MULTIDIMENSIONAL SORTING UNDER RANDOM SEARCH

By

Ilse Lindenlaub, Fabien Postel-Vinay

COWLES FOUNDATION PAPER NO. 1861

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Multidimensional Sorting under Random Search

Ilse Lindenlaub

Yale University, Centre for Economic Policy Research, and National Bureau of Economic Research

Fabien Postel-Vinay

University College London and Institute for Fiscal Studies

We analyze sorting in a frictional labor market when workers and jobs have multidimensional characteristics. We say that matching is positive assortative in dimension (j, k) if workers with higher endowment in skill k are matched to a job distribution with higher values of attribute j in the first-order stochastic dominance sense. Crucial for sorting is a single-crossing property of technology. Sorting is positive between worker-job attributes with strong complementarities but negative in other dimensions. Finally, sorting is based on comparative advantage: workers sort into jobs that suit their skill mix rather than their overall skill level.

I. Introduction

The assignment of heterogenous workers to heterogenous jobs matters for aggregate efficiency if there are complementarities between workers'

We thank Hector Chade and Jan Eeckhout for useful comments as well as Carlos Carrillo-Tudela and Jim Albrecht for their insightful discussions of this paper at the 2016 Cowles Foundation Macro Conference and the 2016 Konstanz Search and Matching Conference. We further thank seminar audiences at Essex, Yale, the Chicago Fed, Einaudi, Konstanz, the National Bureau of Economic Research Summer Institute 2016, Massachusetts Institute of Technology, Simon Fraser University, McGill, the European University Institute, Society for Economic Dynamics 2017, the University of Chicago, Northwestern, and Stanford for their comments. All remaining errors are ours. This paper was edited by James J. Heckman.

Electronically published November 28, 2023

Journal of Political Economy, volume 131, number 12, December 2023.

https://doi.org/10.1086/725362

and firms' productive attributes. Labor market frictions impede the efficient assignment and cause mismatch, which is costly if such complementarities are strong, less so if they are weak.

A growing body of literature focuses on understanding the sign and strength of complementarities in frictional markets—a key question for the design of policies aiming to allocate resources efficiently. With few exceptions, that literature works under the assumption that job and worker heterogeneity can both be captured by scalar indexes, that is, that heterogeneity is one-dimensional.[1] While the restriction to one-dimensional heterogeneity is a natural starting point and convenient for modeling, it is at odds with the fact that typical datasets describe both workers and jobs in terms of many different productive attributes (e.g., cognitive skills, manual skills, or psychometric test scores for workers and numerous task-specific skill requirements for jobs). If agents sort on multiple attributes, then accounting for their multidimensional heterogeneity is necessary to correctly quantify complementarities and mismatch.

In this paper, we develop a theory to help achieve that goal. We incorporate multidimensional (multi-D) worker and firm heterogeneity into a general random-search model. The sorting and mismatch patterns that arise in this multi-D environment are considerably more involved than those in one-dimensional (1-D) models. We propose a notion of assortative matching in this setting and show how model primitives—productive complementarities in particular—shape equilibrium worker-job sorting in complex yet intuitive ways. Workers face sorting trade-offs, whereby they accept mismatch along some skill dimension in order to improve sorting on another, namely, one that is characterized by stronger worker-job complementarities in technology. The resulting worker-job sorting is on comparative rather than absolute advantage.

Our environment is that of the widely used random-search/job ladder model (Burdett and Mortensen 1998; Postel-Vinay and Robin 2002), except that workers and firms (jobs) are endowed with bundles of productive attributes, $\mathbf{x} = (x_1, \cdots, x_X) \in \mathbb{R}^X$ for workers and $\mathbf{y} = (y_1, \cdots, y_Y) \in \mathbb{R}^Y$ for jobs. Employed and unemployed workers sample job offers randomly from an exogenous sampling distribution of job attributes. Utility is transferable: workers and firms are joint surplus maximizers, implying that agents' match formation decisions depend on a scalar value, namely, the match surplus. This makes our multi-D problem tractable.

[1] A recent exception is a paper by Lise and Postel-Vinay (2020), who focus on the accumulation of skills in various dimensions within a model that can otherwise be seen as a special case of ours. Other examples are Roy models with search frictions, as in Moscarini (2001). Beyond these approaches, a growing applied literature takes explicit account of these multiple dimensions of productive heterogeneity. Early examples of influential work are Heckman and Sedlacek (1985) and Heckman and Scheinkman (1987). Recent examples include Yamaguchi (2012), Sanders (2012), and Guvenen et al. (2020).

In our frictional setting, workers with a given skill bundle are matched not to a unique job type (as is often the case in the absence of frictions) but to a whole distribution of job types—there is mismatch. In order to interpret the sorting patterns that arise in equilibrium, we define notions of positive and negative assortative matching (PAM and NAM, respectively) in this environment. Our definition is based on the first-order stochastic dominance (FOSD) ordering of the marginal distributions of job attributes across workers with different skills. If workers with higher skill x_k (e.g., "cognitive skills") are matched to jobs with stochastically "better" attributes y_j (e.g., "cognitive job complexity"), then PAM occurs between (y_j, x_k). Sorting is thus defined dimension by dimension: PAM can arise in one dimension (e.g., between cognitive skills and cognitive job complexity), while NAM occurs in another (between manual skills and cognitive job complexity).

For expositional clarity and ease of interpretation, we focus on a baseline specification with two-dimensional heterogeneity, a bilinear production technology, wage setting by sequential auctions, and positive surplus of all possible matches—all assumptions that we subsequently relax. In this baseline case, we present three main results on multi-D sorting.

Our first result is about the sign of sorting: we provide conditions on the economy's primitives under which PAM or NAM arises in equilibrium. We find that matching in, say, the first dimension (y_1, x_1) is positive assortative if and only if the technology satisfies a single-crossing condition, which implies that the complementarity between worker skill x_1 and job attribute y_1 dominates the complementarity in the competing dimension (y_2, x_1).

Our second result is that sorting cannot be simultaneously positive between all skill and job dimensions. Instead, there are sorting trade-offs. We provide conditions under which PAM arises in the dimension that features relatively strong complementarities in production, while NAM materializes in the dimension characterized by weaker complementarities.

Third, and connected to these sorting trade-offs, our model predicts sorting based on comparative advantage rather than on absolute advantage: workers with uniformly higher skills do not sort into jobs with uniformly higher skill requirements. Rather, they sort into jobs with a higher requirement for the skill in which they are relatively strong, possibly at the cost of a lower requirement for the other skill.

An important insight from our analysis is that the presence of multi-D heterogeneity is crucial for sorting to arise in this setting. What matters to workers is not just to match with a productive job in any component of $\mathbf{y}$. Rather, a worker wants to match with a job that puts much weight on the skill in which he is strong. Thus, workers with different skill bundles accept and reject different types of jobs: they climb different job ladders, which leads to sorting. Multi-D heterogeneity is a new source of sorting in job ladder models, which makes them amenable to the analysis of sorting and mismatch under commonly used assumptions (such as monotone technology and exogenous search effort) that would preclude sorting in a 1-D world.[2]

We generalize our analysis, especially on the sign of sorting, to cases in which (i) not all possible matches generate positive surplus (also implying a sorting-relevant job acceptance decision by the unemployed), (ii) heterogeneity is of dimension higher than two, and (iii) the technology is nonlinear. We show that also in these more general environments the core condition for sorting is a single-crossing property of the technology. We further show that these results do not hinge on our sequential auction wage-setting protocol but hold for other common wage-setting models, such as Nash bargaining, sequential auctions with worker bargaining power, or wage posting.

A natural question is whether multi-D heterogeneity can be collapsed into a single scalar index without loss of generality. If so, then existing 1-D models would provide all the necessary tools for the analysis of sorting. We show that, in the context of our model, a single-index representation is valid only in the special case where the single-crossing condition of technology fails to hold everywhere, thus ruling out sorting in equilibrium. More generally and beyond our model, multiple dimensions cannot be mapped into a single dimension while preserving minimal regularity properties, such as continuity and monotonicity of the surplus function.

We conclude that it will be difficult to ignore multi-D heterogeneity in settings where this heterogeneity affects economic choices, such as the sorting of workers into jobs. Multi-D heterogeneity prompts rich sorting patterns, by combining features of vertical heterogeneity (surplus can be monotone in each attribute) and horizontal heterogeneity (in the sense that there is no common ranking of firms by workers of different multi-D types). As a result, modeling agents with multi-D attributes is particularly attractive, as it can account for both aspects of the data while circumventing the issues associated with single-index representations. We provide a practical framework for the analysis of multi-D sorting and mismatch in applied work.

Related literature.—While much is known about sorting under 1-D heterogeneity either without frictions (Becker 1973; Legros and Newman 2007) or with frictions (Shimer and Smith 2000; Smith 2006; Eeckhout and Kircher 2010), far less is known about sorting on multi-D types. Lindenlaub (2017) studies multi-D sorting in a frictionless assignment game.[3] But, to

[2] If match surplus depends on a scalar (1-D) job type y and is increasing in y for all worker types, workers all rank jobs in the same way. Their common strategy is to accept any job with a higher y than their current one. There is a single job ladder that all workers climb at the same speed, which rules out sorting.

[3] Our restriction on the technology to obtain sorting is most closely related (but not equivalent) to that needed in the multi-D frictionless assignment problem analyzed by

the best of our knowledge, this paper is the first to develop a theory of multi-D sorting under random search—an environment of great importance for applied work, since it carries a well-defined notion of mismatch and allows for policy analysis.

Our work not only shifts the focus to multi-D heterogeneity but also differs in another important way from that theoretical literature on sorting. All of the aforementioned papers analyze *assignment problems*, meaning that agents on either side of the market can be matched with at most one agent from the other side (matching is one to one). While this assumption is certainly appropriate in partnership models, it is less common in analyses of the labor market, where a firm usually employs many workers. By contrast, here we follow the growing applied-search literature that assumes that firms operate constant-return technologies without capacity constraints.

The no-capacity-constraint assumption greatly increases the tractability of structural search models.[4] But it also eliminates one major motive for sorting in assignment models: the scarcity of jobs. As a consequence, job ladder models with 1-D heterogeneity tend to predict a lack of sorting under two common assumptions—monotone technology and exogenous search effort—as workers share a common ranking of firms, climb the same economy-wide job ladder at the same speed, and match with the same distribution of jobs in equilibrium.

Deviating from either of these two assumptions can restore equilibrium sorting in the 1-D analogue of our model. Bagger and Lentz (2019), who build on Lentz (2010), include endogenous search intensity in a 1-D sequential auction model. Under complementarities in production, high-type workers have more to gain from matching with high-type firms, hence search more intensively, and therefore end up in better firms in equilibrium. Another way to introduce sorting is to assume that the technology is nonmonotonic in firm type. Specifically, if workers' optimal firm type differs across skills—that is, if there is horizontal heterogeneity (similar to Gautier, Teulings, and van Vuuren 2006, 2010)—then different worker types climb different job ladders, and sorting arises even under 1-D heterogeneity. A popular example of this technology is of the form $p(x, y) = A - (x - y)^2$,

Lindenlaub (2017), who also has to discipline complementarities across tasks to ensure PAM within tasks. But compared to Lindenlaub, the introduction of random search and no-capacity constraint changes the technical nature of the problem (for instance, we cannot rely on optimal-transport theory here) as well as the insights (in our framework with frictions, we can analyze sorting on the unemployment-to-employment [UE] and employment-to-employment [EE] margins, heterogenous job ladders, mismatch, and policy).

[4] For instance, it obviates the need for a complex existence proof involving a fixed-point problem in the distribution of unmatched agents that is common in one-to-one assignment models under search frictions (Shimer and Smith 2000). Moreover, the lack of capacity constraints also helps simplify the equilibrium value of a job vacancy, which is typically pinned down by a firm optimality (or "free-entry") condition.

$A > 0$, which is a special case of Tinbergen (1956). As we discuss below, however, the assumption of nonmonotonic technology poses challenges for the identification of 1-D types in the data. Moreover, modeling productive heterogeneity as purely horizontal is problematic, as certain worker and job traits have a quintessentially vertical dimension.

Our paper highlights a new source of sorting in search models without capacity constraints that stems from a natural feature of the data: multi-D heterogeneity. Multidimensionality of skills causes different workers to rank jobs differently and climb different job ladders, leading to sorting. Multi-D heterogeneity also creates the incentive to sort in Roy models with search frictions (based on Roy 1951), such as in Moscarini (2001) and Papageorgiou (2014), where workers have bundles of skills and can search for jobs in two sectors.[5] Our model has in common with this literature that sorting is on comparative rather than absolute advantage—a feature of the labor market that has found empirical support already (see, e.g., Heckman and Sedlacek 1985 and Heckman and Scheinkman 1987). Beyond this common feature, our work has a different focus: we characterize sorting patterns between workers' multiple skills and jobs' multiple skill requirements, starting from assumptions on the model primitives.

The rest of the paper is organized as follows. Section II illustrates our main theoretical insights via a simple example. Section III introduces our model. Section IV provides a definition of sorting in multiple dimensions under random search. Section V contains our main results on the sign of sorting, sorting based on comparative advantage, and the arising sorting trade-offs, all established within our baseline model with bilinear technology and two-dimensional heterogeneity. Generalizations are discussed in section VI. Section VII offers a discussion of two important assumptions: lack of capacity constraint and multi-D heterogeneity. Section VIII concludes.

II. An Illustrative Example

We begin by illustrating our main theoretical insights using a simple example, the foundations of which are explored when we introduce our full model in section III. Consider a labor market, in which workers are characterized by two-dimensional skills $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2_{++}$ (e.g., cognitive and manual skills). Workers can be either employed or unemployed, and they face search frictions in that they sample job offers randomly and sequentially. Jobs are also characterized by a two-dimensional vector of attributes $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2_{++}$ (e.g., cognitive and manual skill requirements).

[5] Taber and Vejlin (2020) also introduce search frictions into a Roy model, but they focus on a match output function that does not feature complementarities between worker skills and firm attributes, which precludes the systematic worker-job sorting that is the subject of our analysis.

A match between a worker with skills $\mathbf{x}$ and a job with attributes $\mathbf{y}$ generates surplus $\sigma(\mathbf{x}, \mathbf{y})$.

Further assuming that jobs and workers are joint surplus maximizers, a meeting between a type-$\mathbf{x}$ unemployed worker and a type-$\mathbf{y}$ job will result in a match if and only if $\sigma(\mathbf{x}, \mathbf{y}) \geq 0$. A meeting between a type-$\mathbf{x}$ worker, employed in job $\mathbf{y}$, and an alternative job $\mathbf{y}'$ will result in the worker accepting the type-$\mathbf{y}'$ job if and only if $\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})$.

For illustration, we consider the following bilinear surplus:

$$\sigma(\mathbf{x}, \mathbf{y}) = 2x_1 y_1 + x_2 y_2.$$

(We treat σ as a primitive of the model for now. Below, we spell out conditions and provide a wide class of models for which that is warranted.) First, note that, in this example, $\sigma(\mathbf{x}, \mathbf{y}) \geq 0$ for all pairs $(\mathbf{x}, \mathbf{y})$, implying that all matches out of unemployment will be accepted. Furthermore, complementarities within the first dimension of heterogeneity (the cognitive dimension, say) and within the second dimension (the manual dimension) are positive, while complementarities between dimensions are zero. This implies that complementarities are larger among cognitive attributes—that is, in (y_1, x_1)—than between manual job traits and cognitive skills—that is, between (y_2, x_1)—which ensures that the following single-crossing condition holds:

$$\frac{\partial}{\partial x_1}\left(\frac{\partial \sigma / \partial y_1}{\partial \sigma / \partial y_2}\right) > 0.$$

In words, the marginal rate of substitution between job attributes (y_1, y_2) in the surplus function varies across workers with different skill x_1.

We now examine two situations that illustrate our main insights in intuitive ways. First, consider two unemployed workers, one with skills $(x_1, x_2) = (1, 1)$ and the other with $(x_1, x_2) = (2, 1)$. These two workers have equal levels of manual skills, x_2, but the second worker has more of cognitive skill, x_1. Assume that both workers receive the same sequence of job offers over time, $(y_1, y_2) = \{(2.3, 2.3), (3, 1), (3.4, 0)\}$, which implies the following acceptance decisions based on surplus comparisons across jobs.

