Multidimensional Sorting Under Random Search*

Ilse Lindenlaub† Fabien Postel-Vinay‡

September 2016

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 in assessing sorting, mismatch and policy 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 Carlos Carrillo-Tudela and Jim Albrecht for their insightful discussions of this paper at the 2016 Cowles Foundation Macro Conference and the 2016 Konstanz Search and Matching Conference. We further thank seminar audiences at Essex, Yale, the Chicago Fed, Einaudi, Konstanz and the NBER Summer Institute 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 and IFS. **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 observation that typical data sets describe both workers and jobs in terms of many different productive attributes (e.g. cognitive skills, manual skills, 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 and qualitative errors as well as 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 and $\mathbf{y} = (y_1, \cdots, y_Y)$ 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 the first-order stochastic dominance sense) attributes y_j, then PAM occurs in dimension (x_k, y_j).

[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).

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 four 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 then 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.

Second, we show that this model predicts sorting based on *specialization* rather than on 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, possibly at the cost of a lower requirement in the other skill.

Our third set of results highlights that sorting generally cannot be positive between *all* skill and job dimensions. Instead, there are sorting trade-offs. We provide conditions under which PAM arises in all dimensions except the one that is characterized by the weakest complementarities in production, where forces push towards NAM.

Our fourth set of results is 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 sorting and mismatch to the true two-dimensional model. We find that the misspecified model can suggest fallacious conclusions: It predicts sorting when the true data features none and vice versa. Furthermore, the estimated sorting by the misspecified one-dimensional model is relatively weak (possibly because it produces an average of true sorting which is PAM in some dimensions and NAM in others). As a result, it tends to over-estimate mismatch. Last, and perhaps most importantly, the one-dimensional model suggests a first-best allocation that is very different from the true first-best under multi-dimensional types. Implementing the (fallacious) first-best allocation suggested by the misspecified one-dimensional model can therefore cause sizeable welfare losses.

From those results we draw four conclusions. First, sorting arises in our model *only* due to multi-dimensional heterogeneity. In a comparable one-dimensional model, where match surplus 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.

Finally, 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 develop a theory of *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 essential 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 multiple dimensions 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), and then extend to more general cases (4.2). Section 5 investigates sorting based on absolute advantage vs specialization and Section 6 discusses the interrelation of sorting patterns across all heterogeneity dimensions. Section 7 contains our simulation exercise. Section 8 places the contribution of this paper into the literature and Section 9 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 different skills in the workers' skill bundle), drawn from an exogenous distribution L, with density ℓ. 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" $\mathbf{y} = (y_1, \cdots, y_Y) \in \mathcal{Y} = \times_{j=1}^{Y}[\underline{y}_j, \overline{y}_j]$, 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 Γ,

[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). For multi-dimensional sorting under TU and *without* frictions, see Lindenlaub (2014). We will discuss the related literature in detail below.

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

with density γ. We assume that γ has strictly positive mass over its entire support, $\text{Supp}\,\gamma = \mathcal{Y}$.[5]

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 attibutes $\mathbf{y}$ (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 $b(\mathbf{x})$.

Workers search for jobs at random, both off and on the job. They 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 type-$(\mathbf{x}, \mathbf{y})$ match 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 text that wage contracts are renegotiated sequentially by mutual agreement, as in the sequential auction model of Postel-Vinay and Robin (2002). This is (mainly) to simplify exposition. We show in Appendix B that most of our results extend to other common wage setting rules, as Nash bargaining (Mortensen and Pissarides, 1994; Moscarini, 2001), wage/contract posting (Burdett and Mortensen, 1998; Moscarini and Postel-Vinay, 2013), or sequential auctions *with* worker bargaining power (Cahuc, Postel-Vinay and Robin, 2006).

[5]This assumption ensures that the support of γ is a lattice under the natural (component-wise) partial order in $\mathbb{R}^n$, which is a technical requirement for some of our proofs for cases when $Y \geq 3$ (see Theorems 4 and 5).

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

In the sequential auction model, workers can play 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 leads to the worker ending 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 $W' = \min\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\}$.

It follows that the total value of a type-$(\mathbf{x},\mathbf{y})$ match, $P(\mathbf{x},\mathbf{y})$, 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 entirely based on the comparison of flow surpluses.[8]

Finally note that, in the sequential auction case, the value of unemployment, $U(\mathbf{x})$, is given by $\rho U(\mathbf{x}) = b(\mathbf{x})$, implying $\sigma(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - b(\mathbf{x})$, i.e. σ is entirely determined by technology.

2.3 Steady-state Distribution of Skills and Skill Requirements

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

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

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

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

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 and composed of two groups: unemployed type-$\mathbf{x}$ workers who draw a type-$\mathbf{y}$ job (flow probability $\lambda_0\gamma(\mathbf{y})$) and accept it (which they do if the associated flow surplus is positive $(1\left\{\sigma(\mathbf{x},\mathbf{y})\geq 0\right\})$), and type-$\mathbf{x}$ workers employed in any type-$\mathbf{y}'$ job who draw an offer from a type-$\mathbf{y}$ job (flow probability $\lambda_1\gamma(\mathbf{y})$) and accept it (which they do if the flow surplus with that job exceeds the one with their initial type-$\mathbf{y}'$ job). The measure of type-$\mathbf{x}$ unemployed workers solves 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 Appendix A that the matching distribution $h(\mathbf{x},\mathbf{y})$ has the following closed-form:

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

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

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:

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

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 in a frictionless context is that the Jacobian matrix of the equilibrium matching function 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 — there is mismatch. 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 $\mathbf{y}$ as one varies worker type $\mathbf{x}$.[10]

Formally, let $H_j(y|\mathbf{x})$ denote the marginal c.d.f. of y_j (the jth component of the vector of job attributes $\mathbf{y}$) matched to employed workers with skill bundle $\mathbf{x}$. Using (3), we can express this as

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

We will in fact be interested in $\nabla H_j(y|\mathbf{x}) = \left(\partial H_j(y|\mathbf{x})/\partial x_1, \cdots, \partial H_j(y|\mathbf{x})/\partial x_X\right)^\top$, the gradient of $H_j(y|\mathbf{x})$. A situation of particular interest is when one of the components of this gradient,

[10] We chose to analyze the equilibrium matching distribution of $\mathbf{y}$ given $\mathbf{x}$ and not that of $\mathbf{x}$ given $\mathbf{y}$ for the following reason. While workers sample job types from an exogenous sampling distribution γ, jobs 'sample' workers from an *endogenous* distribution (the distribution of workers across employment statuses and job types), which in itself is a complex equilibrium object. The acceptance decisions of firms would impact the distribution of $\mathbf{x}$ across employment statuses and job types, and the distribution of $\mathbf{x}$, in turn, impacts the acceptance decisions of the firms. Analyzing the matching distributions of $\mathbf{x}$ given $\mathbf{y}$ would therefore require us to deal with a complicated fixed point problem, in spirit similar to that in Shimer and Smith (2000), and proves intractable.

$\partial H_j(y|\mathbf{x})/\partial x_k$, 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 skill is matched to jobs with greater type-j skill requirement (in the FOSD sense) compared to a worker with lower type-k skill. We will thus use the following formal definition of sorting:

Definition 1 (Positive and Negative Assortative Matching)**.** *Matching is positive (negative) assortative in dimension* (y_j, x_k) *at skill bundle* $\mathbf{x}$ *iff.* $\partial H_j(y|\mathbf{x})/\partial x_k$ *is negative (positive) for all* $\mathbf{y}$*.*

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.* Note that Definition 1 is "local" in the sense that it applies at a given skill bundle $\mathbf{x}$. Consequently, all of the following results will be stated for a given $\mathbf{x}$, and we will omit the phrase "at skill bundle $\mathbf{x}$" when talking about the occurrence of PAM or NAM. Finally, to avoid duplication of some results, we focus on positive sorting throughout most of the paper.

