x, y_i)}.$$

Considering $y = y_0$ in (39) and recalling that $r(x, z) = v_z(x, z)$ yields

$$r(\hat{x}, z) - r(x, z) \geq \frac{v(\hat{x}, z) - v(x, z) + v_z(x, z)(m(\hat{x}, y_0) - m(x, y_0)) - (p(\hat{x}) - p(x))}{z - m(\hat{x}, y_0)},$$

so

$$\begin{aligned} \underline{r}_x(x_+, z) &:= \liminf_{\hat{x} \downarrow x} \frac{r(\hat{x}, z) - r(x, z)}{\hat{x} - x} \geq C_0, \\ \bar{r}_x(x_-, z) &:= \limsup_{\hat{x} \uparrow x} \frac{r(\hat{x}, z) - r(x, z)}{\hat{x} - x} \leq C_0. \end{aligned}$$

Similarly, considering $y = y_1$ in (39) yields

$$\begin{aligned} \bar{r}_x(x_+, z) &:= \limsup_{\hat{x} \downarrow x} \frac{r(\hat{x}, z) - r(x, z)}{\hat{x} - x} \leq C_1, \\ \underline{r}_x(x_-, z) &:= \liminf_{\hat{x} \uparrow x} \frac{r(\hat{x}, z) - r(x, z)}{\hat{x} - x} \geq C_1. \end{aligned}$$

Thus,

$$C_0 \leq \underline{r}_x(x_+, z) \leq \bar{r}_x(x_+, z) \leq C_1 \leq \underline{r}_x(x_-, z) \leq \bar{r}_x(x_-, z) \leq C_0,$$

implying that all inequalities hold with equality, so r has a derivative $r_x(x, z)$ at (x, z) , which then, by (39), must satisfy, for all $y \in \text{supp}(\eta)$,

$$v_x(x, z) + v_z(x, z)m_x(x, y) + r_x(x, z)(m(x, y) - z) = p'(x). \quad (40)$$

Taking the expectation of (40) with respect to η gives $p'(x) = v_x(x, z) + v_z(x, z)\mathbb{E}_\eta[m_x(x, y)]$.

Substituting $p'(x)$ back to (40) gives (38).

Q.E.D.

A.7. *Single-Dippedness in the Moment-Measurable Case*

We prove Theorems 8 and 9 only in the single-dipped case; the single-peaked case is symmetric. Moreover, Theorem 8 and its proof remain valid without the assumptions that $v_{xz} > 0$ and that G has full support or a density.

PROOF OF THEOREM 8: We first prove two key lemmas, which do not rely on the assumption that $v_{zz} < 0$.

LEMMA 18: *For every (x_0, η_0) and (x_1, η_1) in Λ such that there exist $y_0 < y_1 < y'_0$ with $y_0, y'_0 \in \text{supp}(\eta_0)$ and $y_1 \in \text{supp}(\eta_1)$, we have $x_0 \geq x_1$.*

PROOF: Denote $z_i = \mathbb{E}_{\eta_i}[m(x_i, y)]$ for $i = 0, 1$. By (37), we have

$$p(x_0) + q(y_0) = v(x_0, z_0) + r(x_0, z_0)(m(x_0, y_0) - z_0), \quad (41)$$

$$p(x_0) + q(y'_0) = v(x_0, z_0) + r(x_0, z_0)(m(x_0, y'_0) - z_0), \quad (42)$$

$$p(x_1) + q(y_1) = v(x_1, z_1) + r(x_1, z_1)(m(x_1, y_1) - z_1). \quad (43)$$

By $(p, q, r) \in \mathcal{D}_M(v, m)$, we have

$$p(x_0) + q(y_1) \geq v(x_0, z_0) + r(x_0, z_0)(m(x_0, y_1) - z_0), \quad (44)$$

$$p(x_1) + q(y_0) \geq v(x_1, z_1) + r(x_1, z_1)(m(x_1, y_0) - z_1), \quad (45)$$

$$p(x_1) + q(y'_0) \geq v(x_1, z_1) + r(x_1, z_1)(m(x_1, y'_0) - z_1). \quad (46)$$