	Job Offer 1 $\mathbf{y} = (2.3, 2.3)$		Job Offer 2 $\mathbf{y} = (3, 1)$		Job Offer 3 $\mathbf{y} = (3.4, 0)$
Worker $\mathbf{x} = (1, 1)$	Accept	$\longrightarrow$	Accept	$\longrightarrow$	Reject
Worker $\mathbf{x} = (2, 1)$	Accept	$\longrightarrow$	Accept	$\longrightarrow$	Accept

Both workers, when employed in job 1, accept job offer 2: They are willing to accept jobs with higher cognitive content y_1 at the expense of lower manual requirements y_2, because the marginal return to higher y_1

in the surplus function is, ceteris paribus, twice as high as the marginal return to higher y_2. But the worker with higher cognitive skill x_1 is even more inclined to do so because of the complementarity between x_1 and y_1. Indeed, he accepts job offer 3—a job with a yet higher cognitive requirement but zero manual content—whereas the worker with lower cognitive skill rejects it.[6] This is why in equilibrium, ceteris paribus (i.e., for given x_2), workers with more cognitive skills x_1 will be matched to a distribution of jobs with "better" cognitive attributes y_1. There is PAM in the cognitive dimension, caused by the relatively strong complementarity in (y_1, x_1) that stems from the single-crossing condition of the surplus function. This is our first insight.

Following from the same intuition, workers with high cognitive skills x_1 will be matched to jobs with "worse" manual content y_2 relative to workers with lower x_1. While PAM arises in the dimension of strong technological complementarities (y_1, x_1), NAM arises in the dimension of relatively weak complementarities (y_2, x_1). This shows that there are sorting trade-offs guided by technology, which is our second insight.

Next, we examine a second scenario, in which we compare the job acceptance choices of two workers who can be strictly ranked in both skill dimensions. Consider a worker with skill bundle $\mathbf{x} = (2, 1.1)$, who is more skilled than a worker with $\mathbf{x} = (1, 1)$ on all accounts. Assume that they start in unemployment and receive the same sequence of job offers over time.

	Job Offer 1 $\mathbf{y} = (2.3, 2.3)$		Job Offer 2 $\mathbf{y} = (3.4, 0)$
Worker $\mathbf{x} = (1, 1)$	Accept	$\longrightarrow$	Reject
Worker $\mathbf{x} = (2, 1.1)$	Accept	$\longrightarrow$	Accept

Here, only the more skilled worker accepts job offer 2, $\mathbf{y} = (3.4, 0)$, which scores higher in the first but lower in the second dimension, compared to the job with $\mathbf{y} = (2.3, 2.3)$. Thus, workers with uniformly better skills do not sort into jobs with uniformly better attributes. Instead, they sort into jobs that best fit their skill mix. Sorting in this multi-D setting is based on comparative rather than absolute advantage—our third insight.

The intuitive reason why these sorting patterns arise in our context is that workers with different skill bundles have different "specialties" and, under the single-crossing condition, rank jobs in different ways. In other words, they climb different job ladders over time. In contrast, if there were no differences in the relative complementarities across dimensions,

[6] The last job offer, with high y_1 but very low y_2, is accepted only by the worker with higher x_1, since his surplus comparison yields $\sigma((2,1),(3.4,0)) = 13.6 > 13 = \sigma((2,1),(3,1))$, while for the worker with lower x_1, $\sigma((1,1),(3.4,0)) = 6.8 < 7 = \sigma((1,1),(3,1))$.

then different workers would share the same ranking of jobs. To illustrate this, change the technology to $\sigma(\mathbf{x},\mathbf{y}) = 2x_1y_1 + 2x_1y_2 + x_2y_1 + x_2y_2$, which implies that single crossing does not hold:

$$\frac{\partial}{\partial x_1}\left(\frac{\partial\sigma/\partial y_1}{\partial\sigma/\partial y_2}\right) = 0.$$

One can easily check that all workers would reject the last job offer $\mathbf{y} = (3.4, 0)$ in this case. This is no coincidence. This surplus function can be written as $\sigma(\mathbf{x},\mathbf{y}) = (2x_1 + x_2)(y_1 + y_2)$, and the model is isomorphic to one with 1-D heterogeneity, where workers have a single skill $x = 2x_1 + x_2$ and jobs a single attribute $y = y_1 + y_2$—the economy has a single-index representation. In such a 1-D world, in which the technology is monotone in y, workers share the same ranking of jobs and climb the same job ladder and there is no sorting (as is the case, e.g., in Postel-Vinay and Robin 2002).

In what follows, we show that the insights from our example are general.

III. The Model

A. *The Environment*

Time t is continuous. The economy is populated by infinitely lived workers and firms. There is a fixed unit mass of workers, each characterized by a time-invariant skill bundle $\mathbf{x} = (x_1, \cdots, x_X) \in \mathcal{X} = \times_{k=1}^{X}[\underline{x}_k, \bar{x}_k]$, where X denotes the number of different skill dimensions. We normalize the lowest worker skill to $\underline{x}_k = 0$ and allow $\bar{x}_k \in \mathbb{R}_+ \cup \{+\infty\}$, with $\bar{x}_k > \underline{x}_k$. Skills are distributed with cumulative distribution function (cdf) L and strictly positive density ℓ.[7] Firms can be thought of as single jobs (possibly vacant) or as collections of independent, perfectly substitutable jobs. Jobs are characterized by a vector of time-invariant productive attributes, or "skill requirements," $\mathbf{y} = (y_1, \cdots, y_Y) \in \mathcal{Y} = \times_{j=1}^{Y}[\underline{y}_j, \bar{y}_j]$, where Y denotes the number of different job attributes, $\underline{y}_j \in \mathbb{R}_+$, $\bar{y}_j \in \mathbb{R}_+ \cup \{+\infty\}$, and $\bar{y}_j > \underline{y}_j$.[8]

Workers can be matched to a job or be unemployed. They search for jobs in both cases. If matched, they lose their job at Poisson rate δ and draw alternative job offers from an exogenous job sampling cdf Γ at rate λ_1. We assume that Γ has a strictly positive, twice continuously differentiable

[7] We adopt the following notational conventions throughout the paper. We denote cdfs with uppercase letters (e.g., L), densities with the associated lowercase letters (e.g., ℓ), and survivor functions with a bar over the cdf (e.g., $\bar{L} = 1 - L$). Also, we state that a function is increasing/decreasing or positive/negative if this is the case in the weak sense. Strict properties are mentioned explicitly.

[8] The restriction that $\mathbf{x}$ and $\mathbf{y}$ are positive is not essential but simplifies the economic interpretation.

density γ over $\mathcal{Y}$.[9] Unemployed workers sample job offers from the same sampling distribution at rate λ_0. The output flow in a match between a worker with skills $\mathbf{x}$ and a job with attributes $\mathbf{y}$ is $p(\mathbf{x}, \mathbf{y})$, where $p : \mathbb{R}^X \times \mathbb{R}^Y \to \mathbb{R}$.[10] We denote the income flow of an unemployed worker with skill $\mathbf{x}$ by $p_0(\mathbf{x})$.

There is no capacity constraint on the firm side (firms are happy to hire any worker with whom they generate positive surplus), and matched jobs do not search for alternate workers. As a result, this setup is really a (partial equilibrium) model of the labor market rather than one of symmetric, one-to-one matching, in which the distributions of unmatched types change endogenously on both sides of the market as matches form. The assumption that firms have no capacity constraint is ubiquitous in the labor-search literature and provides our analysis with the necessary tractability; we discuss it further in section VII.A.

B. *Rent Sharing and Value Functions*

Workers and firms are risk neutral and have the same time discounting rate $\rho > 0$. Under those assumptions, the total present discounted value of a type-$(\mathbf{x}, \mathbf{y})$ match is independent of the way in which it is shared and depends only on match attributes $(\mathbf{x}, \mathbf{y})$. We denote this match value $P(\mathbf{x}, \mathbf{y})$. We further denote the value of unemployment $U(\mathbf{x})$ and the worker's value of being employed under his current wage contract W, where $W \geq U(\mathbf{x})$ (otherwise, the worker would quit into unemployment) and $W \leq P(\mathbf{x}, \mathbf{y})$ (otherwise, the firm would fire the worker). Assuming that the employer's value of a job vacancy is zero, the total surplus generated by a type-$(\mathbf{x}, \mathbf{y})$ match is $P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x})$.

We assume in the main text that wage contracts are set as in the sequential auction model without worker bargaining power of Postel-Vinay and Robin (2002). This is mainly to simplify exposition. We show in appendix OC (apps. OA–OD are available online) that most of our results extend to other common wage-setting rules, such as Nash bargaining (Mortensen and Pissarides 1994; Moscarini 2001), wage/contract posting (Burdett and Mortensen 1998; Moscarini and Postel-Vinay 2013), or sequential auctions with worker bargaining power (Cahuc, Postel-Vinay, and Robin 2006).

In the sequential auction model, firms offer take-it-or-leave-it wage contracts to workers. Wage contracts are long-term contracts specifying a fixed wage that can be renegotiated by mutual agreement only. In particular,

[9] We assume that density γ has strictly positive mass over its entire support, Supp $\gamma = \mathcal{Y}$. This ensures that the support of γ is a lattice under the natural (component-wise) partial order in $\mathbb{R}^n$, which is a technical requirement for some of our proofs for cases when $Y \geq 3$.

[10] We assume that the production function is defined over the entire space $\mathbb{R}^X \times \mathbb{R}^Y$, not just the set $\mathcal{X} \times \mathcal{Y}$ of observed $(\mathbf{x}, \mathbf{y})$. This is to streamline some proofs but can be relaxed.

when an employed worker receives an outside offer, the current and outside employers Bertrand-compete for the worker. Consider a type-$\mathbf{x}$ worker who is employed at a type-$\mathbf{y}$ firm and receives an outside offer from a firm of type $\mathbf{y}'$. Bertrand competition between the type-$\mathbf{y}$ and type-$\mathbf{y}'$ employers results in the worker matching with the employer at which the total match value is higher, while extracting the full surplus from the lower-surplus match. This implies that he stays in his initial job if $P(\mathbf{x}, \mathbf{y}) \geq P(\mathbf{x}, \mathbf{y}')$ and moves to the type-$\mathbf{y}'$ job otherwise. He ends up with a new wage contract worth $W' = \min\{P(\mathbf{x}, \mathbf{y}), P(\mathbf{x}, \mathbf{y}')\}$ (provided that W' exceeds the value of the worker's initial contract, W, as otherwise the worker would not have initiated the contract renegotiation in the first place).

It follows that the total value of a type-$(\mathbf{x}, \mathbf{y})$ match, $P(\mathbf{x}, \mathbf{y})$, solves the equation

$$\rho P(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}, \mathbf{y}) + \delta(U(\mathbf{x}) - P(\mathbf{x}, \mathbf{y})).$$

The annuity value of the match, $\rho P(\mathbf{x}, \mathbf{y})$, equals the output flow $p(\mathbf{x}, \mathbf{y})$ plus the expected capital loss $\delta(U(\mathbf{x}) - P(\mathbf{x}, \mathbf{y}))$ of the firm-worker pair from job destruction.[11]

Given that $U(\mathbf{x})$ is independent of firm type, the optimal mobility choices of workers hinge on the comparison of match surplus $P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x})$ across jobs. It solves $(\rho + \delta)(P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x})) = p(\mathbf{x}, \mathbf{y}) - \rho U(\mathbf{x})$. In what follows, we mostly reason in terms of the match *flow surplus*:

$$\sigma(\mathbf{x}, \mathbf{y}) \equiv p(\mathbf{x}, \mathbf{y}) - \rho U(\mathbf{x}).$$

A worker $\mathbf{x}$ employed in a job $\mathbf{y}$ accepts an offer from a job $\mathbf{y}'$ if and only if $P(\mathbf{x}, \mathbf{y}') - U(\mathbf{x}) > P(\mathbf{x}, \mathbf{y}) - U(\mathbf{x})$. This is equivalent to $\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})$, so that the optimal strategy to accept/reject a job is entirely based on the comparison of flow surpluses. Likewise, an unemployed worker $\mathbf{x}$ accepts a job of type $\mathbf{y}$ if $\sigma(\mathbf{x}, \mathbf{y}) \geq 0$. In turn, the optimal strategy of firm $\mathbf{y}$ is to accept any worker $\mathbf{x}$ if $\sigma(\mathbf{x}, \mathbf{y}) \geq 0$. It follows that the dynamic optimization problem of the agents is solved simply by flow-surplus comparisons (as hinted at in the example of sec. II).

Finally, note that, in the sequential auction case, the value of unemployment, $U(\mathbf{x})$, is given by $\rho U(\mathbf{x}) = p_0(\mathbf{x})$, implying $\sigma(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}, \mathbf{y}) - p_0(\mathbf{x})$. That is, flow surplus σ is pinned down by technology. As a result, optimal mobility decisions are entirely determined by technology.

[11] Note that, under the sequential auction model, the realization of the "other" risk faced by the firm-worker pair, namely, the receipt of an outside job offer by the worker, generates zero capital gain for the match. Either the worker rejects the offer and stays, in which case the continuation value of the match is still $P(\mathbf{x}, \mathbf{y})$, or the worker accepts the offer and leaves, in which case he receives $P(\mathbf{x}, \mathbf{y})$ while his initial employer is left with a vacant job worth 0, so that the initial firm-worker pair's continuation value is again $P(\mathbf{x}, \mathbf{y})$.

C. *Steady-State Distribution of Matches*

In our analysis of sorting below, the key object is the steady-state equilibrium density of type-$(\mathbf{x}, \mathbf{y})$ matches, denoted by $h(\mathbf{x}, \mathbf{y})$, which indicates who matches with whom. It is determined by the following flow-balance equation, derived from optimal mobility decisions:[12]

$$\begin{aligned}&(\delta + \lambda_1 \mathbb{E}_\Gamma[\mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})\}])h(\mathbf{x}, \mathbf{y})\\ &= \lambda_0 \gamma(\mathbf{y}) \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) \geq 0\} u(\mathbf{x}) + \lambda_1 \gamma(\mathbf{y}) \int \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) > \sigma(\mathbf{x}, \mathbf{y}')\} h(\mathbf{x}, \mathbf{y}')\, d\mathbf{y}',\end{aligned} \tag{1}$$

where $u(\mathbf{x})$ is the measure of type-$\mathbf{x}$ unemployed workers in the economy. The left-hand side of equation (1) is the outflow from the stock of type-$(\mathbf{x}, \mathbf{y})$ matches, comprising matches that are destroyed at rate δ and matches that are dissolved because the worker receives a dominating outside offer. The flow probability of this latter event is $\lambda_1 \mathbb{E}_\Gamma[\mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})\}]$, the product of the arrival rate of offers λ_1 and the probability of drawing a job type $\mathbf{y}'$ that yields a higher flow surplus for the worker than the current type-$\mathbf{y}$ job. The right-hand side of equation (1) is the inflow into the stock of type-$(\mathbf{x}, \mathbf{y})$ matches and is composed of two groups: unemployed type-$\mathbf{x}$ workers who draw a type-$\mathbf{y}$ job with flow probability $\lambda_0 \gamma(\mathbf{y})$ and accept it (which they do if the flow surplus is positive), and type-$\mathbf{x}$ workers employed in any type-$\mathbf{y}'$ job who draw a type-$\mathbf{y}$ offer with flow probability $\lambda_1 \gamma(\mathbf{y})$ and accept it (which they do if the flow surplus with that job exceeds the one with their initial type-$\mathbf{y}'$ job). The measure of type-$\mathbf{x}$ unemployed workers solves the following flow-balance equation with similar interpretation:

$$\lambda_0 \mathbb{E}_\Gamma[\mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') \geq 0)\}] u(\mathbf{x}) = \delta \int h(\mathbf{x}, \mathbf{y}')\, d\mathbf{y}'. \tag{2}$$

Finally, note that, consistent with equations (1) and (2), the total measure of workers with skill bundle $\mathbf{x}$ in the economy solves $\ell(\mathbf{x}) = u(\mathbf{x}) + \int h(\mathbf{x}, \mathbf{y}')\, d\mathbf{y}'$.

Note that the job acceptance rule of an employed worker in a type-$(\mathbf{x}, \mathbf{y})$ match hinges on the comparison of two scalars, $\sigma(\mathbf{x}, \mathbf{y}')$ and $\sigma(\mathbf{x}, \mathbf{y})$, despite the underlying multi-D heterogeneity of workers and firms. It is therefore convenient to introduce the conditional sampling cdf $F_{\sigma|\mathbf{x}}$ of flow surplus σ, given $\mathbf{x}$ (with density $f_{\sigma|\mathbf{x}}$). With this notation, the job acceptance probability for an employed worker $\mathbf{x}$ is $\mathbb{E}_\Gamma[\mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})\}] = \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}))$, and that for an unemployed worker is $\mathbb{E}_\Gamma[\mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') \geq 0)\}] = \bar{F}_{\sigma|\mathbf{x}}(0)$.