3.2 A Decomposition Result

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

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

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

All proofs are in Appendix A. 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).[11] 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 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 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, given by the second term of the expression in Theorem 1, operates through selection on the EE margin.

If terms (1) and/or (2) in Theorem 1 are negative, then PAM obtains on the NE and/or EE margin, respectively. The following corollaries, which are immediately implied by Theorem 1, will help us specify conditions under which this is the case. We start with the EE margin:

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

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

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 $\sigma(\mathbf{x}, \mathbf{y}) > 0$

[11] Two important technical notes: first, throughout 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 $\mathbb{E}$. Similarly, $\Pr_\Gamma\{A\}$ is used to denote the probability of A occurring following a random draw of a job type $\mathbf{y}$ from the sampling distribution γ. Second, it may be that the joint event $(\sigma(\mathbf{x}, \mathbf{y}') = s, y'_j \leq y)$ 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.

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, i.e. there is PAM in dimension (y_j, x_k).

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 from 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. We discuss simplifying assumptions below.

The next result concerning sorting on the NE margin also follows from Theorem 1.

Corollary 2. *If Condition* (CMP) *holds with increasing monotonicity and if, in addition,*

$$G'_{\sigma|\mathbf{x}}(0) \times \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = 0 \right] \times \left\{ \int_0^{+\infty} G'_{\sigma|\mathbf{x}}(s) \left[\Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = 0 \right\} - \Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = s \right\} \right] ds \right\} \leq 0 \quad (5)$$

then term (1) in Theorem 1 is negative for all y, i.e. PAM occurs in dimension (y_j, x_k) along the NE margin. If, instead, (CMP) *holds with decreasing monotonicity and* (5) *is positive, then term (1) in Theorem 1 is positive for all y, i.e. NAM occurs in (y_j, x_k) along the NE margin.*

Condition (5) (the expression under the integral in particular) clearly shows that conditions to ensure sorting on the NE margin involve *both* the sampling distribution Γ *and* the technology σ.

4 The Sign of Sorting

While Theorem 1 and its two corollaries afford a clear decomposition of sorting on the NE and EE margins, those two effects involve complex interactions between the technology σ and the sampling distribution of job types γ and cannot easily be signed without further 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 Corollaries 1 and 2 considerably.

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

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

where $\mathbf{Q} = (q_{kj})_{\substack{1 \le k \le X \\ 1 \le j \le Y}}$ 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 $\mathbf{x}$:

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

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

(c) for all $\mathbf{x} \in \mathcal{X}$, there exists $j \in \{1, \cdots, Y\}$ such that $q_j(\mathbf{x}) := \sum_{k=1}^{X} q_{kj}(x_k + a_k) > 0$; 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:

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

The $X \times Y$ technology matrix $\mathbf{Q}$ captures the complementarity structure between job and worker characteristics relating to all types of tasks, $(k, j) \in \{1, \cdots, X\} \times \{1, \cdots, Y\}$, 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 $\left(\sum_{k=1}^{X} a_k q_{kj}\right) dy_j$ 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.[12] Note that we do not impose monotonicity of the production function in *all* job attributes. Nor do we restrict the monotonicity of the production or flow surplus function σ in worker skills $\mathbf{x}$. Next, we consider:

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. We now investigate the sign of sorting along both the EE and the NE margin.

4.1.2 The EE Margin

We begin with the following result on the sign of EE-sorting.

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

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

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.[13]

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

[12] 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.

[13] 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.

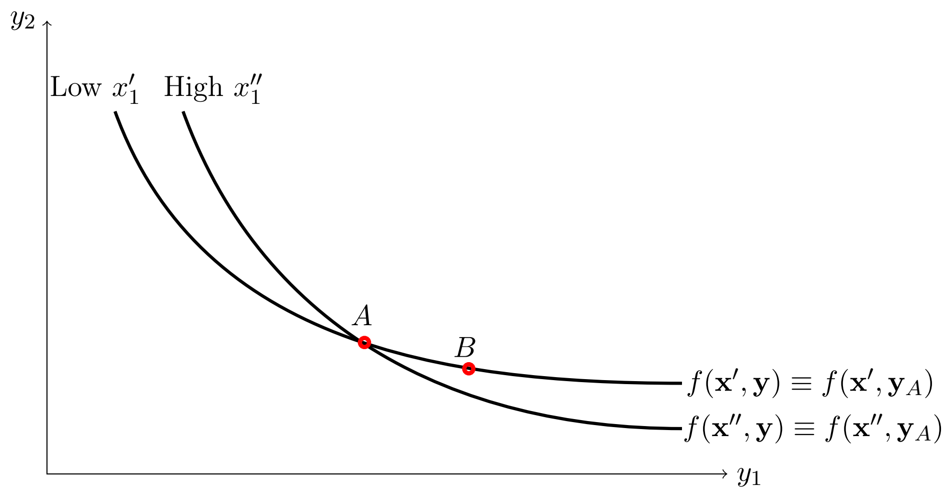


Figure 1: Single Crossing Property

y_1 than to y_2, which is why positive sorting occurs between between x_k and y_1 (but negative sorting between between x_k and y_2 — in the dimension of relatively weak complementarity).

To illustrate the single crossing condition in our setting and its implication graphically, we consider two workers with skill bundles $\mathbf{x}' = (x_1', x_2')$ and $\mathbf{x}'' = (x_1'', x_2'')$ such that $x_1'' > x_1'$ and $x_2'' = x_2'$ (the second has more of x_1 but both have equal amounts of x_2). For each 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. 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 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 other worker (who is more skilled in dimension x_1) *strictly* prefers job B, as 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 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 Corollary 2. This is true even if we restrict our attention to two-dimensional heterogeneity on the job 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 single-crossing condition* (SC-2d) *(Theorem 2), if:*

1. $q_1(\mathbf{x}) > 0$ *(i.e. $f(\mathbf{x}, \mathbf{y})$ is not only increasing in y_2 but also in y_1)*
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 γ, denoted by $\underline{\mathbf{y}} = (\underline{y}_1, \underline{y}_2)$: $\underline{y}_2 \geq b_2$ and $\underline{y}_1 < b_1$*

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

Moreover, (SC-2d) *is also necessary for PAM if we allow for any sampling distribution γ.*

Theorem 3 highlights the importance of single-crossing also for sorting on the NE margin. The last statement, in particular, means that it becomes a necessary condition for sorting when the sampling distribution is unspecified, i.e. if all possible sampling distributions are considered.

However, single crossing alone is not sufficient for PAM on the NE margin: additional restrictions on the sampling distribution are needed. Recall that sorting on the NE margin occurs when a marginal increase in skill x_k affects 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 both single crossing and distributional restrictions in that theorem.

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 compliance with Condition 3 in Theorem 3. The oblique lines are zero level curves of $\sigma(\mathbf{x}, \cdot)$, which under this linear technology

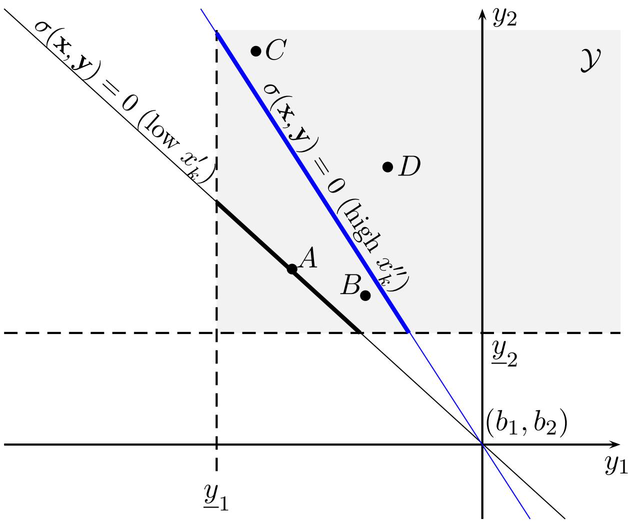


Figure 2: Sorting along the NE margin

(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 emphasized line segment). Note that this boundary lies entirely in the region of $\mathcal{Y}$ where $y_1 < b_1$: because it is assumed that $\underline{y}_2 \geq b_2$, it has to be the case that $y_1 < b_1$ for surplus to equal zero.

