Multidimensional Sorting Under Random Search*

Ilse Lindenlaub† Fabien Postel-Vinay‡

June 2016 — Preliminary

Abstract

We analyze *sorting* in a standard market environment characterized by search frictions and random search, but where both workers and jobs have *multi-dimensional* characteristics. We first offer a definition of multi-dimensional positive (and negative) assortative matching in this frictional environment. According to this notion, matching is positive assortative if a more skilled worker in a certain dimension is matched to a *distribution* of jobs that first-order stochastically dominates that of a less skilled worker. We then provide conditions on the primitives of this economy (technology and distributions) under which positive sorting obtains in equilibrium. We show that in several environments of interest, the main restriction on the primitives is a *single-crossing* condition of the technology, although in general further restrictions on type distributions are needed. Guided by our theoretical framework, we conduct simulation exercises to quantify the errors regarding sorting, mismatch and policy that are made by wrongly assuming that heterogeneity is one-dimensional when it is really multi-dimensional.

Keywords. Multidimensional Heterogeneity, Random Search, Sorting, Assortative Matching

*We would like to thank Jan Eeckhout for useful suggestions at early stages of this project, as well as seminar audiences at Essex, Yale, and the Chicago Fed for insightful comments. All remaining errors are ours.

†Yale University, **Address:** Department of Economics, Yale University, 28 Hillhouse Avenue, New Haven, 06520, US. **Email:** `ilse.lindenlaub@yale.edu`

‡UCL, IFS and Sciences Po. **Address:** Department of Economics, University College London, Drayton House, 30 Gordon Street, London WC1H 0AX, UK. **Email:** `f.postel-vinay@ucl.ac.uk`

1 Introduction

Random search models have become one of the main workhorses of the applied literature on individual wages, job assignment and mismatch between heterogeneous workers and jobs. A standard assumption in these models is that agents are characterized by *one-dimensional* heterogeneity: for instance workers differ in ability and jobs in productivity.[1] This restriction to one-dimensional heterogeneity is at odds with the simple observation that typical data sets describe both workers and jobs in terms of many different productive attributes (e.g. cognitive skills, manual skills, health, or psychometric scores for workers, and task-specific skill requirements for jobs).[2]

The aim of this paper is to analyze the assignment (or *sorting*) of workers into jobs when both workers and jobs differ along several dimensions. Investigating multi-dimensional sorting in the broad class of random search models used in the applied literature is important: as we show in this paper, approximating workers' and jobs' true multi-dimensional characteristics by one-dimensional summary indices when taking those models to the data may lead to sizeable quantitative errors and misguided policy recommendations.

We develop a theoretical framework for the analysis of multi-dimensional sorting under random search. Our environment is that of a standard random search model, except for workers and jobs being endowed with *vectors* of productive attributes, $\mathbf{x} = (x_1, \cdots, x_X)$ for workers ant $\mathbf{y} = (y_1, \cdots, y_Y)$ for jobs. Employed and unemployed workers receive job offers drawn at random from an exogenous sampling distribution of job attributes. Utility is fully transferable: workers and firms are joint surplus maximizers. The fact that agents base their decision whether to accept a job on a *scalar* value (i.e. the match surplus that summarizes all underlying multi-dimensional heterogeneity) is key to the tractability of our multi-dimensional problem.

In order to be able to describe and interpret the assignment patterns that arise in equilibrium, we begin by offering notions of positive assortative matching (PAM) and negative assortative matching (NAM) in this environment. Our proposed notion is based on first-order stochastic dominance ordering of the marginal distributions of job attributes across workers with different skills: if a worker with a higher endowment of some skill x_k is matched to jobs with 'better' (in

[1]A recent exception is the applied paper by Lise and Postel-Vinay (2015), who focus on the accumulation of skills along various dimensions within a model that can otherwise be seen as a special case of the theoretical framework we develop here.

[2]Beyond search models, a growing applied literature takes explicit account of these multiple dimensions of productive heterogeneity. Recent examples include Yamaguchi (2012), Sanders (2012), and Guvenen et al. (2015).

the first-order stochastic dominance sense) attributes y_j, then PAM occurs in dimension (x_k, y_j). It is important to note that sorting is thus defined dimension by dimension, meaning that PAM can arise in one particular dimension while NAM occurs in another.

Using this definition of sorting, we present three main sets of results. The first one is about the *sign of sorting*: we provide conditions on the economy's primitives under which positive (or negative) sorting arises in equilibrium. For ease of exposition and clarity of interpretation, in much of our analysis we focus on a baseline setting that features bilinear technology, two-dimensional heterogeneity on the job side, sequential auction wage setting, and in which employment in any job is always preferable to unemployment (i.e. all possible matches generate positive surplus) — all assumptions that we can and will relax. In this baseline case, we find that matching in, say, dimension (x_1, y_1) is positive assortative if and only if the technology satisfies a *single crossing condition* implying that the complementarity between worker skill x_1 and job attribute y_1 dominates complementarity in the competing dimension (x_1, y_2). This condition is *distribution-free*: it only involves restrictions on the production technology.

We extend the analysis to more general cases where (1) not all possible matches generate positive surplus (implying that there is an active nonemployment-to-employment margin), (2) heterogeneity on the job side is of dimension higher than two, and (3) the technology is monotone in at least one job attribute but not necessarily bilinear. We provide characterizations of sorting in these more general environments but, besides single-crossing of the technology, the conditions for sorting involve complex interactions between the technology and sampling distribution of jobs. We also show that these results do not hinge on the sequential auctions wage setting but hold for several other commonly used wage setting protocols like Nash bargaining, sequential auctions with worker bargaining power, and wage posting.

Our second set of results shows that our model predicts sorting on *specialization* rather than absolute advantage. This arises naturally with multi-dimensional skills because workers with different skill bundles do not rank jobs in the same way. As a consequence, uniformly more skilled workers 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, but with a lower requirement in the other skill.

Our third set of results are quantitative. We simulate data from a two-dimensional model with bilinear technology that complies with our theory. We then fit a misspecified one-dimensional model to those data and compare its predictions on complementarities in technology, sorting and mismatch to the true two-dimensional model. We find that the misspecified model can suggest fallacious conclusions, especially when the true technology features complementarities between worker and job attributes in some dimensions and substitutabilities in others. In such settings, the one-dimensional model produces sign-varying estimates of the complementarities in production as well as sorting patterns that bear little resemblance to the truth. As a consequence, the one-dimensional model over-estimates mismatch (measured by the aggregate output difference between frictional and frictionless first-best allocation) and suggests a first-best allocation that is very far from the true first-best under multi-dimensional types. Implementing the (fallacious) first-best allocation suggested by the misspecified one-dimensional model can cause sizeable welfare losses.

From those results we draw four main conclusions. First, sorting arises in our model *only* due to multi-dimensional heterogeneity. In a comparable one-dimensional model, where match surplus $\sigma(x, y)$ depends on scalar worker type x and job type y and is increasing in y, workers all rank jobs in the same way, regardless of their own type x: their common strategy is to accept any job with a higher y than their current one. This common strategy rules out sorting. In contrast, in the multi-dimensional world where every worker is endowed with a skill *bundle*, what matters is not just to match with a productive job in *any* dimension, but also to match with a job requiring much of the skill in which the worker is relatively strong. Thus, workers with different skill bundles accept and reject different types of jobs, which is why sorting arises.

Second, in all of the environments that we study, the central force toward sorting is an intuitive single-crossing condition on the technology, guaranteeing strong complementarities between worker and job attributes. This holds true independent of whether the conditions for sorting also involve restrictions on the sampling distribution or not.

Third, and contrary to well-known results on one-dimensional sorting both in frictionless and/or frictional environments, the conditions for multi-dimensional sorting are generally *not* distribution-free. In particular, when jobs have more than two characteristics, the conditions for sorting involve interactions between technology and the sampling distribution of jobs.

Last, our results have important implications for applied work. In our simulation exercises, we show that one can make substantial qualitative and quantitative errors by assuming that the data is one-dimensional when it is in fact multi-dimensional.

While much is known about sorting and conditions under which it obtains under *one*-dimensional heterogeneity with and without frictions, little is known about sorting on *multi*-

dimensional types, especially in frictional environments:[3] To the best of our knowledge, this paper is the first to analyze *multi*-dimensional sorting under *random* search — an environment of great importance for applied work. Perhaps most importantly, we show that accounting for multi-dimensional heterogeneity is fundamental to the analysis of sorting: collapsing multi-dimensional job and worker attributes to a single summary index when the true data is multi-dimensional significantly distorts both the qualitative and quantitative conclusions about sorting and mismatch.

The rest of the paper is organized as follows: Section 2 introduces our model. Section 3 provides a definition of sorting in the multi-dimensional space under random search. Section 4 contains our main results on the sign of sorting, which we first establish within our baseline bilinear setting (4.1), then extend to more general cases (4.2). Section 5 investigates sorting on absolute advantage vs specialization. Section 6 contains our simulation exercise. Section 7 places the contribution of this paper into the literature and Section 8 concludes.

2 The Model

2.1 The Environment.

Time is continuous. The economy is populated by infinitely lived, forward looking workers and firms. There is a fixed unit mass of workers that are characterized by time-invariant skill bundles $\mathbf{x} = (x_1, \cdots, x_X) \in \mathcal{X} \subset \mathbb{R}_+^X$ (X denotes the number of skills in the workers' skill bundle), drawn from an exogenous distribution $L(\cdot)$, with density $\ell(\cdot)$. Without loss, we normalize the lower support of worker skills to 0 in every dimension. Firms can either be thought of as single jobs (possibly vacant), or as collections of independent, perfectly substitutable jobs. Jobs are characterized by a vector of productive attributes, or "skill requirements" $\mathbf{y} = (y_1, \cdots, y_Y) \in \mathcal{Y} = \bigtimes_{j=1}^{Y}[\underline{y}_j, \overline{y}_j]$, where Y denotes the number of different job attributes and where $\underline{y}_j \in \mathbb{R}_+$ and $\overline{y}_j \in \mathbb{R}_+ \cup \{+\infty\}$.[4] Job attributes are also time-invariant and are drawn from some distribution $\Gamma(\cdot)$, with density $\gamma(\cdot)$. We assume that γ has strictly positive mass over its entire support, $\operatorname{Supp} \gamma = \mathcal{Y}$.[5]

[3]For frictionless sorting under transferable utility, see Becker (1973), for frictionless sorting under non-transferable utility, see Legros and Newman (2007), for sorting under random search and transferable utility, see Shimer and Smith (2000) and under non-transferable utility Smith (2006). See Lindenlaub (2014) for a TU framework of multi-dimensional sorting *without* frictions. We will discuss the related literature in detail below.

[4]The restriction to $\mathbf{x}$ and $\mathbf{y}$ having nonnegative elements is not strictly necessary for the analysis. It only makes some of the economic interpretations more natural.

[5]The purpose of this assumption is to ensure 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. In particular, this restriction

We denote a generic skill from the worker's skill bundle by x_k where $k \in \{1, \cdots, X\}$ and a generic skill requirement by y_j where $j \in \{1, \cdots, Y\}$. The output flow associated with a match between a worker with skills $\mathbf{x}$ and a job with skill requirements $\mathbf{y}$ (a type-$(\mathbf{x}, \mathbf{y})$ match) is $f(\mathbf{x}, \mathbf{y})$, where $f : \mathbb{R}^X \times \mathbb{R}^Y \longrightarrow \mathbb{R}$.[6] The income flow generated by a nonemployed worker is denoted $b(\mathbf{x})$.

Lindenlaub (2014) develops a multi-dimensional matching model that is static and frictionless. In this paper, we examine the properties of a multi-dimensional dynamic model when the labor market is affected by search frictions and workers search for jobs at random, both off and on the job. Workers can be matched to a job or be unemployed. If matched, they lose their job at rate δ, and sample alternate job offers with requirements drawn from the fixed sampling distribution Γ, at Poisson rate λ_1. Unemployed workers sample job offers from the same sampling distribution at rate λ_0. 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 other workers. As such, this set-up is really a (partial equilibrium) model of the labor market rather than one of a symmetric, one-to-one matching market such as the marriage market.

2.2 Rent Sharing and Value Functions

Preferences are linear and firms and workers have equal time discounting rate ρ. Under those assumptions, the total present discounted value of a match between a type-$\mathbf{x}$ worker and a type-$\mathbf{y}$ firm is independent of the way in which it is shared, and only depends on match attributes $(\mathbf{x}, \mathbf{y})$. We denote this value by $P(\mathbf{x}, \mathbf{y})$. We further denote the value of unemployment by $U(\mathbf{x})$, and the worker's value of his current wage contract by 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 (which would arise under free entry and exit of vacancies on the search market), 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 body of this paper that wage contracts are renegotiated sequentially by mutual agreement, as in the sequential auction model of Postel-Vinay and Robin (2002). This is (mostly) for simplicity of exposition, and we show in Appendix B that the vast majority

is needed for several results when $Y \geq 3$, as will become clear in the proofs of Theorems 6 and 7.

[6] Note that we require that the production function be 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 requirement is there to streamline some of the arguments in the proofs and can be relaxed. Details are available upon request.

of our results extend to other popular wage setting models, notably 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, workers have the possibility of playing off their current employer against any firm from which they receive an outside offer. If they do so, the current and outside employers Bertrand-compete over the worker's services. Consider a type-$\mathbf{x}$ worker employed at a type-$\mathbf{y}$ firm and assume that the worker receives an outside offer from a firm of type $\mathbf{y}'$. Bertrand competition between the type-$\mathbf{y}$ and type-$\mathbf{y}'$ employers implies that the worker ends up in the match that has higher total value — that is, 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 — with a new wage contract worth $W' = \min\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\}$.

Under this rent-sharing protocol, the total value of a match between a type-$\mathbf{x}$ worker and a type-$\mathbf{y}$ firm, $P(\mathbf{x},\mathbf{y})$, solves:

$$\rho P(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) + \delta\left[U(\mathbf{x}) - P(\mathbf{x},\mathbf{y})\right].$$

The annuity value of the match, $\rho P(\mathbf{x},\mathbf{y})$, equals the output flow $f(\mathbf{x},\mathbf{y})$ plus the capital loss $[U(\mathbf{x}) - P(\mathbf{x},\mathbf{y})]$ of the firm-worker pair if the job is destroyed (with flow probability δ).[7]

Match surplus $P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})$ thus solves $(\rho+\delta)\left[P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})\right] = f(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x})$. In what follows, we will mostly reason in terms of the match *flow surplus*:

$$\sigma(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x}).$$

As we just saw, a worker's decision to accept or reject a job offer hinges on comparisons of match surpluses: a type-$\mathbf{x}$ worker employed in a type-$\mathbf{y}$ job accepts an offer from a type-$\mathbf{y}'$ job 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 mobility decisions are based on the comparison of flow surpluses.[8]

Finally note that, in the sequential auction case, the value of unemployment, $U(\mathbf{x})$, is simply

[7] Note that, under the sequential auction model, the realization the "other" risk that the firm-worker pair faces, 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 the match, in which case the continuation value of the match is still $P(\mathbf{x},\mathbf{y})$, or the worker accepts the offer and leaves the match, 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 $P(\mathbf{x},\mathbf{y})$.

[8] Likewise, a type-$\mathbf{x}$ unemployed worker accepts any type-$\mathbf{y}$ offer such that $\sigma(\mathbf{x},\mathbf{y}) \geq 0$.

given by $\rho U(\mathbf{x}) = b(\mathbf{x})$, implying $\sigma(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) - b(\mathbf{x})$.

2.3 Steady-state Distribution of Skills and Skill Requirements.

Let $h(\mathbf{x}, \mathbf{y})$ denote the measure of type-$(\mathbf{x}, \mathbf{y})$ matches. This is determined in steady state by the following flow-balance equation:

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

where $u(\mathbf{x})$ is the measure of type-$\mathbf{x}$ unemployed workers in the economy. The l.h.s. of (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 following receipt of a dominant outside offer by the worker. The flow probability of this latter event is $\lambda_1 \mathbb{E}\left[\mathbf{1}\left\{\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})\right\}\right]$, 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 with the worker than the current type-$\mathbf{y}$ job. The r.h.s. of (1) is the inflow into the stock of type-$(\mathbf{x}, \mathbf{y})$ matches. It is made up of those of the $u(\mathbf{x})$ unemployed type-$\mathbf{x}$ workers who draw a type-$\mathbf{y}$ job (flow probability $\lambda_0 \gamma(\mathbf{y})$) and accept it if the associated flow surplus is positive ($1\left\{\sigma(\mathbf{x}, \mathbf{y}) \geq 0\right\}$), plus those of the $h(\mathbf{x}, \mathbf{y}')$ type-$\mathbf{x}$ workers employed in any type-$\mathbf{y}'$ job who draw an offer from a type-$\mathbf{y}$ job (flow probability $\lambda_1 \gamma(\mathbf{y})$), which they accept 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 a similar (and similarly interpreted) flow-balance equation:

$$\lambda_0 \mathbb{E}\left[\mathbf{1}\left\{\sigma(\mathbf{x}, \mathbf{y}) > 0)\right\}\right] u(\mathbf{x}) = \delta \int h(\mathbf{x}, \mathbf{y}') d\mathbf{y}'. \quad (2)$$

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

The following important remarks can be made at this point: the *acceptance rule* of an offer received by a worker in a type-$(\mathbf{x}, \mathbf{y})$ match hinges on the comparison of two *scalar* random variables, $\sigma(\mathbf{x}, \mathbf{y}')$ and $\sigma(\mathbf{x}, \mathbf{y})$, despite the underlying multi-dimensional heterogeneity of workers and firms. It is therefore convenient to introduce the conditional sampling distribution $F_{\sigma|\mathbf{x}}$ of $\sigma(\mathbf{x}, \mathbf{y})$, given $\mathbf{x}$. With this notation, the acceptance probability for an employed worker writes as $\mathbb{E}\left[\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')>\sigma(\mathbf{x},\mathbf{y})\right\}\right]=\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)$.[9] The acceptance probability of an unemployed worker is similar: $\mathbb{E}\left[\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')>0\right\}\right]=\overline{F}_{\sigma|\mathbf{x}}(0)$.

Substituting these elements into (1), we show in the Appendix that the matching distribution $h(\mathbf{x},\mathbf{y})$ has the following closed-form:

$$\frac{h(\mathbf{x},\mathbf{y})}{\ell(\mathbf{x})\gamma(\mathbf{y})}=\frac{\delta\lambda_0\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y})\geq 0\right\}}{\delta+\lambda_0\overline{F}_{\sigma|\mathbf{x}}(0)}\times\frac{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}.$$

This equation also implies that the equilibrium conditional distribution of job types $\mathbf{y}$ given worker types $\mathbf{x}$ among employed workers is given by:

$$h\left(\mathbf{y}|\mathbf{x}\right)=\frac{\delta\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y})\geq 0\right\}}{\overline{F}_{\sigma|\mathbf{x}}(0)}\times\frac{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]\gamma(\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}=\frac{G'_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}{F'_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}\times\gamma(\mathbf{y}), \tag{3}$$

where for any $s\in\mathbb{R}$:

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

is the steady-state cross-section distribution of flow surplus among employed workers of type $\mathbf{x}$.

3 Equilibrium Sorting

3.1 Measuring Sorting

Lindenlaub's (2014) criterion for multi-dimensional assortative matching is that the Jacobian matrix of the equilibrium matching be a P-matrix. This criterion captures the way in which a worker's job type $\mathbf{y}$ improves or deteriorates as one varies the worker's skill bundle $\mathbf{x}$ when matching is *pure*, i.e. when 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. A natural extension of this measure of sorting to our environment is to consider changes in the quantiles of the conditional *distribution* of job types as one varies worker type $\mathbf{x}$. Formally, let $H_j(y|\mathbf{x})$ denote the c.d.f. of the marginal distribution of y_j (the jth component of the vector of job attributes $\mathbf{y}$) matched to employed workers with

[9]Throughout the paper, we use the upper bar to denote survivor functions: $\overline{F}(\cdot):=1-F(\cdot)$ for any cdf F.

skill bundle $\mathbf{x}$. Using (3), we can express this as

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

To analyze sorting, we will be interested in the gradient of $H_j(y|\mathbf{x})$, i.e.:

$$\nabla H_j(y|\mathbf{x}) = \left(\frac{\partial H_j(y|\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial H_j(y|\mathbf{x})}{\partial x_X}\right)^\top \tag{5}$$

for $j \in \{1, \cdots, Y\}$. A situation of particular interest is when one of the components of this gradient, $\partial H_j(y|\mathbf{x})/\partial x_k$, has a constant sign over the support of γ. If that sign is negative [positive], then $H_j(\cdot|\mathbf{x})$ is increasing [decreasing] in x_k in the sense of first-order stochastic dominance (FOSD): a strong form of positive [negative] assortative matching then occurs in dimension (x_k, y_j), as a worker with higher type-k skills is matched to a jobs with greater type-j skill requirements (in the FOSD sense) than a worker with lower type-k skills. We will thus use the following formal definition of sorting:

Definition 1 (Positive and Negative Assortative Matching)**.** *If $\partial H_j(y|\mathbf{x})/\partial x_k$ is negative (positive), then matching is positive (negative) assortative in dimension (y_j, x_k).*

We will use the acronyms PAM and NAM for positive and negative assortative matching. Alternatively, we will also refer to PAM (or NAM) as *positive (or negative) sorting*. To avoid duplication of some results, we focus on positive sorting throughout the paper, bearing in mind that the results stated below can easily be adjusted to the case of negative sorting.

Finally, for future use, we define the joint distribution of job attribute j and match flow surplus s, conditional on worker type $\mathbf{x}$, in the population of employed workers, whose c.d.f. is:

$$\begin{aligned} K_j(y, s|\mathbf{x}) &= \int \mathbf{1}\{y_j \le y\} \times \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}) \le s\}\, h(\mathbf{y}|\mathbf{x})\, d\mathbf{y} \\ &= \frac{\delta\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \int \frac{\mathbf{1}\{0 \le \sigma(\mathbf{x},\mathbf{y}) \le s\} \times \mathbf{1}\{y_j \le y\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right]^2} \gamma(\mathbf{y}) d\mathbf{y} \end{aligned} \tag{6}$$

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 marginals of $K_j(y, s|\mathbf{x})$, so that $K_j(y, +\infty|\mathbf{x}) = H_j(y|\mathbf{x})$ and $K_j(+\infty, s|\mathbf{x}) = G_{\sigma|\mathbf{x}}(s)$.