[12] Throughout the paper, we use the notation $\mathbb{E}_\Gamma$ to distinguish expectations taken with respect to the sampling distribution Γ from expectations with respect to the equilibrium distribution of matches, which we simply denote by $\mathbb{E}$. Also, we use primes ($'$) in expectations to denote random variables with respect to which expectations are taken.

In appendix A1, we solve equation (1) in closed form for $h(\mathbf{x}, \mathbf{y})$. On the basis of $h(\mathbf{x}, \mathbf{y})$, we can derive the equilibrium conditional density of job types $\mathbf{y}$, given employed worker types $\mathbf{x}$:

$$h(\mathbf{y}|\mathbf{x}) = \frac{\delta \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) \geq 0\}}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \frac{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))\gamma(\mathbf{y})}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y})))^2} = \frac{g_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}))}{f_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}))} \times \gamma(\mathbf{y}), \quad (3)$$

where for any $s \in \mathbb{R}$, $g_{\sigma|\mathbf{x}}(s)$ is the equilibrium density of flow surplus among employed workers of type $\mathbf{x}$, corresponding to the cdf $G_{\sigma|\mathbf{x}}(s)$.[13] We use this conditional density $h(\mathbf{y}|\mathbf{x})$ to define our measure of sorting below.

IV. Measuring and Decomposing Equilibrium Sorting

A. *Measuring Sorting*

We first specify a measure of sorting in this multi-D environment under frictions. A criterion that has been proposed for multi-D PAM in a frictionless context is that the Jacobian matrix of the equilibrium matching function be a *P*-matrix, meaning that all its principal minors are positive (Lindenlaub 2017). This criterion captures the way in which a worker's job type $\mathbf{y}$ improves or deteriorates as one varies the worker's skills $\mathbf{x}$ when matching is pure, that is, when any two workers with the same skill bundle are matched to the exact same type of job. By contrast, in our frictional environment with random search, the equilibrium assignment is generally not pure—there is mismatch. A natural extension of this measure of sorting to our environment is to consider changes in the quantiles of the conditional matching distribution of job types $\mathbf{y}$ as one varies worker type $\mathbf{x}$.[14]

Formally, let $H_j(y|\mathbf{x})$ denote the marginal cdf of y_j (the jth component of job attribute vector $\mathbf{y}$) conditional on workers having skill bundle $\mathbf{x}$. Using equation (3), we can express this as

[13] The equilibrium cdf of flow surplus among employed workers of type $\mathbf{x}$ is

$$G_{\sigma|\mathbf{x}}(s) := 1 - \frac{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0)}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \frac{\bar{F}_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)} = \frac{\delta(F_{\sigma|\mathbf{x}}(s) - F_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s))},$$

an expression that is familiar from 1-D job ladder models. In particular, because of on-the-job search, $G_{\sigma|\mathbf{x}}$ first-order stochastically dominates the sampling cdf of flow surplus $F_{\sigma|\mathbf{x}}$.

[14] We choose to analyze the equilibrium matching distribution of $\mathbf{y}$ given $\mathbf{x}$ and not that of $\mathbf{x}$ given $\mathbf{y}$ for the following reason. While workers sample job types from an exogenous sampling distribution γ, jobs "sample" workers from an endogenous distribution (the distribution of workers across employment statuses and job types), which in itself is a complex equilibrium object. The acceptance decisions of firms would affect and be affected by the distribution of $\mathbf{x}$ across employment statuses and job types. Analyzing the matching distributions of $\mathbf{x}$ given $\mathbf{y}$ would therefore require us to deal with a complicated fixed-point problem, which proved intractable.

$$H_j(y|\mathbf{x}) = \int \mathbf{1}\{y_j' \le y\} h(\mathbf{y}'|\mathbf{x})\, d\mathbf{y}'$$
$$= \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \int \frac{\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') \ge 0\} \times \mathbf{1}\{y_j' \le y\}}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}')))^2} \gamma(\mathbf{y}')\, d\mathbf{y}'. \tag{4}$$

We are interested in signing, for each job attribute j, the elements of the gradient $\nabla H_j(y|\mathbf{x}) = (\partial H_j(y|\mathbf{x})/\partial x_1, \cdots, \partial H_j(y|\mathbf{x})/\partial x_X)^\top$, that is, the Jacobian matrix of $H(\mathbf{y}|\mathbf{x})$. A situation of particular interest is when a component of this matrix, $\partial H_j(y|\mathbf{x})/\partial x_k$, has a constant sign over the support of y_j. If that sign is negative (positive), then $H_j(\cdot|\mathbf{x})$ is increasing (decreasing) in x_k in the sense of FOSD: PAM (NAM) then occurs in dimension (y_j, x_k), as a worker with higher type-k skill (for a given level of skills other than x_k) is matched to jobs with stochastically greater type-j skill requirement, compared to a worker with lower skill x_k.[15] For instance, if $k = j = 1$ indicates the cognitive dimension, then positive (negative) sorting in (y_1, x_1) captures the intuitive notion that workers with more cognitive skill x_1 are matched to jobs with higher (lower) cognitive-skill requirement y_1. Formally, we use the following definition of sorting, which describes the association of skills and job attributes dimension by dimension:[16]

Definition 1 (PAM and NAM). Matching is positive (negative) assortative in dimension (y_j, x_k) if and only if $\partial H_j(y|\mathbf{x})/\partial x_k$ is negative (positive) for all $y \in [\underline{y}_j, \bar{y}_j]$, strictly so on a nonzero measure set of y, and for all $\mathbf{x} \in \mathcal{X}$.

To avoid duplication, we focus on positive sorting throughout most of the paper.

B. A Decomposition Result

We begin our analysis by showing how equilibrium sorting can be decomposed into sorting on the unemployment-to-employment (UE) margin and on the employment-to-employment (EE) margin. As we show in appendix A2, a typical element of the gradient of $H_j(y|\mathbf{x})$, which we use to characterize sorting patterns (definition 1), has two parts, indicating that a marginal increase in the worker's skill x_k affects his equilibrium distribution of job types y_j in two ways:[17]

[15] FOSD has been used to define sorting under frictions and 1-D types (e.g., Chade 2006 and Lentz 2010).

[16] Note that definition 1 is "global," in the sense that it imposes a sign restriction on $\partial H_j(y|\mathbf{x})/\partial x_k$ for all skill bundles $\mathbf{x} \in \mathcal{X}$. Alternatively, we could have opted for a "local" definition by imposing only the weaker condition that $\partial H_j(y|\mathbf{x})/\partial x_k$ be positive or negative at a given skill bundle $\mathbf{x}$. In what follows, we use the global version of this definition, as sorting is commonly envisaged as a global property in the literature.

[17] There are two technical notes. First, $\mathbb{P}_\Gamma\{A\}$ is used to denote the probability of A occurring following a random draw of a job type $\mathbf{y}$ from the sampling distribution γ; second,

$$\frac{\partial H_j(y|\mathbf{x})}{\partial x_k} = \underbrace{C_{\mathrm{UE}}}_{(1)\text{ UE margin}} + \underbrace{C_{\mathrm{EE}}}_{(2)\text{ EE margin}}, \tag{DEC}$$

where components C_{UE} and C_{EE} are given by

$$\begin{aligned} C_{\mathrm{UE}} := g_{\sigma|\mathbf{x}}(0)\Bigg(& \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\,\middle|\,\sigma(\mathbf{x},\mathbf{y}')=0\right] \\ & \times \int_0^{+\infty} g_{\sigma|\mathbf{x}}(s)\left[\mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}')=0\} - \mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}')=s\}\right] ds \\ & + \mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}')=0\}\Bigg\{\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\,\middle|\,\sigma(\mathbf{x},\mathbf{y}')=0, y_j' \le y\right] \\ & - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\,\middle|\,\sigma(\mathbf{x},\mathbf{y}')=0\right]\Bigg\}\Bigg), \end{aligned}$$

$$\begin{aligned} C_{\mathrm{EE}} := \lambda_1 \int_0^{+\infty} & \frac{2f_{\sigma|\mathbf{x}}(s)g_{\sigma|\mathbf{x}}(s)}{\delta+\lambda_1\bar{F}_{\sigma|\mathbf{x}}(s)} \times \mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}')=s\} \\ & \times \left(\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\,\middle|\,\sigma(\mathbf{x},\mathbf{y}')=s, y_j' \le y\right] - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\,\middle|\,\sigma(\mathbf{x},\mathbf{y}')=s\right]\right) ds. \end{aligned}$$

The first term of decomposition (DEC), C_{UE}, reflects selection on the UE margin. A marginal increase in skill x_k can affect the set of job types $\mathbf{y}$ such that $\sigma(\mathbf{x},\mathbf{y}) \ge 0$, for example, by rendering profitable some matches between unemployed workers and jobs that were unprofitable before. Note that, if there are no marginally profitable matches ($g_{\sigma|\mathbf{x}}(0) = 0$) because all potential matches involving type-$\mathbf{x}$ workers have strictly positive surplus to begin with, $\sigma(\mathbf{x},\mathbf{y}) > 0$ for all $\mathbf{y}$, then $C_{\mathrm{UE}} = 0$ and sorting on the UE margin is shut down.

The second term of (DEC), C_{EE}, reflects selection on the EE margin. A marginal increase in x_k affects the job acceptance probability of employed workers of type $\mathbf{x}$ by having an impact on the surplus comparison between any current and incoming job: for any two job types $(\mathbf{y},\mathbf{y}')$, the difference $\sigma(\mathbf{x},\mathbf{y}') - \sigma(\mathbf{x},\mathbf{y})$ generally varies with x_k. This, in turn, affects the reallocation of workers through on-the-job search.

If we shut down sorting on the UE margin—for example, by assuming that $\sigma(\mathbf{x},\mathbf{y}) > 0$ for all $(\mathbf{x},\mathbf{y})$—then $\partial H_j(y|\mathbf{x})/\partial x_k = C_{\mathrm{EE}}$. If we shut down sorting on the EE margin—for example, by setting $\lambda_1 = 0$—then $\partial H_j(y|\mathbf{x})/\partial x_k = C_{\mathrm{UE}}$. In what follows, we say that PAM occurs on the EE margin whenever C_{EE} is negative and that PAM occurs on the UE margin whenever C_{UE} is negative. We thus analyze sorting on the EE margin "as if" sorting

it may be that the joint event $\{\sigma(\mathbf{x},\mathbf{y}') = s, y_j' \le y\}$, on which some of the expectations in (DEC) are conditioned, has zero probability in γ. As explained in the appendix, we set expectations conditional on zero-probability events to zero by convention.

on the UE margin were shut down, and vice versa. This is mainly for exposition. In general, both margins of sorting are present, and the contributions to overall sorting of both terms C_{UE} and C_{EE} must be signed in order to determine the sign of $\partial H_j(y|\mathbf{x})/\partial x_k$, as indicated by (DEC). Yet it is useful, from an analytical and applied standpoint, to consider sorting and mismatch on those two margins separately, and we provide the tools to do so.

To offer some intuition on the drivers of sorting in (DEC), consider the EE margin first. A sufficient condition for PAM (i.e., $C_{\mathrm{EE}} \leq 0$) is that $\mathbb{E}_\Gamma[\partial\sigma(\mathbf{x},\mathbf{y}')/\partial x_k|\sigma(\mathbf{x},\mathbf{y}') = s, y_j' \leq y] - \mathbb{E}_\Gamma[\partial\sigma(\mathbf{x},\mathbf{y}')/\partial x_k|\sigma(\mathbf{x},\mathbf{y}') = s]$ be negative for all (s, y). This will be true if the first conditional expectation is increasing in y, hinting at a form of complementarity between x_k and y_j in surplus function σ. The notion that complementarities between firm and worker attributes are key drivers of sorting patterns is familiar from the analysis of sorting in a variety of settings. The fact that a similar difference in conditional expectations appears in the term C_{UE} (last line) indicates that complementarities in the surplus function drive sorting on the UE margin, too.[18]

Beyond this basic intuition about the driving forces of sorting on the UE and EE margins, those two margins involve complex interactions between the technology σ and the sampling distribution of job types $\boldsymbol{\gamma}$. This implies that terms C_{UE} and C_{EE} in (DEC) cannot easily be signed without further assumptions on the primitives. In order to make progress toward a characterization of the sign of sorting, we focus in the next section on a class of technologies for which we can derive clean and (with one exception) distribution-free conditions for positive sorting, $\partial H_j(y|\mathbf{x})/\partial x_k \leq 0$. We investigate generalizations in the following section and the online appendix.

V. Equilibrium Sorting in the Baseline Model

A. *The Bilinear Technology*

In the main body of this paper, we focus on the case of a bilinear technology. This assumption simplifies decomposition (DEC) considerably and produces easily interpretable results.

Assumption 1.

a) The production function $p(\mathbf{x}, \mathbf{y})$ is bilinear in $(\mathbf{x}, \mathbf{y})$:

$$p(\mathbf{x},\mathbf{y}) = (\mathbf{x} + \mathbf{a})^\top \mathbf{Q}\mathbf{y} = \sum_{k=1}^{X}\sum_{j=1}^{Y} q_{kj}(x_k + a_k)y_j,$$

[18] We note that in the 1-D ($Y = 1$) case, if in addition σ is monotone in the single y, the difference $\mathbb{E}_\Gamma[\partial\sigma(\mathbf{x},y')/\partial x_k|\sigma(\mathbf{x},y') = s, y' \leq y] - \mathbb{E}_\Gamma[\partial\sigma(\mathbf{x},y')/\partial x_k|\sigma(\mathbf{x},y') = s]$ is always zero. This is because σ is invertible with respect to y and thus the conditioning event, $\sigma(\mathbf{x}, y') = s$, pins down a unique y'. This resembles the known result that there is no sorting on the EE margin in 1-D job ladder models of this type.

where $\mathbf{Q} = (q_{kj})_{\substack{1 \le k \le X \\ 1 \le j \le Y}}$ is an $X \times Y$ matrix and $\mathbf{a} = (a_1, \cdots, a_X)^\top \in \mathbb{R}^X_{++}$ is a fixed vector;

b) the nonemployment income function $p_0(\mathbf{x})$ is linear in $\mathbf{x}$:

$$p_0(\mathbf{x}) = (\mathbf{x} + \mathbf{a})^\top \mathbf{Q}\mathbf{b} = \sum_{k=1}^{X}\sum_{j=1}^{Y} q_{kj}(x_k + a_k)b_j,$$

where $\mathbf{b} = (b_1, \cdots, b_Y)^\top \in \mathbb{R}^Y$ is a fixed vector; and

c) there exists $j \in \{1, \cdots, Y\}$ such that $p_j(\mathbf{x}) := \partial p(\mathbf{x}, \mathbf{y})/\partial y_j = \sum_{k=1}^{X} q_{kj}(x_k + a_k) > 0$ for all $\mathbf{x} \in \mathcal{X}$; to fix the notation, we assume without loss of generality that $p_Y(\mathbf{x}) > 0$.

Assumptions 1a and 1b restrict the production technology in such a way that the flow-surplus function $\sigma(\mathbf{x}, \mathbf{y})$ is bilinear in $(\mathbf{x}, \mathbf{y})$. Indeed, they imply that

$$\sigma(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}, \mathbf{y}) - p_0(\mathbf{x}) = (\mathbf{x} + \mathbf{a})^\top \mathbf{Q}(\mathbf{y} - \mathbf{b}).$$

The technology matrix $\mathbf{Q}$ captures the complementarity structure between all job and worker characteristics and will be crucial to our analysis of sorting. We interpret vector $\mathbf{b}$ as the production technology of the unemployed. In turn, we interpret vector $\mathbf{a}$, which is a technological parameter, as the baseline productivity of workers, noting that $\mathbf{a}^\top\mathbf{Q}\mathbf{y}$ is the output of a type-$\mathbf{y}$ job filled with the least skilled worker, $\mathbf{x} = \mathbf{0}_{1\times X}$. We assume that $\mathbf{a} > 0$ (assumption 1a). This ensures that the worker's total input into production, $\mathbf{x} + \mathbf{a}$, is strictly positive in all dimensions. While not strictly necessary for our analysis, this restriction ensures that our sorting results do not change with the sign of $\mathbf{x} + \mathbf{a}$. Finally, assumption 1c ensures that, for any level of worker skills, there is at least one job attribute, here denoted y_Y, that affects output positively. Note that we do not impose monotonicity of the production function in all job attributes. Nor do we restrict the monotonicity of the production or flow-surplus function in worker skills $\mathbf{x}$.