Those zero level curves are drawn for two workers $\mathbf{x}'$ and $\mathbf{x}''$ with $x''_k > x'_k$ (with the same amount of all other skills). The higher-x''_k (blue) curve is 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 high x_k. Therefore, for a given y_2, workers with higher x_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. This 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. After increasing skill k from x'_k to x''_k, a worker no longer breaks even with a job at A. Moreover, jobs around B (with higher y_1 but lower y_2) are also made unprofitable while jobs around C with *lower* y_1 but higher y_2 compared to 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 the density γ be log-supermodular and its marginals log-concave. This class of distributions is quite broad. For instance, any bivariate distribution of independent random variables that has log-concave marginals (such as the uniform distribution with independent random variables) satisfies (NE-2d). 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).[14]

Cases of 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 restriction on f.

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

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 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, different workers rank firms differently and, depending on their skill bundles, may accept and reject different types of jobs, which is why sorting arises. This trade-off of sorting across dimensions is absent from one-dimensional settings.

4.2 Generalizations

In this section we relax Assumptions 1 and 2 in two ways: first, we generalize our results to the case of a nonlinear technology (which is monotone in at least one y) with an unrestricted number of job attributes. We then state the cases of nonlinear 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 nonlinear, non-monotone technology which we call 'separable', again for general Y-dimensional heterogeneity of job attributes.

4.2.1 Monotone Technology with $Y \geq 2$

We begin with sufficient conditions for sorting on the EE margin, which was addressed by Theorem 2 for the case of $Y = 2$ and bilinear technology. The following theorem relaxes both Assumptions 1 and 2, generalizing Theorem 2 to the case of monotone technologies 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* $\lim_{y_Y \to \overline{y}_Y} f(\mathbf{x}, \mathbf{y}) = +\infty$

   (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* $f(\mathbf{x}, \cdot)$*:*

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

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 *less* complementary to x_k than all other y_ℓ.

While Theorem 2 provided distribution-free, necessary and sufficient conditions for EE-sorting in the case of $Y = 2$, Theorem 4 that applies to $Y > 2$ only provides sufficient conditions that are no longer distribution-free (even if we assume bilinear technology, 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 need to restrict the sampling distribution in the case of $Y > 2$ can be explained as follows. 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 in dimensions $(y_\ell, x_k), \ell \neq Y$.

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 nonlinear, monotone surplus functions with $Y = 2$ (Corollary 3) as well as bilinear surplus functions 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 restrict the number of job attributes to two, the sufficient conditions for EE-sorting from Theorem 4 become considerably simpler. Condition (EE-Yd) disappears altogether: the role of that condition was to prevent the sampling distribution from having job attributes y_ℓ ($\ell \neq Y$) associated in a way that countered the force toward PAM in (y_ℓ, x_k) arising from complementarities. This issue of association between $Y - 1$ job attributes becomes moot when $Y = 2$. The reason is that in this case we focus on conditions for PAM in a single dimension ($Y - 1 = 1$). Stronger complementarities between (y_1, x_k) than between (y_2, x_k) are then sufficient to ensure PAM in (y_1, x_k). 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 for bilinear production functions (Assumption 1), which satisfy monotonicity (as, by Assumption 1, $q_Y(\mathbf{x}) > 0$).

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

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

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

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

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.

Condition 1 echoes Condition 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), which is a special case of Condition 3 in Theorem 4. This condition is again satisfied if log-supermodularity of γ in job attribute y_i and any other job attribute, y_j $(i, j \neq Y)$ sufficiently dominates any log-supermodularity in pairs (y_i, y_Y) and (y_j, y_Y) (assuming $q_i(\mathbf{x}) > 0, q_j(\mathbf{x}) > 0$ for all i, j). Like in Theorem 4, this condition limits distributional barriers to PAM in dimensions (y_ℓ, x_k), $\ell \neq Y$, that could arise from negative association among the attributes y_ℓ in γ.

Condition (EE-Yd') is more demanding on γ 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 (EE-Yd') is not empty. For instance, for a (truncated) multivariate normal distribution, (EE-Yd') reads

$$-q_Y(\mathbf{x})^2 \frac{\Sigma_{ij}^{-1} + \Sigma_{ji}^{-1}}{2\,|\Sigma|} - q_i(\mathbf{x}) q_j(\mathbf{x}) \frac{\Sigma_{YY}^{-1}}{|\Sigma|} + q_j(\mathbf{x}) q_Y(\mathbf{x}) \frac{\Sigma_{iY}^{-1} + \Sigma_{Yi}^{-1}}{2\,|\Sigma|} + q_i(\mathbf{x}) q_Y(\mathbf{x}) \frac{\Sigma_{Yj}^{-1} + \Sigma_{jY}^{-1}}{2\,|\Sigma|} > 0 \tag{6}$$

where Σ_{ij}^{-1} is element ij of the inverse of the covariance matrix and $|\Sigma|$ is its determinant. Condition (6) holds if the correlation between y_i and y_j 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.

[15] As an example, consider $Y = 3$ and assume that standard deviations satisfy $\theta_i = \theta_j = \theta$. Then, (6) reads:

$$(\theta^4 / |\Sigma|^2)\left(q_3^2(\tau_{12} - \tau_{13}\tau_{23}) - q_1 q_2 (1 - \tau_{12}^2) - q_2 q_3 (\tau_{13} - \tau_{12}\tau_{23}) - q_1 q_3 (\tau_{23} - \tau_{12}\tau_{13})\right) > 0$$

This inequality holds if, for instance, $\tau_{12} = 1$ and $\tau_{13} = \tau_{23} = 0$ (where $\tau_{ij} = \theta_{ij}/\theta_i\theta_j = \text{corr}_\Gamma(y_i, y_j)$).

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 additional assumptions on the sampling distribution Γ, similar to those we discussed under two-dimensional heterogeneity of jobs. Condition (NE-Yd) (which again imposes some degree of positive association between job attributes in the sampling distribution) has the same interpretation as (NE-2d) in Theorem 3 for $Y = 2$. 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.

4.2.2 Separable Technology with $Y \geq 2$

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_Y. While it 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 $X = Y$, $f(\mathbf{x}, \mathbf{y}) = c_0 - \sum_{i=1}^{X} c_i(x_i - y_i)^2$, where the c_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

[16] Note that there can arise a tension between (NE-Yd) and (EE-Yd') that is also assumed to hold here. Importantly, however, both conditions can be satisfied simultaneously. E.g., in the case of $X = Y = 3$ and where γ is truncated normal, if τ_{12} is sufficiently high, e.g. $\tau_{12} = 1$, and τ_{13} and τ_{23} are low, e.g. $\tau_{13} = \tau_{23} = 0$, then condition (EE-Yd') holds strictly (see Footnote 15) while condition (NE-Yd) holds with equality.

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 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 $\ell \in \{2, \cdots, Y\}$. 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 ($f(\mathbf{x}, \mathbf{y}) = c_0 - \sum_{i=1}^{X} c_i (x_i - y_i)^2$), 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 technologies 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 Theorem 4 are only needed if complementarities in (y_1, x_k) compete with complementarities between x_k and other dimensions y_ℓ, $\ell \neq 1$. Theorem 4 then guarantees PAM in *all but one* dimensions $(y_\ell, x_k), \ell \neq Y$, while Theorem 6 for separable production functions only ensures sorting in *a single* dimension (y_1, x_k).