Further note that:

$$\frac{\partial K_j(y,s|\mathbf{x})}{\partial s} = G'_{\sigma|\mathbf{x}}(s) \times \Pr_\Gamma\left\{y'_j \leq y|\sigma(\mathbf{x},\mathbf{y}') = s\right\},$$

where $\Pr_\Gamma\{A\}$ is used to denote the probability of A occurring following a random draw of a job type $\mathbf{y}$ from the sampling distribution γ.

3.2 A Decomposition Result

We begin our analysis by showing how equilibrium sorting can be usefully decomposed into sorting on the nonemployment-to-employment (NE) margin and on the employment-to-employment (EE) margin.

Theorem 1. *For any $\mathbf{x} \in \mathcal{X}$ and $y \in \mathbb{R}$:*

$$\begin{aligned}\frac{\partial H_j(y|\mathbf{x})}{\partial x_k} = & \underbrace{G'_{\sigma|\mathbf{x}}(0)\Bigg\{\Pr_\Gamma\left\{y'_j \leq y|\sigma(\mathbf{x},\mathbf{y}')=0\right\}\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=0, y'_j \leq y\right] - H_j(y|\mathbf{x})\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=0\right]\Bigg\}}_{\text{(1): NE margin}} \\ & \underbrace{+\int_0^{+\infty}\frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)} \times \frac{\partial K_j(y,s|\mathbf{x})}{\partial s} \times\Bigg\{\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s, y'_j \leq y\right] - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right]\Bigg\}ds.}_{\text{(2): EE margin}}\end{aligned}$$

Theorem 1 offers a decomposition of the typical element of the gradient of $H_j(y|\mathbf{x})$, which we use to characterize sorting patterns (Definition 1).[10] It highlights the fact that a marginal increase in the worker's skill x_k affects the equilibrium distribution of job types to which this worker is matched in two ways.

First, a marginal increase in skill x_k affects the boundary of the set of profitable matches for that worker, i.e. the set of job types $\mathbf{y}$ such that $\sigma(\mathbf{x},\mathbf{y}) \geq 0$. An increase in skill may render some matches between *unemployed* workers and jobs profitable that were unprofitable prior to this change. This is reflected in the first term of the expression above. This first effect

[10] Two important technical notes: first, here and in the rest of the paper, we use the notation $\mathbb{E}_\Gamma$ to distinguish expectations taken w.r.t. the sampling distribution Γ from expectations in the equilibrium distribution of matches, which we simply denote by $\mathbb{E}$. Second, it may be that the joint event $\left(\sigma(\mathbf{x},\mathbf{y}')=s, y'_j \leq y\right)$ on which some of the expectations in Theorem 1 are conditioned have zero probability in γ. As explained in the proof of Theorem 1, we set expectations conditional on zero-probability events to zero by convention.

only works through selection on the NE margin: the first term in Theorem 1 is multiplied by the density of marginally profitable matches for type-$\mathbf{x}$ workers, $G'_{\sigma|\mathbf{x}}(0)$. If the worker's type $\mathbf{x}$ is such that $\sigma(\mathbf{x},\mathbf{y}) > 0$ for all job types $\mathbf{y}$ (i.e. if the worker accepts *any* job type when unemployed), then there are no such marginal matches ($G'_{\sigma|\mathbf{x}}(0) = 0$), and this selection on the NE margin is shut down.

Second, a marginal increase in x_k affects the sampling distribution of match surplus, $F_{\sigma|\mathbf{x}}(\cdot)$ for *employed* workers as well. More specifically, an increase in x_k changes the terms of the comparison between any two potential matches involving the worker: for any two job types $(\mathbf{y},\mathbf{y}')$, the difference $\sigma(\mathbf{x},\mathbf{y}') - \sigma(\mathbf{x},\mathbf{y})$ varies with x_k. This, in turn, changes the way employed workers reallocate between jobs through on-the-job search. This effect, reflected in the second term of the expression in Theorem 1, operates through selection on the EE margin.

4 The Sign of Sorting

While Theorem 1 affords a clear decomposition of sorting on the NE and EE margins, unfortunately those two effects involve complex interactions between the technology $\sigma(\mathbf{x},\mathbf{y})$ and the sampling distribution of job types $\gamma(\mathbf{y})$ and cannot easily be signed under general assumptions. In order to make progress towards a characterization of the sign of sorting, we start the analysis with a special case in which we can derive clean and (with one exception) distribution-free conditions for positive sorting to occur in equilibrium. We then investigate generalizations.

4.1 The Case of Bilinear Technology in Two Dimensions

4.1.1 Assumptions

The following two additional assumptions simplify the decomposition in Theorem 1 considerably and will afford distribution-free and intuitive sorting conditions.

Assumption 1. *(a) The production function $f(\mathbf{x},\mathbf{y})$ is bilinear in worker skills and job attributes:*

$$f(\mathbf{x},\mathbf{y}) = (\mathbf{x}+\mathbf{a})^\top \mathbf{Q}\mathbf{y} = \sum_{i=1}^{X}\sum_{j=1}^{Y} q_{ij}(x_i + a_i) y_j$$

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

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

$$b(\mathbf{x}) = (\mathbf{x} + \mathbf{a})^\top \mathbf{Q} \mathbf{b} = \sum_{i=1}^{X} \sum_{j=1}^{Y} q_{ij}(x_i + a_i) b_j$$

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

(c) for all $\mathbf{x} \in \mathcal{X}$, there exists $j \in \{1, \cdots, Y\}$ such that $q_j(\mathbf{x}) := \sum_{i=1}^{X} q_{ij}(x_i + a_i) > 0$; to fix the notation, we will assume w.l.o.g. that $q_Y(\mathbf{x}) > 0$.

Assumptions 1.a-b 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:

$$\sigma(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) - b(\mathbf{x}) = (\mathbf{x} + \mathbf{a})^\top \mathbf{Q}(\mathbf{y} - \mathbf{b}) = \sum_{i=1}^{X} \sum_{j=1}^{Y} q_{ij}(x_i + a_i)(y_j - b_j)$$

The $X \times Y$ technology matrix $\mathbf{Q}$ captures the complementarity structure between job and worker characteristics relating to all types of tasks, $(i, j) \in \{1, \cdots, X\} \times \{1, \cdots, Y\}$, and will be crucial to our analysis of sorting. We interpret the vector $\mathbf{b}$ as the production technology workers have access to when nonemployed. In turn, the vector $\mathbf{a}$ — or, more precisely, the vector $\mathbf{a}^\top \mathbf{Q}$ — is a technological parameter reflecting the "intrinsic returns" on job attributes $\mathbf{y}$, in the sense that a marginal increase dy_j in any job attribute y_j contributes a fixed amount $\left(\sum_{i=1}^{X} a_i q_{ij}\right) dy_j$ to job productivity regardless of the matched worker type. Alternatively, $\mathbf{a}$ can be interpreted 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 will therefore refer to $\mathbf{a}$ as the *baseline productivity vector*, which we assume to be nonnegative (Assumption 1.a). This ensures that the worker's total input into production, $\mathbf{x} + \mathbf{a}$, is nonnegative in all dimensions (remember that $\mathbf{x} \in \mathbb{R}_+^X$). 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 1.c ensures that, for any level of worker skills, there is at least one job attribute that impacts output positively.[11] Note that we do not impose increasing monotonicity of the production function in *all* of the job attributes. Nor do we restrict the monotonicity of the production or flow surplus function $\sigma(\mathbf{x}, \mathbf{y})$ in worker skills $\mathbf{x}$.

Next, we consider:

[11] Because all of the results stated below are "local" (in the sense that they hold in a neighborhood of a given skill bundle $\mathbf{x}$), the set of j's such that $q_j(\mathbf{x}) > 0$ needs not be the same for all $\mathbf{x}$. Yet, for notational convenience, we relabel job attributes such that $j = Y$ is always in that set.

Assumption 2. *Each job has $Y = 2$ attributes, i.e.* $\mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^2_+$.

The sorting results established in this subsection will rely on Assumptions 1 and 2. In the next Subsection, we provide generalizations of our results on the sign of sorting to other surplus functions and to higher dimensions of job heterogeneity. These generalizations, however, come at the cost of more involved conditions for sorting.

We now investigate the sign of sorting along both the EE and the NE margin. If terms (1) and (2) in Theorem 1 are negative, then PAM obtains on the NE and EE margin, respectively.

4.1.2 The EE Margin

The following statement is immediately implied by Theorem 1:

Corollary 1. *If for all $s \geq 0$ and y' such that* $\Pr_\Gamma \{y_j = y' | \sigma(\mathbf{x}, \mathbf{y}) = s\} > 0$,

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

then term (2) in Theorem 1 is negative [positive] for all y, i.e. positive [negative] assortative matching occurs in the (y_j, x_k) dimension along the EE margin.

Corollary 1 implies in particular that, if the NE margin is shut down (i.e. if $\sigma(\mathbf{x}, \mathbf{y}) > 0$ 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.

Condition (CMP) can be loosely 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. Indeed, in the one-dimensional case ($Y = 1$), condition (CMP) collapses to a restriction on the sign of $\partial^2 \sigma / \partial x_k \partial y = \partial^2 f / \partial x_k \partial y$, which is the familiar super- (or sub) modularity condition on the production function encountered in one-dimensional frictionless models. Beyond this simple intuitive interpretation, Condition (CMP) is not easy to work with in the multi-dimensional case. Loosely speaking, it imposes that supermodularity hold along all level curves of $\sigma(\mathbf{x}, \mathbf{y})$, which amounts to a complex restriction involving not only the technology, but also the sampling distribution of job types. Yet Assumptions 1 and 2 simplify condition (CMP) considerably and allow us to obtain the following result on the sign of EE-sorting.

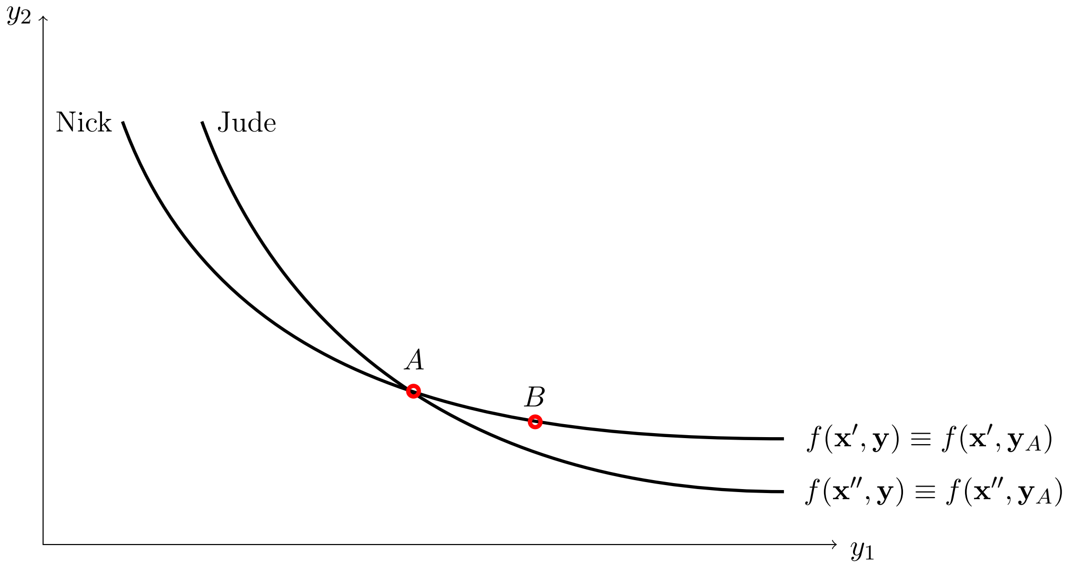


Figure 1: Single Crossing Property

Theorem 2 (EE-Sorting, $Y = 2$, bilinear technology)**.** *Under Assumptions 1 and 2, PAM occurs in the (y_1, x_k) dimension along the EE margin if and only if, for all $\mathbf{y} \in \mathcal{Y}$:*

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f(\mathbf{x}, \mathbf{y})/\partial y_1}{\partial f(\mathbf{x}, \mathbf{y})/\partial y_2}\right) > 0 \;\Leftrightarrow\; \frac{\partial}{\partial x_k}\left(\frac{q_1(\mathbf{x})}{q_2(\mathbf{x})}\right) > 0 \tag{SC-2d}$$

All proofs are in the Appendix. Condition (SC-2d) has a natural interpretation and is well-known in matching problems. It is a single-crossing property of the production function (also known as Spence-Mirrlees condition, in its differential form) that was shown to guarantee positive sorting in several one-dimensional matching problems.[12]

The analysis of our *multi*-dimensional matching model with search frictions and transferable utility further highlights the importance of single crossing as a driving force toward positive sorting. Condition (SC-2d) states that the marginal rate of substitution between (y_1, y_2) is increasing in worker skill x_k. This implies that skill x_k is a stronger complement to job attribute y_1 than to y_2, which is why positive sorting occurs between between x_k and y_1.

To illustrate the single crossing condition in our setting and its implication graphically, we consider two workers (Nick and Jude) with skill bundles $\mathbf{x}' = (x_1', x_2')$ and $\mathbf{x}'' = (x_1'', x_2'')$ such that $x_1'' > x_1'$ and $x_2'' = x_2'$ (Jude has more of x_1 but both have equal amounts x_2). For each

[12]In an important paper, Legros and Newman (2007) show that a single crossing property is sufficient to guarantee PAM in frictionless one-dimensional problems with *non*-transferable utility (NTU). Chade, Eeckhout and Smith (2014) then demonstrate that several one-dimensional matching problems with *transferable* utility both in environments with and without frictions can be recast as NTU, frictionless matching problems. After finding the associated NTU problem, the Legros-Newman-condition can be applied and guarantees PAM.

worker, we plot the locus of job attributes with which that worker produces the same output as when matched to the job with attribute bundle A. The single-crossing condition (SC-2d) implies that these isoquants cross only once (at point A). Moreover, because the marginal rate of substitution is increasing in x_1, the curve of the more skilled worker (Jude) is steeper. Consider point A as a benchmark with no sorting (both workers are matched to the same job). Condition (SC-2d) says the following: if the lower-skilled worker weakly prefers job B over job A where B has lower y_2 but higher y_1, then the higher-skilled worker (who is more skilled in dimension x_1) *strictly* prefers job B, which is the case in the graph.

We end this first analysis of sorting on the EE margin with two important remarks. First, the characterization of sorting patterns in Theorem 2 is independent of the sampling distribution. In particular, the restriction to two-dimensional job heterogeneity ($Y = 2$) allows us to circumvent condition (CMP). This independence result partially generalizes to nonlinear technologies under two-dimensional job heterogeneity, but generally not to dimensions higher than two. Second, Theorem 2 provides a necessary and sufficient condition for assortative matching. This, in turn, does not generalize to nonlinear technologies or to more than two dimensions of job heterogeneity. We discuss generalizations below in Subsection 4.2.

4.1.3 The NE Margin

While the conditions for sorting on the EE margin presented in Theorem 2 are distribution-free, this is not the case for sorting along the NE margin, as established by the following Corollary that also follows from Theorem 1.

Corollary 2. *If condition (CMP) holds and if, in addition,*

$$G'_{\sigma|\mathbf{x}}(0) \times \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = 0 \right] \times \left\{ \int_0^{+\infty} G'_{\sigma|\mathbf{x}}(s) \left[\Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = 0 \right\} - \Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = s \right\} \right] ds \right\} \leq 0 \quad (7)$$

then term (1) in Theorem 1 is negative [positive] for all y, i.e. positive [negative] assortative matching occurs in the (y_j, x_k) dimension along the NE margin.

Condition (7) (the expression under the integral in particular) clearly shows that sorting on the NE margin involves the sampling distribution Γ. This is true even if we restrict our attention to two-dimensional heterogeneity on the demand side. The next result establishes conditions on the sampling distribution that, together with sufficient complementarities in production, guarantee positive sorting along the NE margin.

Theorem 3 (NE-Sorting, $Y = 2$, bilinear technology)**.** *Under Assumptions 1 and 2 and under the single-crossing condition* (SC-2d) *(Theorem 2), if:*

1. $q_1(\mathbf{x}) > 0$ *(i.e. $f(\mathbf{x}, \mathbf{y})$ is increasing in both y_1 and y_2)*

2. *the following condition holds along all level curves of $f(\mathbf{x}, \cdot)$ (i.e. at all $\mathbf{y}$ such that $f(\mathbf{x}, \mathbf{y}) = C$ for some fixed $C \geq 0$):*

$$q_2(\mathbf{x})\frac{\partial^2 \ln \gamma}{\partial y_2 \partial y_1} - q_1(\mathbf{x})\frac{\partial^2 \ln \gamma}{\partial y_2^2} \geq 0 \tag{NE-2d}$$

3. *at the lower support of $\mathbf{y}$ ($\underline{\mathbf{y}} = (\underline{y}_1, \underline{y}_2)$): $\underline{y}_2 \geq b_2$ and $\underline{y}_1 < b_1$*

then PAM occurs in the (y_1, x_k) dimension along the NE margin.

As mentioned earlier, sorting on the NE margin results from the impact of a marginal increase in skill x_k on the boundary of the set of profitable matches, i.e. the locus of $\mathbf{y}$'s such that $\sigma(\mathbf{x}, \mathbf{y}) = 0$. Figure 2 shows how this boundary shifts with x_k under the assumptions of Theorem 3, and helps visualize the role of each condition in that theorem.

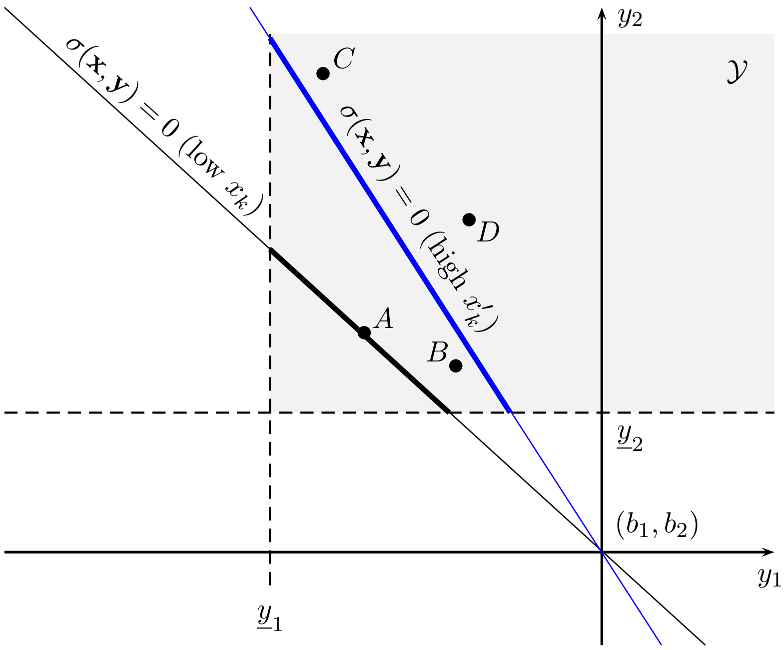


Figure 2: Sorting along the NE margin

Figure 2 represents the (y_1, y_2) plane, where the origin is placed at $\mathbf{b} = (b_1, b_2)$. The shaded area 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 accordance with Condition 3 in Theorem 3. The oblique lines are zero level curves of $\sigma(\mathbf{x}, \cdot)$, which under this linear technology (Assumption 1) are given by $y_2 = b_2 - \frac{q_1(\mathbf{x})}{q_2(\mathbf{x})}(y_1 - b_1)$. By Condition 1 in Theorem 3, 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 shaded rectangle on the figure). 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 surplus to equal zero.

Those zero level curves are represented for two values $x'_k > x_k$ of skill k, the higher-x'_k (blue) curve being steeper than the lower-x_k (black) one, meaning that for a given y_2, the more skilled worker needs a higher y_1 to generate non-negative surplus. The reason is as follows: by the single crossing property (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 in particular for those workers with higher skill x'_k. Therefore, for a given y_2, workers with higher skill k need jobs with higher y_1 to generate non-negative surplus, which is clearly a force towards PAM. In the figure, this means that all job types between the black and the blue line can be profitably matched with the low skilled (x_k) worker, but produce negative surplus with the high-skilled (x'_k) worker, which is why all those jobs with relatively low attribute y_1 drop out of his equilibrium matching set.

However, this force alone (which relies on complementarities in production) is not enough to ensure PAM on the NE margin. To see this, consider points A, B, C and D on the figure. By increasing skill k from x_k to x'_k, the 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 A remain profitable. Therefore, if the sampling distribution γ has most of its mass concentrated around points A, B and C then workers with *higher* x'_k will tend to 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 towards NAM. To prevent this, one must 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. Notice that the potential 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 $q_2(\mathbf{x})$, which makes the zero-surplus lines flatter).

Summing up, to ensure positive sorting along the NE margin, we not only have to assume sufficient complementarities in production but also sufficient positive association of job attributes in the sampling distribution.

How likely is condition (NE-2d) to hold? Sufficient conditions on the sampling distribution are that density γ be both log-supermodular and log-concave. This class of distributions is quite broad. For instance, any bivariate distribution of independent random variables that has log-concave marginals satisfies (NE-2d) (e.g. the uniform distribution with independent random variables). Another bivariate distribution that is both log-supermodular (for positive covariance) and log-concave is the bivariate normal distribution. In fact, if γ is a (truncated) normal distribution with covariance $\Sigma = \begin{pmatrix} \theta_1^2 & \theta_{12} \\ \theta_{12} & \theta_2^2 \end{pmatrix}$:

$$\frac{\partial^2 \ln \gamma(\mathbf{y})}{\partial y_1 \partial y_2} = \frac{\theta_{12}}{\theta_1^2\theta_2^2 - \theta_{12}^2} \quad \text{and} \quad \frac{\partial^2 \ln \gamma(\mathbf{y})}{\partial y_2^2} = -\frac{\theta_1^2}{\theta_1^2\theta_2^2 - \theta_{12}^2}$$

and condition (NE-2d) becomes equivalent to $\theta_{12}q_1(\mathbf{x}) + \theta_1^2 q_2(\mathbf{x}) \geq 0$, which is always true if the covariance of (y_1, y_2) in γ is positive. Yet another example of a density that is both log-supermodular and log-concave is the multivariate Gamma distribution (which is defined by a linear combination of independent random variables that have standard gamma distribution).[13]

Cases involving log-convex distributions are more complex. For example, if γ is bivariate Pareto

$$\gamma(\mathbf{y}) = \frac{\alpha(\alpha+1)}{\tau_1\tau_2}\left(\sum_{j=1}^{2}\frac{y_j - 1}{\tau_j} + 1\right)^{-\alpha-2}, \ \tau_j > 0, \ y_j \geq 1$$

then (NE-2d) holds iff $\tau_1 q_1(\mathbf{x}) \leq \tau_2 q_2(\mathbf{x})$, which places an additional joint restriction on the technology and support of $\mathbf{x}$.