Furthermore, in our baseline model, we focus on two-dimensional heterogeneity:

Assumption 2. Each job has $Y = 2$ attributes, $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^2_+$, and each worker has $X = 2$ skills, $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^2_+$.

The results in the next subsections are established under assumptions 1 and 2 (our baseline model). We chose to focus on this 2×2 bilinear specification in the text to facilitate the exposition and provide clear intuitions. In section VI and the online appendix, we provide generalizations of our results to other production functions and to higher dimensions of heterogeneity.

B. The Sign of Sorting in a Given Dimension

We now investigate the sign of sorting along both the EE and UE margins, based on decomposition result (DEC). We first focus on sorting "dimension by dimension," or "ceteris paribus," and analyze the sign of sorting between a given skill x_k and job attribute y_j, keeping all other skills fixed.

1. The EE Margin

We begin with the following result on the conditions for positive sorting on the EE margin.

THEOREM 1. Under assumptions 1 and 2, PAM occurs in dimension (y_1, x_k) along the EE margin if and only if, for all $\mathbf{y} \in \mathcal{Y}$,

$$\frac{\partial}{\partial x_k}\left(\frac{\partial p(\mathbf{x},\mathbf{y})/\partial y_1}{\partial p(\mathbf{x},\mathbf{y})/\partial y_2}\right) > 0,\ \text{or, equivalently,}\ \frac{\partial}{\partial x_k}\left(\frac{p_1(\mathbf{x})}{p_2(\mathbf{x})}\right) > 0. \qquad \text{(SC-2D)}$$

Condition (SC-2D) is a single-crossing property of the production function (the Spence-Mirrlees condition, in this differential form). Technically, we prove that (SC-2D) is sufficient for the conditional expectation of the marginal surplus $\mathbb{E}_\Gamma[\partial\sigma(\mathbf{x},\mathbf{y}')/\partial x_k|\sigma(\mathbf{x},\mathbf{y}') = s, y_1' = y]$ in term C_{EE} of (DEC)—which decomposes the Jacobian element $\partial H_1(y|\mathbf{x})/\partial x_k$ that we aim to sign—to be increasing in y (see app. B2). More specifically, the worker-job complementarities in technology p, assumed in (SC-2D), give rise to worker-job complementarities in surplus function σ. This renders C_{EE} negative, implying PAM on the EE margin. Intuitively, single-crossing property (SC-2D) captures stronger complementarities in dimension (y_1, x_k), relative to (y_2, x_k). This induces workers with higher levels of skill x_k to match with jobs that have higher levels of y_1, as we also illustrated with our example in section II.

To further demonstrate the intuition behind our single-crossing condition and its implications, we consider two workers with skill bundles $\mathbf{x}' = (x_1', x_2')$ and $\mathbf{x}'' = (x_1'', x_2'')$, such that $x_1'' > x_1'$ and $x_2'' = x_2'$. To fix ideas, we refer to dimension 1 as "cognitive" and dimension 2 as "manual," so the second worker has more cognitive skills but the two have equal amounts of manual skills. In figure 1, we plot for each worker the locus of jobs that render the same output as the job with attribute bundle A. Single-crossing condition (SC-2D) implies that the marginal rate of substitution between (y_1, y_2) is increasing in worker skill x_1. Thus, the isoquant of the more skilled worker is steeper, and the two isoquants cross only once (at point A). Consider point A a benchmark with no sorting (both workers are matched to the same type of job). Then, under condition (SC-2D), if the worker with fewer cognitive skills weakly prefers job B over job A, where B has a lower manual skill requirement y_2 but higher cognitive content y_1, then the worker with more cognitive skills strictly prefers job B. This is depicted

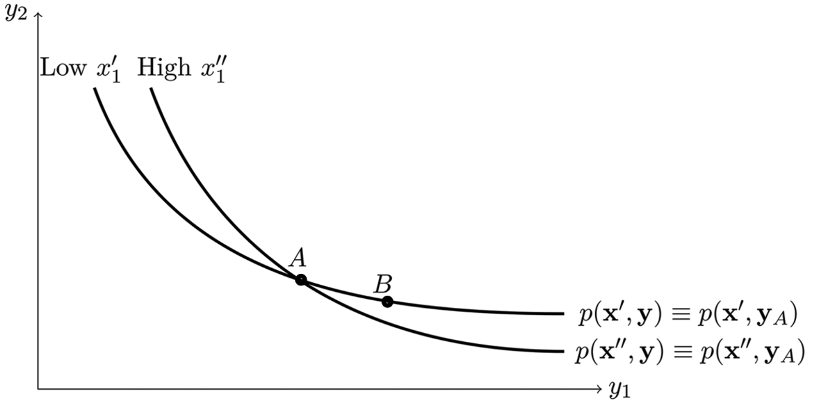

FIG. 1.—Single-crossing property.

in the figure: for the worker with more cognitive skills, the job with attribute bundle *B* lies on a higher isoquant than job *A*. As a result, workers with more cognitive skills tend to be matched to jobs with greater cognitive-skill requirements (in the FOSD sense).

Single-crossing properties have been shown to guarantee positive sorting in a variety of 1-D matching problems.[19] The analysis of our multi-D matching model with search frictions further highlights the importance of single crossing as a driving force toward positive sorting.

2. The UE Margin

The next result establishes conditions for positive sorting along the UE margin.

THEOREM 2. Under assumptions 1 and 2, if, for all $\mathbf{x} \in \mathcal{X}$,

1. single-crossing condition (SC-2D) is satisfied;
2. $p(\mathbf{x}, \cdot)$ is increasing in all elements of $\mathbf{y}$;
3. along all level curves of $p(\mathbf{x}, \cdot)$ (i.e., at all $\mathbf{y}$ such that $p(\mathbf{x}, \mathbf{y}) = C$ for some fixed $C \geq 0$),

$$p_2(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_1 \partial y_2} - p_1(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_2^2} \geq 0; \tag{UE-2D}$$

[19] In an important paper, Legros and Newman (2007) show that a single-crossing property is sufficient to guarantee PAM in frictionless 1-D problems with imperfectly transferable utility (ITU). Chade, Eeckhout, and Smith (2017) then demonstrate that several 1-D matching problems with transferable utility in both environments with and environments without frictions can be recast as ITU, frictionless matching problems. After finding the associated ITU problem, the Legros-Newman condition can be applied and guarantees PAM.

4. at the lower endpoint of the support of γ (denoted $\underline{\mathbf{y}} = (\underline{y}_1, \underline{y}_2)$), $\underline{y}_2 \geq b_2$ and $\underline{y}_1 < b_1$,

then PAM occurs in dimension (y_1, x_k) along the UE margin. Moreover, (SC-2D) is also necessary for PAM to occur generically under any sampling distribution γ.

Theorem 2 also highlights the importance of single crossing for positive sorting on the UE margin. Technically, single-crossing condition (SC-2D) ensures that the conditional expectation $\mathbb{E}_\Gamma[\partial\sigma(\mathbf{x}, \mathbf{y}')/\partial x_k | \sigma(\mathbf{x}, \mathbf{y}') = 0, y'_1 = y]$ increases in y, making the second line of C_{UE} in (DEC) negative. In turn, condition (UE-2D) ensures that the first line of C_{UE} is negative, while the boundary restriction in condition 4 of the theorem ensures that $g_{\sigma|\mathbf{x}}(0) > 0$, that is, that there are some marginally profitable matches with zero surplus, $\sigma(\mathbf{x}, \mathbf{y}) = 0$, to begin with.[20] Together, these conditions guarantee that C_{UE} is negative, implying PAM on the UE margin. Details of the proof are in appendix B3, but it is evident from this brief sketch that, contrary to the EE margin, single crossing alone is not sufficient for PAM on the UE margin (it is not sufficient for signing the first line of C_{UE}). Additional restrictions on the sampling distribution γ are needed.

To provide some economic intuition, we go back to our interpretation of $k = 1$ as the cognitive dimension and $k = 2$ as the manual one. As above, we compare two workers $\mathbf{x}'$ and $\mathbf{x}''$, who are characterized by cognitive skills $x''_1 > x'_1$ and the same amount of manual skills $x''_2 = x'_2$. They have different boundaries of marginally profitable (zero-surplus) jobs, given by $y_2 - b_2 + (p_1(\mathbf{x})/p_2(\mathbf{x}))(y_1 - b_1) = 0$, with $y_2 > b_2$ and $y_1 < b_1$ (by condition 4). Workers with different skills break even with different types of jobs. The reason is that under (SC-2D), whereby $p_1(\mathbf{x})/p_2(\mathbf{x})$ is increasing in x_1, productive complementarities are stronger between x_1 and y_1 than between x_1 and y_2. Therefore, for a given manual skill requirement y_2, workers with higher cognitive skills x_1 need jobs with higher cognitive attributes y_1 to generate nonnegative surplus. Similar to the EE margin, single crossing is a force toward PAM.

However, complementarities in production alone are not enough to ensure PAM on the UE margin. The main reason is that there may be distributional obstructions to positive sorting. If job attributes y_1 and y_2 are strongly negatively correlated, then workers with higher cognitive skill x_1 may prefer matching with jobs of lower cognitive content y_1, if cognitive-intensive jobs have too little manual content y_2 for the surplus to be

[20] Recall that sorting on the UE margin occurs only when an increase in skill x_k affects the boundary of the set of profitable matches, which can be the case only if there are some matches on that boundary, i.e., if $g_{\sigma|\mathbf{x}}(0) > 0$. Because, under condition 4 of theorem 2, $\underline{y}_2 \geq b_2$, it has to be the case that $y_1 < b_1$ for surplus to equal zero.

positive. To rein in this force toward NAM, we assume a sufficient degree of positive association between y_1 and y_2 in γ. This is the content of the distributional restriction in condition 3 of theorem 2, which supports PAM.[21]

Figure B1 supplements our explanations with an in-depth analysis of how the boundary of marginally profitable matches shifts with x_1 under the assumptions of theorem 2, and helps visualize the role of both single crossing and distributional restrictions in that theorem.

3. Taking Stock

Our first main insight is that in this multi-D setting, sufficiently strong complementarities in a certain dimension (e.g., between (y_1, x_1)), captured by a single-crossing property of technology, are the key driver of positive sorting. They induce workers with higher skill x_1 to match with jobs that have stochastically better attribute y_1—on both the UE and EE margins.

That sorting arises in a model without firm capacity constraints may seem surprising at first sight. It is well known that in a comparable model with 1-D heterogeneity, there is no sorting along the EE margin. The strategy of firms is to accept any worker who generates positive (flow) surplus, while the strategy of workers is to accept all jobs that yield a higher surplus than their current one. The assumption that the flow surplus is increasing in y (the 1-D version of assumption 1c) then implies that all workers rank firms in the same way. They all climb a single economy-wide job ladder, which rules out sorting. Moreover, regarding the UE margin, there would again be no sorting in the 1-D model, since any match in which the job productivity is too low ($y < b$) would not form, independent of the worker's skill.

With multi-D heterogeneity, however, workers are not just looking for jobs that are more productive in any dimension. Instead, a given worker aims for jobs that require much of the skill in which he is particularly strong. Thus, workers with different skill bundles rank jobs differently and therefore accept and reject different types of jobs. In short, they climb different job ladders. The heterogeneity of job ladders induced by the single-crossing condition in the presence of multiple skill dimensions is the reason why sorting arises in our setting.

C. Interrelation of Sorting Patterns across Dimensions

In the previous section, we approached the question of sorting dimension by dimension, focusing on the equilibrium relation between a given skill

[21] Sufficient conditions on the sampling distribution for (UE-2D) are provided in "Remarks" in app. B3.

x_k and a given job attribute y_j. In this section, we investigate the interdependence of sorting patterns across dimensions of worker skills and job attributes. Formally, we show that the signs of the entries of the conditional matching distribution's Jacobian matrix,

$$\frac{\partial H(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}^\top} = \begin{pmatrix} \frac{\partial H_1(y|\mathbf{x})}{\partial x_1} & \frac{\partial H_1(y|\mathbf{x})}{\partial x_2} \\ \frac{\partial H_2(y|\mathbf{x})}{\partial x_1} & \frac{\partial H_2(y|\mathbf{x})}{\partial x_2} \end{pmatrix}, \tag{5}$$

are systematically interrelated across rows and columns. This will complete our analysis of equilibrium sorting and generate several empirically testable predictions.

1. Sorting on Comparative Advantage and Sorting Trade-Offs

We present two results, one about sorting on absolute versus comparative advantage and the other about sorting trade-offs in a multi-D world, as illustrated with our example in section II. Both can be derived from theorem 3, which links sorting patterns between a given job attribute y_j and different skills and is based on the simultaneous expansion of all skills:

THEOREM 3. Under assumptions 1 and 2, $\forall j \in \{1, 2\}$:

$$(\mathbf{x} + \mathbf{a})^\top \nabla H_j(y|\mathbf{x}) = 0;$$

that is, the function $(\mathbf{x} + \mathbf{a}) \mapsto H_j(y|\mathbf{x})$ is homogeneous of degree 0 in $(\mathbf{x} + \mathbf{a})$ for all j.

We use this result to shed light on the nature of sorting in our setting, that is, whether workers sort on comparative or absolute advantage. To this end, consider a generic skill expansion in $\mathbf{x}$, namely, an increase of a worker's skills from (x_1, x_2) to $(x_1 + \Delta x_1, x_2 + \Delta x_2)$, where we impose assumptions on primitives that guarantee PAM in dimension (y_1, x_1), $\partial H_1(y|\mathbf{x})/\partial x_1 \leq 0$.[22] Using theorem 3, we compute the change in jobs

[22] Assumptions that ensure PAM are that $\mathbf{Q}$ is a positive matrix with $\det \mathbf{Q} > 0$; that γ is bivariate normal and truncated over $\mathcal{Y}$ with positive covariance; and that $\underline{y}_2 - b_2 \geq 0 > \underline{y}_1 - b_1$. Condition (SC-2D), from theorems 1 and 2, reads $(x_2 + a_2) \det \mathbf{Q} > 0$, which holds here by assumption. Hence, the example has PAM on the EE margin. Next, because $\mathbf{x} + \mathbf{a}$ is a positive vector and $\mathbf{Q}$ a positive matrix, $p(\mathbf{x}, \mathbf{y})$ is increasing in both y_1 and y_2, satisfying condition 2 of theorem 2. The truncated normal with positive covariance satisfies condition 3 (see app. B3), and condition 4 from theorem 2 is satisfied by assumption. PAM thus also occurs on the UE margin.

In contrast to this generic expansion in $\mathbf{x}$, theorem 3 considers an expansion of all skills such that the sum $\mathbf{x} + \mathbf{a}$ is scaled up; i.e., it considers an expansion in the direction of $\mathbf{x} + \mathbf{a}$. The theorem says that if two workers $\mathbf{x}$ and $\mathbf{x}'$ are such that one is twice as productive as the other in all jobs, $\mathbf{x}' + \mathbf{a} = 2(\mathbf{x} + \mathbf{a})$, then both workers are matched to the same distribution of jobs in equilibrium, irrespective of the complementarities in production.

held in response to this skill expansion. The resulting change in the distribution of the first job attribute, $H_1(y|\mathbf{x})$, is given by

$$\Delta H_1(y|\mathbf{x}) \simeq (x_1 + a_1)\frac{\partial H_1(y|\mathbf{x})}{\partial x_1}\left(\frac{\Delta x_1}{x_1 + a_1} - \frac{\Delta x_2}{x_2 + a_2}\right). \tag{6}$$

A major implication of equation (6) is that under a proportional increase in both skills ($\Delta x_1/x_1 = \Delta x_2/x_2$), $\Delta H_1(y|\mathbf{x}) < 0$ if and only if $x_1/x_2 > a_1/a_2$. In words, scaling up all skills leads to a stochastically better distribution of job matches in the first dimension, in which the worker is specialized relative to the baseline productivity vector **a**. By contrast, scaling up all skills leads to a deterioration of the distribution of job matches in the second dimension, $\Delta H_2(y|\mathbf{x}) > 0$, which can be obtained from the analogue of equation (6) for job attribute y_2. Scaling up all of a worker's skills simultaneously thus has a nonuniform effect on his distribution of jobs across dimensions, which depends on his specialization. Our interpretation is that this multi-D model does not feature any hierarchical sorting based on absolute advantage but instead features sorting based on specialization or comparative advantage.[23]

The result that uniformly better workers select into jobs that are not uniformly more productive suggests that there are sorting trade-offs in our multi-D setting. We now explore those trade-offs, starting with the following corollary of theorem 3, which addresses the interrelation of sorting patterns between a given job attribute y_j and different skills. That is, we focus on signing each row j of Jacobian matrix (5).