5 Sorting on Absolute Advantage vs Sorting on Specialization

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

[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 f.

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.*

The 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": 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, i.e. $\mathbf{x}' = -\mathbf{a} + 2(\mathbf{x} + \mathbf{a})$, 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*.*

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 in terms of matching distributions. Workers with different skill bundles can be matched to the same distribution of jobs.

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), γ 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 dimension (x_1, y_1),[19] and the statement in Theorem 7 writes as:

$$\frac{\partial H_1(y|\mathbf{x})}{\partial x_2} = -\frac{x_1 + a_1}{x_2 + a_2} \frac{\partial H_1(y|\mathbf{x})}{\partial x_1} > 0. \tag{7}$$

Theorem 7 addresses the case of a simultaneous expansion of all skills such that the *sum*

[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).

[19] Condition (SC-2d) from Theorem 2 writes as $(x_2 + a_2) \det \mathbf{Q} > 0$, which holds here by assumption. 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, satisfying Condition 1 of Theorem 3. 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 thus also occurs on the NE margin in this example.

$\mathbf{x} + \mathbf{a}$ 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 $(x_1 + \Delta x_1, x_2 + \Delta x_2)$. Using (7), the resulting change in $H_1(y|\mathbf{x})$ is:

$$\Delta H_1(y|\mathbf{x}) \simeq (x_1 + a_1)\frac{\partial H_1(y|\mathbf{x})}{\partial x_1}\left[\frac{\Delta x_1}{x_1 + a_1} - \frac{\Delta x_2}{x_2 + a_2}\right] \tag{8}$$

Equation (8) 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, (8) 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 if and only if $\frac{\Delta x_1}{x_1+a_1} > \frac{\Delta x_2}{x_2+a_2}$. 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 cause the worker to sort into jobs with stochastically higher y_1 (but with *lower* y_2).

A third implication of (8) is that in the case of a *proportional* increase in both skills ($\frac{\Delta x_1}{x_1} = \frac{\Delta x_2}{x_2}$), $\Delta H_1(y|\mathbf{x}) < 0$ if and only if $x_1/x_2 > a_1/a_2$. In 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.[20]

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

[20] It is important to note that this result is *not* due to our assumption of no capacity constraint on the firm side but appears in models of multi-dimensional sorting more generally. The absence of sorting based on absolute advantage can also be obtained in a multi-dimensional model *with* capacity constraints (e.g. Lindenlaub, 2014).

$\partial H_j(y|x)/\partial x = 0$ between x and any $y_j, j \in \{1, \cdots, Y\}$. In other words, x and y_j 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, Condition 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.

6 Interrelation of Sorting Patterns Across Dimensions

As the preceding analysis made clear, changes in worker skill along some dimension will in general affect the assignment. Section 4 gave conditions under which sorting is positive in dimensions (y_j, x_k) for all $j \in \{1, \cdots, Y-1\}$ and $Y \geq 2$, providing a good understanding of sorting between job characteristics y_j and *a given skill* x_k. In this section, we aim to complete the picture of sorting in this economy by exploring the link between sorting patterns across all dimensions. We first highlight the implications of our theory for sorting between different skills x_k and *a given job attribute* y_j. Second, we discuss some implications for sorting in the "remaining dimension" from Theorem 4, i.e. in (y_Y, x_k). We begin with a result that follows directly from Theorem 7:

Corollary 5. *Under Assumption 1, if PAM occurs between y_j and x_k for all worker skill dimensions but one, i.e. for $k \in \{1, \cdots, X\} \backslash \{K\}$, then NAM occurs between y_j and x_K.*

The root cause of Corollary 5 is easy to understand in the special case $X = Y = 2$ with the NE margin shut down. Suppose for example that the single-crossing condition (SC-2d), $\partial \left(q_1(\mathbf{x})/q_2(\mathbf{x}) \right) / \partial x_1 > 0$, holds, which is equivalent to $\det \mathbf{Q} > 0$, implying PAM in dimension

(y_1, x_1). However, $\det \mathbf{Q} > 0$ also implies $\partial (q_1(\mathbf{x})/q_2(\mathbf{x})) / \partial x_2 < 0$, leading to NAM in dimension (x_2, y_1). By the same argument, $\det \mathbf{Q} > 0$ implies PAM in dimension (x_2, y_2) but NAM in dimension (x_1, y_2). This illustrates that PAM cannot arise simultaneously across those two dimensions because the (necessary and sufficient) single-crossing conditions cannot hold simultaneously for x_1 and x_2. Put differently, $\det \mathbf{Q} > 0$ says that complementarities in (x_1, y_1) and (x_2, y_2) "dominate" complementarities in (x_2, y_1) and (x_1, y_2), which is why PAM occurs *within* task dimensions 1 and 2 but NAM occurs *between* those dimensions. Again in this case of $Y = 2$, this is for purely technological reasons, independent of the sampling distribution.[21]

Next we investigate the interrelation of sorting patterns on the EE margin across job heterogeneity dimensions *for a given skill* x_k. The following theorem implies that the way conditional distributions of job attributes $H_j(y|\mathbf{x})$ (and thus the conditional expectations) co-vary with x_k 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 complex manner. Yet, if we focus on the EE margin, Theorem 8 shows that the co-movement of the conditional *means* of the matching distribution triggered by a skill change is entirely determined by technology.

For the sake of illustration, 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, so the slope of the $\mathbb{E}(\mathbf{y}|\mathbf{x})$ locus is entirely determined by the technology. Moreover, to fix ideas, Figure 3 was drawn under the assumption that f is non-decreasing in all components of $\mathbf{y}$ (all components of $\mathbf{Q}^\top(\mathbf{x} + \mathbf{a})$ are nonnegative), so that the slope of the $\mathbb{E}(\mathbf{y}|\mathbf{x})$ locus is negative. We consider again our two workers with skill bundles $\mathbf{x}' = (x_1', x_2')$ and $\mathbf{x}'' = (x_1'', x_2'')$ such that $x_1'' > x_1'$ and $x_2'' = x_2'$. 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 x_1'' will be matched on average to a job with

[21] Things are much more involved in higher dimensions $Y \geq 3$, where the underlying cause of Corollary 5 is less clear. Yet, the result still holds: under Assumption 1, sorting cannot be simultaneously positive between a given y_j and all x_k, $k \in \{1, \cdots, X\}$. More details are available upon request.

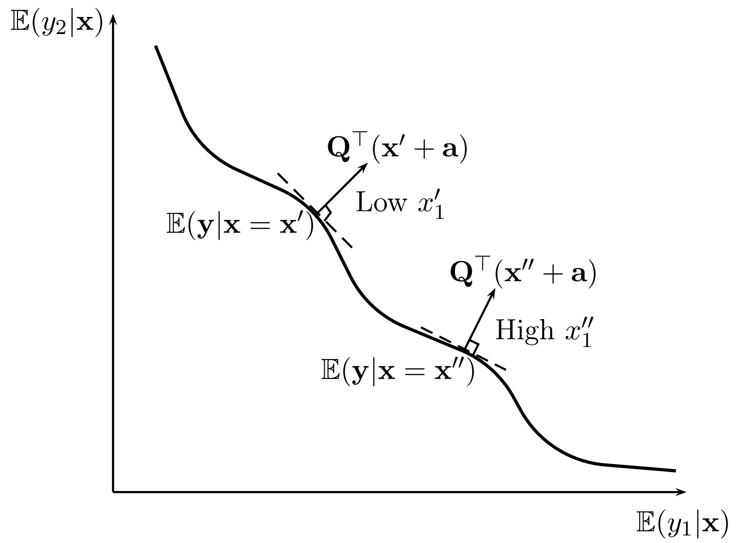


Figure 3: Sorting Across Dimensions

higher type-1 requirements than the lower-skilled worker, as shown on Figure 3. However, the average job to which the higher-x_1'' is matched will have *lower* requirements in type-2 skills than the low-x_1' worker's average job. Theorem 8 also says by how much the type-1 (type-2) attribute of the higher-x_1'' worker's job will be higher (lower) than that of the lower-x_1' worker's job, which solely depends on the components of $\mathbf{Q}^\top(\mathbf{x}+\mathbf{a})$.