Substituting $q(y_0)$, $q(y'_0)$, and $q(y_1)$ from (41), (42), and (43) into (45), (46), and (44) gives

$$p(x_0) - p(x_1) + v(x_1, z_1) + r(x_1, z_1)(m(x_1, y_1) - z_1) \geq v(x_0, z_0) + r(x_0, z_0)(m(x_0, y_1) - z_0),$$

$$p(x_1) - p(x_0) + v(x_0, z_0) + r(x_0, z_0)(m(x_0, y_0) - z_0) \geq v(x_1, z_1) + r(x_1, z_1)(m(x_1, y_0) - z_1),$$

$$p(x_1) - p(x_0) + v(x_0, z_0) + r(x_0, z_0)(m(x_0, y'_0) - z_0) \geq v(x_1, z_1) + r(x_1, z_1)(m(x_1, y'_0) - z_1).$$

Summing up these inequalities multiplied by $(m(x_1, y'_0) - m(x_1, y_0))$, $(m(x_1, y'_0) - m(x_1, y_1))$, $(m(x_1, y_1) - m(x_1, y_0))$, and then dividing by $r(x_0, z_0) = v_z(x_0, z_0) > 0$ (see Lemma 17) gives

$$0 \leq (m(x_0, y'_0) - m(x_0, y_1))(m(x_1, y_1) - m(x_1, y_0))$$

$$\begin{aligned}
& -(m(x_0, y_1) - m(x_0, y_0))(m(x_1, y'_0) - m(x_1, y_1)) \\
&= \int_{y_1}^{y'_0} \int_{y_0}^{y_1} (m_y(x_0, y')m_y(x_1, y) - m_y(x_0, y)m_y(x_1, y')) dy dy'.
\end{aligned}$$

This inequality implies that $x_0 \geq x_1$, as otherwise the integrand would be strictly negative by strict log-supermodularity of m_y , given that $x_0 < x_1$ and $y < y'$. *Q.E.D.*

LEMMA 19: For μ -almost every $x \in X$ and every $\eta \in \Delta(Y)$ such that $(x, \eta) \in \Lambda$, we have $|\text{supp}(\eta)| \leq 2$.

PROOF: By Lemma 17, for Lebesgue almost every $x \in X$ (and thus for μ -almost every $x \in X$ since μ has a density), for every nondegenerate $\eta \in \Delta(Y)$ such that $(x, \eta) \in \Lambda$, for every $y \in \text{supp}(\eta)$, and for $z = \mathbb{E}_\eta[m(x, y)]$, there exists $r_x(x, z)$ such that (38) holds. Fix any such x and suppose by contradiction that $\text{supp}(\eta)$ contains $y_0 < y_1 < y_2$. Summing up (38) for $i = 0$ multiplied by $m(x, y_2) - m(x, y_1)$, for $i = 1$ multiplied by $m(x, y_0) - m(x, y_2)$, and for $i = 2$ multiplied by $m(x, y_1) - m(x, y_0)$ yields a contradiction:

$$\begin{aligned}
0 &= v_z(x, z)(m_x(x, y_2) - m_x(x, y_1))(m(x, y_1) - m(x, y_0)) \\
&\quad - v_z(x, z)(m_x(x, y_1) - m_x(x, y_0))(m(x, y_2) - m(x, y_1)) \\
&= v_z(x, z) \left[\int_{y_1}^{y_2} \int_{y_0}^{y_1} (m_{xy}(x, y')m_y(x, y) - m_{xy}(x, y)m_y(x, y')) dy dy' \right] \\
&> v_z(x, z) \frac{m_{xy}(x, y_1)}{m_y(x, y_1)} \left[\int_{y_1}^{y_2} \int_{y_0}^{y_1} (m_y(x, y')m_y(x, y) - m_y(x, y)m_y(x, y')) dy dy' \right] = 0,
\end{aligned}$$

where the inequality holds because m_y is strictly log-supermodular, so

$$\frac{m_{xy}(x, y')}{m_y(x, y')} > \frac{m_{xy}(x, y_1)}{m_y(x, y_1)} > \frac{m_{xy}(x, y)}{m_y(x, y)}, \quad \text{for } y' > y_1 > y. \quad \text{Q.E.D.}$$