COROLLARY. 1. Under assumptions 1 and 2, if PAM occurs in dimension (y_j, x_k), then NAM occurs in $(y_j, x_{k'})$, $k \neq k'$.

The intuition behind corollary 1 is most transparent when the UE margin is shut down. We first focus on job dimension y_1 (i.e., the cognitive job content) and how cognitive skills x_1 and manual skills x_2 relate to it. Suppose that single-crossing condition (SC-2D) holds, $\partial(p_1(\mathbf{x})/p_2(\mathbf{x}))/\partial x_1 > 0 \Leftrightarrow \det \mathbf{Q} > 0$, which implies PAM in the cognitive dimension (y_1, x_1). Now, $\det \mathbf{Q} > 0$ also implies that $\partial(p_1(\mathbf{x})/p_2(\mathbf{x}))/\partial x_2 < 0$, which leads to NAM between the cognitive job trait and manual skill, (y_1, x_2). By the same argument, we can analyze how the second (manual) job dimension y_2 comoves with the two skills. The single-crossing condition, $\det \mathbf{Q} > 0$, implies PAM within the manual dimension, (y_2, x_2) but NAM "between" dimensions, (y_2, x_1).

This illustrates that for a given job attribute y_j, PAM cannot arise across both skill dimensions because the (necessary and sufficient) single-crossing conditions cannot hold simultaneously for both x_1 and x_2. Intuitively,

[23] To fix ideas, we focus in the text on a proportional increase in all skills. But from eq. (6) it is clear that sorting on comparative advantage also materializes under a nonproportional expansion of skills.

det $\mathbf{Q} > 0$ says that productive complementarities in (y_1, x_1) and (y_2, x_2) dominate complementarities in (y_1, x_2) and (y_2, x_1). This is why PAM occurs within the cognitive task (dimension 1) and within the manual task (dimension 2) but NAM occurs between those dimensions. These sorting trade-offs occur for purely technological reasons, independent of the sampling distribution.

Note that in this 2×2 case, corollary 1 also characterizes the sorting trade-offs across job attributes y_1 and y_2, for a given skill x_k (in each column k of matrix [5]). In particular, sorting cannot be simultaneously positive between a given skill and all job attributes, echoing the sorting trade-offs across skills for a given job attribute, described above. The reason why corollary 1 pins down the sign pattern of the entire Jacobian matrix (5) is that the characterization of sorting across both rows of the Jacobian automatically determines the sorting patterns across both columns. Beyond the two-dimensional case, however, corollary 1 does not fully describe the sorting trade-offs across job attributes (for a given skill). We address the general case in theorem O4 and corollary O3, in appendix OA.3. See also the generalizations of the baseline model in section VI.

2. Taking Stock

This section conveys two main predictions, both of which derive from theorem 3 and are empirically testable. First, sorting cannot be simultaneously positive between a given job dimension and all skill dimensions. Instead, there are sorting trade-offs. Agents need to decide which skill dimension to "sacrifice" and base that decision on the relative strength of complementarities in the technology. PAM arises between a job attribute and a skill with relatively strong complementarities, but NAM arises in the remaining dimension, where complementarities are weaker. In a similar way, sorting cannot be simultaneously positive between a given skill dimension and all job dimensions.

Second, these trade-offs play an important role for the sorting patterns that arise when we vary all skills simultaneously. We show that uniformly better workers match not with uniformly better jobs but instead with jobs that suit their skill mix. An improvement in the job dimension in which the worker is relatively strong goes hand in hand with a deterioration of the job dimension in which he is relatively weak. Our model thus predicts that multi-D sorting under random search is based on comparative advantage instead of absolute advantage.[24]

[24] The content of these results is again different when skills are 1-D. Consider theorem 3 for the 1-D case $X = 1$, so that $\mathbf{x}$ and $\mathbf{a}$ are scalars x and a, $\mathbf{Q}$ is a 1×2 row vector, $\tilde{y} = \mathbf{Q}(\mathbf{y} - \mathbf{b})$ is a scalar, and the flow-surplus function is $\sigma(x, \mathbf{y}) = (x + a)\tilde{y}$. In this case, theorem 3 again echoes the known result that there cannot be sorting, $\partial H_j(y|x)/\partial x = 0$: traits x and y_j, $j \in \{1, 2\}$ are independent in the population of worker-job matches.

Similar to section V.B, the underlying cause of these sorting patterns is that workers with different skill bundles rank jobs differently and climb different job ladders. Here, we illustrate more clearly how a worker's entire skill bundle matters for sorting.

VI. Equilibrium Sorting in More General Environments

We now show that our results on equilibrium sorting generalize to a wide range of environments: arbitrary numbers of job and worker attributes, various broad classes of technology, and all commonly used wage-setting protocols. Departing from the baseline case of two-dimensional heterogeneity and bilinear technology complicates the analysis because, with few exceptions, the characterization of sorting requires restrictions on the sampling distribution. Yet across all of our environments, a single-crossing condition on the production technology that guarantees sufficiently strong complementarities is the linchpin of positive sorting under random search.

These generalizations are technically involved. To keep the text focused on the baseline model that conveys the full intuition, we state our generalized theorems, along with a detailed discussion, in appendix OA and give only a brief overview here.

A. Higher Dimensions of Heterogeneity with a Bilinear Technology

We begin by noting that almost none of the results depend on there being only two dimensions of worker skills, $X = 2$. Our results extend verbatim to the case of $X > 2$ (with $Y = 2$).[25] Additional complications arise only for more than two job attributes, $Y > 2$.

Regarding the sign of sorting, in appendix OA we provide sufficient conditions for positive sorting under the bilinear technology for an unrestricted number of heterogeneity dimensions in corollary O2 (EE margin) and theorem O2 (UE margin), generalizing theorems 1 and 2. Similar to our baseline model, a generalized single-crossing condition is at the core of these results. Single crossing implies stronger complementarities between job attribute–worker skill pairs (y_j, x_k) for all $j < Y$ than between (y_Y, x_k), pushing toward positive sorting in all pairs (y_j, x_k) except (y_Y, x_k). But in this more general environment, assortative matching further requires restrictions on the sampling distribution γ. This is to prevent distributional obstructions to positive sorting that could arise from negatively correlated job attributes. Intuitively, sorting skill x_k positively along a certain

[25] The only complication from $X > 2$ is that single-crossing condition (SC-2D) is no longer equivalent to $\det \mathbf{Q} > 0$ and now depends on $\mathbf{x}$.

job dimension y_j may not be beneficial for a worker if y_j is negatively correlated with some other job attribute y_ℓ, which would cause negative sorting between (y_ℓ, x_k). This type of distributional barrier to sorting is irrelevant under 1-D heterogeneity, which is why distribution-free conditions are more commonly obtained in those contexts.

Regarding the interrelation of sorting patterns across dimensions, theorem 3 on sorting across skills readily generalizes to higher dimensions. Also, corollary 1 generalizes to the statement that sorting cannot be simultaneously positive between a given y_j and all x_k, $k \in \{1, \cdots, X\}$: there are sorting trade-offs. Finally, theorem O4 and corollary O3, in appendix OA, establish additional results about the interrelation of sorting across job attributes y_j, $j \in \{1, \cdots, Y\}$, for a given skill x_k, once again highlighting how agents trade off sorting across various dimensions.

B. Non-Bilinear Technology

Regarding the sign of sorting, theorem O1, in appendix OA, establishes sufficient conditions for positive sorting on the EE margin, generalizing theorem 1 to the case of nonlinear technologies that are strictly monotone in at least one job attribute y_j, with arbitrary dimensions of $\mathbf{x}$ and $\mathbf{y}$. Once again, the key ingredient of our results is a generalized single-crossing condition. Theorem O1 is our most general result on EE sorting and nests several special cases of interest. It nests not only theorem 1 but also the conditions for sorting under nonlinear surplus functions with $Y = 2$ (corollary O1, in app. OA) as well as under bilinear surplus functions with $Y > 2$ (corollary O2, in app. OA, discussed above).

These generalized results highlight that under higher-dimensional job heterogeneity ($Y > 2$), the conditions for sorting generally involve not only single crossing of the technology but also restrictions on the sampling distribution γ, echoing the case of bilinear technologies with $Y > 2$ above. Yet we can provide distribution-free results for the broad class of separable technologies, where skill x_k complements only a single y_j and is neither complement nor substitute for any other job attribute $y_{j'}$ (formally, $\partial^2 p(\mathbf{x}, \mathbf{y})/\partial x_k \partial y_j > 0$, but $\partial^2 p(\mathbf{x}, \mathbf{y})/\partial x_k \partial y_{j'} = 0$ for $j' \neq j$).[26] In appendix OA, theorem O3, we show that PAM arises on the EE margin in the pair (y_j, x_k) for such technologies.

Regarding the interrelation of sorting patterns across dimensions, we show in the appendix (see "Remarks" in sec. B4) that theorem 3 generalizes to technologies beyond the bilinear one.

[26] The separable class includes the popular specification of Tinbergen (1956), $p(\mathbf{x}, \mathbf{y}) = c_0 - \Sigma_{i=1}^{X} c_i (x_i - y_i)^2$, where $X = Y$ and where the c_i are strictly positive numbers. In this example, each job has an ideal skill bundle, given by $\mathbf{x} = \mathbf{y}$, and output is a decreasing function of the distance between the worker's skill bundle $\mathbf{x}$ and that ideal skill bundle. Another example is the bilinear technology with only "within" complementarities.

C. Alternative Wage-Setting Protocols

Our focus on surplus splitting via sequential auctions without worker bargaining power is for expositional clarity only. We show in appendix OC that our results on the sign of sorting hold for other wage-setting rules, such as Nash bargaining (Mortensen and Pissarides 1994; Moscarini 2001), wage/contract posting (Burdett and Mortensen 1998; Moscarini and Postel-Vinay 2013), or sequential auctions with worker bargaining power (Cahuc, Postel-Vinay, and Robin 2006).

VII. Discussion

We now discuss two important assumptions of our model: the lack of capacity constraints on the firm side and multi-D heterogeneity of workers and jobs.

A. Lack of Capacity Constraint

The assumption that firms lack capacity constraints—essentially, an assumption of constant returns to labor with free entry of firms on the search market—is widely used in the quantitative macro-labor literature. One of the main reasons is tractability.

Tractability is also a big motive for us to assume away capacity constraints. The lack of capacity constraints affords two simplifications in our analysis. First, our measure of sorting (FOSD monotonicity of the job type distribution conditional on worker types, $H_j(y|\mathbf{x})$) is tractable because workers always sample jobs from an exogenous and invariant distribution γ. The invariant job-sampling distribution naturally arises under no capacity constraint because filling a job at a firm does not change the availability of that job to other workers. If instead each firm had a fixed number of jobs, then workers would sample job types from an endogenous distribution, on which we could not impose any restrictions. Second, in the absence of capacity constraints, a firm's vacancy value is zero, and the match surplus depends on job characteristics $\mathbf{y}$ only through the technology. As a consequence, the comparison of match surpluses between two different jobs $\mathbf{y}$ and $\mathbf{y}'$ for a given worker of type $\mathbf{x}$ boils down to comparing flow surpluses (determined by technology). Instead, with capacity constraints, match surplus depends on the value of a vacant job, which is positive (as there is an option value of rejecting a suboptimal worker and waiting for a better one) and depends on $\mathbf{y}$. This creates a wedge between surplus and flow-surplus comparisons and makes it difficult to characterize sorting in terms of primitives.

The first issue (endogenous sampling distribution) could be circumvented by means of a "cloning" assumption, according to which any vacant job that is filled is immediately replaced by an identical one, essentially rendering γ exogenous.[27] But the second issue (positive value of a vacant job) would still be present in this case.

While the absence of firm capacity constraints provides us with the necessary tractability, we note that most existing studies of sorting—in both frictionless and frictional settings—heavily rely on capacity constraints. They induce scarcity, making agents more selective, which strengthens the desire to sort. Even if all workers produce the most output with the best firm, if the best firm has only one job, then productive workers can outbid less productive ones in the competition for that job (given that there are worker-job complementarities).

The scarcity mechanism is shut down if there are no capacity constraints, making it more difficult to obtain sorting. Indeed, as discussed above, under 1-D heterogeneity and in settings that otherwise resemble ours (notably, in settings with monotonicity of the technology in agents' 1-D characteristics and without endogenous worker search effort), the lack of capacity constraints rules out sorting. The reason is that all workers, regardless of their 1-D type, move up the same job ladder toward more productive firms at the same speed.[28]

By contrast, under multi-D heterogeneity, sorting can arise in equilibrium even with monotone technologies and exogenous search effort and without capacity constraints. The reason is that, in general, different multi-D worker types rank multi-D job types differently. Heterogenous rankings of firms across workers induce their job acceptance decisions to differ, generating sorting even if all firms accept all workers.

B. Multi-D Heterogeneity

One natural question is whether multi-D heterogeneity can be mapped into 1-D heterogeneity without loss of generality. If that were the case, 1-D models would provide all the tools we need for the analysis of sorting. However, we now show that generically, this cannot be done in a way that preserves minimal regularity properties of the surplus function.

We formalize that question in the context of our model by asking whether there exist (twice-differentiable) functions $\tilde{\sigma} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, $I : \mathbb{R}_+^X \to \mathbb{R}_+$, and $J : \mathbb{R}_+^Y \to \mathbb{R}_+$, where I and J are the worker and firm single indices, such that for all $(\mathbf{x}, \mathbf{y})$, $\sigma(\mathbf{x}, \mathbf{y}) = \tilde{\sigma}(I(\mathbf{x}), J(\mathbf{y}))$. In words, there exists a single-index representation if workers' skill bundles and jobs'

[27] We thank an anonymous referee for this suggestion.

[28] Things differ in the 1-D job ladder model by Lentz (2010) and Bagger and Lentz (2019), featuring endogenous search intensity and sorting, because differently skilled workers climb the same job ladder (meaning that all workers have the same ranking of firms) at different speeds; see our discussion of the literature in sec. I.

attribute bundles can be collapsed into 1-D indices without changing the surplus of any match (i.e., without changing the mapping from worker and job inputs to match surplus). In appendix B6, we prove the following result, which can be used as the base of a simple theory-based test to assess the single-index representation in the data (see Lindenlaub and Postel-Vinay 2022):

THEOREM 4. Under assumption 1, a single-index representation of the multi-D heterogeneity exists if and only if $\mathbf{Q}$ has rank (at most) one.

In the two-dimensional case, rank $\mathbf{Q} \leq 1$ is equivalent to $\det\mathbf{Q} = 0$, which in turn is equivalent to the failure of the single-crossing condition. Thus, in our baseline model, the single-index representation is valid only in rare cases, in which $\det\mathbf{Q} = 0$. This knife-edge result arises despite imposing few restrictions on indices I and J; for example, we allow them to be noninjective (and indeed, they fail to be injective in our model if the single-index representation exists).

More generally, and beyond the context of our model, multiple dimensions cannot be mapped into a single dimension while preserving basic regularity properties. When agents' attributes are continuously distributed (as in this paper), there exists no continuous bijective map of multi-D heterogeneity to 1-D heterogeneity. We base this statement on well-known results from topology about space-filling curves. A space-filling curve is a continuous function from the unit interval into the two-dimensional unit square (or, more generally, to an N-dimensional unit hypercube).[29]

The concept of a space-filling curve is closely related to the question whether multi-D heterogeneity can be collapsed to 1-D heterogeneity, since its inverse—if it exists—is what we are interested in: a continuous function that maps the entire multi-D to the 1-D space. Although it is well known that bijections between these spaces exist (Cantor 1878), a continuous bijection does not.[30] Continuity is one of the minimal regularity properties a dimension-reduction mapping should have. Points that are close together in the multi-D space should be associated with points that are close together on a line.