4.1.4 Taking Stock

Both theorems on the sign of sorting show that sorting under multi-dimensional job heterogeneity is fundamentally different from a comparable model with one-dimensional heterogeneity. In such a model, there is *no sorting* on the EE margin (Postel-Vinay and Robin, 2002): the strategy of firms is to accept any worker that yields positive surplus while the strategy of workers is to accept all jobs that yield a higher (flow) surplus than the current one. Under the assumption that the flow surplus is increasing in y (the one-dimensional version of Assumption 1), this implies that *all* workers tend to move into higher-y jobs over time, which rules out sorting. Moreover, when

[13] Log-supermodularity of the multivariate Gamma distribution is implied by Proposition 3.8 in Karlin and Rinott (1980) and log-concavity of the Gamma distribution is implied by Theorem 4.26 in Shapiro, Dentcheva and Ruszczynski (2009).

the NE-margin operates, there is positive sorting in our multi-dimensional setting under the specified conditions. But there would again be *no sorting* in the model with one-dimensional heterogeneity since any match in which the job productivity is too low ($y < b$) would not form, independent of the worker's skill.

Why does sorting arise only in the multi-dimensional model? Compared to the one-dimensional case, what matters here is not only to match with a productive job in *any* dimension. Instead it is important to obtain a job that requires much of the skill in which the worker is particularly strong. Thus, depending on their skill bundles, different workers may accept and reject different types of jobs, which is why sorting arises. This trade-off of sorting across dimensions is absent in a one-dimensional setting, which is why all workers share the same ranking of jobs and end up in similar jobs over time.[14]

4.2 Generalizations

In this section we provide partial generalizations of the results on the sign of sorting presented in Section 4.1. We relax Assumptions 1 and 2 in two ways: first, we generalize our results to the case of a monotone technology with an unrestricted number of job attributes. We then state the cases of monotone technology with two job attributes and of bilinear technology with unrestricted number of job attributes as corollaries of the general theorem. Second, we investigate a specific form of non-linear, non-monotone technology which we call 'separable', again for general Y-dimensional heterogeneity of job attributes.

4.2.1 Monotone Technology in Y Dimensions

We begin with sufficient conditions for sorting on the EE margin, which was addressed, in the case of $Y = 2$ and bilinear technology, by Theorem 2. The following theorem relaxes both Assumptions 1 and 2, generalizing Theorem 2 to the case of a monotone technology for $Y \geq 2$:

Theorem 4 (EE-Sorting, $Y \geq 2$, monotone technology)**.** *If:*

1. $f(\mathbf{x}, \mathbf{y})$ *is three times differentiable in* $\mathbf{y}$
2. (a) $f(\mathbf{x}, \mathbf{y})$ *is strictly increasing in* y_Y *(monotonicity)*

   (b) *for all* $\mathbf{y} \in \mathcal{Y}$, $\lim_{y_Y \to \underline{y}_Y} f(\mathbf{x}, \mathbf{y}) < b(\mathbf{x})$ *and* $\lim_{y_Y \to \overline{y}_Y} f(\mathbf{x}, \mathbf{y}) = +\infty$

[14] See Lindenlaub (2014) for the sorting-trade off across dimensions in a friction*less* environment with multi-dimensional heterogeneity.

(c) *for all* $\ell \in \{1, \cdots, Y-1\}$,

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f(\mathbf{x},\mathbf{y})/\partial y_\ell}{\partial f(\mathbf{x},\mathbf{y})/\partial y_Y}\right) > 0 \tag{SC-Yd}$$

3. *if* $Y \geq 3$, *then for all* $(i,j) \in \{1, \cdots, Y-1\}^2$, $i \neq j$, *and along all level curves of* $f(\mathbf{x}, \cdot)$:

$$\begin{aligned}
&\frac{\partial f}{\partial y_Y}\left[\left(\frac{\partial f}{\partial y_Y}\right)^2 \frac{\partial^2 \ln\gamma}{\partial y_i \partial y_j} + \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\frac{\partial^2 \ln\gamma}{\partial y_Y^2} - \frac{\partial f}{\partial y_j}\frac{\partial f}{\partial y_Y}\frac{\partial^2 \ln\gamma}{\partial y_i \partial y_Y} - \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_Y}\frac{\partial^2 \ln\gamma}{\partial y_j \partial y_Y}\right] \\
&\quad - \frac{\partial \ln\gamma}{\partial y_Y}\left[\left(\frac{\partial f}{\partial y_Y}\right)^2 \frac{\partial^2 f}{\partial y_i \partial y_j} + \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\frac{\partial^2 f}{\partial y_Y^2} - \frac{\partial f}{\partial y_j}\frac{\partial f}{\partial y_Y}\frac{\partial^2 f}{\partial y_i \partial y_Y} - \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_Y}\frac{\partial^2 f}{\partial y_j \partial y_Y}\right] \\
&\quad - \left(\frac{\partial f}{\partial y_Y}\right)^2 \frac{\partial^3 f}{\partial y_i \partial y_j \partial y_Y} + \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_Y}\frac{\partial^3 f}{\partial y_j \partial y_Y^2} + \frac{\partial f}{\partial y_j}\frac{\partial f}{\partial y_Y}\frac{\partial^3 f}{\partial y_i \partial y_Y^2} - \frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\frac{\partial^3 f}{\partial y_Y^3} \\
&\qquad + \frac{\partial f}{\partial y_Y}\left[\frac{\partial^2 f}{\partial y_j \partial y_Y}\frac{\partial^2 f}{\partial y_i \partial y_Y} - \frac{\partial^2 f}{\partial y_Y^2}\frac{\partial^2 f}{\partial y_i \partial y_j}\right] > 0
\end{aligned} \tag{EE-Yd}$$

then PAM occurs in all dimensions (y_ℓ, x_k) *other than* $\ell = Y$ *along the EE margin.*

Conditions 2a (monotonicity) and 2b (y_Y is an "essential" input) are technical and ensure that $f(\mathbf{x}, \mathbf{y})$ is invertible w.r.t. one of the y's (which w.l.o.g. we take to be y_Y). Conditions 2a and 2b further ensure that the support of γ keeps a lattice structure under a change of variables (see proof). Condition 2c is central to the theorem and generalizes the single-crossing condition (SC-2d) from Theorem 2. It states that there exists a job attribute y_Y, satisfying Conditions 2a and 2b, that is among those which are most substitutable to x_k.

While we proved in Theorem 2 that for EE-sorting in the case of $Y = 2$, sufficient conditions for sorting are *distribution-free*, our results from this section show that this is no longer the case for $Y > 2$ (even if we assume bilinear surplus, see Corollary 4): Condition 3 — or, equivalently, equation (EE-Yd) — restricts both the sampling distribution and its interaction with the production function in a complicated way. In essence, equation (EE-Yd) places restrictions on the way the job attributes are associated in the sampling distribution γ.

The reason for why the sampling distribution needs to be restricted in the case of $Y > 2$ can be explained in light of the sorting trade-off that is prevalent in multi-dimensional matching problems: In our setting with multiple dimensions of heterogeneity and frictions, it is not feasible that workers match positively to firms on all dimensions. Instead, they must match negatively in at least one dimension. As a result, the agents need to choose in which dimension negative sorting is most tolerable. Theorem 4 says that this choice depends on the *relative* complementarities in production. Because complementarities between skill k with *all* job attributes ℓ ($\ell \neq Y$) are stronger than with attribute Y (Condition 2c), a worker with higher x_k would ideally move towards jobs with higher levels of y_ℓ *for all* $\ell \neq Y$ (and possibly with lower level of y_Y). Yet, if the attributes y_ℓ are too strongly negatively associated (and/or there is a positive correlation between y_ℓ and y_Y), then such moves may not be feasible. The role of condition (EE-Yd) is to prevent these distributional obstructions to positive sorting.

Theorem 4 is our most general result on EE-sorting and is useful because it nests several special cases of interest. We show in the following corollaries that the sufficient conditions for sorting simplify considerably for monotone surplus function with $Y = 2$ (Corollary 3) as well as bilinear surplus function with $Y > 2$ (Corollary 4).

Corollary 3. *(EE-Sorting, $Y = 2$, monotone technology) Under Assumption 2 ($Y = 2$), if $f(\mathbf{x}, \mathbf{y})$ is twice continuously differentiable and quasi-concave in $\mathbf{y}$, strictly increasing in y_2 with $\min_{y_2 \in \mathbb{R}} f(\mathbf{x}, (y_1, y_2)) < b(\mathbf{x})$ for all $y_1 \in \left[\underline{y}_1, \overline{y}_1\right]$, and such that:*

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f(\mathbf{x}, \mathbf{y})/\partial y_1}{\partial f(\mathbf{x}, \mathbf{y})/\partial y_2}\right) > 0$$

then PAM occurs in the (y_1, x_k) dimension along the EE margin.

If we consider only two job attributes, the sufficient conditions for EE-sorting from Theorem 4 become considerably simpler. First, Condition (EE-Yd) disappears altogether: the role of that condition was to prevent the $Y-1$ job attributes y_ℓ, $\ell \neq Y$, from being associated in Γ in a way that countered the force toward PAM in $Y-1$ dimensions arising from technology. This issue of association between $Y-1$ job attributes becomes moot when $Y = 2$. The reason is that if complementarities are stronger along (y_1, x_k) than along (y_2, x_k), then sorting will be PAM in (y_1, x_k) and NAM in (y_2, x_k). Hence, assuring a positive degree of association between y_1, y_2 in the sampling distribution is not needed here to resolve the sorting trade-off.

Second, the generalized single crossing condition (SC-Yd) collapses to the condition that already appeared in our result on bilinear technology with $Y = 2$ (Theorem 2). This result again highlights that as long as we restrict our attention to two job attributes, we can specify distribution-free conditions for EE-sorting.

Another useful special case on EE-sorting that emerges from Theorem 4 is the bilinear production function (Assumption 1), which satisfies monotonicity (as, by Assumption 1, $q_Y(\mathbf{x}) > 0$).

Corollary 4 (EE-Sorting, $Y > 2$, bilinear technology)**.** *Under Assumption 1, if:*

1. $\overline{y}_Y = +\infty$*, and, for all* $\mathbf{y} \in \mathcal{Y}$*,* $\lim_{y_Y \to \underline{y}_Y} f(\mathbf{x}, \mathbf{y}) < b(\mathbf{x})$

2. *for all* $\ell \in \{1, \cdots, Y-1\}$*,*

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f(\mathbf{x},\mathbf{y})/\partial y_\ell}{\partial f(\mathbf{x},\mathbf{y})/\partial y_Y}\right) > 0 \;\Leftrightarrow\; \frac{\partial}{\partial x_k}\left(\frac{q_\ell(\mathbf{x})}{q_Y(\mathbf{x})}\right) > 0 \tag{SC-Yd}$$

3. *for all* $(i,j) \in \{1, \cdots, Y-1\}^2$*,* $i \neq j$*, and along all level curves of* $f(\mathbf{x}, \cdot)$*:*

$$q_Y(\mathbf{x})^2 \frac{\partial^2 \ln \gamma}{\partial y_i \partial y_j} + q_i(\mathbf{x}) q_j(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_Y^2} - q_j(\mathbf{x}) q_Y(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_i \partial y_Y} - q_i(\mathbf{x}) q_Y(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_j \partial y_Y} > 0 \tag{EE-Yd'}$$

then PAM occurs in all dimensions (y_ℓ, x_k) *other than* $\ell = Y$ *along the EE margin.*

Focusing on $Y > 2$ is innocuous, as the case $Y = 2$ is covered in Theorem 2. Condition 1 echoes Conditions 2a and 2b of Theorem 4 and is of the same technical nature. Condition 2 parallels the generalized single crossing condition from Theorem 4. In contrast to Theorem 2 that focuses on sorting under two-dimensional job heterogeneity, the case with general Y-dimensional heterogeneity requires restrictions on the sampling distribution (Condition 3). For instance, condition (EE-Yd') is satisfied if log-supermodularity of γ in the job attribute under consideration, y_1, and any other job attribute, y_j ($j \neq Y$), overall dominates any log-concavity of the sampling distribution in y_Y and any log-supermodularity in pairs (y_1, y_Y) and (y_j, y_Y) (assuming $q_i(\mathbf{x}) > 0$ for all i). Like in Theorem 4, this condition limits distributional barriers to PAM in dimension (x_k, y_ℓ) for any $\ell \neq Y$.

Condition (EE-Yd') is more demanding on the distributions than condition (NE-2d) since it involves the relative strength of log-supermodularity in the various dimensions as well as subtle interactions with the technology. Nevertheless we can show that the set of models satisfying this condition is not empty. For instance, it can be satisfied by a (truncated) multivariate normal distribution, for which (EE-Yd') reads

$$q_Y(\mathbf{x})^2 \frac{\Sigma_{ij}^{-1} + \Sigma_{ji}^{-1}}{2\,|\Sigma|} + q_1(\mathbf{x}) q_j(\mathbf{x}) \frac{-\Sigma_{YY}^{-1}}{|\Sigma|} - q_j(\mathbf{x}) q_Y(\mathbf{x}) \frac{\Sigma_{iY}^{-1} + \Sigma_{Yi}^{-1}}{2\,|\Sigma|} - q_i(\mathbf{x}) q_Y(\mathbf{x}) \frac{\Sigma_{Yj}^{-1} + \Sigma_{jY}^{-1}}{2\,|\Sigma|} \geq 0 \tag{8}$$

where Σ_{ij}^{-1} is element ij of the inverse of the covariance matrix and $|\Sigma|$, its determinant. Condition (8) holds if, for instance, the technology is symmetric ($q_j(\mathbf{x}) = q_{j'}(\mathbf{x})$ for all $j \neq j'$) and if the correlation between y_1 and y_ℓ is sufficiently strong.[15]

We finally generalize Theorem 3 on NE-sorting to Y-dimensional heterogeneity of job attributes. Similar to Corollary 4 on EE-sorting, in this more general setting sorting requires complex restrictions on the sampling distribution Γ.

Theorem 5 (NE-Sorting, $Y > 2$, bilinear technology)**.** *Under Assumption 1 and Conditions 1-3 from Corollary 4, if:*

1. $q_j(\mathbf{x}) > 0$ *for all* $j \in \{1, \cdots, Y\}$ *(i.e.* $f(\mathbf{x}, \mathbf{y})$ *is increasing in all job attributes)*

2. *the following condition holds along all level curves of* $f(\mathbf{x}, \cdot)$ *and for all* $j = \{1, ..., Y-1\}$*:*

$$q_Y(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_Y \partial y_j} - q_j(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_Y^2} \geq 0 \tag{NE-Yd}$$

3. *denoting the lower support of* $\mathbf{y}$ *by* $\underline{\mathbf{y}} = \left(\underline{y}_1, \cdots, \underline{y}_Y\right)$*:*

$$\sum_{j=1}^{Y} \left[\frac{q_{kj}}{q_j(\mathbf{x})} - \max_{j'} \left\{ \frac{q_{kj'}}{q_{j'}(\mathbf{x})} \right\} \right] q_j(\mathbf{x}) \left(\underline{y}_j - b_j \right) \leq 0$$

then PAM occurs in all dimensions (y_ℓ, x_k) *other than* $\ell = Y$ *along the NE margin.*

Theorem 5 shows that signing the sorting patterns on the NE margin is even more involved than signing EE sorting. It requires stronger assumptions on the sampling distribution Γ, similar to the case of two-dimensional heterogeneity of job requirements. Condition (NE-Yd) parallels Condition (NE-2d) in Theorem 3 for $Y = 2$. Condition (NE-Yd) (which again heavily relies on the positive association between job attributes in the sampling distribution) has the same interpretation as (NE-2d). It prevents distributional obstructions to PAM that could occur despite sufficient complementarities in production.[16] Lastly, Condition 3 restricts the lower support of the sampling distribution and echoes Condition 3 from Theorem 3, extended to Y-dimensional heterogeneity in job attributes.

[15] As an example, consider $Y = 3$. Then (8) holds for symmetric technology (note that in this case, condition (SC-Yd) holds with equality) if τ_{12} is large relative to τ_{13} and τ_{23} (where $\tau_{ij} = \theta_{ij}/\theta_i\theta_j = \text{corr}_\Gamma(y_i, y_j)$):

$$\frac{q(\mathbf{x})^2}{|\Sigma|} \left(\frac{\tau_{12} - \tau_{13}\tau_{23}}{\theta_1\theta_2} - \frac{1 - \tau_{12}}{\theta_3^2} - \frac{\tau_{13} - \tau_{12}\tau_{23}}{\theta_1\theta_3} - \frac{\tau_{23} - \tau_{12}\tau_{13}}{\theta_3\theta_2} \right) \geq 0$$

[16] Note that there is a tension between (NE-Yd) and Condition (EE-Yd') that is also assumed to hold. Importantly, both conditions can be satisfied simultaneously. For instance, in the case of the truncated normal with positive correlation with $X = Y = 3$, both conditions will be satisfied if the technology is symmetric (note that in this case, condition (SC-Yd) holds with equality) if τ_{12} is high, $\tau_{12} \to 1$ and τ_{13} and τ_{23} are low, $\tau_{13}, \tau_{23} \to 0$.

4.2.2 Nonlinear Technology in Y Dimensions

Arguably the most substantive restriction placed by Theorem 4 on the production technology is strict monotonicity of f w.r.t. at least one job attribute y_J. While it still covers a wide range of applications, it excludes some important special cases, such as the popular "bliss point" specification. An example (which goes back to, at least, Tinbergen, 1956) would be, in the case $X = Y$, $f(\mathbf{x}, \mathbf{y}) = a_0 - \sum_{i=1}^{X} a_i (x_i - y_i)^2$, where the a_i's are strictly positive numbers. In this example, each job has an "ideal" skill bundle given by $\mathbf{y}$, and output is a decreasing function of the distance between the worker's skill bundle $\mathbf{x}$ and that ideal skill bundle. This and other related specifications are covered in the following theorem:

Theorem 6. *(EE-Sorting, $Y \geq 2$, separable technology) If:*

1. $f(\mathbf{x}, \mathbf{y})$ *is continuously differentiable w.r.t.* $\mathbf{y}$
2. *for a given k and for all $\ell \in \{2, \cdots, Y\}$, $\partial^2 f(\mathbf{x}, \mathbf{y}) / \partial x_k \partial y_\ell = 0$ (separability)*
3. *for all $\mathbf{y} \in \mathbb{R}_+^Y$, $\partial^2 f(\mathbf{x}, \mathbf{y}) / \partial x_k \partial y_1 > 0$*

then, for $k \in \{1, \cdots, X'\}$, PAM occurs in the (y_1, x_k) dimension along the EE margin.

The key restriction imposed in Theorem 6 is Condition 2, which states that there are components of the worker's skill bundle that are only relevant to perform task 1, i.e. that are neither complement nor substitute with any other task $j \in \{2, \cdots, Y\}$. The sign of sorting on the EE margin *between task 1 and the skills that are only relevant to task 1* can then be determined: it has the sign of $\partial^2 f(\mathbf{x}, \mathbf{y}) / \partial x_k \partial y_1$.[17] Assumption 2 is easily satisfied, for instance, by production functions that only feature within-complementarities of skills and job attributes but no complementarities across tasks. For example, under the "bliss point" specification mentioned above ($f(\mathbf{x}, \mathbf{y}) = a_0 - \sum_{i=1}^{X} a_i (x_i - y_i)^2$), Theorem 6 establishes that there will be positive *within-skill* sorting along the EE margin (i.e. $H_1 (\cdot|\mathbf{x})$ increases in x_1 in the FOSD sense), but says nothing about *between-skill* sorting (i.e. the monotonicity of $H_\ell (\cdot|\mathbf{x})$ w.r.t. x_1 for $\ell \geq 2$).

Note that the case of separable technology is special: the characterization of sorting in Theorem 6 is independent of the sampling distribution Γ irrespective of the dimensionality of job heterogeneity. In particular, the restrictions on the sampling distribution (EE-Yd) in

[17] Note that, under separability Assumption 2 in Theorem 6, the sign of $\partial^2 f / \partial x_k \partial y_1$ is the same as the sign of $\frac{\partial}{\partial x_k} \left(\frac{\partial f / \partial y_1}{\partial f / \partial y_{j'}} \right), 1 \neq j'$. In that sense, the characterization of sorting in Theorem 6 parallels the characterization provided by Theorems 2 and 4 under different assumptions on S.

Theorem 4 are only needed if complementarities in (x_k, y_1) compete with complementarities between x_k and other dimensions y_ℓ, $\ell \neq 1$. Theorem 4 then guarantees PAM in all but one dimensions while Theorem 6 for separable production functions only ensures sorting in a single dimension (x_k, y_1).

5 Simultaneous Sorting in Multiple Dimensions

Theorems 4 and 5 and their corollaries show that sorting arises in our model under familiar assumptions on complementarities in the production technology combined with assumptions on the sampling distribution, where we focused on the effect of an increase in *a single skill* on the matching distribution, keeping all other skills fixed. In this section we move away from the ceteris paribus setting and investigate the effect of a simultaneous expansion of all skills.

5.1 Absolute Advantage vs. Specialization

Results in this section are established under Assumption 1, i.e. for a bilinear technology.[18] Our first result is that there is no sorting on "absolute advantage", in that if two workers with skills $\mathbf{x}$ and $\mathbf{x}'$ are such that the type-$\mathbf{x}'$ worker produces twice as much output than the type $\mathbf{x}$ worker in all jobs, both workers are matched to the same distribution of job types in equilibrium, irrespective of the complementarities in production.