To complete the proof of Theorem 8, we now use the assumption that $v_{zz} < 0$. Suppose for contradiction that there exist $(x, \eta_0), (x, \eta_1) \in \Lambda$ with $z_0 = \mathbb{E}_{\eta_0}[m(x, y)] < \mathbb{E}_{\eta_1}[m(x, y)] = z_1$. Then, denoting $\eta = (\eta_0 + \eta_1)/2$ and $z = (z_0 + z_1)/2$, we have

$$p(x) + \int_Y q(y) \eta(dy) \geq v(x, z) > \frac{1}{2}v(x, z_0) + \frac{1}{2}v(x, z_1)$$

$$= \frac{1}{2} \left(p(x) + \int_Y q(y) \eta_0(dy) \right) + \frac{1}{2} \left(p(x) + \int_Y q(y) \eta_1(dy) \right) = p(x) + \int_Y q(y) \eta(dy),$$

where the first inequality is by $(p, q) \in \mathcal{D}(V)$ and $V(x, \eta) = v(x, z)$, the second inequality is by $v_{zz} < 0$, the first equality is by $(x, \eta_0), (x, \eta_1) \in \Lambda$ and the second equality is by the rearrangement. Hence $\mathbb{E}_{\eta_0}[m(x, y)] = \mathbb{E}_{\eta_1}[m(x, y)]$ for all $(x, \eta_0), (x, \eta_1) \in \Lambda$.

By Lemma 19, there exists a full-measure set $X^* \subset X$ such that $|\text{supp}(\eta)| \leq 2$ for all $(x, \eta) \in \Lambda$ with $x \in X^*$. Fix $(x, \eta_0), (x, \eta_1) \in \Lambda$ with $x \in X^*$. Denoting $\eta = (\eta_0 + \eta_1)/2$ and recalling that $\mathbb{E}_{\eta_0}[m(x, y)] = \mathbb{E}_{\eta_1}[m(x, y)] = \mathbb{E}_\eta[m(x, y)] = z$, we have

$$p(x) + \int_Y q(y) \eta(dy) = \frac{1}{2} \left(p(x) + \int_Y q(y) \eta_0(dy) \right) + \frac{1}{2} \left(p(x) + \int_Y q(y) \eta_1(dy) \right) = v(x, z),$$

showing that $(x, \eta) \in \Lambda$, so $|\text{supp}(\eta)| = |\text{supp}(\eta_0) \cup \text{supp}(\eta_1)| \leq 2$, and thus $\eta_0 = \eta_1$, by $\mathbb{E}_{\eta_0}[m(x, y)] = \mathbb{E}_{\eta_1}[m(x, y)]$ and $m_y > 0$. So, there exist functions $\rho : X^* \rightarrow [0, 1]$ and $T_d, T_u : X^* \rightarrow Y$, with $T_d(x) \leq T_u(x)$ for all $x \in X^*$, such that $\eta = (1 - \rho(x))\delta_{T_d(x)} + \rho(x)\delta_{T_u(x)}$ for all $(x, \eta) \in \Lambda$ with $x \in X^*$. Moreover, $T_d(x'), T_u(x') \notin (T_d(x), T_u(x))$ for all $x < x'$ in X^* , by Lemma 18. Q.E.D.

PROOF OF THEOREM 9: By Theorem 8, each γ_n is unique and is strictly single-dipped. Since $\mathcal{P}(\mu, \nu)$ is compact, up to a subsequence, $\|\gamma_n - \gamma^*\|_{KR} \rightarrow 0$ for some $\gamma^* \in \mathcal{P}(\mu, \nu)$.