More generally, we can use Theorem 8 to obtain some insights about sorting in dimension (y_Y, x_k), the one dimension in which the sign of sorting is not determined by Theorem 4.

Corollary 6. *Under Assumption 1, assuming $q_j(\mathbf{x}) \geq 0$ for all $j \in \{1, \cdots, Y\}$ and $\sigma(\mathbf{x}, \mathbf{y}) > 0$ for all $\mathbf{y} \in \mathcal{Y}$, then if PAM occurs in dimensions $(y_j, x_k), j \in \{1, \cdots, Y-1\}$, there cannot be PAM in dimension (y_Y, x_k).*

The statement of Corollary 6 brings about two remarks. First, Corollary 6 implies that if PAM occurs in dimensions (y_j, x_k), $j \in \{1, \cdots, Y-1\}$, and *if there is sorting in dimension* (y_Y, x_k) — i.e. if $\partial H_Y(y|\mathbf{x})/\partial x_k$ has a constant sign across the entire support of $\mathbf{y}$ — then sorting in this last dimension can only be negative.[22] But it may well be that there is no particular sorting pattern in dimension (y_Y, x_k). Second, the requirement that production is increasing in all y_j, $j \in \{1, \cdots, Y\}$, is essential for Corollary 6: if $q_j(\mathbf{x}) < 0$ for some j, then any sorting pattern (PAM, NAM, or neither) can arise in the residual dimension (y_Y, x_k) (example available on request).

[22] A stronger result than Corollary 6 holds in the two-dimensional case $Y = 2$. In this case, it follows directly from Theorem 2 that, under Assumption 1 and with $\sigma(\mathbf{x}, \mathbf{y}) > 0$ for all $\mathbf{y} \in \mathcal{Y}$, if PAM arises in dimension (y_1, x_k), then there must be NAM in dimension (y_2, x_k), and vice versa. This is the case depicted on Figure 3.

The results in this section convey the following insight about overall sorting patterns in this economy: in our setting with multi-dimensional heterogeneity and random search, sorting cannot be simultaneously positive between all skill and job dimensions (at least when the technology is bilinear in job and worker attributes). Instead, there are trade-offs. We provide conditions under which PAM arises in all dimensions except the one that is characterized by the weakest complementarities, where forces push towards NAM.

7 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 and mismatch. We begin with a discussion of identification and estimation of the misspecified one-dimensional model.

7.1 Estimation of a Misspecified 1D Model

7.1.1 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 σ. Importantly, we do not assume monotonicity of σ in x and y, which may impact the ability of the one-dimensional model to produce sorting. We only make the following identifying assumptions:

Assumption 3. *The flow surplus function σ 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*

Below we will estimate these one-dimensional indices of worker and firm characteristics along with all other parameters of the misspecified model. We therefore follow the large literature on identifying sorting based on *unobserved* (to economists) scalar attributes of workers and firms.

7.1.2 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 $J_{it} = 0$).[23]

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 3, 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 is the quantile function in the population of workers.

[23]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 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.

Firm types and surplus. The flow surplus $\sigma_{i,J_{it},t}$ generated by a match between worker i and firm J_{it} equals $\sigma\left(x_i, y_{J_{it}}\right)$. 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 (by Assumption 3) 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 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 non-parametric estimation of σ over the set of viable type-(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.

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

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

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

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

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 γ must integrate to 1, affords estimates of the sampling distribution γ 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:

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

7.2 Simulation Exercises

7.2.1 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 $\delta = 0.025$.[24] The population density of skills ℓ is 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 γ 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. Specifically, 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 each y_j, $j = 1, 2$, on both x_1 and x_2 in the cross-section of employed workers, for each example.[25]

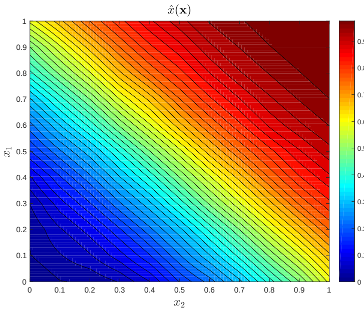
We next fit a misspecified model with one-dimensional heterogeneity to the simulated data generated from our three examples, following the procedure explained above. Figures 4 and 5 show show contour plots of the estimated scalar attributes of the estimated scalar attributes $\widehat{x}$ and $\widehat{y}$ as functions of the underlying true (vector) attributes $\mathbf{x}$ and $\mathbf{y}$, respectively. Figure 6 plots

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

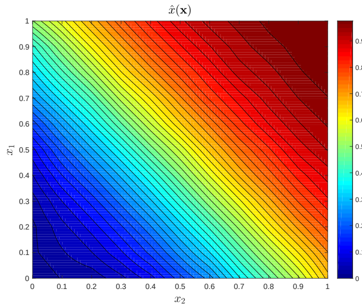
[25] Note that this is not a test of FOSD of the matching distributions in worker skills but FOSD *implies* the monotonicity of the conditional means that we report here.

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		$(0,0)^\top$	

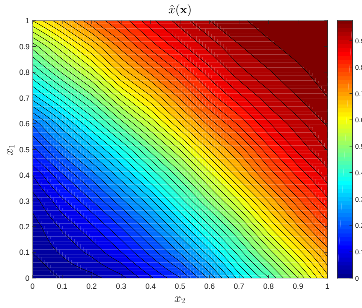
Table 1: Examples of Technologies



(a) Example 1

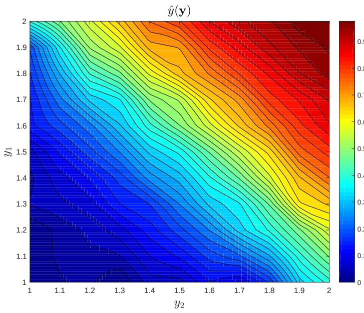


(b) Example 2

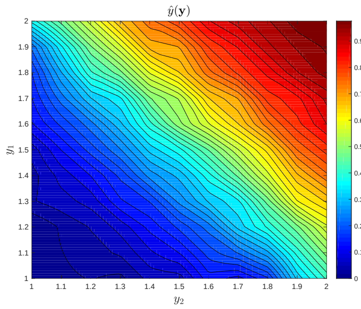


(c) Example 3

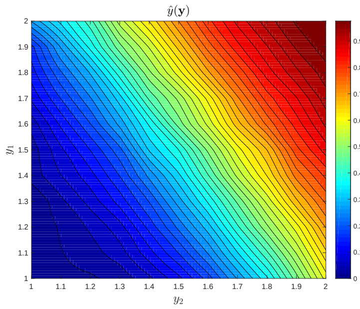
Figure 4: True and Estimated Worker Types