Theorem 7. *Under Assumption 1,* $\forall j \in \{1, \cdots, Y\}$*:*

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

i.e. the function $(\mathbf{x} + \mathbf{a}) \mapsto H_j\left(y|\mathbf{x}\right)$ *is homogeneous of degree* 0 *in* $(\mathbf{x} + \mathbf{a})$ *for all* j*.*

Theorem 7 implies that a worker with skills $\mathbf{x}' = -\mathbf{a} + 2(\mathbf{x} + \mathbf{a})$ produces twice as much as a worker with skills $\mathbf{x}$, but is matched to the same distribution of jobs. One obvious consequence of this result is that the mappings $\mathbf{x} \mapsto H_j\left(\cdot|\mathbf{x}\right)$, $j \in \{1, \cdots, Y\}$ are not one-to-one: contrary to the multi-dimensional matching model *without* frictions (Lindenlaub, 2014), in our frictional environment sorting is not even pure as far as matching distributions are concerned: workers with different skill bundles can be matched to the same distribution of jobs in equilibrium.

[18] Those results are straightforward to generalize to the case of a homogeneous technology, i.e. a technology such that $\sigma\left(-\mathbf{a} + t(\mathbf{x} + \mathbf{a}), \mathbf{y}\right) = t^\alpha \cdot \sigma(\mathbf{x}, \mathbf{y})$ for any positive scaling factor t (see the appendix).

To further illustrate the implications of Theorem 7, we consider the case $X = Y = 2$, so that $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$, and $\mathbf{a} = (a_1, a_2)$. To fix ideas, we consider an example where $\mathbf{Q}$ is a positive matrix (so that $\mathbf{x}$ and $\mathbf{y}$ are complements both within and across skill dimensions), $\det \mathbf{Q} > 0$ (so that within-dimension complementarities "dominate" between-dimension complementarities), $\gamma(\cdot)$ is a bivariate normal distribution truncated over $\mathcal{Y}$ with positive covariance, and $\underline{y}_2 - b_2 \geq 0 > \underline{y}_1 - b_1$. Under these assumptions, Theorems 2 and 3 imply PAM in the (x_1, y_1) dimension,[19] and the statement in Theorem 7 writes as:

$$\frac{\partial H_1(y|\mathbf{x})}{\partial x_2} = -\frac{x_1 + a_1}{x_2 + a_2} \frac{\partial H_1(y|\mathbf{x})}{\partial x_1}. \tag{9}$$

Theorem 7 thus implies is that sorting in the (y_1, x_2) dimension is *negative*, which again reflects the sorting trade-off across dimensions typical of settings with multi-dimensional heterogeneity.

Theorem 7 addresses the case of a simultaneous 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 $\mathbf{x} + \mathbf{a}$. Since $\mathbf{a}$ is only a productivity parameter that affects all workers alike (see Assumption 1), there is no obvious reason why workers' skills should expand along this particular direction. We thus now consider a generic marginal skill expansion by letting a worker increase his skills marginally from (x_1, x_2) to $(x_1 + \Delta x_1, x_2 + \Delta x_2)$. The resulting change in $H_1(y|\mathbf{x})$ is:

$$\Delta H_1(y|\mathbf{x}) = (x_1 + a_1) \frac{\partial H_1(y|\mathbf{x})}{\partial x_1} \left[\frac{\Delta x_1}{x_1 + a_1} - \frac{\Delta x_2}{x_2 + a_2} \right] \tag{10}$$

Equation (10) has three implications: First, it confirms our earlier finding that $\Delta H_1(y|\mathbf{x}) = 0$ if $\frac{\Delta x_1}{x_1 + a_1} = \frac{\Delta x_2}{x_2 + a_2}$ (i.e. no change in sorting if the worker improves his skills in the direction of $\mathbf{x} + \mathbf{a}$).

Second, (10) shows more generally that a marginal (but not necessarily proportional) improvement in both skills will cause the worker to match with jobs with (stochastically) higher y_1 attributes iff. $\frac{\Delta x_1}{x_1 + a_1} > \frac{\Delta x_2}{x_2 + a_2}$. This can be interpreted as follows. A worker's contribution to production consists of two different inputs: his individual skills $\mathbf{x}$, plus the baseline productivity $\mathbf{a}$. The condition $\frac{\Delta x_1}{x_1 + a_1} > \frac{\Delta x_2}{x_2 + a_2}$ states that his total input is increased proportionately more in dimension 1 than in dimension 2. As a result, the simultaneous improvement in both skills will

[19] Condition (SC-2d) from Theorem 2 writes as $(x_2 + a_2) \det \mathbf{Q} > 0$, which is true in our example because $\det \mathbf{Q}$ is positive by assumption, and so is $x_2 + a_2$ by normalization to zero of the lower support of $\mathbf{x}$ and Assumption 1.d. Therefore, the example has PAM on the EE margin. Next, because $\mathbf{x} + \mathbf{a}$ is a positive vector and $\mathbf{Q}$ a positive matrix, $q_k(\mathbf{x}) > 0$ for $k = 1, 2$ and $f(\mathbf{x}, \mathbf{y})$ is therefore increasing in both y_1 and y_2. The truncated normal with positive covariance satisfies Condition (NE-2d), as shown in Subsection 4.1, and Condition 3 from Theorem 3 is satisfied by assumption. PAM therefore also occurs on the NE margin in this example.

cause the worker to sort into jobs with stochastically higher y_1 (but with lower y_2).

A third implication of (10) is based on a *proportional* increase in both skills: $\frac{\Delta x_1}{x_1} = \frac{\Delta x_2}{x_2}$. It is easy to see that $\Delta H_1(y|\mathbf{x}) < 0$ iff. $x_1/x_2 > a_1/a_2$. In other words, scaling up all skills leads to a stochastically better distribution of job matches in the dimension where the worker is *specialized* relative to the baseline productivity vector $\mathbf{a}$. By contrast, scaling up all skills leads to a deterioration of the distribution of job matches in the second dimension, i.e. $\Delta H_2(y|\mathbf{x}) > 0$. Scaling up all of a worker's skills simultaneously thus has a non-uniform effect on the worker's distribution of job matches across dimensions, which depends on the worker's specialization. Our interpretation is that this multi-dimensional model does not feature any hierarchical sorting based on absolute advantage but instead features sorting based on specialization.

The content of these results is quite different when a worker's skill is one-dimensional. Consider Theorem 7 for the one-dimensional case. $X = 1$ — i.e., $\mathbf{x}$ and $\mathbf{a}$ are scalars x and a, $\mathbf{Q}$ is a $1 \times Y$ row vector so that $\tilde{y} = \mathbf{Q}(\mathbf{y} - \mathbf{b})$ is a scalar, and the flow surplus function $\sigma(x, \mathbf{y}) = (x + a)\tilde{y}$. In this case, Theorem 7 echoes a known result: there cannot be sorting, in the sense that $\partial H_j(y|x)/\partial x = 0$ for all x, y and $j \in \{1, \cdots, Y\}$. In other words, x and y are independent in the population of job-worker matches (Postel-Vinay and Robin, 2002). Note that, contrary to our multi-dimensional setting, the no-sorting result under one-dimensional heterogeneity holds even if we only scale up x without changing a.

Our results have important implications for empirical measures of sorting on absolute versus comparative advantage. In their setting with scalar heterogeneity, Hagedorn, Law and Manovskii (2014) propose a test based on monotonicity of output in firm attribute y: if $f_y > 0$, then the interpretation is that sorting is based on absolute advantage. In turn, if output is not increasing in firm type then sorting is based on comparative advantage. Our results in this section show that this test may be problematic. First, if we assume scalar heterogeneity, there is no sorting in our model (on absolute or comparative advantage) despite $f_y > 0$. Second, and most importantly, in multi-dimensional settings sorting on comparative advantage (or on specialization) naturally arises, *especially* if the output is increasing in each firm attribute (see Assumption 1, and Theorem 3, point 1). Multi-dimensional heterogeneity thus breaks the link between the monotonicity of technology in firm attributes and hierarchical sorting that is based on absolute advantage.

5.2 The Strength of Sorting

As the preceding analysis made clear, changes in worker skill along any dimension will in general affect the assignment, i.e. $\partial H_j(y|\mathbf{x})/\partial x_k$ is in general nonzero for any given skill k. The following theorem implies that the way individual conditional distributions of job attributes $H_j(y|\mathbf{x})$ (and thus the conditional expectations) co-vary follows a pattern, which is driven by the technology $\mathbf{Q}$:

Theorem 8. *Under Assumption 1, if $\sigma(\mathbf{x}, \mathbf{y}) > 0$ for all $\mathbf{y} \in \mathcal{Y}$ (NE margin is shut down), then:*

$$(\mathbf{x} + \mathbf{a})^\top \mathbf{Q} \frac{\partial \mathbb{E}(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}^\top} = \mathbf{0}_{1\times X}.$$

In general, the way in which the distribution of job types a worker is matched to varies with that worker's skills depends upon both the production technology $\mathbf{Q}$ and the sampling distribution Γ in a seemingly complex manner. Yet, if we shut down the NE margin and focus on the EE margin, Theorem 8 shows that the *mean* of that matching distribution changes with worker skills in a way that is entirely determined by technology. As such, Theorem 8 permits a direct assessment of the strength of sorting on the EE margin, both within and between tasks.

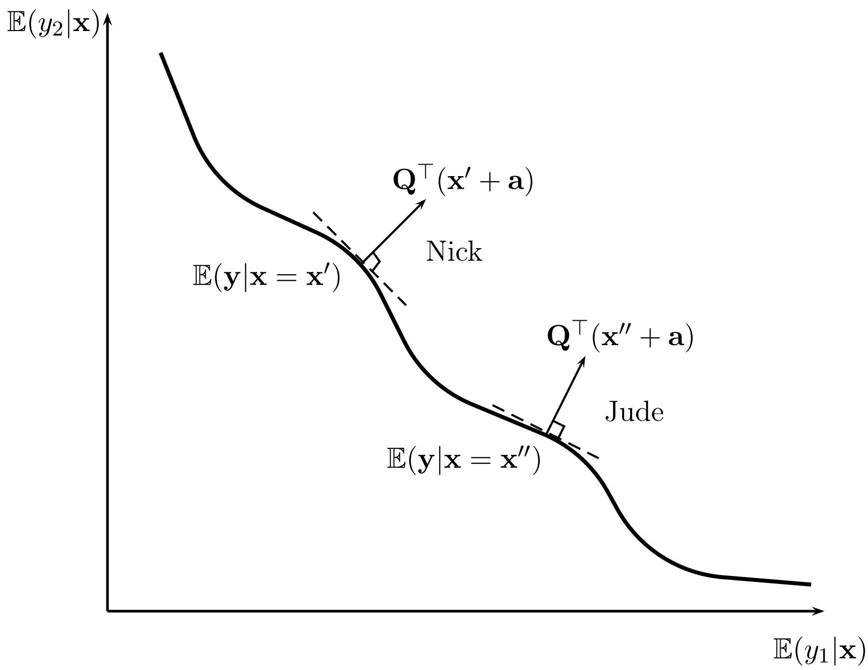


Figure 3: Sorting in two dimensions

To see those patterns more clearly, we consider again the two-dimensional case ($X = Y = 2$). Figure 3 shows the locus of the pair $(\mathbb{E}(y_1|\mathbf{x}), \mathbb{E}(y_2|\mathbf{x}))$ when one of the components of $\mathbf{x}$ (say x_1) varies. Theorem 8 implies that the normal vector to this locus is $\mathbf{Q}^\top(\mathbf{x}+\mathbf{a})$ at all points. In other words, the slope of the locus of $\mathbb{E}(\mathbf{y}|\mathbf{x})$ is entirely determined by the technology, independently of the distributions of $\mathbf{x}$ and $\mathbf{y}$. Moreover, to fix ideas, Figure 3 was drawn under the assumptions laid out in the previous subsection, as per which $\mathbf{Q}$ is a positive matrix. In this case, $f(\mathbf{x}, \mathbf{y})$ is non-decreasing in all components of $\mathbf{y}$ (all components of $\mathbf{Q}^\top(\mathbf{x} + \mathbf{a})$ are nonnegative), and the slope of the $\mathbb{E}(\mathbf{y}|\mathbf{x})$ locus is negative. We consider again our two workers (Nick and Jude) with skill bundles $\mathbf{x}' = (x'_1, x'_2)$ and $\mathbf{x}'' = (x''_1, x''_2)$ such that $x''_1 > x'_1$ and $x''_2 = x'_2$. If there is PAM within skill x_1 and job attribute y_1, then Theorem 2 implies $\partial\mathbb{E}(y_1|\mathbf{x})/\partial x_1 > 0$, and the worker with higher type-1 skills (Jude) will be matched on average to a job with higher type-1 requirements than Nick, as shown on Figure 3. What Theorem 8 further implies is that the average job to which Jude is matched will have *lower* requirements in type-2 skills than Nick's average job. That is, positive sorting *within* dimensions implies negative sorting *between* dimensions. Theorem 8 also says by how much the type-1 (type-2) attribute of Jude's job will be higher (lower) than that of Nick's job, which solely depends on the components of $\mathbf{Q}^\top(\mathbf{x}+\mathbf{a})$.

Theorem 8 and this example again highlight one of the major differences between one and multi-dimensional sorting. When workers and jobs differ in more than one dimension, there is a *sorting trade-off* between the various dimensions: increased sorting in one dimension comes along with reduced sorting in another dimension. Compared to Nick, Jude is matched to a job that requires more of his relatively strong skill, x_1, but this comes at the cost of a worse job match regarding his relatively weak skill, x_2.

6 Numerical Application

Having shown that the equilibrium sorting patterns under multi-dimensional heterogeneity are theoretically different and more complex than those occurring under scalar heterogeneity, we now investigate the quantitative importance of those differences from the following practical angle. We simulate an economy based on our multi-dimensional model, which we consider to be the true data generating process for the purpose of this exercise. We then fit a (misspecified) one-dimensional model to these data and assess the "errors" arising from the one-dimensional approximation by comparing the predictions of these two models about sorting patterns, complementarities in the surplus function and mismatch. We begin with a discussion of identification and estimation of the misspecified one-dimensional model.

6.1 Estimation of a Misspecified 1D Model

6.1.1 The model

The one-dimensional heterogeneity model that we consider in this section is essentially the multidimensional model presented in the previous section in which we set the dimensionality of heterogeneity to $X = Y = 1$. The only other departure from our multidimensional model is that we do not impose any functional form restriction on the flow surplus function $\sigma(x, y)$, other than the following identifying assumptions:

Assumption 3. *The flow surplus function $\sigma(x, y)$ has the following properties:*

1. $x \to \max_{y \in \mathcal{Y}} \sigma(x, y)$ *is well-defined and strictly increasing in* x
2. $y \to \max_{x \in \mathcal{X}} \sigma(x, y)$ *is well-defined and strictly increasing in* y

As we show below (and as has been established in the literature for more sophisticated versions of the same model), this model is non-parametrically identified up to strictly increasing transforms of x and y. We therefore normalize worker and firm productive attributes by specifying the model in terms of the ranks of said attributes:

Assumption 4.

1. x *is uniformly distributed over* $\mathcal{X} = [0, 1]$ *in the population of workers*
2. y *is uniformly distributed over* $\mathcal{Y} = [0, 1]$ *in the population of firms*

6.1.2 Identification and Estimation

The simulated sample on which we estimate the misspecified 1D model is a panel of N workers, indexed $i \in \{1, \cdots, N\}$, sorting themselves into M firms, $j \in \{1, \cdots, M\}$. Time is discretized for the purposes of simulation, and workers are followed over T periods, $t \in \{1, \cdots, T\}$. A typical observation is described as a vector $\{J_{it}, \sigma_{i,J_{it},t}\}$, where $J_{it} \in \{1, \cdots, M\}$ is the identity of the worker's employer at date t (we further normalize $J_{it} = 0$ if worker i is unemployed at date t so that J_{it} also indicates the worker's employment status), and $\sigma_{i,J_{it},t}$ is the flow surplus achieved in the match between worker i and firm J_{it} (missing when $J_{it} = 0$).[20]

[20] The assumption that the flow surplus is observed in the data is obviously a shortcut. What is typically observed in practice are wages, not match output or surplus. However, in the context of the family of search models considered in this paper, it has been shown elsewhere in the literature that the flow output from a match between a worker i and a firm j is identified from the maximum wage earned in firm j by workers with the same

The steps below establish identification of the 1D model based on this sample. The identification proof is a constructive one, and therefore also provides a practical estimation protocol.

Worker types. The maximum attainable surplus for a type-x worker is $\max_{y \in \mathcal{Y}} \sigma(x, y)$ in this model which, by Assumption 4, is a strictly increasing function of x. Any worker i's type x_i can thus be estimated as:

$$\widehat{x}_i = Q_W \left(\max \{ \sigma_{i,J_{it},t} \, : \, t = 1, \cdots, T \} \right)$$

where $Q_W(\cdot)$ is the quantile function in the population of workers.

Firm types and surplus. The flow surplus $\sigma_{i,J_{it},t}$ generated by the match between worker i and firm J_{it} is equal to $\sigma \left(x_i, y_{J_{it}} \right)$. Flow surplus is thus observed for any viable match involving any firm j in the sample, i.e. for matches between firm j and any type-x worker such that $\sigma(x, y_j) \geq 0$:

$$\widehat{\sigma(x, y_j)} = \text{mean} \left\{ \sigma_{i,J_{it},t} \, : \, \widehat{x}_i = x, \; J_{it} = j, \; t = 1, \cdots, T \right\}.$$

Knowledge of $\sigma(x, y_j)$ then allows estimation of any firm j's type, y_j:

$$\widehat{y}_j = Q_F \left(\max \left\{ \widehat{\sigma(x, y_j)} \, : \, x \in [0, 1] \right\} \right)$$

where $Q_F(\cdot)$ is the quantile function in the population of firms.

At this stage we have estimates $(\widehat{x}_i, \widehat{y}_j)$ of the types of every worker and firm in the sample, as well as estimates $\widehat{\sigma(x_i, y_j)}$ of the surplus of every viable match. Together those allow nonparametric estimation of the technology $\sigma(\cdot)$ over the set of viable (x, y) matches, as well as the construction of the equilibrium conditional distribution of employer types y amongst employed workers of any given type x, which in turn permits the analysis of equilibrium sorting patterns.

Sampling distribution and offer arrival rates. A type-x worker exiting unemployment draws her/his employer type from the following conditional density:

$$\gamma_x \left(y \mid x_i = x, e_{i,t-1} = 0, e_{it} = 1 \right) = \frac{\gamma(y)}{\gamma \left\{ y' \, : \, \sigma(x, y') \geq 0 \right\}}.$$

type as i, while unemployment income is identified from the wages earned by workers hired from unemployment — see for example Lamadon et al. (2015) for details. Because these particular identification issues are peripheral to the question addressed in this section (namely the distinction between one- vs multidimensional heterogeneity), we bypass them by assuming that flow surplus is directly observed.

Moreover, the job finding rate of any unemployed worker of type x is $\lambda_0 \gamma \{y' : \sigma(x, y') \geq 0\}$. Combining those two properties, one obtains the following estimator of $\lambda_0 \gamma(y_j)$ for any employer j in the sample:

$$\widehat{\lambda_0 \gamma(y_j)} = \Pr\{J_{it} = j \mid \widehat{x}_i = x, e_{i,t-1} = 0, e_{it} = 1\} \times \widehat{JF(x_i)}$$

where $\widehat{JF(x_i)}$ is the empirical unemployment exit rate of type-x_i workers. Note that the above estimator is valid for any worker type in the sample. This, together with the constraint that $\gamma(\cdot)$ must integrate to 1, affords estimates of the sampling distribution $\gamma(\cdot)$ and the unemployed offer arrival rate λ_0.

Finally, the probability of an employer change occurring for a type-x in a type-y firm is $\lambda_1 \gamma \{y' : \sigma(x, y') > \sigma(x, y)\} = \lambda_1 \overline{\Gamma}(y)$ for all $x \in [0, 1]$. Therefore:

$$\widehat{\lambda_1 \overline{\Gamma}(y_j)} = \Pr\{J_{it} \neq J_{i,t-1} \mid \widehat{x}_i = x, J_{i,t-1} = j, e_{i,t-1} = e_{it} = 1\}$$

so that:

$$\widehat{\lambda}_1 = \text{mean}\left[\frac{\Pr\{J_{it} \neq J_{i,t-1} \mid \widehat{x}_i = x, J_{i,t-1} = j, e_{i,t-1} = e_{it} = 1\}}{1 - \widehat{\Gamma(y_j)}}\right].$$

6.2 Simulation Exercises

6.2.1 Parameterization and Basic Estimation Results

All of the examples shown below are based on simulations of 100,000 workers with two dimensions of skills and (monthly) parameter values of $\lambda_0 = 0.3$, $\lambda_1 = 0.1$, and $\delta = 0.025$.[21] The population density of skills $\ell(\mathbf{x})$ is constructed such that $\mathbf{x}$ follows a normal distribution with mean $(0, 0)$ and correlation $\text{corr}_\ell(x_1, x_2) = -0.5$, truncated over $[0, 1]^2$. The sampling density of job attributes $\gamma(\mathbf{y})$ is constructed similarly, with $\mathbf{y}$ following a normal with mean $(1, 1)$ and correlation $\text{corr}_\Gamma(y_1, y_2) = 0.33$, truncated over $[1, 2]^2$.

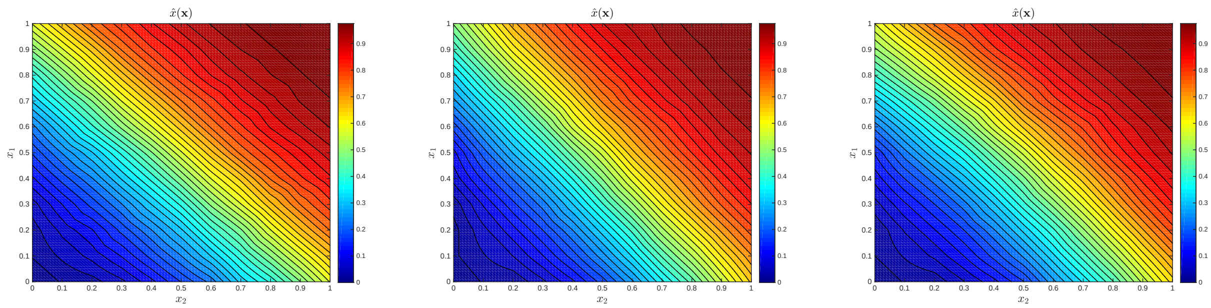
We consider three different examples of the bilinear technology (parameters $\mathbf{Q}$, $\mathbf{a}$ and $\mathbf{b}$), shown in Table 1. All three parameterizations have $\mathbf{a} = (1, 1)^\top$ and $\mathbf{b} = (0, 0)^\top$. Therefore, under all three parameterizations, unemployed workers accept all job offers and there is no sorting on the NE margin. The three different $\mathbf{Q}$ matrices are designed to feature different patterns of complementarities that, according to Theorem 2, produce certain sorting patterns. Specifi-

[21] All samples are simulated for 480 months (40 years), after a 100-year burn-in period to reach steady-state.