We now show that γ^* solves (P) under m . Fix arbitrary $\tilde{\gamma} \in \mathcal{P}(\mu, \nu)$ and $\delta > 0$. Define $V_n(x, \eta) = v(x, \mathbb{E}_\eta[m_n(x, y)])$ and $V(x, \eta) = v(x, \mathbb{E}_\eta[y])$. There exists $N \geq 1$ such that, for all $n \geq N$, we have $|V_n(x, \eta) - V(x, \eta)| \leq \delta$ for all $(x, \eta) \in X \times \Delta(Y)$, because v is Lipschitz and m_n converges uniformly to m , and $|\int V d\gamma^* - \int V d\gamma_n| \leq \delta$, because V is Lipschitz, $X \times \Delta(Y)$ is compact, and $\|\gamma_n - \gamma^*\|_{KR} \rightarrow 0$. Thus,

$$\int V d\gamma^* \geq \int V d\gamma_n - \delta \geq \int V_n d\gamma_n - 2\delta \geq \int V_n d\tilde{\gamma} - 2\delta \geq \int V d\tilde{\gamma} - 3\delta,$$

showing that γ^* solves (P).

Next, we show that $\mathbb{E}_\eta[y] = T(x)$ for all $(x, \eta) \in \text{supp}(\gamma^*)$. By Theorems 1 and 3, there exists $(p, q) \in \mathcal{D}_L(v)$ such that, for all $(x, \eta) \in \text{supp}(\gamma^*)$, we have $p(x) + q(\mathbb{E}_\eta[y]) = v(x, \mathbb{E}_\eta[y])$, so $\mathbb{E}_\eta[y] = T(x)$ by Lemma 8.

Next, we show that γ^* is *single-dipped*, meaning that there do not exist $(x_0, \eta_0), (x_1, \eta_1) \in \text{supp}(\gamma^*)$, with $y_0, y'_0 \in \text{supp}(\eta_0)$ and $y_1 \in \text{supp}(\eta_1)$, such that $x_0 < x_1$ and $y_0 < y_1 < y'_0$.

Suppose by contradiction that such (x_0, η_0) and (x_1, η_1) exist. Since $\|\gamma_n - \gamma^*\|_{KR} \rightarrow 0$, lower hemicontinuity of the support correspondence (Aliprantis and Border, 2006, Theorem 17.14), with $\mu(X_n^*) = 1$ (where X_n^* is the full μ -measure set from Lemma 19 for γ_n), implies that, for all large n , there exist $(x_{0,n}, \eta_{0,n}), (x_{1,n}, \eta_{1,n}) \in \text{supp}(\gamma_n)$, with $y_{0,n}, y'_{0,n} \in \text{supp}(\eta_{0,n})$ and $y_{1,n} \in \text{supp}(\eta_{1,n})$, such that $x_{0,n}, x_{1,n} \in X_n^*$, $x_{0,n} < x_{1,n}$, and $y_{0,n} < y_{1,n} < y'_{0,n}$, contradicting that γ_n is strictly single-dipped.

Suppose by contradiction that γ^* is not strictly single-dipped. Since γ^* is single-dipped, $\mathbb{E}_\eta[y] = T(x)$, for all $(x, \eta) \in \text{supp}(\gamma^*)$, and μ has a density, there must exist an uncountable set $\tilde{X} \subset X$ such that, for each $x \in \tilde{X}$, we have $|\cup_{\eta:(x,\eta) \in \text{supp}(\gamma^*)} \text{supp}(\eta)| \geq 3$. By the pigeonhole principle (and $\mathbb{E}_\eta[y] = T(x)$ for all $(x, \eta) \in \text{supp}(\gamma^*)$), there exist (rational) $y < y'$ in Y and $(x_0, \eta_0), (x_1, \eta_1) \in \text{supp}(\gamma^*)$ with $y_0, y'_0 \in \text{supp}(\eta_0)$ and $y_1 \in \text{supp}(\eta_1)$, such that $x_0 < x_1$ and $y_0 < y < y_1 < y' < y'_0$, contradicting that γ^* is single-dipped.