(a) Example 1

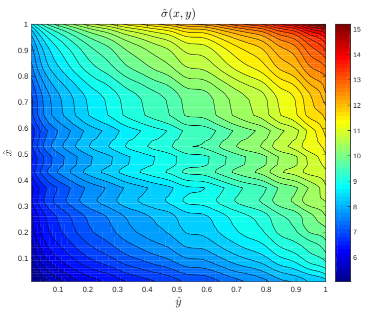


(b) Example 2

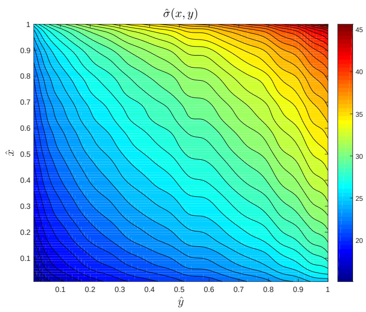


(c) Example 3

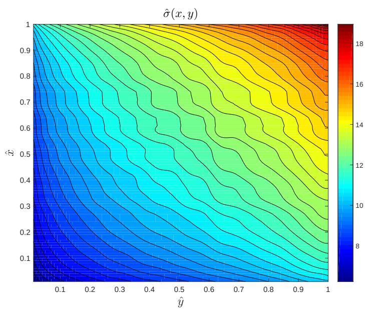
Figure 5: True and Estimated Firm Types



(a) Example 1



(b) Example 2



(c) Example 3

Figure 6: Estimated Surplus Function

the estimated flow surplus function $\widehat{\sigma}$ as a function of estimated scalar types $(\widehat{x}, \widehat{y})$. 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. In turn, Panels (b) and (c) on Figures 4 and 5 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 have 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 1D estimation more weight is put on skill x_1 than skill x_2, possibly owing to the relatively stronger complementarities within dimension 1 compared to dimension 2, featured by the true 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(c)). Figure 6 finally shows that the estimated surplus function increases in $\widehat{x}$ and $\widehat{y}$ (up to some estimation noise) in all three examples, which implies that no sorting should occur in any of these examples, were the 1D model correctly specified.

We now aim to compare the true 2D model to the estimated misspecified 1D model with focus on two issues: sorting and mismatch.

7.2.2 Estimating Sorting in 1D

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 right column of Table 2 and 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) in the data, the 1D model predicts positive sorting: the coefficient of the regression of $\widehat{y}$ on $\widehat{x}$ is positive, sizeable, and statistically significant. Similar results hold for Example 2: the 1D model predicts a positive correlation between $\widehat{y}$ and $\widehat{x}$, whereas the true sorting patterns are positive within skill dimensions and negative between. Only in Example 3, where the true 2D model features in fact the strongest positive sorting within heterogeneity dimensions (as measured by the regression coefficients in the middle column of Table 2), the 1D model fails to predict any clear sorting pattern: the regression coefficient of $\widehat{y}$ on $\widehat{x}$ is very small

Specification:	True 2D model	Misspecified 1D model
1	$\mathbb{E}(y_1 \vert \mathbf{x}) = 1.688 \underset{[-0.011,0.002]}{-0.005} x_1 \underset{[-0.006,0.006]}{+0.000} x_2$ $\mathbb{E}(y_2 \vert \mathbf{x}) = 1.686 \underset{[-0.006,0.006]}{+0.000} x_1 \underset{[-0.007,0.005]}{-0.001} x_2$	$\mathbb{E}(\widehat{y} \vert \widehat{x}) = 0.670 \underset{[0.004,0.016]}{+0.010} \widehat{x}$
2	$\mathbb{E}(y_1 \vert \mathbf{x}) = 1.687 \underset{[0.018,0.030]}{+0.024} x_1 \underset{[-0.033,-0.020]}{-0.027} x_2$ $\mathbb{E}(y_2 \vert \mathbf{x}) = 1.686 \underset{[-0.035,-0.022]}{-0.028} x_1 \underset{[0.020,0.032]}{+0.026} x_2$	$\mathbb{E}(\widehat{y} \vert \widehat{x}) = 0.671 \underset{[0.003,0.015]}{+0.009} \widehat{x}$
3	$\mathbb{E}(y_1 \vert \mathbf{x}) = 1.709 \underset{[0.079,0.092]}{+0.085} x_1 \underset{[-0.100,-0.088]}{-0.094} x_2$ $\mathbb{E}(y_2 \vert \mathbf{x}) = 1.650 \underset{[-0.137,-0.124]}{-0.130} x_1 \underset{[0.120,0.133]}{+0.127} x_2$	$\mathbb{E}(\widehat{y} \vert \widehat{x}) = 0.667 \underset{[-0.003,0.009]}{+0.003} \widehat{x}$

Table 2: Sorting Patterns

and not statistically significant.

These results suggest that the 1D sorting estimates are largely uninformative about the true underlying sorting patterns. Worse, relying on a 1D model to assess sorting when the true data is 2D can be severely misleading: remember that, because the 1D surplus function was estimated to be increasing in $\widehat{y}$ (Figure 6), no sorting should occur in equilibrium between $\widehat{y}$ and $\widehat{x}$ if agents actually behaved as per the 1D model. The fact that $\widehat{y}$ and $\widehat{x}$ are found to be positively correlated in Examples 1 and 2 should therefore alert the researcher to some form of model misspecification. In Example 3, however, things are even worse: The 1D estimates look perfectly consistent with the estimated surplus function ($\widehat{y}$ and $\widehat{x}$ are uncorrelated), even though they suggest very different sorting patterns than the true model.

7.2.3 Mismatch

In this second 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 in a world where worker and job types are really 2D. Results are reported in Table 3.

In this entire exercise we take a short-run perspective where the *distribution of active jobs is fixed*. The exercise is then to redistribute workers across existing jobs in an output-maximizing way, as if there were no search frictions (i.e. if $\lambda_0 \to \infty$ and $\lambda_1 \to \infty$).[26]

[26] The allocation and sorting of this frictionless benchmark with fixed multivariate distributions of workers and jobs is analyzed in Lindenlaub (2014): Examples 2 and 3 feature PAM while Example 1 yields a matching

Specification:	**1**	**2**	**3**
	True 2D model		
$\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]$	12.298	36.902	15.446
$\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]$	13.458	40.531	17.234
$\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]$	12.415	37.277	15.550
Surplus Loss from Mismatch	-0.086	-0.090	-0.104
Surplus Gain from 1D Optimal Allocation rel. to 2D Allocation	0.010	0.010	0.007
Surplus Loss from 1D Optimal Allocation rel. to 2D Optimum	-0.077	-0.080	-0.098
	Misspecified 1D model		
$\mathbb{E}[\sigma(\widehat{x},\widehat{y})]$	12.209	36.830	15.358
$\mathbb{E}[\sigma^*(\widehat{x},\widehat{y})]$	13.482	40.564	17.016
Surplus Loss from Mismatch	-0.094	-0.092	-0.097

Table 3: Expected Surplus and Mismatch

For the 2D model in the top part of the table, the rows display, in this 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 surplus 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 instead of the *actual* 2D allocation $\frac{\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]-\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]}{\mathbb{E}[\sigma(\mathbf{x},\mathbf{y})]}$, and the percentage surplus loss from implementing the optimal 1D allocation instead of the *optimal* 2D allocation $\frac{\mathbb{E}[\sigma^*_{1D}(\mathbf{x},\mathbf{y})]-\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]}{\mathbb{E}[\sigma^*(\mathbf{x},\mathbf{y})]}$. Similar notation is used for the 1D model in the lower panel of Table 3.

Four features stand out. First, the surplus loss from mismatch can be sizeable in both models (around 10%). 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 the "wrong" optimal allocation (the one suggested by the 1D model), relative to implementing the correct one can be substantial (between 7.7% and 9.8% of aggregate surplus). Third, implementing the 1D frictionless allocation generates essentially no welfare gain compared to the *equilibrium* (2D, frictional) allocation. Last, and importantly, comparison across model specifications suggests that surplus losses from implementing the frictionless 1D allocation tend to be 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 this latter case, the 1D model predicts sorting patterns that are particularly far from the truth, causing a significantly distorted

function that in not one-to-one.

first-best allocation compared to the true first best under multidimensional types.