Specification:	**1**	**2**	**3**
Q	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 4.5 & 2 \\ 1.5 & 4 \end{pmatrix}$	$\begin{pmatrix} 4.5 & -2 \\ -1.5 & 4 \end{pmatrix}$
a		$(1,1)^\top$	
b		$(0,0)^\top$	

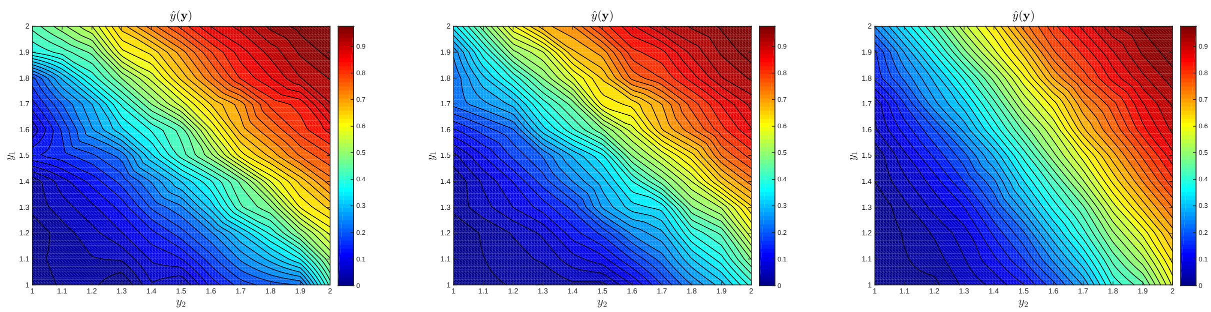
Table 1: Examples of Technologies

cally, Example 1 is the case of no sorting, within or between occupations. Both example 2 and 3 have positive sorting within occupation, and negative between occupations but differ in the complementarity structure. While the technology in Example 2 features skill-job attribute complementarities both within and between dimensions, in Example 3, there are complementarities within but substitutabilities between dimensions. The predicted sorting patterns are confirmed in the middle column of Table 2, which shows regressions of y_j on both x_i (for $(i,j) \in \{1,2\}^2$) in the cross-section of employed workers, for each example.[22]



(a) Example 1 (b) Example 2 (c) Example 3

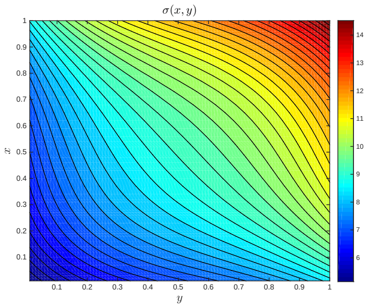
Figure 4: True and Estimated Worker Types



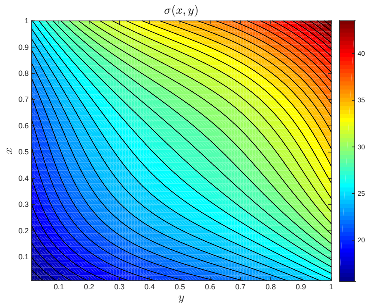
(a) Example 1 (b) Example 2 (c) Example 3

Figure 5: True and Estimated Firm Types

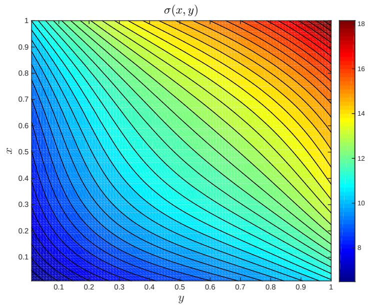
[22]Note that this is not a test of FOSD of the matching distributions in worker skills but FOSD *implies* the monotonicity of the conditional means that we report here.



(a) Example 1



(b) Example 2



(c) Example 3

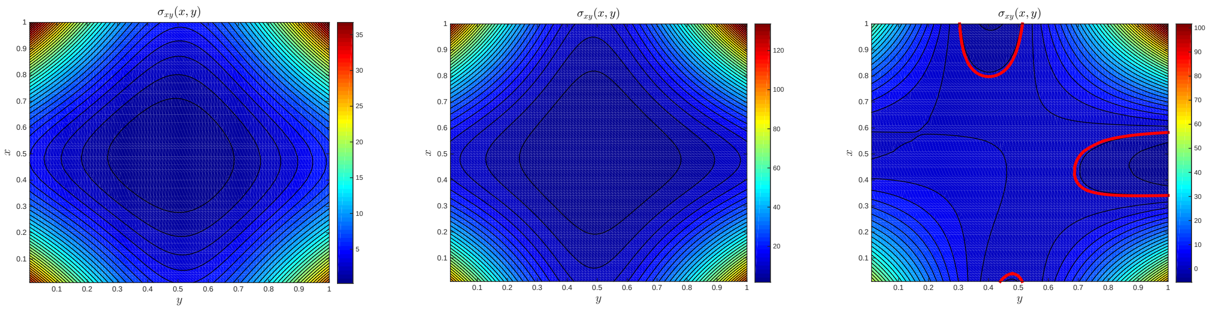
Figure 6: Estimated Surplus Function

We next fit a misspecified model with one-dimensional heterogeneity to the simulated data produced from our three examples, following the procedure explicated in Subsection 6.1.1. Figures 4 and 5 show show contour plots of the estimated scalar attributes of the estimated scalar attributes $\widehat{x}$ and $\widehat{y}$ as functions of the underlying true (vector) attributes $\mathbf{x}$ and $\mathbf{y}$, and Figure 6 plots the estimated flow surplus function $\sigma(\widehat{x}, \widehat{y})$. Panel (a) on each figure relates to Example 1, the case with no 'real' sorting in the two-dimensional world. Figure 4(a) suggests that the "iso-$\widehat{x}$" curves (the contour lines on the graph) roughly coincide with lines of slope -1 in the (x_1, x_2) plane, as was intuitively expected. Figure 5(a) suggests the same pattern for firm attributes, albeit with slightly more noise. Figure 6(a) shows that the estimated surplus function increases in $\widehat{x}$ and $\widehat{y}$, which occurs by construction, but does so in a way that suggests strong complementarity between $\widehat{x}$ and $\widehat{y}$. In turn, Panels (b) and (c) on Figures 4-6 plot these results for Examples 2 and 3 from Table 1 that feature varying complementarities across dimensions. It is no longer true in those two examples that the "iso-type" curves coincide with lines of slope -1 in the (x_1, x_2) and (y_1, y_2) planes. For instance, in Example 2, the "iso-$\widehat{x}$" curves are steeper than in Example 1. This means that in the estimation the 1D model is putting more weight on skill x_1 than skill x_2, possibly due to the relatively stronger complementarities within dimension 1 compared to dimension 2, featured by the technology matrix $\mathbf{Q}$ in Example 2. A similar pattern arises in Example 3, here more pronounced for the estimates of firm types (Figure 5, panel c).

We aim to compare the true 2D model to the estimated misspecified 1D model with focus on three issues: complementarities in the surplus function, sorting and mismatch. In each exercise, we compare the three examples of technology specified above.

6.2.2 Exercise 1: Estimating Complementarities in 1D

We first estimate the cross-partial derivative of the estimated surplus function from the 1D misspecified model, $\partial^2\sigma(\widehat{x},\widehat{y})/\partial\widehat{x}\partial\widehat{y}$, for all three model specifications (see Figure 7).[23] Two patterns stand out: first, while in the true 2D model complementarities are constant over the support, the estimated complementarities from the misspecified 1D model vary widely across the domain of the surplus function. Second, while the estimated complementarities are varying in magnitude but are always *positive* for Examples 1 and 2, this is no longer the case for Example 3. Here the cross-partial derivative of the estimated surplus function changes its sign as $\widehat{x}$ and $\widehat{y}$ vary (the zero-level contour lines are highlighted in red on Figure 7, panel c). It is striking (but also quite intuitive) that the true 2D technology from Example 3, which features both super and submodular parts, generates sign-varying cross-partials in the estimated 1D model, whereas Examples 1 and 2 with purely positive complementarities in production result in only positive estimates of cross-partials in 1D.



(a) Example 1 (b) Example 2 (c) Example 3

Figure 7: 1D Estimated Complementarities

6.2.3 Exercise 2: Estimating Sorting in 1D

We now analyze sorting in the the misspecified 1D model and try to understand the link between the estimated complementarities from Exercise 1 and the resulting sorting patterns.

We first examine the one-dimensional model's predictions in terms of sorting between the estimated (scalar) job and worker attributes, denoted $\widehat{y}$ and $\widehat{x}$. To this end, we simply regress $\widehat{y}$ on $\widehat{x}$ in a cross section of job-worker matches, as we did in each dimension of heterogeneity for the correctly specified (two-dimensional) model. The results are in the bottom panel of Table 2.

[23] This estimate of $\partial^2\sigma(\widehat{x},\widehat{y})/\partial\widehat{x}\partial\widehat{y}$ must be considered with caution, though. It was constructed based on a smooth approximation of the estimated surplus function using third-order Chebyshev polynomials. As such, it is only indicative of the "true" cross-partial. Yet, both panels (c) and (d) suggest very clearly that $\widehat{x}$ and $\widehat{y}$ are strong complements towards the edges of the domain of $(\widehat{x},\widehat{y})$.

Those results suggest that relying on the (misspecified) one-dimensional model for inference on sorting patterns is strongly misleading. In Example 1, in which there is truly no sorting in any pair of dimensions (x_k, y_j), the one-dimensional model predicts positive sorting: the coefficient of the regression of $\widehat{y}$ on $\widehat{x}$ is positive, sizeable, and statistically significant. This suggests that the estimated positive complementarities in Example 1 (see Exercise 1) are then picked up by sorting estimates that are positive as well.

Similar results hold for Example 2, where the one-dimensional model produces positive assortative matching, seemingly picking up the estimated complementarities of the surplus function that are positive throughout the entire domain.

Only in Example 3, where the true 2D model features in fact the most sizeable (and statistically significant) positive sorting within heterogeneity dimensions, the 1D model fails to predict PAM: the sorting coefficient is of negligible size and not statistically significant. This reflects the sign-varying cross-partial estimates of the surplus function that dampen equilibrium sorting compared to technologies with pure complementarities.

These results suggest the following: First, 1D sorting estimates are relatively small (both Examples 2 and 3 feature sorting of different signs within and between skills, something that a one-dimensional model is bound to miss since it can only produce some average). Second, the 1D sorting estimates can be completely misleading in the sense that they suggest positive and statistically significant sorting in cases where sorting in the true 2D data is absent. Last, our results from Exercises 1 and 2 suggest a tight link between the 1D estimates of complementarities in the surplus function and those of sorting: estimated *positive* cross-partials produce significant *positive* sorting whereas cross-partials of varying signs result in weak or even no sorting.

6.2.4 Exercise 3: Mismatch

In this third exercise, we focus on *mismatch* and have two objectives. The first is to compare measures of mismatch, constructed as the gap between actual and optimal average flow surplus, from the misspecified 1D versus the true 2D model. Second, we aim to understand how big the welfare loss/gain would be if we implemented the optimal allocation suggested by the misspecified 1D model into the true world where both worker and job types are 2D. The results are reported in Table 3.

Note that in this entire exercise we take a short or medium-run perspective where the *distribution of active jobs is fixed.* The exercise is then how to redistribute workers across given jobs

Specification:	True 2D model	Misspecified 1D model
1	$\mathbb{E}(y_1\vert\mathbf{x}) = 1.70 \underset{[-0.010,0.008]}{-0.001} x_1 \underset{[-0.013,0.004]}{-0.005} x_2$ $\mathbb{E}(y_2\vert\mathbf{x}) = 1.68 \underset{[-0.008,0.009]}{-0.001} x_1 \underset{[-0.006,0.011]}{+0.002} x_2$	$\mathbb{E}(\widehat{y}\vert\widehat{x}) = 0.68 \underset{[0.002,0.018]}{+0.010} \widehat{x}$
2	$\mathbb{E}(y_1\vert\mathbf{x}) = 1.69 \underset{[0.019,0.037]}{+0.028} x_1 \underset{[-0.040,-0.023]}{-0.031} x_2$ $\mathbb{E}(y_2\vert\mathbf{x}) = 1.69 \underset{[-0.037,-0.020]}{-0.028} x_1 \underset{[0.020,0.037]}{+0.029} x_2$	$\mathbb{E}(\widehat{y}\vert\widehat{x}) = 0.68 \underset{[0.002,0.020]}{+0.010} \widehat{x}$
3	$\mathbb{E}(y_1\vert\mathbf{x}) = 1.73 \underset{[0.079,0.096]}{+0.088} x_1 \underset{[-0.108,-0.091]}{-0.099} x_2$ $\mathbb{E}(y_2\vert\mathbf{x}) = 1.65 \underset{[-0.138,-0.121]}{-0.130} x_1 \underset{[0.122,0.140]}{+0.131} x_2$	$\mathbb{E}(\widehat{y}\vert\widehat{x}) = 0.67 \underset{[-0.006,0.011]}{+0.002} \widehat{x}$

Table 2: Sorting Patterns

in an optimal way if there are no search frictions (i.e. if $\lambda_0 \to 0$ and $\lambda_1 \to 0$).

For the 2D model in the top part of the table, the rows display in that order: average actual surplus, $\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]$, average optimal surplus, $\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]$, average surplus when implementing the 1D optimal allocation, $\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]$, the percentage output loss from mismatch $\frac{\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]-\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]}{\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]}$, the percentage surplus gain/loss from implementing the optimal 1D allocation relative to the *actual* 2D surplus $\frac{\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]-\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]}{\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]}$, and the percentage surplus loss from implementing the optimal 1D allocation relative to the *optimal* 2D surplus $\frac{\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]-\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]}{\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]}$. For the 1D model in the lower part of the Table 3, we use analogous notation.

Four features stand out: First, the surplus loss from mismatch can be sizable in both models, roughly ranging between 9-11%. Taking the surplus loss as a measure of mismatch, the 1D model overestimates mismatch in two out of three specifications relative to the 2D model. Second, the results indicate that welfare losses from implementing an erroneous optimal allocation (here given by the 1D frictionless allocation) in the 2D economy can be substantial. The economy loses between 7.8% and 10.5% of surplus when implementing the optimal 1D allocation instead of the optimal 2D allocation. Third, implementing the 1D frictionless (i.e. optimal) allocation generates basically no welfare gains when compared to the *actual* 2D allocation *with* frictions. Last, and importantly, the comparison across model specifications suggests that surplus losses from implementing the optimal 1D allocation tend to become larger in environments where the true production function features not only asymmetries in the technology matrix $\mathbf{Q}$ (Example 2) but especially where it contains both super- and sub-modular elements (Example 3). In such settings, the 1D model produces estimates of cross-partials of the surplus function as well as sorting patterns that are particularly off compared to the truth, giving rise to a significantly distorted first best allocation compared to the true first best under multidimensional types.

Specification:	**1**	**2**	**3**
True 2D model			
$\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]$	9.20	27.59	11.58
$\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]$	10.11	30.45	13.04
$\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]$	9.33	27.95	11.67
Surplus Loss from Mismatch	-0.09	-0.09	-0.11
Surplus Gain from 1D Optimal Allocation rel. to 2D Allocation	0.01	0.01	0.01
Surplus Loss from 1D Optimal Allocation rel. to 2D Optimum	-0.08	-0.08	-0.11
Misspecified 1D model			
$\mathbb{E}[\sigma(\widehat{x},\widehat{y})]$	9.06	27.52	11.44
$\mathbb{E}[\sigma^*(\widehat{x},\widehat{y})]$	10.13	30.49	12.84
Surplus Loss from Mismatch	-0.11	-0.10	-0.11

Table 3: Expected Surplus and Mismatch

6.2.5 Exercise 4: Sorting on Specialization

In Section 5, we have shown that the multi-dimensional model predicts sorting based on *specialization* rather than on absolute advantage and, moreover, that sorting depended on the balance between a worker's different skills relative to the baseline productivity vector $\mathbf{a}$. In the extreme case where $a_1 = a_2$, there is no re-sorting of "generalist" workers (with $x_1 = x_2$) at all in response to a uniform improvement of their skills, whereas "specialist" workers (with $x_1 \neq x_2$) will respond to such a uniform improvement by sorting into jobs with higher attributes in the skill dimension they are relatively strong in, e.g. dimension 1 if $x_1 > x_2$. We will show that the one-dimensional model fails to generate these patterns: first, it picks up one-dimensional sorting patterns in line with sorting on absolute advantage. Second, due to scalar heterogeneity it cannot distinguish between specialists and generalists.

For this exercise, we fix $\mathbf{a} = (0.1, 0.1)$ and focus on the technology $\mathbf{Q}$ from Example 2 (neither choice is crucial). We focus on the way the mean job types a worker is matched to in equilibrium change as we scale up workers' skills. We compute these conditional means from the simulated data and report some examples in Table 4. Specifically, for a worker with skill bundle $\mathbf{x}$ and estimated one-dimensional skill index $\widehat{x}$, we report $\mathbb{E}(\mathbf{y}|t\mathbf{x})$ and $\mathbb{E}(\widehat{y}|t\widehat{x})$ as the scaling factor t takes on values $t = 1$, 1.2, 1.3. Table 4 shows results for two workers: a "generalist" worker with

	Generalist: $\mathbf{x} = (0.3, 0.3)$			Specialist: $\mathbf{x} = (0.1, 0.5)$		
$t =$	1	1.2	1.3	1	1.2	1.3
$\mathbb{E}(y_1 \mid t\mathbf{x}) =$	1.6860	1.6864	1.6866	1.7090	1.7140	1.7165
$\mathbb{E}(y_2 \mid t\mathbf{x}) =$	1.6843	1.6837	1.6833	1.6621	1.6570	1.6545
$\mathbb{E}(\widehat{y} \mid t\widehat{x}) =$	0.6741	0.6743	0.6744	0.6741	0.6743	0.6744

The corresponding estimated 1D skill types are $\widehat{x} = 0.21$ and $\widehat{x} = 0.23$, respectively.

Table 4: Sorting on Comparative versus Absolute Advantage

$\mathbf{x} = (0.3, 0.3)$ and a "specialist" worker with $\mathbf{x} = (0.5, 0.1)$. Those two workers look very similar under the lens of the 1D model, with $\widehat{x} = 0.21$ and $\widehat{x} = 0.23$, respectively.

As the multi-dimensional model predicts, for a generalist there is no significant change in sorting on either dimension when scaling up the skill vector. Yet for the specialist, who in this example has comparative strength in skill 1 (as $x_1 > x_2$), a homothetic increase in all skills cause a discernible increase in mean job attribute *in the dimension where he is stronger (y_1)*, but to a decrease in mean job attribute in the other dimension (y_2): as predicted by our theoretical model, sorting is based on comparative advantage. Scaling up all skills does *not* lead to better job attributes across the board, which would be the case if sorting were based on absolute advantage.

The misspecified one-dimensional model misses those differences completely. The two workers in this example look almost the same to the one-dimensional model, so that the predicted path of $\mathbb{E}(\widehat{y} \mid t\widehat{x})$ as t increases it the same for both workers. Moreover, the one-dimensional model fails to pick up any sorting in this case, reflecting the tension between the underlying true improvement in y_1 and deterioration in y_2.

6.3 Taking Stock

Our various exercises highlight the parts of the estimation and its implications that differ considerably between the true multi-dimensional model and the misspecified one-dimensional model. The one-dimensional model collapses the multiple dimensions into a single index, making it difficult to correctly estimate the cross-partial derivative of the surplus function and leading to sorting patterns that differ substantially from the true multi-dimensional ones. As a consequence, the one-dimensional model tends to over-estimate mismatch and generates substantial welfare losses when the one-dimensional optimal matching is implemented as opposed to the

	True Value	Misspecified 1D model 1	2	3
λ_0	0.3	0.2998	0.2998	0.2998
λ_1	0.1	0.0967	0.1020	0.1018

Table 5: Estimated transition parameters

optimal two-dimensional allocation. Our exercises further suggest that the mistakes of the one-dimensional approximation become more severe when the cross-partials of the true surplus function are sign-varying. We conclude that multi-dimensional heterogeneity is crucial for estimating complementarities, sorting and mismatch.

It is important to note, however, that there are other parts of the estimation on which the one-dimensional (misspecified) model performs well. Table 5 shows, the values of λ_0 and λ_1 as estimated from the one-dimensional model (along with their true values, for comparison).[24] Clearly, the one-dimensional model gets the transition parameters right. This is unsurprising, as the estimators of λ_0 and λ_1 presented in Section 6.1.2 primarily exploit job transition data, which are independent of any assumption on the dimensionality of job or worker heterogeneity. Note that this result relies on our specific stepwise estimation protocol, which estimates λ_0 and λ_1 separately from the rest of the model. Application of a (more efficient) one-step method, such as indirect inference, would probably result in the misspecification of the dimensionality of $\mathbf{x}$ and $\mathbf{y}$ "polluting" the estimates of the transition parameters.

7 The Literature

Our work relates to a vast literature on partial equilibrium models with search frictions and random search as well as to the literature on conditions for sorting in a variety of environments.

Random Search Models. Our environment closely resembles that of a standard (partial equilibrium) search model with random search on and off the job (e.g. Postel-Vinay and Robin, 2002). The only departure from the standard model is that we introduce *multi-dimensional heterogeneity* of jobs and workers. This simple change drastically alters the model's predictions on sorting. While in both settings, the strategy of firms is to accept any worker that yields a

[24]The job loss rate δ is not estimated in this exercise: it is set to its true value when performing estimation. This is just to save time and space: a straightforward and consistent estimator of δ is the empirical job loss rate, which is independent of any assumption on the dimensionality of job or worker heterogeneity.

positive surplus, workers' incentives to sort differ across models. In the one-dimensional model, if the technology is monotone in job type, all worker types share a common ranking of firms and want be employed in the most productive firm. This implies that all workers tend to gradually select into more productive jobs over time in the exact same way, ruling out sorting (Postel-Vinay and Robin, 2002).[25] This result is independent of the complementarities in production.