By the subsequence criterion for convergence, to complete the proof of the theorem, it suffices to show that there is a unique single-dipped assignment γ that induces T . Suppose by contradiction that there exist single-dipped (and thus strictly single-dipped) assignments $\gamma \neq \hat{\gamma}$ that induce T . Let ψ and $\hat{\psi}$ be the distributions of $(T(x), \eta)$ induced by γ and $\hat{\gamma}$. By part 1 of Theorem 4, T is continuous and strictly increasing, so $\psi \neq \hat{\psi}$ are both strictly single-dipped, contradicting Theorem 1.5 of Beiglböck and Juillet (2016). *Q.E.D.*

A.8. Uniform Price Equilibrium

PROOF OF THEOREM 10: *Part 4.* By definition, if $(\phi; p, q)$ is a UPE, then

$$p(x) = W(x, S(x), \delta_{S(x)}) - q(S(x)) = \max_{y \in Y} \{W(x, y, \delta_{S(x)}) - q(y)\},$$

so, by the envelope theorem (Milgrom and Segal, 2002, Theorem 3) and continuity of W_x , W_z , and S' , p has a derivative $p'(x) = W_x(x, S(x), \delta_{S(x)}) + W_z(x, S(x), \delta_{S(x)})S'(x)$ at each $x \in (\underline{x}, \bar{x})$, yielding (14); and (15) follows from $p(S^{-1}(y)) + q(y) = W(S^{-1}(y), y, \delta_y)$.

Part 1. PAS is a UPE if and only if, for all $(x, y) \in X \times Y$,

$$q(y) - q(S(x)) \geq W(x, y, \delta_{S(x)}) - W(x, S(x), \delta_{S(x)}),$$

which is equivalent to (9), since q is given by (15).

Part 2. Suppose that (10) and (11) hold. Then, for all $(x, y) \in X \times Y$,

$$\begin{aligned} \int_{S(x)}^y \int_x^{S^{-1}(\tilde{y})} \frac{d^2 W(\tilde{x}, \tilde{y}, \delta_{S(\tilde{x})})}{d\tilde{x} d\tilde{y}} d\tilde{x} d\tilde{y} &= \int_{S(x)}^y (W_y(S^{-1}(\tilde{y}), \tilde{y}, \delta_{\tilde{y}}) - W_y(x, \tilde{y}, \delta_{S(x)})) d\tilde{y} \\ &\geq \int_{S(x)}^y (W_y(x, \tilde{y}, \delta_{\tilde{y}}) - W_y(x, \tilde{y}, \delta_{S(x)})) d\tilde{y} = \int_{S(x)}^y \int_{S(x)}^{\tilde{y}} W_{yz}(x, \tilde{y}, \delta_{\tilde{z}}) d\tilde{z} d\tilde{y} \geq 0, \end{aligned}$$

so PAS is a UPE.

Conversely, suppose that PAS is a UPE for all continuous, positive densities f and g . First, fix $x < x'$ in (\underline{x}, \bar{x}) , $y \in (y, \bar{y})$, and small $\varepsilon > 0$. There exist continuous, positive densities such that $S(\tilde{x}) = y + \varepsilon(\tilde{x} - x)$ for $\tilde{x} \in [x, x' - \varepsilon - \varepsilon^2]$ and $S(\tilde{x}) = y + \varepsilon(x' - \varepsilon - x) + (\tilde{x} - x' + \varepsilon)$ for $\tilde{x} \in [x' - \varepsilon, x']$. Since PAS is a UPE, (9) holds. Taking $\varepsilon \downarrow 0$ yields $W_y(x', y, \delta_y) \geq W_y(x, y, \delta_y)$, so, by continuity of W_{xy} , (10) holds. Next, fix $x \in (\underline{x}, \bar{x})$, $y > z$ in (y, \bar{y}) (the case $y < z$ is analogous and omitted), and small $\varepsilon > 0$. There exist continuous, positive densities such that $S(\tilde{x}) = z + (y - z)(\tilde{x} - x)/\varepsilon$ for $\tilde{x} \in [x, x + \varepsilon]$. Since PAS is a UPE, (9) holds. Taking $\varepsilon \downarrow 0$ yields $\int_z^y \int_z^{\tilde{y}} W_{yz}(x, \tilde{y}, \delta_{\tilde{z}}) d\tilde{z} d\tilde{y} \geq 0$, so, by continuity of W_{yz} , (11) holds.