7.2.4 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. In contrast, "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. in dimension 1 if $x_1 > x_2$. We illustrate the fact 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 focus on Example 2 (not a crucial choice) and examine the way in which the mean job types a worker is matched to in equilibrium change as we scale up his 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 $t = 1,\ 1.4,\ 1.8$. Table 4 shows results for two workers: a "generalist" worker with $\mathbf{x} = (0.3, 0.3)$ and a "specialist" worker with $\mathbf{x} = (0.5, 0.06)$. Those two workers look very similar under the lens of the 1D model, with $\widehat{x} = 0.212$ and $\widehat{x} = 0.209$, respectively.

As the multi-dimensional model predicts, for the generalist there is very little change in sorting on either dimension when scaling up his 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 causes a discernible *increase* in mean job attribute *in the dimension where he is stronger (y_1)*, but a *decrease* 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 1D model misses those differences completely. The two workers in this example look almost the same to the 1D model, so that the predicted path of $\mathbb{E}(\widehat{y}|t\widehat{x})$ as t increases is the same for both workers. Moreover, the 1D model picks up very little sorting, reflecting the tension between the underlying true improvement in y_1 and deterioration in y_2.

	Generalist: $\mathbf{x} = (0.3, 0.3)$			**Specialist:** $\mathbf{x} = (0.5, 0.06)$		
$t =$	1	1.4	1.8	1	1.4	1.8
$\mathbb{E}(y_1 \| t\mathbf{x}) =$	1.6857	1.6853	1.6850	1.6968	1.7010	1.7051
$\mathbb{E}(y_2 \| t\mathbf{x}) =$	1.6856	1.6853	1.6850	1.6737	1.6687	1.6636
$\mathbb{E}(\widehat{y} \| t\widehat{x}) =$	0.6728	0.6736	0.6743	0.6727	0.6735	0.6743

The corresponding estimated 1D skill types are $\widehat{x} = 0.212$ and $\widehat{x} = 0.209$, respectively.

Table 4: Sorting on Absolute Advantage versus Specialization

7.3 Taking Stock

This section highlights some important differences between the implications of the true multi-dimensional model and those obtained from estimating a misspecified one-dimensional model. The 1D model collapses the multiple dimensions into a single index, leading to estimated sorting patterns that are largely uninformative about the true, multi-dimensional ones. As a further consequence, the 1D model tends to overestimate mismatch and is liable to suggest policies that would generate severe welfare losses compared to implementing the optimal two-dimensional allocation. Our exercises further suggest that the 1D approximation becomes even more misleading when the cross-partials of the true technology are sign-varying. We conclude that multi-dimensional heterogeneity is crucial for estimating sorting and mismatch, and for specifying effective policies to reduce allocative inefficiencies.

It is important to note, however, that there are other parts of the estimation on which the misspecified 1D model performs well. Table 5 shows the values of λ_0 and λ_1 as estimated from the 1D model (along with their true values, for comparison).[27] Clearly, the 1D model gets the transition parameters right. This is unsurprising, as the estimators of λ_0 and λ_1 presented in Section 7.1.2 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 as well.

[27]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.

	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

8 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 TU 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 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). 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. 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.

The literature has also analyzed environments affected by search frictions, as well as frictionless environments where agents are characterized by multi-dimensional heterogeneity.

Lindenlaub (2014) develops a framework for the analysis of multi-dimensional sorting under TU but in a context *without* search frictions and *with* capacity constraint on the firm side.[28] 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 reflects a pure matching with positive sorting within tasks. PAM occurs in equilibrium if the production function exhibits worker-firm complementarities within tasks (and no complementarities between tasks). This is a distribution-free technological requirement and most closely resembles our condition for sorting under separable production functions (Theorem 6). But it is quite different from our other general results on the sign of sorting that also involve restrictions on the sampling *distribution*.

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

[28]Note that there also is a 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.

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

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 the 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.

In this environment, we establish conditions for sorting to obtain in equilibrium. Our central condition for PAM takes the form of an intuitive *single-crossing* condition on the technology, confirming the well-known importance of complementarities for sorting. Note, however, that 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 a single dimension, we find that the conditions for PAM in multiple dimensions under random search are generally not distribution-free: they involve not only sufficient complementarities in types but also restrictions on the sampling distribution.

Our restriction on the technology is similar (but not equivalent) to that needed in the multi-dimensional, friction*less* assignment problem by Lindenlaub (2014), who also has to discipline complementarities across tasks to ensure PAM within tasks. But compared to Lindenlaub (2014), the introduction of random search opens up the possibility to study multi-dimensional sorting on the NE and EE margin, as well as multi-dimensional mismatch and policy implications — thereby enlarging the domain of economic applications of multi-dimensional matching.

9 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 dis-

[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 ($(\ln f_{xy})_{xy} > 0$).

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

tribution of a more skilled worker compared to a less skilled worker. We then provide conditions on the primitives of this economy 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.

Our theory yields rich implications for sorting patterns across different dimensions of worker-firm heterogeneity and for sorting on absolute vs comparative advantage. It is important to note that sorting arises here *only* due to multi-dimensional heterogeneity: in a comparable one-dimensional model, all workers share the same ranking of firms, which prevents sorting.

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 a consequence, policy recommendations that are based on the estimated (misspecified) one-dimensional model and that aim to reduce allocative inefficiencies may be severely misguided.

References

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

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

[3] 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.

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

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

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

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

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

[9] 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.

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

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

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

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

[14] 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.

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

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

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

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

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

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

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

[22] 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:

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

This ODE solves as:

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

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:

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

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:

$$\begin{aligned} \frac{\partial}{\partial x_j}\left[1 - F_{\sigma|\mathbf{x}}(s)\right] &= \int \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = s\} \frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_j}\gamma(\mathbf{y}')d\mathbf{y}' \\ &= \mathbb{E}_\Gamma\left[\frac{\partial \sigma(\mathbf{x},\mathbf{y}')}{\partial x_j} \mid \sigma(\mathbf{x},\mathbf{y}') = s\right] \times F'_{\sigma|\mathbf{x}}(s). \end{aligned} \tag{9}$$

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:

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

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

$$\begin{aligned} K_j(y,s|\mathbf{x}) &= \int \mathbf{1}\{y_j \leq y\} \times \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}) \leq s\} h(\mathbf{y}|\mathbf{x})\, d\mathbf{y} \\ &= \frac{\delta\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \int \frac{\mathbf{1}\{0 \leq \sigma(\mathbf{x},\mathbf{y}) \leq s\} \times \mathbf{1}\{y_j \leq y\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right]^2}\gamma(\mathbf{y})d\mathbf{y} \end{aligned} \tag{10}$$

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 $K_j(+\infty,s|\mathbf{x}) = G_{\sigma|\mathbf{x}}(s)$.

Note that:

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

Now on to term (3). Again from (9), we have that:

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

Substituting into term (3):

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

which can be recast as:[32]

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

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

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, $Y \geq 2$)

The objective is to find conditions for PAM in dimension (y_j, x_k). Based on Corollary 1 this means we want to specify conditions under which $\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k}\mid \sigma(\mathbf{x},\mathbf{y})=s, y_j=y'\right]$ is *increasing* in y'.

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 $\mathcal{Y}_{-Y}\in \bigtimes_{j=1}^{Y-1}\left[\underline{y}_j,\overline{y}_j\right]$.

[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, $\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}')=s\right\}\times\mathbf{1}\left\{y_j'\leq y\right\}\gamma(\mathbf{y}')d\mathbf{y}'=0$. The formula thus remains correct with $[0,+\infty)$ as integration bounds if we adopt the convention that any expectation conditioned on a zero probability event is equal to zero.