This contrasts starkly with our multi-dimensional setting, in which what matters to a worker is to obtain a job requiring much of the specific skill in which he is particularly strong. This causes different workers to rank jobs differently, which is why sorting arises. This trade-off across skill dimensions is absent by design from the one-dimensional model. It is important to note that our model's predictions on sorting differ from the standard model *only* because we introduce multi-dimensional heterogeneity.

Conditions for Sorting. Becker (1973) established the first results on sorting in frictionless environments with TU: matching is positive assortative if the match payoff function is supermodular, highlighting the crucial role of *complementarities* for sorting.[26] Legros and Newman (2007) subsequently extended this sorting framework to the case of imperfectly TU (where utility cannot be transferred at a constant rate) and showed that PAM obtains if the Pareto frontier exhibits *generalized increasing differences*, which is essentially a single-crossing property that nests the TU case when utility is linear.

More recently, the literature has moved on to environments affected by search frictions, or to frictionless environment where agents are characterized by multi-dimensional heterogeneity.

Lindenlaub (2014) develops a framework for the analysis of multi-dimensional sorting in the TU, frictionless context.[27] Since workers and firms match on *bundles* of attributes, the one-dimensional Beckerian notion of PAM, given by strict monotonicity of the (real-valued) matching function, needs to be modified. Assuming that types are continuously distributed, she defines PAM as the Jacobian matrix of the matching function being a *P-matrix* which, like in the one-dimensional case, reflects a pure matching with positive sign on the 'natural' sorting dimensions.

[25] Moreover, in the one-dimensional model there cannot be sorting on the NE margin either since any job with $y < b$ yields a negative surplus independent of worker type.

[26] Becker also shows that in the case of strict NTU, the monotonicity of payoffs in partners' types is sufficient for sorting. Legros and Newman (2010) show that a necessary and sufficient condition for sorting in this environment is that preferences exhibit *co-ranking*.

[27] Note that there is an entire mathematical literature concerned with assignment problems under (possibly) multi-dimensional heterogeneity in frictionless TU settings (see Villani, 2009 for an extensive review on *optimal transport*), which is however not concerned with *sorting*. Also note that contrary to optimal transport problems, our problem is strictly speaking not an assignment problem since firms face no capacity constraint.

PAM occurs in equilibrium if the production function exhibits complementarities in types within these natural dimensions (and no complementarities across the natural dimensions).[28]

There has also been growing interest in sorting in frictional environments, not least because of their empirical relevance. In settings with search frictions and *directed* search, the definition of sorting essentially remains the same as in the frictionless case. The reason is that under strong enough complementarities, the directed-search equilibrium is characterized by *pure* assignments, generating perfect segmentation of types just as in the frictionless cases.[29]

However, when agents face a *random* search technology, matching is generically not pure: instead, each worker type matches with a distribution of job types in equilibrium. Shimer and Smith (2000) were the first to analyze sorting in this context under TU. They did so in a dynamic, one-to-one (i.e. both sides face a capacity constraint) equilibrium matching model where agents randomly search for partners. Surplus is split by Nash-bargaining and there is no on-the-match search. Shimer and Smith define positive sorting by the requirement that the boundaries of matching sets be (weakly) increasing in types. They show that the occurrence of PAM again hinges on complementarity of match output in types, although the complementarities needed for PAM to occur are stronger in this environment than in the frictionless case.[30]

To our knowledge, ours is the first analysis combining random search and multi-dimensional heterogeneity — two features that are critical to empirical applications. Our definition of positive sorting is based on dimension-by-dimension first-order stochastic ordering of matching distributions.[31] This definition is not equivalent to the Shimer-Smith definition: strict PAM can occur according to our FOSD-based criterion even when the matching sets of workers are invariant to their types (which is the case, for example, when the NE margin is shut down and all workers accept to match with *any* job). In that sense, FOSD is weaker than the Shimer-Smith criterion of sorting: in fact, increasing matching sets implies FOSD.

We establish conditions on the economy's primitives under which sorting obtains in this environment. In line with the existing literature, we find that complementarity between worker and job attributes are crucial to generate PAM. In our framework, those complementarity require-

[28]Formally, if the matrix of cross-partials is a diagonal dominant P-matrix then sorting in equilibrium is PAM. This condition is distribution-free. Note that the *natural* sorting dimensions have to be identified ex-ante.

[29]See Eeckhout and Kircher (2010) for sorting under directed search with one-dimensional heterogeneity and Lindenlaub (2014) for an extension to multi-dimensional heterogeneity.

[30]Their conditions require supermodularity of match output $f(x, y)$ ($f_{xy} > 0$, where x and y are types), log-supermodularity of its marginal product ($(\ln f_x)_{xy} > 0$) and of its cross-partial ($(\ln f_{xy})_{xy} > 0$). For conditions for sorting under random search and NTU, see e.g. Burdett and Coles (1997).

[31]Note that first-order stochastic dominance has been used to characterize sorting in frictional settings with one-dimensional heterogeneity (e.g. Chade, 2006).

ments take the form of an intuitive *single-crossing* condition on the technology. Our restriction on the technology (single-crossing) most closely resembles the condition in Lindenlaub (2014) who also studies multi-dimensional sorting and needs to discipline complementarities in competing dimensions. But our conditions are quite distinct from those needed under one-dimensional heterogeneity and random search (Shimer and Smith, 2000): In contrast to this literature on sorting in one dimension, we find that the conditions for PAM in multiple dimensions under random search are not distribution-free: they involve not only sufficient complementarities in types but also restrictions on the sampling distribution.

8 Conclusion

This paper analyzes sorting in a standard market environment characterized by search frictions and random search, but where both workers and jobs have *multi-dimensional* characteristics. We first offer a definition for multi-dimensional positive (and negative) assortative matching in this frictional environment that is based on first-order stochastic dominance of the matching distribution of a more skilled worker compared to a less skilled worker. We then provide conditions on the primitives of this economy (technology and distributions) under which positive sorting obtains in equilibrium, where we distinguish sorting on the nonemployment-to-employment and the employment-to-employment margin. In all the environments we consider, the central restriction on the primitives for PAM to obtain is a *single-crossing* condition of the technology that guarantees sufficient complementarities between skills and productivities. But, contrary to well-known results on one-dimensional sorting, our conditions for multi-dimensional sorting are generally *not* distribution-free. *Negative correlation* between different job attributes in the sampling distribution can become a barrier to PAM in a certain worker-job dimension if the worker's skills are complement to various job characteristics.

Our theory has important implications for applied work: We show in a series of simulation exercises that approximating workers' and jobs' true multi-dimensional characteristics by one-dimensional summary indices in empirical applications may lead to quantitatively and even qualitatively mistaken conclusions regarding the sign and extent of sorting and mismatch, as well as to misguided policy recommendations.

References

[1] Becker, G. S. (1973) "A Theory of Marriage: Part I", *Journal of Political Economy*, **81**(4), 813-46.

[2] Burdett, K. and M. G. Coles (1997) "Marriage and Class", *The Quarterly Journal of Economics*, **112**(1), 141-168.

[3] Burdett, K. and D. T. Mortensen (1998) "Wage Differentials, Employer Size and Unemployment", *International Economic Review*, **39**, 257-73.

[4] Cahuc, P., F. Postel-Vinay and J.-M. Robin (2006) "Wage Bargaining with On-the-job Search: Theory and Evidence", *Econometrica*, **74**(2), 323-64.

[5] Chade, H. (2006) "Matching with Noise and the Acceptance Curse", *Journal of Economic Theory*, **129**, 81-113.

[6] Chade, H., J. Eeckhout and L. Smith (2016) "Sorting Through Search and Matching Models in Economics", *manuscript*.

[7] Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2015) "Multidimensional Skill Mismatch", *manuscript*, University of Minnesota.

[8] Eeckhout, J. and P. Kircher (2010) "Sorting and Decentralized Price Competition", *Econometrica*, **78**, 539-74.

[9] Hagedorn, M., T. Law and I. Manovksii (2014) "Identifying Equilibrium Models of Labor Market Sorting", NBER Working Paper No. 18661.

[10] Karlin, S. and Y. Rinott (1980) "Classes of Orderings of Measures and Related Correlation Inequalities. I - Multivariate Totally Positive Distributions", *Journal of Multivariate Analysis*, **10**, 467-98.

[11] Lamadon, T., J. Lise, C. Meghir and J.-M. Robin (2015) "Matching, Sorting, and Wages", *manuscript*, University of Chicago.

[12] Legros, P. and A. F. Newman (2007) "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities", *Econometrica*, **75**(4), 1073-102.

[13] Lindenlaub, I. (2014) “Sorting Multidimensional Types: Theory and Application”, *manuscript*, Yale University.

[14] Lise, J., and F. Postel-Vinay (2015) “Multidimensional Skills, Sorting, and Human Capital Accumulation”, *manuscript*, University College London.

[15] Mortensen, D. T. and C. A. Pissarides (1994) “Job Creation and Job Destruction in the Theory of Unemployment”, *Review of Economic Studies*, **61**(3), 397-415.

[16] Moscarini, G. (2001) “Excess Worker Reallocation”, *Review of Economic Studies*, **69**(3), 593-612.

[17] Moscarini, G. and F. Postel-Vinay (2013) “Stochastic Search Equilibrium”, *Review of Economic Studies*, **80**(4), 1545-81.

[18] Postel-Vinay, F. and J.-M. Robin (2002) “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity”, *Econometrica*, **70**(6), 2295-350.

[19] Sanders, C. (2012) “Skill Uncertainty, Skill Accumulation, and Occupational Choice”, *manuscript*, Washington University in St. Louis.

[20] Shapiro, A., D. Dentcheva and A. Ruszczyński (2009) *Lectures on Stochastic Programming - Modeling and Theory*, MPS-SIAM series on optimization, **9**.

[21] Shimer, R., and L. Smith (2000) “Assortative Matching and Search”, *Econometrica*, **68**(2), 343-69.

[22] Villani, C. (2009) *Optimal Transport: Old and New*. Springer, Berlin.

[23] Yamaguchi, S. (2012) “Tasks and Heterogeneous Human Capital”, *Journal of Labor Economics*, **30**(1), 1?53.

APPENDIX

A Proofs and Derivations

A.1 Derivation of $h(\mathbf{x},\mathbf{y})$

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

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

This ODE solves as:

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

In other words, by differentiation:

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

Finally, remembering from the flow-balance equations that $\lambda_0 \overline{F}_{\sigma|\mathbf{x}}(0) u(\mathbf{x}) = \delta\left(\ell(\mathbf{x}) - u(\mathbf{x})\right)$ and substituting yields the expression of $h(\mathbf{x},\mathbf{y})$ in the main body of the paper.

A.2 Proof of Theorem 1

Recalling equation (4):

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

and differentiating yields:

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

We examine those three terms in turn.

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

$$\begin{aligned}
\frac{\partial}{\partial x_j}\left[1 - F_{\sigma|\mathbf{x}}(s)\right] &= \int \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\}\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_j}\gamma(\mathbf{y}')d\mathbf{y}' \\
&= \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_j} \mid \sigma(\mathbf{x},\mathbf{y}') = s\right] \times F'_{\sigma|\mathbf{x}}(s). \qquad (11)
\end{aligned}$$

Replacing into term (1) yields:

$$(1) = -\frac{\delta F'_{\sigma|\mathbf{x}}(0)}{\overline{F}_{\sigma|\mathbf{x}}(0)\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]} \times \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}') = 0\right] \times H_j(y|\mathbf{x}).$$

Next, term (2) can be rewritten as:

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

Now on to term (3). Again from (11), we have that:

$$\frac{d}{dx_k}\left[1 - F_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}')\right)\right] = F'_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}')\right) \times \left\{\mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}'')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}'') = \sigma(\mathbf{x},\mathbf{y}')\right] - \frac{\partial \sigma \mathbf{x},\mathbf{y}')}{\partial x_k}\right\}.$$

Substituting into term (3):

$$\begin{aligned}
(3) = \frac{\delta\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int & \frac{2\lambda_1 \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') \geq 0\} \times \mathbf{1}\left\{y'_j \leq y\right\} \times F'_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}')\right)}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}')\right)\right]^3} \\
& \times \left\{\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}'')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}'') = \sigma(\mathbf{x},\mathbf{y}')\right]\right\}\gamma(\mathbf{y}')d\mathbf{y}'
\end{aligned}$$

which can be recast as:[32]

$$\begin{aligned}
(3) = \frac{\delta\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty} & \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(s)\right]} \times \int \frac{\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\} \times \mathbf{1}\left\{y'_j \leq y\right\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(s)\right]^2}\gamma(\mathbf{y}')d\mathbf{y}' \\
& \times \left\{\mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}') = s, y'_j \leq y\right] - \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}') = s\right]\right\}ds
\end{aligned}$$

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

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

rather than $[0,+\infty)$. To avoid cluttering the formula with these unwieldy integration bounds, we write it as an integral over all $s \geq 0$. As a consequence, it may be that the joint event $\left(\sigma(\mathbf{x},\mathbf{y}') = s, y'_j \leq y\right)$ on which some of the expectations are conditioned has zero probability for some values of (s,y). Yet in those cases, $\int \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\} \times \mathbf{1}\left\{y'_j \leq y\right\}\gamma(\mathbf{y}')d\mathbf{y}' = 0$. The formula thus remains correct with $[0,+\infty)$ as integration bounds if adopt the convention that any expectation conditioned on a zero probability event is equal to zero.

$$= \int_0^{+\infty} \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(s)\right]} \times \frac{\partial K_j(y, s|\mathbf{x})}{\partial s}$$
$$\times \left\{ \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = s, y'_j \leq y \right] - \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = s \right] \right\} ds.$$

Combining terms (1), (2) and (3) and substituting the definitions of $G'_{\sigma|\mathbf{x}}(0)$ and $\partial K_j(y, 0|\mathbf{x})/\partial s$ proves the theorem. □

A.3 EE-Sorting: Proofs of Theorems 2 and 4 and Corollaries 3 and 4

To avoid duplication, we first prove the most general result, Theorem 4, then return to the proofs of the special cases (bilinear surplus and/or $Y = 2$).

A.3.1 Proof of Theorem 4 (Monotone Technology, $Y \geq 2$)

Let $\mathbf{y}_{-Y} = (y_1, \cdots, y_{Y-1})$ denote the $(Y-1)$-dimensional vector formed of all components of $\mathbf{y}$ except y_Y. Note that $\mathcal{Y}_{-Y} \in \bigtimes_{j=1}^{Y-1} \left[\underline{y}_j, \overline{y}_j\right]$.

Fix any $\mathbf{x} \in \mathcal{X}$ and any $s \geq 0$, and consider the equation

$$\sigma\left(\mathbf{x}, (\mathbf{y}_{-Y}, y_Y)\right) = s \Leftrightarrow f\left(\mathbf{x}, (\mathbf{y}_{-Y}, y_Y)\right) = b(\mathbf{x}) + s \tag{12}$$

Then strict monotonicity of $y_Y \mapsto f\left(\mathbf{x}, (\mathbf{y}_{-Y}, y_Y)\right)$ (Assumption 2a) guarantees that at most one value of $y_Y \in \left[\underline{y}_Y, \overline{y}_Y\right]$ solves (12). In turn, assumption 2b ensures that there *always* exists a unique $y_Y \in \left[\underline{y}_Y, \overline{y}_Y\right]$ that solves (12). The equation $\sigma(\mathbf{x}, \mathbf{y}) = s$ is therefore equivalent to $y_Y = R(s, \mathbf{y}_{-Y})$, where $R(s, \cdot)$ is a well-defined function over $\bigtimes_{j=1}^{Y-1} \left[\underline{y}_j, \overline{y}_j\right]$.[33] Application of the Implicit Function Theorem further implies that $R(\cdot)$ is differentiable over its domain, with, for all $j \in \{1, \cdots, Y-1\}$:

$$\frac{\partial R(S, \mathbf{y}_{-Y})}{\partial y_j} = -\frac{\partial f/\partial y_j}{\partial f/\partial y_Y}\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right) \tag{13}$$

and:

$$\frac{\partial R(s, \mathbf{y}_{-Y})}{\partial s} = \frac{1}{\partial f/\partial y_Y}\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right) \tag{14}$$

We next notice that, for $j \in \{1, \cdots, Y-1\}$:

$$\frac{d}{dy_j}\left(\frac{\partial \sigma\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right)}{\partial x_k}\right) = \frac{\partial^2 f}{\partial x_k \partial y_j} - \frac{\partial f/\partial y_j}{\partial f/\partial y_Y} \times \frac{\partial^2 f}{\partial x_k \partial y_Y} \tag{15}$$

where (13) was used and where the arguments of $f\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right)$ were omitted in the r.h.s. to reduce notational clutter. Condition (SC-Yd) in the theorem then implies, together with (13) that $\partial \sigma\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right)/\partial x_k$ is increasing in all elements of $\mathbf{y}_{-Y} \in \bigtimes_{j=1}^{Y-1} \left[\underline{y}_j, \overline{y}_j\right]$.

[33]The dependence of $R(\cdot)$ on $\mathbf{x}$, which is fixed for this proof, is omitted.

Our objective is to exhibit conditions under which condition (CMP) in Corollary 1 holds. Fixing $\ell \neq Y$, notice that (CMP) can be expressed as

$$\mathbb{E}_{\Gamma}\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}) = s, y_\ell = y\right] = \mathbb{E}_{\mu_{s,\mathbf{x}}}\left(\frac{\partial \sigma\left(\mathbf{x},(\mathbf{y}_{-Y}, R(s,\mathbf{y}_{-Y}))\right)}{\partial x_k} \mid y_\ell = y\right).$$

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

$$\mu_{s,\mathbf{x}}(\mathbf{y}_{-Y}) = \frac{\gamma\left(\mathbf{y}_{-Y}, R(s,\mathbf{y}_{-Y})\right) \partial R(s,\mathbf{y}_{-Y})/\partial s}{\int \gamma\left(\mathbf{y}'_{-Y}, R(s,\mathbf{y}'_{-Y})\right) \partial R(s,\mathbf{y}'_{-Y})/\partial s \, d\mathbf{y}'_{-Y}}$$

with $\partial R(s,\mathbf{y}_{-Y})/\partial s$ given in (14). We will now derive conditions under which

$$y \mapsto \mathbb{E}_{\mu_{s,\mathbf{x}}}\left(\frac{\partial \sigma\left(\mathbf{x},(\mathbf{y}_{-Y}, R(s,\mathbf{y}_{-Y}))\right)}{\partial x_k} \mid y_\ell = y\right) \tag{16}$$

is increasing in y and proceed in two steps.

First, we show that the support of $\mu_{s,\mathbf{x}}$ is a lattice. The equation $\sigma(\mathbf{x},\mathbf{y}) = s$ has one unique solution $y_Y = R(s,\mathbf{y}_{-Y}) \in \left[\underline{y}_Y, \overline{y}_Y\right]$ for all $\mathbf{y}_{-Y} \in \bigtimes_{j=1}^{Y-1}\left[\underline{y}_j, \overline{y}_j\right]$. Thus, for any two points $(\mathbf{y}'_{-Y}, \mathbf{y}''_{-Y}) \in \left(\bigtimes_{j=1}^{Y-1}\left[\underline{y}_j, \overline{y}_j\right]\right)^2$, $R(s,\mathbf{y}'_{-Y} \wedge \mathbf{y}''_{-Y})$ and $R(s,\mathbf{y}'_{-Y} \vee \mathbf{y}''_{-Y})$ both exist and are both elements of $\left[\underline{y}_Y, \overline{y}_Y\right]$, since both $\mathbf{y}'_{-Y} \wedge \mathbf{y}''_{-Y}$ and $\mathbf{y}'_{-Y} \vee \mathbf{y}''_{-Y}$ are elements of $\bigtimes_{j=1}^{Y-1}\left[\underline{y}_j, \overline{y}_j\right]$. This establishes that

$$\begin{aligned}
\left(\mathbf{y}'_{-Y} \wedge \mathbf{y}''_{-Y}, R(s,\mathbf{y}'_{-Y} \wedge \mathbf{y}''_{-Y})\right) &\in \bigtimes_{j=1}^{Y}\left[\underline{y}_j, \overline{y}_j\right] \\
\left(\mathbf{y}'_{-Y} \vee \mathbf{y}''_{-Y}, R(s,\mathbf{y}'_{-Y} \vee \mathbf{y}''_{-Y})\right) &\in \bigtimes_{j=1}^{Y}\left[\underline{y}_j, \overline{y}_j\right],
\end{aligned}$$

so that $\operatorname{Supp}\mu_{s,\mathbf{x}}$ is a lattice, implying that $\mu_{s,\mathbf{x}}\left(\mathbf{y}'_{-Y} \wedge \mathbf{y}''_{-Y}\right)$ and $\mu_{s,\mathbf{x}}\left(\mathbf{y}'_{-Y} \vee \mathbf{y}''_{-Y}\right)$ are strictly positive.