Part 3. Suppose that $W(x, y, \eta) = w(x, y, \mathbb{E}_\eta[m(x, \tilde{y})])$, and (12) and (13) hold. Let γ be a UPE. Fix $(x, \eta), (x', \eta') \in \text{supp}(\gamma)$, $y \in \text{supp}(\eta)$, and $y' \in \text{supp}(\eta')$. By the intermediate value theorem, there exist $z, z' \in Y$ such that $\mathbb{E}_\eta[m(x, \tilde{y})] = m(x, z)$ and $\mathbb{E}_{\eta'}[m(x', \tilde{y})] = m(x', z')$. Then

$$\begin{aligned} \int_y^{y'} \int_x^{x'} W_{xy}(\tilde{x}, \tilde{y}, \delta_{\tilde{y}}) d\tilde{x} d\tilde{y} &= W(x', y', \delta_{z'}) - W(x', y, \delta_{z'}) - W(x, y', \delta_z) + W(x, y, \delta_z) \\ &= p(x') + q(y) - W(x', y, \delta_{z'}) + p(x) + q(y') - W(x, y', \delta_z) \geq 0, \end{aligned}$$

where the first equality holds because W_y is independent of z , and the second equality and the inequality hold because γ is a UPE. So, since $W_y(\tilde{x}, \tilde{y}, \delta_{\tilde{y}})$ is strictly increasing in \tilde{x} , we have $(x' - x)(y' - y) \geq 0$. That is, for all $(x, \eta), (x', \eta') \in \text{supp}(\gamma)$ with $x < x'$, we have $\bar{S}(x) = \max\{\text{supp}(\eta)\} \leq \min\{\text{supp}(\eta')\} = \underline{S}(x')$, so, for almost all $x \in X$, $\underline{S}(x) = \bar{S}(x) = S(x)$, implying that $\gamma = \phi$. *Q.E.D.*

PROOF OF COROLLARY 2: Since w is affine in y , w_y is independent of y . Thus, (9) holds if and only if $w_y(x, y, S(x)) = w_y(x, S(x), S(x))$ is increasing in x , (10) holds

if and only if $w_y(x, y, z) = w_y(x, z, z)$ is increasing in x , and (11) holds if and only if $w_y(x, y, z) = w_y(x, z, z)$ is increasing in z .

Define the negative assortative matching (NAS) as the matching where each student x goes to school $N(x) = G^{-1}(1 - F(x))$. Note that NAS is a UPE if and only if

$$\int_{N(x)}^y \int_x^{N^{-1}(\tilde{y})} \frac{d^2 w(\tilde{x}, \tilde{y}, N(\tilde{x}))}{d\tilde{x} d\tilde{y}} d\tilde{x} d\tilde{y} \geq 0, \quad \text{for all } (x, y) \in X \times Y,$$

which holds if and only if $w_y(x, y, N(x))$ is decreasing in x , when w is affine in y .

Suppose that PAS is the unique UPE for all densities f and g . Then $w_y(x, z, z)$ is increasing in x and z , and NAS is not a UPE for any f and g . First, suppose for contradiction that there exist $x \in (\underline{x}, \bar{x})$ and $z \in (\underline{y}, \bar{y})$ such that $w_{yz}(x, z, z) > 0$. Fix small $\varepsilon > 0$, and let f and g be the uniform densities on $[x, x + \varepsilon^2]$ and $[z, z + \varepsilon]$. Then $w_y(\tilde{x}, z, N(\tilde{x}))$ is strictly decreasing in x on $[x, x + \varepsilon^2]$, so NAS is a UPE. A contradiction. Thus, w_y is independent of z . Next, suppose for contradiction that there exist $x < x'$ such that $w_y(x, y, z) = w_y(x', y, z)$. Let f be the uniform density on $[x, x']$. Then $w_y(x, y, N(x))$ is constant for all $x \in [x, x']$, so NAS is a UPE. A contradiction. Thus, w_y is strictly increasing in x . The converse follows from the proof of Theorem 10, which is valid if $w_y(x, y, y)$ is strictly increasing in x , rather than $w_{xy}(x, y, y) > 0$ for all (x, y) . *Q.E.D.*