Thus, the property of continuity makes the distinction between multi-D and 1-D heterogeneity meaningful, regardless of any other property of the surplus function. In addition, another desirable property—monotonicity of the surplus function in types—cannot be guaranteed when mapping

[29] Peano (1890) was the first to construct a space-filling curve, the "Peano curve," using an iterative process until a single line fills the entire square, thus proving existence of a surjective continuous function $[0, 1] \rightarrow [0, 1]^2$.

[30] This result goes back to Netto (1879), who proves that a continuous surjective map $f:[0, 1] \rightarrow [0, 1]^2$ cannot be injective. In particular, continuous space-filling curves cannot be inverted. Dispensing with surjectivity does not help. There also exists no continuous one-to-one map from multi-D to 1-D heterogeneity, which follows from the Invariance of Domain Theorem by Brouwer (1912) and its corollary, the Topological Invariance of Dimension; see, for instance, Terence Tao's (2011) discussion.

multi-D to 1-D (for an illustrative example, see appendix OD). The lack of monotonicity would pose a challenge for the literature that aims to identify models with 1-D unobserved heterogeneity (which is generally used as a single index that proxies the underlying multi-D heterogeneity): Hagedorn, Law, and Manovksii (2017) assume monotonicity of output in worker and firm types, and so do Lamadon et al. (2014), as well as Bagger and Lentz (2019). In turn, Abowd, Kramarz, and Margolis (1999) assume monotonicity of wages in the firms' unobserved type; Sorkin (2018) assumes monotonicity of the workers' value of a job in firm type. We are not aware of identification arguments for unobserved scalar heterogeneity in the absence of any monotonicity assumption.[31]

We conclude that generically, both in the context of our model and beyond, it is challenging to map multi-D heterogeneity into 1-D heterogeneity in a meaningful way.

VIII. Conclusion

We generalize one of the workhorse models for labor market analysis—a job ladder model with random search—by incorporating the empirically relevant feature of multidimensional (multi-D) heterogeneity of workers and jobs. The goal of our analysis is to understand sorting between workers and jobs in this environment.

To describe the possibly complex sorting patterns that can arise, we first define notions of multi-D PAM and NAM in this frictional environment, based on FOSD monotonicity of the equilibrium matching distribution.

Using this notion of sorting, we highlight three main results. First, in all the environments we study, the key restriction on primitives for positive sorting between a given worker skill and a given job attribute is a single-crossing condition on the technology. It guarantees that the skill and job attribute under consideration are sufficiently complementary in production relative to other skills and job attributes, triggering the desire to sort.

Second, sorting patterns across the various dimensions of heterogeneity are interrelated. There are significant sorting trade-offs, in the sense that positive sorting occurs along the dimensions of skills and job attributes that are strong complements in production, relative to other dimensions. But negative sorting tends to arise in the dimension of weakest complementarity.

[31] An example of an application with discrete types is Bonhomme, Lamadon, and Manresa (2019), who assign firms to a fixed number of classes using *k*-means clustering based on similarity in within-firm wage distributions. Their identification relies on the assumption that any two firm classes have different wage distributions, something that tends to fail with horizontal heterogeneity, where both production and surplus are nonmonotone in firm types and the optimal firm type varies with worker types (see, e.g., Gautier, Teulings, and van Vuuren 2006, 2010).

Third, and related to these trade-offs, sorting in multiple dimensions is based on comparative rather than absolute advantage. Workers with uniformly higher skills do not sort into jobs with uniformly higher skill requirements. Rather, they sort into jobs with a higher requirement for the skill in which they are relatively strong, at the cost of lower requirements for the other skills.

That sorting arises in our setting, in which jobs face no capacity constraints, is surprising at first glance. It is well known that in comparable 1-D models, all workers share the same ranking of firms and thus face the same job ladder. They all climb this ladder at the same speed, which prevents sorting. By contrast, the sorting patterns that we describe arise because workers with different skill specializations accept different kinds of jobs, meaning that they move up and down different job ladders. Multi-D heterogeneity is itself a source of sorting.

Our theory provides useful tools for the analysis of multi-D sorting and mismatch in applied work, as illustrated in a companion paper, Lindenlaub and Postel-Vinay (2022). In that paper, we build on our multi-D framework to develop a theory-based, empirical protocol that detects the number and types of surplus-relevant worker and job characteristics in the data. Implementing this method on US data, we find that both worker heterogeneity and job heterogeneity are multi-D and that workers with different skill bundles climb different job ladders.

Our framework, along with these relevant worker and job characteristics, can be used to quantify multi-D mismatch and to assess the errors in the measurement of mismatch that are caused by imposing the 1-D assumption when the data-generating process is really multi-D. This is something we leave for future work.

Appendix A

Derivations

A1. Derivation of h*(x, y)*

Substituting the definition of $F_{\sigma|\mathbf{x}}$ (i.e., $\bar{F}_{\sigma|\mathbf{x}}(s) = \mathbb{E}[\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') > s\}]$) into equation (1), we see that $h(\mathbf{x},\mathbf{y})$ can be written as $h(\mathbf{x},\mathbf{y}) = \chi(\mathbf{x},\sigma(\mathbf{x},\mathbf{y}))\gamma(\mathbf{y})$, where the function χ solves

$$(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s))\chi(\mathbf{x},s)f_{\sigma|\mathbf{x}}(s) = \lambda_0 f_{\sigma|\mathbf{x}}(s)\mathbf{1}\{s \geq 0\}u(\mathbf{x}) + \lambda_1 f_{\sigma|\mathbf{x}}(s)\int_0^s \chi(\mathbf{x},s')\,dF_{\sigma|\mathbf{x}}(s').$$

This ordinary differential equation solves as

$$(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s))\int_0^s \chi(\mathbf{x},s')\,dF_{\sigma|\mathbf{x}}(s') = \lambda_0 \mathbf{1}\{s \geq 0\}u(\mathbf{x})(F_{\sigma|\mathbf{x}}(s) - F_{\sigma|\mathbf{x}}(0)).$$

In other words, by differentiation,

$$h(\mathbf{x},\mathbf{y}) = \lambda_0 \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}) \geq 0\} u(\mathbf{x}) \frac{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0)}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y})))^2} \gamma(\mathbf{y}).$$

Finally, recalling from the flow-balance equations that $\lambda_0 \bar{F}_{\sigma|\mathbf{x}}(0) u(\mathbf{x}) = \delta(\ell(\mathbf{x}) - u(\mathbf{x}))$ and substituting out $u(\mathbf{x})$ yields the expression of $h(\mathbf{x}, \mathbf{y})$, from which we derive $h(\mathbf{y}|\mathbf{x})$ (eq. [3]) in the text.

A2. Derivation of Decomposition (DEC)

Recall equation (4):

$$H_j(y|\mathbf{x}) = \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \int \frac{\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') \geq 0\} \times \mathbf{1}\{y'_j \leq y\}}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}')))^2} \gamma(\mathbf{y}')\, d\mathbf{y}'.$$

Differentiating yields

$$\begin{aligned}\frac{\partial H_j(y|\mathbf{x})}{\partial x_k} = & \underbrace{-\frac{\delta(\partial \bar{F}_{\sigma|\mathbf{x}}(0)/\partial x_k)}{\bar{F}_{\sigma|\mathbf{x}}(0)(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))} H_j(y|\mathbf{x})}_{(1)} \\ & \underbrace{+ \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \int \frac{(\partial\sigma(\mathbf{x},\mathbf{y}')/\partial x_k) \times \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = 0\} \times \mathbf{1}\{y'_j \leq y\}}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}')))^2} \gamma(\mathbf{y}')\, d\mathbf{y}'}_{(2)} \\ & \underbrace{- \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \int \frac{2\lambda_1 \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') \geq 0\} \times \mathbf{1}\{y'_j \leq y\}}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}')))^3} \times \frac{\partial}{\partial x_k}(1 - F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}')))\gamma(\mathbf{y}') d\mathbf{y}'}_{(3)}.\end{aligned}$$

We examine those three terms in turn.

First, the definition $\bar{F}_{\sigma|\mathbf{x}}(s) := 1 - F_{\sigma|\mathbf{x}}(s) = \int \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') \geq s\}\gamma(\mathbf{y}')\, d\mathbf{y}'$ implies

$$\begin{aligned}\frac{\partial}{\partial x_k}(1 - F_{\sigma|\mathbf{x}}(s)) &= \int \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\} \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \gamma(\mathbf{y}')\, d\mathbf{y}' \\ &= \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s\right] \times f_{\sigma|\mathbf{x}}(s).\end{aligned} \tag{7}$$

Replacing into term 1 yields

$$\begin{aligned}(1) &= -\frac{\delta f_{\sigma|\mathbf{x}}(0)}{\bar{F}_{\sigma|\mathbf{x}}(0)(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))} \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0\right] \times H_j(y|\mathbf{x}) \\ &= -g_{\sigma|\mathbf{x}}(0) \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0\right] \times H_j(y|\mathbf{x}),\end{aligned}$$

where we used the density $g_{\sigma|\mathbf{x}}(s)$, corresponding to the cdf

$$G_{\sigma|\mathbf{x}}(s) := 1 - \frac{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0)}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \frac{\bar{F}_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)}.$$

Next, term 2 can be rewritten as

$$
\begin{aligned}
(2) &= \frac{\delta}{\bar{F}_{\sigma|\mathbf{x}}(0)(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))} \times \int \frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \times \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') = 0\} \times \mathbf{1}\{y_j' \le y\} \gamma(\mathbf{y}')\, d\mathbf{y}' \\
&= \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \int \frac{\mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') = 0\} \times \mathbf{1}\{y_j' \le y\}}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))^2} \gamma(\mathbf{y}')\, d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} | \sigma(\mathbf{x}, \mathbf{y}') = 0, y_j' \le y\right] \\
&= \frac{\partial K_j(y, 0|\mathbf{x})}{\partial s} \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} | \sigma(\mathbf{x}, \mathbf{y}') = 0, y_j' \le y\right],
\end{aligned}
$$

where $K_j(y, s|\mathbf{x})$ is the joint cdf of job attribute y_j and match flow surplus s, conditional on worker type $\mathbf{x}$, in the population of employed workers, given by

$$
\begin{aligned}
K_j(y, s|\mathbf{x}) &= \int \mathbf{1}\left\{y_j' \le y\right\} \times \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') \le s\} h(\mathbf{y}'|\mathbf{x})\, d\mathbf{y}' \\
&= \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \int \frac{\mathbf{1}\{0 \le \sigma(\mathbf{x}, \mathbf{y}') \le s\} \times \mathbf{1}\left\{y_j' \le y\right\}}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}')))^2} \gamma(\mathbf{y}')\, d\mathbf{y}',
\end{aligned}
\tag{8}
$$

which is the probability that a randomly chosen type-$\mathbf{x}$ employed worker is in a job whose jth attribute is less than y and generates a flow surplus less than s. Note that $H_j(y|\mathbf{x})$ and $G_{\sigma|\mathbf{x}}(s)$ are the marginal cdfs of $K_j(y, s|\mathbf{x})$, so that $K_j(y, +\infty|\mathbf{x}) = H_j(y|\mathbf{x})$ and $K_j(+\infty, s|\mathbf{x}) = G_{\sigma|\mathbf{x}}(s)$, and, moreover,

$$
\frac{\partial K_j(y, s|\mathbf{x})}{\partial s} = g_{\sigma|\mathbf{x}}(s) \times \mathbb{P}_\Gamma\left\{y_j' \le y | \sigma(\mathbf{x}, \mathbf{y}') = s\right\}.
$$

Therefore,

$$
(2) = g_{\sigma|\mathbf{x}}(0) \times \mathbb{P}_\Gamma\left\{y_j' \le y | \sigma(\mathbf{x}, \mathbf{y}') = 0\right\} \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} | \sigma(\mathbf{x}, \mathbf{y}') = 0, y_j' \le y\right].
$$

Now on to term 3. Again from equation (7), we have that

$$
\begin{aligned}
\frac{\partial}{\partial x_k}(1 - F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}'))) &= f_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}')) \\
&\quad \times \left\{\mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}'')}{\partial x_k} \middle| \sigma(\mathbf{x}, \mathbf{y}'') = \sigma(\mathbf{x}, \mathbf{y}')\right] - \frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k}\right\}.
\end{aligned}
$$

Substituting into term 3,

$$
\begin{aligned}
(3) = \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} &\times \int \frac{2\lambda_1 \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') \ge 0\} \times \mathbf{1}\left\{y_j' \le y\right\} \times f_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}'))}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}')))^3} \\
&\times \left\{\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}'')}{\partial x_k} \middle| \sigma(\mathbf{x}, \mathbf{y}'') = \sigma(\mathbf{x}, \mathbf{y}')\right]\right\} \gamma(\mathbf{y}')\, d\mathbf{y}',
\end{aligned}
$$

which can be recast as[32]

[32] A technical note: strictly speaking, the correct integration bounds in the following formula are

$$
s \in \left[\max\left\{0, \min_{\mathbf{y}' \in \mathcal{Y}, y_j' \le y} \sigma(\mathbf{x}, \mathbf{y}')\right\}, \max_{\mathbf{y}' \in \mathcal{Y}, y_j' \le y} \sigma(\mathbf{x}, \mathbf{y}')\right].
$$

rather than $[0, +\infty)$. To avoid cluttering the formula with these unwieldy integration bounds, we write it as an integral over all $s \ge 0$. As a consequence, it may be that the joint event

$$(3) = \frac{\delta(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(0))}{\bar{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty} \frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s))} \times \int \frac{\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\} \times \mathbf{1}\{y'_j \le y\}}{[\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)]^2} \gamma(\mathbf{y}')\, d\mathbf{y}'$$
$$\times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s, y'_j \le y \right] - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s \right] \right\} ds$$
$$= \int_0^{+\infty} \frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)}{(\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s))} \times \frac{\partial K_j(y, s|\mathbf{x})}{\partial s}$$
$$\times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y})}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s, y'_j \le y \right] - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s \right] \right\} ds.$$

Combining terms 1–3 and substituting the definitions of $g_{\sigma|\mathbf{x}}(0)$ and $\partial K_j(y, 0|\mathbf{x})/\partial s$ proves that

$$\frac{\partial H_j(y|\mathbf{x})}{\partial x_k} = g_{\sigma|\mathbf{x}}(0) \times \left\{ \mathbb{P}_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\} \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0, y'_j \le y \right] \right.$$
$$\left. - H_j(y|\mathbf{x}) \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0 \right] \right\}$$
$$+ \int_0^{+\infty} \frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)} \times g_{\sigma|\mathbf{x}}(s) \times \mathbb{P}_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = s\}$$
$$\times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s, y'_j \le y \right] - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s \right] \right\} ds,$$

where we incorporated the identity $\partial K_j(y, s|\mathbf{x})/\partial s = g_{\sigma|\mathbf{x}}(s) \times \mathbb{P}_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = s\}$. Further, add and subtract

$$g_{\sigma|\mathbf{x}}(0) \times \mathbb{P}_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\} \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0 \right],$$

to obtain

$$\frac{\partial H_j(y|\mathbf{x})}{\partial x_k} = g_{\sigma|\mathbf{x}}(0) \times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0 \right] \times \left(\Pr_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\} - H_j(y|\mathbf{x}) \right) \right.$$
$$+ \Pr_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\}$$
$$\left. \times \left(\mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0, y'_j \le y \right] - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = 0 \right] \right) \right\}$$
$$+ \int_0^{+\infty} \frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)} \times g_{\sigma|\mathbf{x}}(s) \times \mathbb{P}_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = s\}$$
$$\times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s, y'_j \le y \right] - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \middle| \sigma(\mathbf{x},\mathbf{y}') = s \right] \right\} ds.$$

Finally, use the identity $H_j(y|\mathbf{x}) = \int_0^{+\infty} g_{\sigma|\mathbf{x}}(s) \mathbb{P}_\Gamma\{y'_j \le y|\sigma(\mathbf{x},\mathbf{y}') = s\}\, ds$ to reformulate the term in the first two lines as

$(\sigma(\mathbf{x},\mathbf{y}') = s, y'_j \le y)$, on which some of the expectations are conditioned, has zero probability for some values of (s, y). Yet in those cases, $\int \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\} \times \mathbf{1}\{y'_j \le y\}\gamma(\mathbf{y}')\, d\mathbf{y}' = 0$. The formula thus remains correct with $[0, +\infty)$ as integration bounds if we adopt the convention that any expectation conditioned on a zero-probability event is equal to zero.

$$\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\middle|\sigma(\mathbf{x},\mathbf{y}') = 0\right]$$

$$\times\int_0^{+\infty} g_{\sigma|\mathbf{x}}(s)\left[\mathbb{P}_\Gamma\left\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\right\} - \mathbb{P}_\Gamma\left\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}') = s\right\}\right] ds,$$

which gives (DEC).