Second, given increasing monotonicity of $\partial\sigma\left(\mathbf{x},(\mathbf{y}_{-Y}, R(s,\mathbf{y}_{-Y}))\right)/\partial x_k$ in all elements of $\mathbf{y}_{-Y} \in \bigtimes_{j=1}^{Y-1}\left[\underline{y}_j, \overline{y}_j\right]$ and given that $\operatorname{Supp}\mu_{s,\mathbf{x}}$ is a lattice (both established above), a sufficient condition for (16) to be an increasing function of y is that the density $\mu_{s,\mathbf{x}}$ be such that (the proof is a simple adaptation of Theorem 4.1 in Karlin and Rinott, 1980):

$$\forall (i,j) \in \{1,\cdots,Y-1\}^2 : i \neq j, \quad \frac{\partial^2 \ln \mu_{s,\mathbf{x}}}{\partial y_i \partial y_j} \geq 0$$

which translates into condition (EE-Yd):

$$\frac{\partial f}{\partial y_Y}\left[\left(\frac{\partial f}{\partial y_Y}\right)^2\frac{\partial^2\ln\gamma}{\partial y_i\partial y_j}+\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\frac{\partial^2\ln\gamma}{\partial y_Y^2}-\frac{\partial f}{\partial y_j}\frac{\partial f}{\partial y_Y}\frac{\partial^2\ln\gamma}{\partial y_i\partial y_Y}-\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_Y}\frac{\partial^2\ln\gamma}{\partial y_j\partial y_Y}\right]$$
$$-\frac{\partial\ln\gamma}{\partial y_Y}\left[\left(\frac{\partial f}{\partial y_Y}\right)^2\frac{\partial^2 f}{\partial y_i\partial y_j}+\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\frac{\partial^2 f}{\partial y_Y^2}-\frac{\partial f}{\partial y_j}\frac{\partial f}{\partial y_Y}\frac{\partial^2 f}{\partial y_i\partial y_Y}-\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_Y}\frac{\partial^2 f}{\partial y_j\partial y_Y}\right]$$
$$-\left(\frac{\partial f}{\partial y_Y}\right)^2\frac{\partial^3 f}{\partial y_i\partial y_j\partial y_Y}-\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_j}\frac{\partial^3 f}{\partial y_Y^3}+\frac{\partial f}{\partial y_i}\frac{\partial f}{\partial y_Y}\frac{\partial^3 f}{\partial y_j\partial y_Y^2}+\frac{\partial f}{\partial y_j}\frac{\partial f}{\partial y_Y}\frac{\partial^3 f}{\partial y_i\partial y_Y^2}$$
$$+\frac{\partial f}{\partial y_Y}\left[\frac{\partial^2 f}{\partial y_j\partial y_Y}\frac{\partial^2 f}{\partial y_i\partial y_Y}-\frac{\partial^2 f}{\partial y_Y^2}\frac{\partial^2 f}{\partial y_i\partial y_j}\right]\geq 0.$$

A.3.2 Proof of Corollary 3 (Monotone Technology, $Y=2$)

When $Y=2$, equation (12) writes as $f\left(\mathbf{x},(y_1,y_2)\right)=b(\mathbf{x})+s$. Strict monotonicity of $y_2\mapsto f\left(\mathbf{x},(\mathbf{y}_1,y_2)\right)$ still guarantees that at most one value of $y_2\in\left[\underline{y}_2,\overline{y}_2\right]$ solves (12). Moreover, the set of y_1 such that (12) has one solution is an interval (possibly empty). To see this, suppose there exist two distinct values $y_1'<y_1''$ such that (12) has a solution. Denote these solutions by y_2' and y_2'', respectively. Consider a number $t\in(0,1)$ and define $y_1(t)=ty_1'+(1-t)y_1''$. Quasi-concavity of $f(\mathbf{x},\cdot)$ implies that $f\left(\mathbf{x},(y_1(t),ty_2'+(1-t)y_2'')\right)\geq b(\mathbf{x})+s$. Moreover, by assumption, $\min_{y_2\in\mathbb{R}}f\left(\mathbf{x},(y_1(t),y_2)\right)<b(\mathbf{x})\leq b(\mathbf{x})+s$. By continuity, (12) has a solution $R\left(s,y_1(t)\right)\leq ty_2'+(1-t)y_2''$ when $y_1=y_1(t)$. Note that this solution can be smaller than $\underline{y}_2$: in this case, $\gamma\left(y_1(t),R\left(s,y_1(t)\right)\right)=0$.

Denote the interval of values of y_1 for which (12) has one solution by $\mathcal{I}_1(s)$. Application of the Implicit Function Theorem (as in the proof of Theorem 4 above) then implies that equation (15) holds for all $y\in\mathcal{I}_1(s)$:

$$\frac{d}{dy_1}\left(\frac{\partial\sigma\left(\mathbf{x},(y,R(s,y))\right)}{\partial x_k}\right)=\frac{\partial^2 f}{\partial x_k\partial y_1}-\frac{\partial f/\partial y_1}{\partial f/\partial y_2}\times\frac{\partial^2 f}{\partial x_k\partial y_2}$$

which is positive by the single-crossing condition in Corollary 3.

Consider $s\geq 0$ and two values $(y',y'')\in\left[\underline{y}_1,\overline{y}_1\right]^2$ such that $\Pr_\Gamma\{y_1=y'|\sigma(\mathbf{x},\mathbf{y})=s\}>0$ and $\Pr_\Gamma\{y_1=y''|\sigma(\mathbf{x},\mathbf{y})=s\}>0$, which implies in particular that y' and y'' are both in $\mathcal{I}_1(s)$. Assume w.l.o.g. that $y''>y'$. Then, by the results derived above:

$$\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k}\mid\sigma(\mathbf{x},\mathbf{y})=s,y_1=y''\right]-\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k}\mid\sigma(\mathbf{x},\mathbf{y})=s,y_1=y'\right]$$
$$=\frac{\partial\sigma\left(\mathbf{x},(y'',R(s,y''))\right)}{\partial x_k}-\frac{\partial\sigma\left(\mathbf{x},(y',R(s,y'))\right)}{\partial x_k}>0.$$

This proves that Condition (CMP) holds for monotone production functions and $Y=2$, i.e. there is PAM along (y_1,x_k). □

Note in passing that Condition (EE-Yd) from Theorem 4 becomes irrelevant in the two-dimensional case $Y=2$, which obviates the need to impose conditions on the behavior of f over the support of γ: Condition 2b in Theorem 4 is replaced by the twofold requirement of quasi-concavity and $\min_{y_2\in\mathbb{R}}f(\mathbf{x},\mathbf{y})<$

$b(\mathbf{x})$.

A.3.3 Proof of Corollary 4 (Bilinear Technology, $Y \geq 2$)

Theorem 4 also nests the case of a bilinear technology from Corollary 4. The bilinear technology is C^3 (Condition 1 from Theorem 4). The condition $q_Y(\mathbf{x}) > 0$ (imposed in Assumption 1) ensures that $f(\mathbf{x}, \mathbf{y})$ is strictly increasing in y_Y (Condition 2a from Theorem 4). Moreover, with bilinear technology, we can explicitly solve the equation $\sigma(\mathbf{x}, \mathbf{y}) = s$, thus circumventing the need to appeal to the Implicit Function Theorem. Together, the requirements that $\overline{y}_Y = +\infty$ and $\lim_{y_Y \to \underline{y}_Y} f(\mathbf{x}, \mathbf{y}) < b(\mathbf{x})$ parallel Condition 2b from Theorem 4 and ensure that supp$\mu_{s,\mathbf{x}}$ is a lattice so that we can apply Theorem 4.1 in Karlin and Rinott (1980) to derive conditions under which (16) is increasing in y. Next, Condition 2 in the corollary is the generalized single crossing condition (which echoes Condition 2c in Theorem 4). Finally, Condition (EE-Yd') in point 3 of the corollary is a rewrite of Condition (EE-Yd) in Theorem 4 in the case of a bilinear production function: bilinearity implies that the last three lines in (EE-Yd) vanish (as all the partial derivatives of order greater than one are zero in this case) and that the derivatives of the flow surplus (or of the production) function can be expressed explicitly. □

A.3.4 Proof of Theorem 2 (Bilinear Technology, $Y = 2$)

Sufficiency. Sufficiency follows immediately from Theorem 4 and Corollary 4, where Assumption 3 from Corollary 4 vanishes for $Y = 2$. Thus, in the case of a bilinear technology with $Y = 2$, the sufficient condition for sorting (single-crossing) is distribution-free and guarantees that condition (CMP) holds.

Necessity. Assumption 1 (bilinear technology) implies $\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k = \sum_{j=1}^{2} q_{kj}(y_j - b_j)$. Assumption 1 further implies that, for any $s \geq 0$:

$$\sigma(\mathbf{x}, \mathbf{y}) = s \Leftrightarrow y_2 - b_2 = \frac{s}{q_2(\mathbf{x})} - \frac{q_1(\mathbf{x})}{q_2(\mathbf{x})}(y_1 - b_1).$$

Taken together, those implications yield the following expression for (CMP) in Corollary 1:

$$\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x}, \mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}) = s, y_1 = y\right] = \frac{q_{k2}}{q_2(\mathbf{x})}s + \frac{q_{k1}q_2(\mathbf{x}) - q_{k2}q_1(\mathbf{x})}{q_2(\mathbf{x})}(y - b_1).$$

Again defining $\mu_{s,\mathbf{x}}(y_1) = \gamma\left(y_1, b_2 + \frac{s}{q_2(\mathbf{x})} - \frac{q_1(\mathbf{x})}{q_2(\mathbf{x})}(y_1 - b_1)\right) / \int \gamma\left(y_1', b_2 + \frac{s}{q_2(\mathbf{x})} - \frac{q_j(\mathbf{x})}{q_2(\mathbf{x})}(y_1' - b_1)\right) dy_1'$, we have that:

$$\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x}, \mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}) = s, y_1 \leq y\right] = \frac{q_{k2}}{q_2(\mathbf{x})}s + \frac{q_{k1}q_2(\mathbf{x}) - q_{k2}q_1(\mathbf{x})}{q_2(\mathbf{x})}\mathbb{E}_{\mu_{s,\mathbf{x}}}\left[y_1 - b_1 | y_1 \leq y\right]$$

and term (2) in Theorem 1 is equal to:

$$-\frac{q_{k1}q_2(\mathbf{x})-q_{k2}q_1(\mathbf{x})}{q_2(\mathbf{x})}\int_0^{+\infty}\frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)}\frac{\partial K_j(y,s|\mathbf{x})}{\partial s}\left(\mathbb{E}_{\mu_{s,\mathbf{x}}}(y_1-b_1)-\mathbb{E}_{\mu_{s,\mathbf{x}}}\left[y_1-b_1|y_1\leq y\right]\right)ds.$$

Both $q_2(\mathbf{x})$ (by assumption) and the difference in expectations (by construction) are positive. Condition (SC-2d), which is equivalent in the case at hand to $q_{k1}q_2(\mathbf{x})-q_{k2}q_1(\mathbf{x})>0$, is therefore necessary and sufficient for term (2) in Theorem 1 to be negative. No condition on the sampling density γ is required.

□

A.4 NE-Sorting: Proof of Corollary 2 and Theorems 3 and 5

A.4.1 Proof of Corollary 2

The first term in Theorem 1 (which, as explained in the main text, reflects sorting along the NE margin) can be re-expressed as follows:

$$\Bigg[\Pr_\Gamma\left\{y'_j\leq y|\sigma(\mathbf{x},\mathbf{y}')=0\right\}\times\left\{\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\mid\sigma(\mathbf{x},\mathbf{y}')=0,y'_j\leq y\right]-\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\mid\sigma(\mathbf{x},\mathbf{y}')=0\right]\right\}$$
$$+\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\mid\sigma(\mathbf{x},\mathbf{y}')=0\right]\times\left[\Pr_\Gamma\left\{y'_j\leq y|\sigma(\mathbf{x},\mathbf{y}')=0\right\}-H_j(y|\mathbf{x})\right]\Bigg]\times G'_{\sigma|\mathbf{x}}(0)\quad(17)$$

The first term in curly brackets is negative under Condition (CMP). We now focus on the second term of (17). Noticing that

$$G'_{\sigma|\mathbf{x}}(0)\times\left[\Pr_\Gamma\left\{y'_j\leq y|\sigma(\mathbf{x},\mathbf{y}')=0\right\}-H_\ell(y|\mathbf{x})\right]$$
$$=G'_{\sigma|\mathbf{x}}(0)\times\int_0^{+\infty}G'_{\sigma|\mathbf{x}}(s)\left[\Pr_\Gamma\left\{y'_j\leq y|\sigma(\mathbf{x},\mathbf{y}')=0\right\}-\Pr_\Gamma\left\{y'_j\leq y|\sigma(\mathbf{x},\mathbf{y}')=s\right\}\right]ds$$

and multiplying by $\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\mid\sigma(\mathbf{x},\mathbf{y}')=0\right]$ yields the condition in the Corollary. □

A.4.2 Proof of Theorem 5 (Bilinear Technology, $Y\geq 2$)

To avoid duplication, we first prove the more general Theorem 5, which covers $Y=2$ as a special case. We have to sign the first term in Theorem 1 (which, as explained in the main text, reflects sorting along the NE margin) which is spelled out in (17). By the same arguments as were used in the proofs of Theorems 2 and 4, the first term in curly brackets in (17) is negative under Condition (EE-Yd'). We thus focus on the second term of (17).

First, we need to find conditions under which $\Pr_\Gamma\left\{y'_\ell\leq y|\sigma(\mathbf{x},\mathbf{y}')=s\right\}$ is decreasing in s for any fixed $\ell \in \{1, \cdots, Y-1\}$. In particular, we want to derive conditions under which, for any $s_H > s_L \geq 0$,

$$\frac{\int_{\underline{y}_\ell}^{y} \int \gamma \left(y'_\ell, \mathbf{y}'_{-(\ell,Y)}, \frac{s_H}{q_Y(\mathbf{x})} + b_Y - \sum_{j=1}^{Y-1} \frac{q_j(\mathbf{x})}{q_Y(\mathbf{x})}(y'_j - b_j) \right) d\mathbf{y}'_{-(\ell,Y)} dy'_\ell}{\int_{\underline{y}_\ell}^{\overline{y}_\ell} \int \gamma \left(y'_\ell, \mathbf{y}'_{-(\ell,Y)}, \frac{s_H}{q_Y(\mathbf{x})} + b_Y - \sum_{j=1}^{Y-1} \frac{q_j(\mathbf{x})}{q_Y(\mathbf{x})}(y'_j - b_j) \right) d\mathbf{y}'_{-(\ell,Y)} dy'_\ell}$$
$$\leq \frac{\int_{\underline{y}_\ell}^{y} \int \gamma \left(y'_\ell, \mathbf{y}'_{-(\ell,Y)}, \frac{s_L}{q_Y(\mathbf{x})} + b_Y - \sum_{j=1}^{Y-1} \frac{q_j(\mathbf{x})}{q_Y(\mathbf{x})}(y'_j - b_j) \right) d\mathbf{y}'_{-(\ell,Y)} dy'_\ell}{\int_{\underline{y}_\ell}^{\overline{y}_\ell} \int \gamma \left(y'_\ell, \mathbf{y}'_{-(\ell,Y)}, \frac{s_L}{q_Y(\mathbf{x})} + b_Y - \sum_{j=1}^{Y-1} \frac{q_j(\mathbf{x})}{q_Y(\mathbf{x})}(y'_j - b_j) \right) d\mathbf{y}'_{-(\ell,Y)} dy'_\ell}$$

where $\mathbf{y}'_{-(\ell,Y)} = \left(y'_1, \cdots, y'_{\ell-1}, y'_{\ell+1}, \cdots, y'_{Y-1}\right)$. Defining

$$g(y, s) = \int_{\underline{y}_\ell}^{y} \int \gamma \left(y'_\ell, \mathbf{y}'_{-(\ell,Y)}, \frac{s}{q_Y(\mathbf{x})} + b_Y - \sum_{j=1}^{Y-1} \frac{q_j(\mathbf{x})}{q_Y(\mathbf{x})}(y'_j - b_j) \right) d\mathbf{y}'_{-(\ell,Y)} dy'_\ell$$

and rearranging the previous inequality gives $g(\overline{y}_\ell, s_L) g(y, s_H) \leq g(y, s_L) g(\overline{y}_\ell, s_L)$. Since $y \leq \overline{y}_\ell$, this inequality holds if g is log-supermodular in (y, s). To show when this is the case, define the joint distribution of $\mathbf{y}_{-Y}$ and s (conditional on $\mathbf{x}$) as

$$\mu_{\mathbf{x}}(\mathbf{y}_{-Y}, s) = \gamma \left(\mathbf{y}_{-\mathbf{Y}}, \frac{s}{q_Y(\mathbf{x})} + b_Y - \sum_{j=1}^{Y-1} \frac{q_j(\mathbf{x})}{q_Y(\mathbf{x})}(y_j - b_j) \right)$$

and rewrite $g(y, s) = \int \mathbf{1}\{y'_\ell < y\} \mu_{\mathbf{x}} \left(\mathbf{y}'_{-Y}, s\right) d\mathbf{y}_{-\mathbf{Y}}'$. Note that

1. the support of $\mu_{\mathbf{x}}(\mathbf{y}_{-Y}, s)$ is a lattice;[34]

2. the joint distribution $\mu_{\mathbf{x}}(\mathbf{y}_{-Y}, s)$ is log-supermodular in $(\mathbf{y}_{-Y}, s)$ if it is log-supermodular in all pairs of arguments. This is the case if:

$$\forall j = \{1, ..., Y-1\} : \qquad q_Y(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_Y \partial y_j} - q_j(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_Y^2} \geq 0 \tag{18}$$

$$\forall (i, j) \in \{1, \cdots, Y-1\}^2, i \neq j :$$
$$q_Y(\mathbf{x})^2 \frac{\partial^2 \ln \gamma}{\partial y_i \partial y_j} + q_i(\mathbf{x}) q_j(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_Y^2} - q_j(\mathbf{x}) q_Y(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_i \partial y_Y} - q_i(\mathbf{x}) q_Y(\mathbf{x}) \frac{\partial^2 \ln \gamma}{\partial y_j \partial y_Y} \geq 0 \tag{19}$$

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

Therefore, the product $1\{y'_\ell < y\} \mu_{\mathbf{x}}(\mathbf{y}'_{-Y}, s)$ is log-supermodular in $(y, \mathbf{y}'_{-Y}, s)$ since the product of log-supermodular functions is log-supermodular). This implies in turn that $g(\cdot)$ is log-supermodular in (y, S)

[34] This follows from the proof of Theorem 4 with one extra step to deal with the joint distribution of $(\mathbf{y}_{-Y}, s)$ (instead of the conditional distribution of $\mathbf{y}_{-Y}$ given s). Note that we have proven that $y_Y = R(s, \mathbf{y}_{-Y}) \in \left[\underline{y}_Y, \overline{y}_Y\right]$ for *any* $s \geq 0$. Hence, for two points $s \geq 0$ and $s' \geq 0$, $s \wedge s' \geq 0$ and $s \vee s' \geq 0$ and

$$(\mathbf{y}_{-Y} \wedge \mathbf{y}'_{-Y}, R(s \wedge s', \mathbf{y}_{-Y} \wedge \mathbf{y}'_{-Y})) \in \bigtimes_{j=1}^{Y} \left[\underline{y}_j, \overline{y}_j\right] \quad \text{and} \quad (\mathbf{y}_{-Y} \vee \mathbf{y}'_{-Y}, R(s \vee s', \mathbf{y}_{-Y} \vee \mathbf{y}'_{-Y})) \in \bigtimes_{j=1}^{Y} \left[\underline{y}_j, \overline{y}_j\right].$$

Hence the support of $\mu_{\mathbf{x}}$ is a lattice and $\mu_{\mathbf{x}}(\mathbf{y}_{-Y} \wedge \mathbf{y}'_{-Y}, s \wedge s')$ and $\mu_{\mathbf{x}}(\mathbf{y}_{-Y} \vee \mathbf{y}'_{-Y}, s \vee s')$ are strictly positive. This is why we can take $\ln \mu_{\mathbf{x}}$ in the next step.

since log-supermodularity is preserved under integration.

We have thus proven that if (18) (stated as condition (NE-Yd) in Theorem 5) and (19) (stated as condition (EE-Yd') in Corollary 4) hold, then $\Pr_\Gamma \{y'_\ell \leq y | \sigma(\mathbf{x}, \mathbf{y}') = s\}$ is decreasing in s.

Second, we will provide a condition guaranteeing that $\mathbb{E}_\Gamma \left[\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k \mid \sigma(\mathbf{x}, \mathbf{y}) = 0\right] \leq 0$. First note that $\sigma(\mathbf{x}, \mathbf{y}) = 0$ is equivalent to $y_Y = b_Y - \frac{1}{q_Y(\mathbf{x})} \sum_{j=1}^{Y-1} q_j(\mathbf{x})(y_j - b_j)$. Thus:

$$\sigma(\mathbf{x}, \mathbf{y}) = 0 \implies \frac{\partial\sigma(\mathbf{x}, \mathbf{y})}{\partial x_k} = \sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} (y_j - b_j).$$

Therefore, a sufficient condition for $\mathbb{E}_\Gamma \left[\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k \mid \sigma(\mathbf{x}, \mathbf{y}) = 0\right] \leq 0$ is that the value of the linear program

$$\begin{aligned} & \max_{\mathbf{y}} \sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} (y_j - b_j) \\ \text{subject to:} \quad & \frac{1}{q_Y(\mathbf{x})} \sum_{j=1}^{Y-1} q_j(\mathbf{x})(y_j - b_j) \leq b_Y - \underline{y}_Y \\ & \underline{y}_j \leq y_j \;\; j = 1, \cdots, Y-1 \end{aligned}$$

be nonpositive. This is a sufficient condition which ensures that $\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k \leq 0$ over the *entire* set of $\mathbf{y}$'s such that $\sigma(\mathbf{x}, \mathbf{y}) = 0$. Although strong, this condition is the minimal 'distribution-free' one. This program can be rewritten in standard form as:

$$\begin{aligned} \sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} \left(\underline{y}_j - b_j\right) + & \max_{\mathbf{Y}} \sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} Y_j \\ \text{subject to:} \quad & \sum_{j=1}^{Y-1} q_j(\mathbf{x}) Y_j \leq -\sigma\left(\mathbf{x}, \underline{\mathbf{y}}\right) \\ & 0 \leq Y_j \;\; j = 1, \cdots, Y-1 \end{aligned} \tag{20}$$

where $Y_j = y_j - \underline{y}_j$. The first thing to note about program (20) is that if there exists a $j \in \{1, \cdots, Y-1\}$ such that $q_j(\mathbf{x}) < 0$, then (20) is clearly unbounded: in that case, one can set $Y_j \to +\infty$ whenever $q_j(\mathbf{x}) < 0$ and $Y_{j'} = 0$ when $q_{j'}(\mathbf{x}) \geq 0$, which satisfies the constraints and gives (20) an infinite value. We thus assume that $q_j(\mathbf{x}) \geq 0$ for all $j \in \{1, \cdots, Y-1\}$.