Appendix B

Proofs

B1. An Ancillary Lemma

We begin by stating a lemma to better understand the nature of decomposition (DEC). This will help us specify conditions on primitives under which sorting arises.

LEMMA 1. If, for all $s \ge 0$ and y such that $\mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}') = s\} > 0$,

$$y \mapsto \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\middle|\sigma(\mathbf{x},\mathbf{y}') = s, y_j' = y\right] \tag{CMP}$$

is strictly increasing (decreasing), then term 2 in (DEC) is strictly negative (positive) for all y; that is, PAM (NAM) occurs in dimension (y_j, x_k) along the EE margin.

The proof follows directly from inspecting term 2 in (DEC). QED

Lemma 1 implies that, if the UE margin is shut down (i.e., if $\sigma(\mathbf{x},\mathbf{y}) \ge 0$ for all $\mathbf{y}$) and if condition (CMP)—our label for complementarity—holds, then the marginal distribution of job attribute y_j of employed workers of type $\mathbf{x}$, $H_j(\cdot|\mathbf{x})$, is monotone with respect to worker skill x_k in the FOSD sense; that is, there is PAM in dimension (y_j, x_k).

Condition (CMP) can be interpreted as imposing a strong form of complementarity (or substitutability, in the decreasing case) between job attribute j and worker skill k, as is typical of models of sorting. It imposes, loosely speaking, that $\sigma(\mathbf{x},\mathbf{y})$ be supermodular along all of its level sets, which amounts to a restriction involving not only the technology but generally also the sampling distribution of job types.

B2. Proof of Theorem 1

The objective is to find conditions for PAM in dimension (y_1, x_k). Generally, in order to obtain sufficient conditions we want to specify conditions under which condition (CMP) in lemma 1 holds, that is, under which $\mathbb{E}_\Gamma[(\partial\sigma(\mathbf{x},\mathbf{y}')/\partial x_k)|\sigma(\mathbf{x},\mathbf{y}') = s, y_j' = y]$ is strictly increasing in y. In the case of bilinear technology with $Y = 2$, however, we can circumvent this sufficient condition and compute term 2 in expression (DEC) explicitly. This allows us to state necessary conditions as well.

First, note that assumption 1 implies invertibility of the surplus function, so that

$$\sigma(\mathbf{x},\mathbf{y}) = s \Leftrightarrow y_2 - b_2 = \frac{s}{p_2(\mathbf{x})} - \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})}(y_1 - b_1) \coloneqq R(s, y_1).$$

Then, we can express

$$\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\middle|\sigma(\mathbf{x},\mathbf{y}') = s, y_1' \le y\right] = \mathbb{E}_{\mu_{s,\mathbf{x}}}\left[\frac{\partial\sigma(\mathbf{x},(y_1', R(s,y_1')))}{\partial x_k}\middle|y_1' \le y\right], \quad (9)$$

where $\mu_{s,\mathbf{x}}(y_1)$ is the sampling density of y_1 conditional on $\mathbf{x}$ and on $\sigma(\mathbf{x},\mathbf{y}) = s$:

$$\mu_{s,\mathbf{x}}(y_1) = \frac{\gamma(y_1, R(s,y_1))\partial R(s,y_1)/\partial s}{\int \gamma(y_1', R(s,y_1'))\partial R(s,y_1')/\partial s\, dy_1'}.$$

Note that $\partial R(s,y_1)/\partial s$ is a constant when p is bilinear and thus drops from $\mu_{s,\mathbf{x}}$. Finally, assumption 1 implies $\partial\sigma(\mathbf{x},\mathbf{y})/\partial x_k = \Sigma_{j=1}^2 q_{kj}(y_j - b_j)$, so that

$$\frac{\partial\sigma(\mathbf{x},(y_1, R(s,y_1)))}{\partial x_k} = \frac{q_{k2}}{p_2(\mathbf{x})}s + \frac{q_{k1}p_2(\mathbf{x}) - q_{k2}p_1(\mathbf{x})}{p_2(\mathbf{x})}(y_1 - b_1).$$

It follows that

$$\mathbb{E}_{\mu_{s,\mathbf{x}}}\left[\frac{\partial\sigma(\mathbf{x},(y_1', R(s,y_1')))}{\partial x_k}|y_1' \le y\right] = \frac{q_{k2}}{p_2(\mathbf{x})}s + \frac{q_{k1}p_2(\mathbf{x}) - q_{k2}p_1(\mathbf{x})}{p_2(\mathbf{x})}\mathbb{E}_{\mu_{s,\mathbf{x}}}[y_1' - b_1|y_1' \le y].$$

Hence, term 2 in expression (DEC) is equal to

$$-\frac{q_{k1}p_2(\mathbf{x}) - q_{k2}p_1(\mathbf{x})}{p_2(\mathbf{x})}\int_0^{+\infty}\frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)g_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1\bar{F}_{\sigma|\mathbf{x}}(s)} \times \mathbb{P}_\Gamma\{y_1' \le y|\sigma(\mathbf{x},\mathbf{y}') = s\} \times (\mathbb{E}_{\mu_{s,\mathbf{x}}}[y_1' - b_1] - \mathbb{E}_{\mu_{s,\mathbf{x}}}[y_1' - b_1|y_1' \le y])\, ds. \quad (10)$$

Both $p_2(\mathbf{x})$ (by assumption) and the difference in expectations (by construction) are positive. Condition (SC-2D), which here implies $q_{k1}p_2(\mathbf{x}) - q_{k2}p_1(\mathbf{x}) > 0$ (or $\det\mathbf{Q} > 0$), is therefore necessary and sufficient for equation (10) (which is equivalent to term 2 in [DEC]) to be negative, that is, it is a necessary and sufficient for PAM in (y_1, x_k). QED

Remark.—Note that, even though we stated theorem 1 for the 2×2 case to ease exposition, nothing in the proof hinges on $X = 2$, and thus the theorem and proof readily apply to the case of $X \ge 2$.

B3. *Proof of Theorem 2*

B3.1. Sufficiency

We have to sign term 1 in decomposition (DEC), which, as explained in the main text, reflects sorting along the UE margin. Whenever $g_{\sigma|\mathbf{x}}(0) > 0$, said term 1 has the sign of

$$\mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\} \times \left(\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}|\sigma(\mathbf{x},\mathbf{y}') = 0, y_j' \le y\right] - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}|\sigma(\mathbf{x},\mathbf{y}') = 0\right]\right) + \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}|\sigma(\mathbf{x},\mathbf{y}') = 0\right] \times \int_0^{+\infty} g_{\sigma|\mathbf{x}}(s)\left[\mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}') = 0\} - \mathbb{P}_\Gamma\{y_j' \le y|\sigma(\mathbf{x},\mathbf{y}') = s\}\right] ds. \quad (11)$$

In the case of $Y = 2$, the first line in expression (11) is negative under the assumed condition (SC-2D) (see proof of theorem 1). We thus focus on showing that the second line is negative. We proceed in two steps.

First, we find conditions under which $\mathbb{P}_\Gamma\{y_1' \le y | \sigma(\mathbf{x}, \mathbf{y}') = s\}$ is decreasing in s, implying that $[\mathbb{P}_\Gamma\{y_1' \le y | \sigma(\mathbf{x}, \mathbf{y}') = 0\} - \mathbb{P}_\Gamma\{y_1' \le y | \sigma(\mathbf{x}, \mathbf{y}') = s\}] \ge 0$. Equivalently, we want to derive conditions under which, for any $s_H > s_L \ge 0$,

$$\frac{\int_{\underline{y}_1}^{y} \gamma(y_1', (s_H/p_2(\mathbf{x})) + b_2 - (p_1(\mathbf{x})/p_2(\mathbf{x}))(y_1' - b_1))\, dy_1'}{\int_{\underline{y}_1}^{\bar{y}_1} \gamma(y_1', (s_H/p_2(\mathbf{x})) + b_2 - (p_1(\mathbf{x})/p_2(\mathbf{x}))(y_1' - b_1))\, dy_1'}$$
$$\le \frac{\int_{\underline{y}_1}^{y} \gamma(y_1', (s_L/p_2(\mathbf{x})) + b_2 - (p_1(\mathbf{x})/p_2(\mathbf{x}))(y_1' - b_1))\, dy_1'}{\int_{\underline{y}_1}^{\bar{y}_1} \gamma(y_1', (s_L/p_2(\mathbf{x})) + b_2 - (p_1(\mathbf{x})/p_2(\mathbf{x}))(y_1' - b_1))\, dy_1'}.$$

Defining

$$g(y, s) := \int_{\underline{y}_1}^{y} \gamma\left(y_1', \frac{s}{p_2(\mathbf{x})} + b_2 - \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})}(y_1' - b_1)\right) dy_1'$$

and rearranging the previous inequality gives $g(\bar{y}_1, s_L)g(y, s_H) \le g(y, s_L)g(\bar{y}_1, s_H)$. Since $y \le \bar{y}_1$, this inequality holds if g is log-supermodular in (y, s). To show when this is the case, define—similar to the proof of theorem 1—the joint distribution of y_1 and s (conditional on $\mathbf{x}$) as

$$\mu_{\mathbf{x}}(y_1, s) = \gamma\left(y_1, \frac{s}{p_2(\mathbf{x})} + b_2 - \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})}(y_1 - b_1)\right)$$

and rewrite $g(y, s) = \int \mathbf{1}\{y_1' < y\}\mu_{\mathbf{x}}(y_1', s)\, dy_1'$. Note that

1. the support of $\mu_{\mathbf{x}}(y_1, s)$ is a lattice (for more details, see the proof of theorem O2 [and in particular footnote 4] in appendix OB.4);
2. the joint distribution $\mu_{\mathbf{x}}(y_1, s)$ is log-supermodular in (y_1, s) if

$$p_2(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_2 \partial y_1} - p_1(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_2^2} \ge 0; \tag{12}$$

3. the indicator function, $\mathbf{1}\{y_1' < y\}$, is log-supermodular in (y, y_1').

Therefore, the product $\mathbf{1}\{y_1' < y\}\mu_{\mathbf{x}}(y_1', s)$ is log-supermodular in (y, y_1', s), since the product of log-supermodular functions is log-supermodular. In turn, this implies that g is log-supermodular in (y, s), since log-supermodularity is preserved under integration.

Thus, if condition (12) holds (stated as condition [UE-2D] in theorem 2), then $\mathbb{P}_\Gamma\{y_1' \le y | \sigma(\mathbf{x}, \mathbf{y}') = s\}$ is decreasing in s.

Second, we derive conditions such that $\mathbb{E}_\Gamma[(\partial\sigma(\mathbf{x}, \mathbf{y}')/\partial x_k) | \sigma(\mathbf{x}, \mathbf{y}') = 0] \le 0$. Note that under assumptions 1 and 2,

$$\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\middle|\sigma(\mathbf{x},\mathbf{y}') = 0\right] = p_1(\mathbf{x})\left(\frac{q_{k1}}{p_1(\mathbf{x})} - \frac{q_{k2}}{p_2(\mathbf{x})}\right)\mathbb{E}_\Gamma[y_1' - b_1|\sigma(\mathbf{x},\mathbf{y}') = 0].$$

The term outside the expectation is strictly positive by single-crossing condition (SC-2D), and the expectation is negative if $y_2 \geq b_2$ and $p_1(\mathbf{x}) > 0$ (stated as conditions 2 and 4 in theorem 2), since

$$\mathbb{E}_\Gamma[y_1' - b_1|\sigma(\mathbf{x},\mathbf{y}') = 0] = \mathbb{E}_\Gamma\left[y_1'\middle|y_1' = -(y_2' - b_2)\frac{p_2(\mathbf{x})}{p_1(\mathbf{x})} + b_1\right] - b_1.$$

Thus, $\mathbb{E}_\Gamma[(\partial\sigma(\mathbf{x},\mathbf{y}')/\partial x_k)|\sigma(\mathbf{x},\mathbf{y}') = 0] \leq 0$ under the conditions from theorem 2, proving that those conditions are sufficient for expression (11) (and thus, term 1 in decomposition [DEC]) to be negative and hence sufficient for PAM on the UE margin.

B3.2. Necessity

To show that single-crossing condition (SC-2D) is also necessary for PAM when considering all possible sampling distributions γ, recall that there is PAM in dimension (y_1, x_k) on the UE margin if and only if expression (11) is negative. It thus suffices to show that there exists a sampling distribution γ under which condition (SC-2D) is necessary for expression (11) to be negative.

We again focus on bilinear technology and $Y = 2$. First, note that the first line in expression (11) is negative if and only if condition (SC-2D) holds (by an argument analogous to theorem 1). Second, note that if γ is log-supermodular with log-concave marginals, then $[\mathbb{P}_\Gamma\{y_1' \leq y|\sigma(\mathbf{x},\mathbf{y}') = 0\} - \mathbb{P}_\Gamma\{y_1' \leq y|\sigma(\mathbf{x},\mathbf{y}') = s\}] \geq 0$; see section B3.1. Hence, if γ is uniform with independent marginals then $\partial^2 \ln\gamma/\partial y_1\partial y_2 = 0$ and $\partial^2\ln\gamma/\partial y_2^2 = 0$, so that $[\mathbb{P}_\Gamma\{y_1' \leq y|\sigma(\mathbf{x},\mathbf{y}') = 0\} - \mathbb{P}_\Gamma\{y_1' \leq y|\sigma(\mathbf{x},\mathbf{y}') = s\}] = 0$ for all (y, s). Only the first line in expression (11) remains, which is negative only if condition (SC-2D) holds. QED

Remarks.—First, note that, even though we stated theorem 2 for the 2×2 case to ease exposition, nothing in the proof hinges on $X = 2$, and thus the theorem and proof readily apply to the case of $X \geq 2$.

Second, sufficient conditions on the sampling distribution for condition (UE-2D) to hold are that the density γ be log-supermodular and its marginals log-concave. This class of distributions is quite broad. For instance, any bivariate distribution of independent random variables that has log-concave marginals (e.g., the uniform distribution with independent random variables) satisfies condition (UE-2D). Other examples of log-supermodular densities with log-concave marginals are the (truncated) bivariate Gaussian with positive covariance and the multivariate Gamma distribution. The multivariate Gamma distribution is that of a linear combination of independent standard Gamma-distributed random variables. Log-supermodularity of the multivariate Gamma density is implied by Karlin and Rinott (1980, proposition 3.8) and log-concavity by Shapiro, Dentcheva, and Ruszczyński (2009, theorem 4.26).

B3.3. Supplementary Graphical Analysis

Figure B1 supplements our explanations from the main text with an in-depth analysis of how the boundary of marginally profitable matches shifts with x_k under the assumptions of theorem 2. It helps visualize the role of both single crossing and distributional restrictions in that theorem.

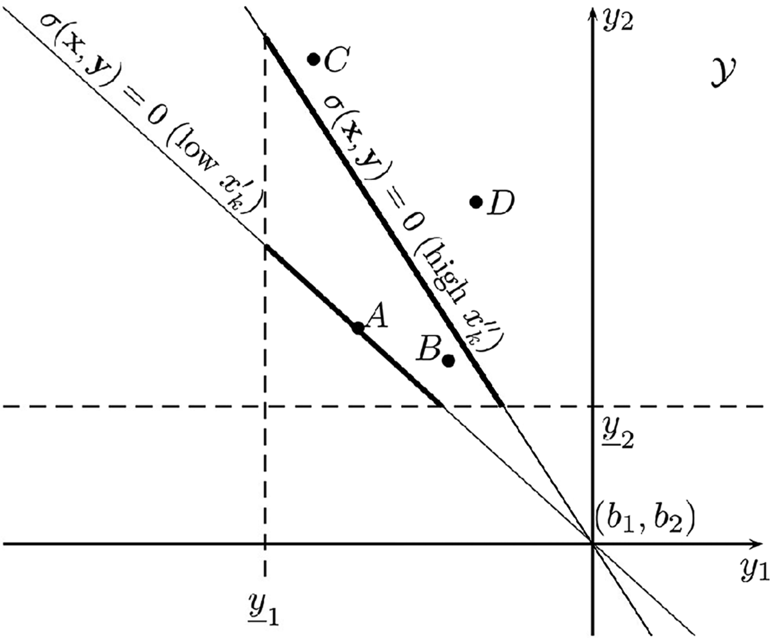

FIG. B1.—Sorting along the UE margin.

Figure B1 represents the (y_1, y_2) plane, where the origin is placed at $\mathbf{b} = (b_1, b_2)$. The area above $\underline{y}_2$ and to the right of $\underline{y}_1$ materializes $\mathcal{Y}$, the support of γ: the (lower) boundaries of $\mathcal{Y}$ are the horizontal line at $y_2 = \underline{y}_2$ and the vertical line at $y_1 = \underline{y}_1$, which are placed in compliance with condition 4 in theorem 2. The oblique lines are zero-level curves of $\sigma(\mathbf{x}, \cdot)$, which under the assumed linear technology are given by $y_2 = b_2 - (p_1(\mathbf{x})/p_2(\mathbf{x}))(y_1 - b_1)$. By condition 2 in theorem 2, such lines are downward sloping and go through point $\mathbf{y} = \mathbf{b}$. The boundary of feasible matches for a given skill bundle $\mathbf{x}$ is at the intersection between the zero-level curve of $\sigma(\mathbf{x}, \cdot)$ and $\mathcal{Y}$ (the thickened line segment). Note that this boundary lies entirely in the region of $\mathcal{Y}$ where $y_1 < b_1$: because it is assumed that $\underline{y}_2 \geq b_2$, it has to be the case that $y_1 < b_1$ for the surplus to equal zero.

Those zero-level curves are drawn for two workers $\mathbf{x}'$ and $\mathbf{x}''$ with $x''_k > x'_k$ (and with the same amount of the other skill). The higher-x''_k curve is steeper than the lower-x'_k one, meaning that for a given y_2, the more skilled worker needs a higher y_1 to generate nonnegative surplus. The reason is as follows: by the single-crossing property condition (SC-2D), complementarities in production are stronger between x_k and y_1 than between x_k and y_2. Thus, the jobs under consideration (with $y_1 < b_1$) are prone to generate surplus losses, especially for those workers with high x_k. In the figure, this means that all job types between the two solid lines can be profitably matched with the low-skilled (x_k') worker. But they produce negative surplus with the high-skilled (x_k'') worker, and therefore the jobs with relatively low attribute y_1 drop out of his equilibrium matching set. In conclusion, for a given y_2, workers with higher x_k need jobs with higher y_1 to generate nonnegative surplus, which is clearly a force toward PAM.

However, complementarities in production alone are not enough to ensure PAM on the UE margin. To see this, consider points *A*–*D* on the figure. After increasing skill *k* from x_k' to x_k'', a worker no longer breaks even with a job at *A*. Moreover, jobs around *B* (with higher y_1 but lower y_2) are also made unprofitable, while jobs around *C*, with lower y_1 but higher y_2 compared to those around *A*, remain profitable. Therefore, if sampling distribution γ has most of its mass concentrated around points *A*–*C* (i.e., there is a negative correlation between y_1 and y_2), then workers with higher x_k will be matched to jobs with lower y_1 (since jobs around *B* with higher y_1 have too little of y_2, leading to negative surplus)—a force toward NAM. To prevent this, we assume a sufficient degree of positive association between y_1 and y_2 in γ to ensure that more mass is concentrated around points *A* and *D*. This is the content of condition 3 of theorem 2. Note that the distributional barrier to PAM arising from a negative association of (y_1, y_2) becomes more severe the larger is the positive impact of y_2 on the surplus (i.e., the larger is $p_2(\mathbf{x})$, which makes the zero-surplus lines flatter).

B4. Proof of Theorem 3

To economize on notation, we use the identity $\partial K_j(y, s|\mathbf{x})/\partial s = g_{\sigma|\mathbf{x}}(s) \times \mathbb{P}_\Gamma\{y_j' \leq y| \sigma(\mathbf{x}, \mathbf{y}') = s\}$ throughout the proof (see eq. [8] for its derivation).

From decomposition (DEC) applied to the case of bilinear production function, we obtain

$$\begin{aligned}
(\mathbf{x}+\mathbf{a})^\top \nabla H_j(y|\mathbf{x}) &= \sum_{k=1}^{2}(x_k + a_k)\frac{\partial H_j(y|\mathbf{x})}{\partial x_k} \\
&= \mathbb{E}_\Gamma\left[(\mathbf{x}+\mathbf{a})^\top \nabla_\mathbf{x}\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = 0, y_j' \leq y\right]\frac{\partial K_j(y,0|\mathbf{x})}{\partial s} \\
&\quad - \mathbb{E}_\Gamma[(\mathbf{x}+\mathbf{a})^\top \nabla_\mathbf{x}\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = 0]H_j(y|\mathbf{x})g_{\sigma|\mathbf{x}}(0) \\
&\quad + \int_0^{+\infty}\frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)} \times \frac{\partial K_j(y,s|\mathbf{x})}{\partial s} \\
&\quad \times \left\{\mathbb{E}_\Gamma\left[(\mathbf{x}+\mathbf{a})^\top \nabla_\mathbf{x}\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = s, y_j' \leq y\right] - \mathbb{E}_\Gamma[(\mathbf{x}+\mathbf{a})^\top \nabla_\mathbf{x}\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = s]\right\} ds.
\end{aligned}$$

But then linearity of the flow-surplus function σ in $(\mathbf{x} + \mathbf{a})$ implies that $(\mathbf{x} + \mathbf{a})^\top \nabla_\mathbf{x}\sigma(\mathbf{x}, \mathbf{y}) = \sigma(\mathbf{x}, \mathbf{y})$. Substitution into the latter equation yields the following equation, whose right-hand-side terms are all equal to zero:

$$\begin{aligned}
(\mathbf{x}+\mathbf{a})^\top \nabla H_j(y|\mathbf{x}) &= \mathbb{E}_\Gamma\left[\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = 0, y_j' \leq y\right] \times \frac{\partial K_j(y,0|\mathbf{x})}{\partial s} \\
&\quad - \mathbb{E}_\Gamma[\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = 0]H_j(y|\mathbf{x})g_{\sigma|\mathbf{x}}(0) \\
&\quad + \int_0^{+\infty}\frac{2\lambda_1 f_{\sigma|\mathbf{x}}(s)}{\delta + \lambda_1 \bar{F}_{\sigma|\mathbf{x}}(s)} \times \frac{\partial K_j(y,s|\mathbf{x})}{\partial s} \times \left\{\mathbb{E}_\Gamma[\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = s, y_j' \leq y] - \mathbb{E}_\Gamma[\sigma(\mathbf{x},\mathbf{y}')|\sigma(\mathbf{x},\mathbf{y}') = s]\right\} ds.
\end{aligned}$$

QED

Remarks.—First, even though theorem 3 is stated for the case of $Y = X = 2$ to streamline the exposition, both the theorem and the proof readily apply more generally if $Y \geq 2$ and $X \geq 2$. Second, note that the proof above is virtually unchanged if, instead of assuming that $\sigma(\cdot)$ is linear in $(\mathbf{x}+\mathbf{a})$, one assumes only that it is homogeneous in $(\mathbf{x}+\mathbf{a})$. In that case, $(\mathbf{x}+\mathbf{a})^{\top}\nabla_{\mathbf{x}}\sigma(\mathbf{x},\mathbf{y}) = \alpha\sigma(\mathbf{x},\mathbf{y})$, where α is the degree of homogeneity (a constant), and the proof goes through as above.

B5. Proof of Corollary 1

The proof follows from theorem 3. If for a given y_j, $\partial H_j(y|\mathbf{x})/\partial x_k < 0$ for all $k \in \{1,\cdots,X\}\backslash\{K\}$ (PAM), then in order for $(\mathbf{x}+\mathbf{a})^{\top}\nabla H_j(y|\mathbf{x}) = 0$ to hold it must be that $\partial H_j(y|\mathbf{x})/\partial x_K > 0$ (NAM). QED

Remark.—In order to streamline our results in the text, we stated this corollary for the case of $Y = X = 2$. But it readily applies to the case of $Y \geq 2$, $X \geq 2$, as the proof shows.

B6. Proof of Theorem 4

Suppose that there exist three twice-differentiable functions $\tilde{\sigma}:\mathbb{R}^2 \to \mathbb{R}$, $I:\mathbb{R}_+^X \to \mathbb{R}_+$, and $J:\mathbb{R}_+^Y \to \mathbb{R}_+$ such that for all $(\mathbf{x},\mathbf{y}) \in \mathcal{X}\times\mathcal{Y}$, $\sigma(\mathbf{x},\mathbf{y}) = \tilde{\sigma}(I(\mathbf{x}),J(\mathbf{y}))$. Then, differentiating the latter identity with respect to (x_k, y_j) and recalling that the surplus function σ is a bilinear form, one has

$$\frac{\partial^2\tilde{\sigma}}{\partial I\partial J}\cdot\frac{\partial I}{\partial x_k}\cdot\frac{\partial J}{\partial y_j} = q_{kj}.$$

This implies that for all $k \neq k'$ and $j \neq j'$, $q_{kj}q_{k'j'} - q_{k'j}q_{kj'} = 0$; that is, the determinants of all 2×2 submatrices of $\mathbf{Q}$ are zero, which in turn implies that the determinant of the largest square submatrix of $\mathbf{Q}$ is zero. Hence, under the single-index representation the matrix $\mathbf{Q}$ has rank of at most 1.

Conversely, if $\mathbf{Q}$ has rank 1, then it can be written as $\mathbf{Q} = \mathbf{q}_{\mathbf{x}}\mathbf{q}_{\mathbf{y}}^{\top}$, where $\mathbf{q}_{\mathbf{x}} \in \mathbb{R}^X$ and $\mathbf{q}_{\mathbf{y}} \in \mathbb{R}^Y$ are column vectors, and the surplus function becomes $\sigma(\mathbf{x},\mathbf{y}) = [\mathbf{q}_{\mathbf{x}}^{\top}(\mathbf{x}+\mathbf{a})]^{\top}(\mathbf{q}_{\mathbf{y}}^{\top}\mathbf{y})$, a function of the scalar indices $x = \mathbf{q}_{\mathbf{x}}^{\top}\mathbf{x}$ and $y = \mathbf{q}_{\mathbf{y}}^{\top}\mathbf{y}$. QED

Remark.—In the two-dimensional case, $\mathbf{Q}$ having rank at most 1 is equivalent to $\det\mathbf{Q} = 0$, which in turn is equivalent to the single-crossing condition failing to hold.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. “High Wage Workers and High Wage Firms.” *Econometrica* 67 (2): 251–333.

Bagger, Jesper, and Rasmus Lentz. 2019. “An Empirical Model of Wage Dispersion with Sorting.” *Rev. Econ. Studies* 86 (1): 153–90.

Becker, Gary S. 1973. “A Theory of Marriage: Part I.” *J.P.E.* 81 (4): 813–46.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. “A Distributional Framework for Matched Employer Employee Data.” *Econometrica* 87 (3): 699–739.

Brouwer L. E. J. 1912. “Beweis der Invarianz des *n*-dimensionalen Gebiets.” *Mathematische Annalen* 71 (3): 305–15.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size and Unemployment." *Internat. Econ. Rev.* 39 (2): 257–73.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence." *Econometrica* 74 (2): 323–64.

Cantor, G. 1878. "Ein Beitrag zur Mannigfaltigkeitslehre." *J. reine und angewandte Math.* 84:242–58.

Chade, Hector. 2006. "Matching with Noise and the Acceptance Curse." *J. Econ. Theory* 129:81–113.

Chade, Hector, Jan Eeckhout, and Lones Smith. 2017. "Sorting through Search and Matching Models in Economics." *J. Econ. Literature* 55 (2): 493–544.

Eeckhout, Jan, and Philipp Kircher. 2010. "Sorting and Decentralized Price Competition." *Econometrica* 78 (2): 539–74.

Gautier, Pieter A., Coen N. Teulings, and Aico van Vuuren. 2006. "Labor Market Search with Two-Sided Heterogeneity: Hierarchical versus Circular Models." In *Structural Models of Wage and Employment Dynamics* (Contributions to Economic Analysis, vol. 275), edited by Henning Bunzel, Bent Christensen, George R. Neumann, and Jean-Marc Robin, 117–32. Bingley: Emerald.

———. 2010. "On-the-Job Search, Mismatch and Efficiency." *Rev. Econ. Studies* 77 (1): 245–72.

Guvenen, Fatih, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. 2020. "Multidimensional Skill Mismatch." *American Econ. J. Macroeconomics* 12 (1): 210–44.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovksii. 2017. "Identifying Equilibrium Models of Labor Market Sorting." *Econometrica* 85 (1): 29–65.

Heckman, James, and Jose Scheinkman. 1987. "The Importance of Bundling in a Gorman-Lancaster Model of Earnings." *Rev. Econ. Studies* 54 (2): 243–55.

Heckman, James J., and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *J.P.E.* 93 (6): 1077–125.

Karlin, Samuel, and Yosef Rinott. 1980. "Classes of Orderings of Measures and Related Correlation Inequalities. I. Multivariate Totally Positive Distributions." *J. Multivariate Analysis* 10 (4): 467–98.

Lamadon, Thibaut, Jeremy Lise, Costas Meghir, and Jean-Marc Robin. 2014. "Matching, Sorting, Firm Productivity and Wages." Manuscript, Univ. Coll. London.

Legros, Patrick, and Andrew F. Newman. 2007. "Beauty Is a Beast, Frog Is a Prince: Assortative Matching with Nontransferabilities." *Econometrica* 75 (4): 1073–102.

Lentz, Rasmus. 2010. "Sorting by Search Intensity." *J. Econ. Theory* 145 (4): 1436–52.

Lindenlaub, Ilse. 2017. "Sorting Multidimensional Types: Theory and Application." *Rev. Econ. Studies* 84 (2): 718–89.

Lindenlaub, Ilse, and Fabien Postel-Vinay. 2022. "The Worker-Job Surplus." Manuscript, Univ. Coll. London and Yale Univ.

Lise, Jeremy, and Fabien Postel-Vinay. 2020. "Multidimensional Skills, Sorting, and Human Capital Accumulation." *A.E.R.* 110 (8): 2328–76.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Rev. Econ. Studies* 61 (3): 397–415.

Moscarini, Giuseppe. 2001. "Excess Worker Reallocation." *Rev. Econ. Studies* 68 (3): 593–612.

Moscarini, Giuseppe, and Fabien Postel-Vinay. 2013. "Stochastic Search Equilibrium." *Rev. Econ. Studies* 80 (4): 1545–81.

Netto, E. 1879. "Beitrag zur Mannigfaltigkeitslehre." *J. reine und angewandte Math.* 86:263–68.

Papageorgiou, Theodore. 2014. “Learning Your Comparative Advantages.” *Rev. Econ. Studies* 81 (3): 1263–95.

Peano, G. 1890. “Sur une courbe, qui remplit toute une aire plane.” *Mathematische Annalen* 36 (1): 157–60.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity.” *Econometrica* 70 (6): 2295–350.

Roy, A. D. 1951. “Some Thoughts on the Distribution of Earnings.” *Oxford Econ. Papers* 3 (2): 135–46.

Sanders, Carl. 2012. “Skill Uncertainty, Skill Accumulation, and Occupational Choice.” Manuscript, Washington Univ., St. Louis.

Shapiro, A., Darinka Dentcheva, and Andrzej Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*, MPS-SIAM (Math. Programming Soc.–Soc. Indus. and Appl. Math.) series on optimization, no. 9. Philadelphia: SIAM.

Shimer, Robert, and Lones Smith. 2000. “Assortative Matching and Search.” *Econometrica* 68 (2): 343–69.

Smith, Lones. 2006. “The Marriage Model with Search Frictions.” *J.P.E.* 114 (6): 1124–46.

Sorkin, Isaac. 2018. “Ranking Firms Using Revealed Preference.” *Q.J.E.* 133 (3): 1331–93.

Taber, Christopher, and Rune Vejlin. 2020. “Estimation of a Roy/Search/Compensating Differential Model of the Labor Market.” *Econometrica* 88 (3): 1031–69.

Tao, Terence. 2011. “Brouwer’s Fixed Point and Invariance of Domain Theorems, and Hilbert’s Fifth Problem.” https://terrytao.wordpress.com/tag/invariance-of-domain/.

Tinbergen, Jan. 1956. “On the Theory of Income Distribution.” *Weltwirtschaftliches Archiv* 77:155–75.

Yamaguchi, Shintaro. 2012. “Tasks and Heterogeneous Human Capital.” *J. Labor Econ.* 30 (1): 1–53.