The dual of (20) is simply:

$$\begin{aligned} \sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} \left(\underline{y}_j - b_j\right) + & \min_{Z} \left\langle -Z\sigma\left(\mathbf{x}, \underline{\mathbf{y}}\right)\right\rangle \\ \text{subject to:} \quad & q_j(\mathbf{x}) Z \geq \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})}, \;\; j = 1, \cdots, Y-1 \\ & Z \geq 0 \end{aligned} \tag{21}$$

Assuming that $\sigma\left(\mathbf{x}, \underline{\mathbf{y}}\right) < 0$, the solution to the latter program is then to set:

$$Z = \max_{j'} \left\{ \frac{q_{kj'}}{q_{j'}(\mathbf{x})} \right\} - \frac{q_{kY}}{q_Y(\mathbf{x})}$$

and the value of the linear program (20) (which equals that of its dual) is:

$$\begin{aligned} \sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} \left(\underline{y}_j - b_j\right) - \sum_{j=1}^{Y} \left[\max_{j'} \left\{ \frac{q_{kj'}}{q_{j'}(\mathbf{x})} \right\} - \frac{q_{kY}}{q_Y(\mathbf{x})} \right] q_j(\mathbf{x}) \left(\underline{y}_j - b_j\right) \\ = \sum_{j=1}^{Y} \left[\frac{q_{kj}}{q_j(\mathbf{x})} - \max_{j'} \left\{ \frac{q_{kj'}}{q_{j'}(\mathbf{x})} \right\} \right] q_j(\mathbf{x}) \left(\underline{y}_j - b_j\right). \quad (22) \end{aligned}$$

The requirement that this be negative yields the condition in the theorem. □

A.4.3 Proof of Theorem 3 (Bilinear Technology, $Y = 2$)

In the special case of $Y = 2$, treated in Theorem 3, the first line in (17) is negative under the assumed condition (SC-2d) from Theorem 2. Moreover, the second line is nonpositive if (18) holds for $j = 1$ and $Y = 2$ (as stated by Assumption 2 in Theorem 3; note that condition (19) vanishes entirely for $Y = 2$), and if condition (22) is negative for $Y = 2$ which collapses to $\underline{y}_2 \geq b_2$ (Assumption 3 in Theorem 3). □

A.5 Sorting on Absolute Advantage vs Specialization: Proof of Theorem 7

From Theorem 1 applied in the case of a bilinear production function:

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

But then linearity in $(\mathbf{x}+\mathbf{a})$ of the flow surplus function $\sigma(\cdot)$ 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:

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

all terms of which are equal to zero. □

Note that the proof above is virtually unchanged if, instead of assuming that $\sigma(\cdot)$ is linear in $(\mathbf{x}+\mathbf{a})$, one only assumes 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.

A.6 Strength of Sorting: Proof of Theorem 8.

We first note that similar steps as were taken in the proof of Theorem 1 can be used to establish the following result about the way in which the marginal *density* of the jth job attribute in the population of firm-worker matches responds to a change in the kth worker attribute:

$$
\begin{aligned}
\frac{\partial H'_j(y|\mathbf{x})}{\partial x_k} &= G'_{\sigma|\mathbf{x}}(0) \Bigg\{ \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=0, y'_j = y\right] \Pr_\Gamma\left\{y'_j = y | \sigma(\mathbf{x},\mathbf{y})=0\right\} \\
&\qquad - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=0\right] H'_j(y|\mathbf{x}) \Bigg\} \\
&+ \int_0^{+\infty} \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}(s)} \times \frac{\partial^2 K_j(y,s|\mathbf{x})}{\partial y \partial s} \\
&\qquad \times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s, y'_j = y\right] - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right] \right\} ds.
\end{aligned}
$$

Assuming, as is done in the theorem, that the NE margin is shut down ($\sigma(\mathbf{x},\mathbf{y}) > 0$ for all $\mathbf{y}$), implies that the first two terms in the equation above are both equal to zero. Next consider:

$$
(\mathbf{x}+\mathbf{a})^\top \mathbf{Q}\partial_{x_k}\mathbb{E}(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i+a_i)q_{ij}\int_{\underline{y}_j}^{\overline{y}_j} y_j \frac{\partial H'_j(y_j|\mathbf{x})}{\partial x_k} dy_j
$$

for all $k \in \{1,\cdots,X\}$, where $\partial_{x_k}\mathbb{E}(\mathbf{y}|\mathbf{x}) = (\partial\mathbb{E}(y_1|\mathbf{x})/\partial x_k, \cdots, \partial\mathbb{E}(y_Y|\mathbf{x})/\partial x_k)^\top$. Using the expression for $\partial H'_j(y_j|\mathbf{x})/\partial x_k$ derived above, the definition of the bilinear flow surplus function from Assumption

1, and the definition of $K_j(y,s|\mathbf{x})$ from equation (6), we have that

$$
\begin{aligned}
(\mathbf{x}+\mathbf{a})^\top \mathbf{Q}\partial_{x_k}\mathbb{E}(\mathbf{y}|\mathbf{x}) &= \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i+a_i)q_{ij}\int_{\underline{y}_j}^{\overline{y}_j} y_j \int_0^{+\infty} \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)} \times \frac{\partial^2 K_j(y_j,s|\mathbf{x})}{\partial y \partial s} \\
&\quad \times \left\{ \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s, y'_j=y_j\right] - \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right]\right\} ds dy_j \\
&= \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i+a_i)q_{ij}\frac{\delta\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty}\frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3} \\
&\quad \times \int_{\underline{y}_j}^{\overline{y}_j}\left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\}\mathbf{1}\left\{y'_j=y_j\right\} y'_j \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}' \right.\\
&\quad \left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\}\mathbf{1}\left\{y'_j=y_j\right\} y'_j\gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right]\right\} dy_j ds \\
&= \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i+a_i)q_{ij}\frac{\delta\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty}\frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3} \\
&\quad \times \left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\} y'_j \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}' \right.\\
&\quad \left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\} y'_j\gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right]\right\} ds \\
&= \frac{\delta\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty}\frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3} \\
&\quad \times \left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\} \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i+a_i)q_{ij}y'_j \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}' \right.\\
&\quad \left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\} \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i+a_i)q_{ij}y'_j\gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right]\right\} ds \\
&= \frac{\delta\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty}\frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3} \times \left(s+(\mathbf{x}+\mathbf{a})^\top\mathbf{Q}\,\mathbf{b}\right) \\
&\quad \times \left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\} \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}' \right.\\
&\quad \left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\} \gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}')=s\right]\right\} ds = 0,
\end{aligned}
$$

as the term in curly brackets inside the integral is equal to zero. □

A.7 Proof of Theorem 6

Under the separability assumption 2 in the Theorem, $f(\mathbf{x},\mathbf{y}) = f_1(\mathbf{x},y_1)+f_2(\mathbf{x}_{-k},\mathbf{y})$ where $\mathbf{x}_{-k}$ includes all components of $\mathbf{x}$ but x_k. Hence:

$$\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} = \frac{\partial f_1(\mathbf{x},y_1)}{\partial x_k} - \frac{\partial b(\mathbf{x})}{\partial x_k}$$

which only depends on the first job attribute y_1. Then:

$$\mathbb{E}_{\Gamma}\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y})=s, y_1=y\right]=\frac{\partial f_1(\mathbf{x},y)}{\partial x_k}-\frac{\partial b(\mathbf{x})}{\partial x_k}$$

which is increasing in y under Assumption 3. Condition (CMP) thus holds, hence the result. $\square$

B Alternative Wage Setting Protocols

In this appendix, we explore generalizations of our results to wage setting protocols other than the pure Sequential Auction model of Postel-Vinay and Robin (2002).

For notational brevity, we first denote the surplus of a match as $S(\mathbf{x},\mathbf{y})=P(\mathbf{x},\mathbf{y})-U(\mathbf{x})$. Under any wage setting rule, so long as workers and firms have linear preferences over wages, the surplus and unemployment value functions are defined by:

$$(\rho+\delta)S(\mathbf{x},\mathbf{y})=f(\mathbf{x},\mathbf{y})-\rho U(\mathbf{x})+\lambda_1\int m(\mathbf{x},\mathbf{y},\mathbf{y}')\left[\Omega^m(\mathbf{x},\mathbf{y},\mathbf{y}')-S(\mathbf{x},\mathbf{y})\right]\gamma(\mathbf{y}')d\mathbf{y}' \tag{23}$$

$$\rho U(\mathbf{x})=b(\mathbf{x})+\lambda_0\int m(\mathbf{x},\mathbf{0}_Y,\mathbf{y}')\Omega^m(\mathbf{x},\mathbf{0}_Y,\mathbf{y}')\gamma(\mathbf{y}')d\mathbf{y}' \tag{24}$$

where $\Omega^m(\mathbf{x},\mathbf{y},\mathbf{y}')$ is the surplus (over the value of unemployed search) achieved by the *worker* if s/he moves from employer $\mathbf{y}$ to employer $\mathbf{y}'$ and $m(\mathbf{x},\mathbf{y},\mathbf{y}')$ is the worker's mobility decision: $m(\mathbf{x},\mathbf{y},\mathbf{y}')=1$ if the worker chooses to move from $\mathbf{y}$ to $\mathbf{y}'$ and 0 otherwise.

We now consider three alternative wage setting models: sequential auction with worker bargaining power (Cahuc et al., 2006), fixed-share surplus-splitting (Mortensen and Pissarides, 1994; Moscarini, 2001), and wage posting (Burdett and Mortensen, 1998). In each case, we refer the reader to the relevant reference for details about the wage setting model at hand.

B.1 Sequential Auctions with Worker Bargaining Power (Cahuc et al., 2006)

In this case, $\Omega^m(\mathbf{x},\mathbf{y},\mathbf{y}')=S(\mathbf{x},\mathbf{y})+\beta\left[S(\mathbf{x},\mathbf{y}')-S(\mathbf{x},\mathbf{y})\right]$ and the mobility decision rule is $m(\mathbf{x},\mathbf{y},\mathbf{y}')=\mathbf{1}\left\{S(\mathbf{x},\mathbf{y}')>S(\mathbf{x},\mathbf{y})\right\}$. Thus:

$$\begin{aligned}\rho U(\mathbf{x})&=b(\mathbf{x})+\lambda_0\beta\int \mathbf{1}\left\{S(\mathbf{x},\mathbf{y}')\geq 0\right\}S(\mathbf{x},\mathbf{y}')\gamma(\mathbf{y}')d\mathbf{y}' \\ &=b(\mathbf{x})+\lambda_0\beta\int_0^{+\infty}SdF_{S|\mathbf{x}}(S)=b(\mathbf{x})+\lambda_0\beta\int_0^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS\end{aligned} \tag{25}$$

where $F_{S|\mathbf{x}}(S)=\int \mathbf{1}\left\{S(\mathbf{x},\mathbf{y}')\leq S\right\}\gamma(\mathbf{y}')d\mathbf{y}'$ is the sampling cdf of match surplus faced by a type-$\mathbf{x}$ worker, and where the last equality above obtains from integration by parts. Following similar steps:

$$(\rho+\delta)S(\mathbf{x},\mathbf{y})=f(\mathbf{x},\mathbf{y})-\rho U(\mathbf{x})+\lambda_1\beta\int_{S(\mathbf{x},\mathbf{y})}^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS$$

Substituting (25) into the latter equation:

$$(\rho+\delta)S(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \left[b(\mathbf{x}) + (\lambda_0-\lambda_1)\beta\int_0^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS\right] - \lambda_1\beta\int_0^{S(\mathbf{x},\mathbf{y})}\overline{F}_{S|\mathbf{x}}(S)dS$$

Letting $\widetilde{b}(\mathbf{x}) = b(\mathbf{x}) + (\lambda_0-\lambda_1)\beta\int_0^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS$ and $\sigma(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \widetilde{b}(\mathbf{x})$, we thus have:

$$(\rho+\delta)S(\mathbf{x},\mathbf{y}) = \sigma(\mathbf{x},\mathbf{y}) - \lambda_1\beta\int_0^{S(\mathbf{x},\mathbf{y})}\overline{F}_{S|\mathbf{x}}(S)dS \tag{26}$$

As is clear from (26), $S(\mathbf{x},\mathbf{y})$ only depends on $\mathbf{y}$ through $\sigma(\mathbf{x},\mathbf{y})$ (and in a differentiable way). With a slight abuse of notation, we can thus define $dS/d\sigma$ which, from (26), is given by:

$$\frac{dS(\mathbf{x},\mathbf{y})}{d\sigma(\mathbf{x},\mathbf{y})} = \frac{1}{\rho+\delta+\lambda_1\beta\overline{F}_{S|\mathbf{x}}(S(\mathbf{x},\mathbf{y}))} > 0.$$

This proves that, for any $\mathbf{y}_1 \neq \mathbf{y}_2$, $S(\mathbf{x},\mathbf{y}_2) > S(\mathbf{x},\mathbf{y}_1) \iff \sigma(\mathbf{x},\mathbf{y}_2) > \sigma(\mathbf{x},\mathbf{y}_1)$. Moreover, it is also clear from (26) that $S(\mathbf{x},\mathbf{y}) = 0 \iff \sigma(\mathbf{x},\mathbf{y}) = 0$. Hence, the job acceptance rule is equivalent to $m(\mathbf{x},\mathbf{y},\mathbf{y}') = \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') > \sigma(\mathbf{x},\mathbf{y})\}$ for employed workers and $m(\mathbf{x},\mathbf{0}_Y,\mathbf{y}') = \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') > 0\}$ for unemployed workers: the acceptance rule is the same as in the pure sequential auction case seen in the main text, up to the redefinition of $\sigma(\mathbf{x},\mathbf{y})$ from $f(\mathbf{x},\mathbf{y}) - b(\mathbf{x})$ to $f(\mathbf{x},\mathbf{y}) - \widetilde{b}(\mathbf{x})$.

Crucially, this redefinition of σ preserves monotonicity and linearity of σ w.r.t. $\mathbf{y}$, as well as supermodularity w.r.t. $(\mathbf{x},\mathbf{y})$. Therefore, any result relying on those properties alone will continue to hold in this modified model. We take stock of what those are in the last paragraph of this appendix.

B.2 Fixed-share Surplus-splitting (Moscarini, 2001)

In this case, $\Omega^m(\mathbf{x},\mathbf{y},\mathbf{y}') = \beta S(\mathbf{x},\mathbf{y}')$. But also, the worker's value in the incumbent match is $\beta S(\mathbf{x},\mathbf{y})$, implying that the mobility decision rule is again $m(\mathbf{x},\mathbf{y},\mathbf{y}') = \mathbf{1}\{S(\mathbf{x},\mathbf{y}') > S(\mathbf{x},\mathbf{y})\}$. We thus still have $\rho U(\mathbf{x}) = b(\mathbf{x}) + \lambda_0\beta\int_0^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS$, and now:

$$\begin{aligned}(\rho+\delta)S(\mathbf{x},\mathbf{y}) &= f(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x}) + \lambda_1\int \mathbf{1}\{S(\mathbf{x},\mathbf{y}') > S(\mathbf{x},\mathbf{y})\}\left[\beta S(\mathbf{x},\mathbf{y}') - S(\mathbf{x},\mathbf{y})\right]\gamma(\mathbf{y}')d\mathbf{y}' \\ &= f(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x}) + \lambda_1\int_{S(\mathbf{x},\mathbf{y})}^{+\infty}\left[\beta S - S(\mathbf{x},\mathbf{y})\right]dF_{S|\mathbf{x}}(S) \\ \iff \left[\rho+\delta+\lambda_1(1-\beta)\overline{F}_{S|\mathbf{x}}(S(\mathbf{x},\mathbf{y}))\right]S(\mathbf{x},\mathbf{y}) &= f(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x}) + \lambda_1\beta\int_{S(\mathbf{x},\mathbf{y})}^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS\end{aligned}$$

Substituting $U(\mathbf{x})$:

$$\left[\rho+\delta+\lambda_1(1-\beta)\overline{F}_{S|\mathbf{x}}(S(\mathbf{x},\mathbf{y}))\right]S(\mathbf{x},\mathbf{y}) = \sigma(\mathbf{x},\mathbf{y}) - \lambda_1\beta\int_0^{S(\mathbf{x},\mathbf{y})}\overline{F}_{S|\mathbf{x}}(S)dS \tag{27}$$

where $\widetilde{b}(\mathbf{x}) = b(\mathbf{x}) + (\lambda_0 - \lambda_1)\beta \int_0^{+\infty} \overline{F}_{S|\mathbf{x}}(S)dS$ and $\sigma(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \widetilde{b}(\mathbf{x})$ are defined as before. Then, using a similar argument as in the Cahuc et al. (2006) case:

$$\left(\rho + \delta + \lambda_1 \overline{F}_{S|\mathbf{x}}\left(S(\mathbf{x},\mathbf{y})\right) - \lambda_1(1-\beta)S(\mathbf{x},\mathbf{y})F'_{S|\mathbf{x}}\left(S(\mathbf{x},\mathbf{y})\right)\right) dS(\mathbf{x},\mathbf{y}) = d\sigma(\mathbf{x},\mathbf{y})$$

Thus, a sufficient condition for $S(\mathbf{x},\mathbf{y})$ to be in a one-to-one increasing relationship with $\sigma(\mathbf{x},\mathbf{y})$ is that the hazard rate of $F_{S|\mathbf{x}}$ be small enough:

$$\frac{SF'_{S|\mathbf{x}}(S)}{\overline{F}_{S|\mathbf{x}}(S)} \leq \frac{\rho+\delta+\lambda_1}{\lambda_1(1-\beta)} \iff \frac{S\int \mathbf{1}\left\{S(\mathbf{x},\mathbf{y}') = S\right\}\gamma(\mathbf{y}')d\mathbf{y}'}{\int \mathbf{1}\left\{S(\mathbf{x},\mathbf{y}') \geq S\right\}\gamma(\mathbf{y}')d\mathbf{y}'} \leq \frac{\rho+\delta+\lambda_1}{\lambda_1(1-\beta)}. \tag{28}$$

This is an unwieldy condition on γ, but still a condition on the primitives. Note that the reason we need this condition is that, under this particular rent-sharing rule, a share $1-\beta$ of the surplus from the initial match is lost to a third party (the new employer) when the worker changes jobs. More precisely, when the worker changes jobs, the initial firm-worker collective "gains" $\beta\left[S(\mathbf{x},\mathbf{y}') - S(\mathbf{x},\mathbf{y})\right]$ (a share β of the rent supplement brought about by the new match, although all of these gains accrue to the workers), and "loses" $(1-\beta)S(\mathbf{x},\mathbf{y})$ to the new employer. Thus, if there is a high concentration of potential matches with equal surplus (if $F'_{S|\mathbf{x}}\left(S(\mathbf{x},\mathbf{y})\right)$ is large), the initial match stands to lose a lot and gain very little in case the worker accepts an outside offer. As a result, the surplus may be higher in a slightly less productive match but with fewer similar potential matches available in the economy. Condition (28) prevents that from happening.

B.3 Wage Posting (Burdett and Mortensen, 1998)

Assuming that firms post wages and are allowed to make offers contingent on worker type $\mathbf{x}$, the model becomes one of segmented wage-posting markets (one market for each $\mathbf{x}$), where workers are homogeneous within each market and firms in market $\mathbf{x}$ are heterogeneous with (scalar) productivity $f(\mathbf{x},\mathbf{y})$. We then know from Burdett and Mortensen (1998) — or indeed from standard monotone comparative statics — that firms with higher $f(\mathbf{x},\mathbf{y})$ will post higher wages for type-$\mathbf{x}$ workers (and thus offer higher values to those workers).

Adjusting the notation from Burdett and Mortensen (1998), posted wages are given by:

$$w(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \left[\delta + \lambda_1 \overline{F}_{f|\mathbf{x}}\left(f(\mathbf{x},\mathbf{y})\right)\right]^2 \left\{\int_{\widetilde{b}(\mathbf{x})}^{f(\mathbf{x},\mathbf{y})} \frac{dt}{\left[\delta + \lambda_1 \overline{F}_{f|\mathbf{x}}(t)\right]^2} + C(\mathbf{x})\right\}$$

where $F_{f|\mathbf{x}}(\cdot)$ is the sampling distribution of match productivity conditional on $\mathbf{x}$, $\widetilde{b}(\mathbf{x})$ is the lowest productivity amongst viable matches on the market for type-$\mathbf{x}$ workers, and $C(\mathbf{x})$ is the profit of the least productive match employing a type-$\mathbf{x}$ worker.[35] Again defining $\sigma(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \widetilde{b}(\mathbf{x})$, it is

[35] Calling the reservation wage of a type-$\mathbf{x}$ unemployed worker $R(\mathbf{x})$ (see Burdett and Mortensen, 1998 for a derivation), we have that $\widetilde{b}(\mathbf{x}) = \max\left\{R(\mathbf{x}); \min_{\mathbf{y}'\in\mathcal{Y}} f(\mathbf{x},\mathbf{y}')\right\}$. Then, the integration constant $C(\mathbf{x})$ is zero if

straightforward to check that $w(\mathbf{x},\mathbf{y})$ is a strictly increasing function of $\sigma(\mathbf{x},\mathbf{y})$, so that employed workers move up the $\sigma(\mathbf{x},\mathbf{y})$-ladder. Moreover, by construction, unemployed workers accept offers iff. $\sigma(\mathbf{x},\mathbf{y}) \geq 0$.

B.4 Taking Stock

In all the cases reviewed above, worker mobility is governed by comparisons of $\sigma(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \widetilde{b}(\mathbf{x})$, where $\widetilde{b}(\mathbf{x})$ is a (potentially very complex) function of $\mathbf{x}$ only. So any theorem that only uses monotonicity of σ in $\mathbf{y}$, linearity of σ in $\mathbf{y}$, or supermodularity σ in $(\mathbf{x},\mathbf{y})$ goes through. What fails to go through, though, is any property that relies on the linearity of σ in $\mathbf{x}$: even assuming linearity in $\mathbf{x}$ of f and b, the function $\widetilde{b}(\mathbf{x})$ is generically nonlinear. What this means in terms of the results in this paper is that the only properties that are specific to the pure sequential auction model are Theorems 7 and 8. All of the conditions ensuring PAM in equilibrium are preserved under any of the three alternative wage setting models covered in this appendix.

$\widetilde{b}(\mathbf{x}) = R(\mathbf{x})$ and some positive function of $\mathbf{x}$ otherwise.