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

$$\sigma\left(\mathbf{x},\left(\mathbf{y}_{-Y}, y_Y\right)\right)=s \Leftrightarrow f\left(\mathbf{x},\left(\mathbf{y}_{-Y}, y_Y\right)\right)=b(\mathbf{x})+s \tag{11}$$

Then strict monotonicity of $y_Y \mapsto f\left(\mathbf{x},\left(\mathbf{y}_{-Y}, y_Y\right)\right)$ (Assumption 2a) guarantees that at most one value of $y_Y \in \left[\underline{y}_Y, \overline{y}_Y\right]$ solves (11). In turn, assumption 2b ensures that there *always* exists a unique $y_Y \in \left[\underline{y}_Y, \overline{y}_Y\right]$ that solves (11). 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 $j \in\{1, \cdots, Y-1\}$:

$$\frac{\partial R(S, \mathbf{y}_{-Y})}{\partial y_j}=-\frac{\partial f / \partial y_j}{\partial f / \partial y_Y}\left(\mathbf{x},\left(\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y})\right)\right) \tag{12}$$

and:

$$\frac{\partial R(s, \mathbf{y}_{-Y})}{\partial s}=\frac{1}{\partial f / \partial y_Y}\left(\mathbf{x},\left(\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y})\right)\right) \tag{13}$$

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

$$\frac{d}{d y_j}\left(\frac{\partial \sigma\left(\mathbf{x},\left(\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y})\right)\right)}{\partial x_k}\right)=\frac{\partial^2 f}{\partial x_k \partial y_j}-\frac{\partial f / \partial y_j}{\partial f / \partial y_Y} \times \frac{\partial^2 f}{\partial x_k \partial y_Y} \tag{14}$$

where (12) was used and where the arguments of $f\left(\mathbf{x},\left(\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y})\right)\right)$ were omitted in the r.h.s. to reduce notational clutter. Condition (SC-Yd) in the theorem then implies, together with (12) that $\partial \sigma\left(\mathbf{x},\left(\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y})\right)\right) / \partial x_k$ is increasing in all elements of $\mathbf{y}_{-Y} \in \bigtimes_{j=1}^{Y-1}\left[\underline{y}_j, \overline{y}_j\right]$.

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

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

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

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

with $\partial R(s, \mathbf{y}_{-Y}) / \partial s$ given in (13). We will now derive conditions under which

$$y \mapsto \mathbb{E}_{\mu_{s, \mathbf{x}}}\left(\frac{\partial \sigma\left(\mathbf{x},\left(\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y})\right)\right)}{\partial x_k} \mid y_\ell=y\right) \tag{15}$$

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

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

$y_Y = R(s, \mathbf{y}_{-Y}) \in \left[\underline{y}_Y, \overline{y}_Y\right]$ 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

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

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 that $\partial\sigma\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right)/\partial x_k$ is increasing in all elements of $\mathbf{y}_{-Y} \in \bigtimes_{j=1}^{Y-1} \left[\underline{y}_j, \overline{y}_j\right]$ and given that $\operatorname{Supp} \mu_{s,\mathbf{x}}$ is a lattice (both established above), a sufficient condition for (15) 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):

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

□

Remark: The proof for sufficient conditions for NAM is similar. In this case, we need to find conditions under which $\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}) = s, y_j = y'\right]$ is *decreasing* in y' (see Corollary 1), or, equivalently, under which $\mathbb{E}_\Gamma\left[-\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}) = s, y_j = y'\right]$ is *increasing* in y'. To apply the results from Karlin and Rinott (1980) as in the proof for PAM, we thus need a condition under which (using the same change of variable as above) $-\partial\sigma\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right)/\partial x_k$ is increasing in all $\mathbf{y}_{-Y}$ or, equivalently, under which $\partial\sigma\left(\mathbf{x}, (\mathbf{y}_{-Y}, R(s, \mathbf{y}_{-Y}))\right)/\partial x_k$ is *decreasing* in all y_j. This is the case if (14) is *negative* for all $j \in \{1, \cdots, Y-1\}$, i.e. if the generalized signal crossing condition (SC-Yd) reverses its sign for all $j \in \{1, \cdots, Y-1\}$

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f(\mathbf{x},\mathbf{y})/\partial y_j}{\partial f(\mathbf{x},\mathbf{y})/\partial y_Y}\right) = \frac{\partial^2 f}{\partial x_k \partial y_j} - \frac{\partial f/\partial y_j}{\partial f/\partial y_Y} \times \frac{\partial^2 f}{\partial x_k \partial y_Y} \leq 0. \tag{16}$$

This is the only change to Theorem 4 if we want to provide sufficient conditions for NAM. If (16) holds along with the remaining conditions from the theorem, then sorting is NAM in dimensions (y_j, x_k), $j \in \{1, \cdots, Y-1\}$, on the EE-margin.

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

When $Y = 2$, equation (11) writes as $f(\mathbf{x}, (y_1, y_2)) = b(\mathbf{x}) + s$. Strict monotonicity of $y_2 \mapsto f(\mathbf{x}, (\mathbf{y}_1, y_2))$ still guarantees that at most one value of $y_2 \in \left[\underline{y}_2, \overline{y}_2\right]$ solves (11). Moreover, the set of y_1 such that (11) has one solution is an interval (possibly empty). To see this, suppose there exist two distinct values $y_1' < y_1''$ such that (11) 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 $f(\mathbf{x}, (y_1(t), ty_2' + (1-t)y_2'')) \geq b(\mathbf{x}) + s$. Moreover, by assumption, $\min_{y_2 \in \mathbb{R}} f(\mathbf{x}, (y_1(t), y_2)) < b(\mathbf{x}) \leq b(\mathbf{x}) + s$. By continuity, (11) has a solution $R(s, y_1(t)) \leq ty_2' + (1-t)y_2''$ when $y_1 = y_1(t)$. Note that this solution can be smaller than $\underline{y}_2$: in this case, $\gamma(y_1(t), R(s, y_1(t))) = 0$.

Denote the interval of values of y_1 for which (11) 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 (14) holds for all $y \in \mathcal{I}_1(s)$:

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

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:

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

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 (15) 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.3.5 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. □

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 $\Pr_\Gamma\{y'_\ell \leq y|\sigma(\mathbf{x},\mathbf{y}') = s\}$ 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-

[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

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

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

supermodular functions is log-supermodular). This implies in turn that $g(\cdot)$ is log-supermodular in (y, S) 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 $\Pr_\Gamma \{y'_\ell \le y | \sigma(\mathbf{x}, \mathbf{y}') = s\}$ is decreasing in s.

Second, we will provide a condition guaranteeing that $\mathbb{E}_\Gamma [\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k \mid \sigma(\mathbf{x}, \mathbf{y}) = 0] \le 0$. 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 $\mathbb{E}_\Gamma [\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k \mid \sigma(\mathbf{x}, \mathbf{y}) = 0] \le 0$ is that the value of the linear program

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

be nonpositive. This is a sufficient condition which ensures that $\partial\sigma(\mathbf{x}, \mathbf{y})/\partial x_k \le 0$ 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:

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

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 $q_{j'}(\mathbf{x}) \ge 0$, which satisfies the constraints and gives (20) an infinite value. We thus assume that $q_j(\mathbf{x}) \ge 0$ for all $j \in \{1, \cdots, Y-1\}$.

The dual of (20) is simply:

$$\sum_{j=1}^{Y-1} \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})} \left(\underline{y}_j - b_j\right) + \min_Z \left\langle -Z\sigma\left(\mathbf{x}, \underline{\mathbf{y}}\right)\right\rangle \tag{21}$$

$$\text{subject to:} \quad q_j(\mathbf{x}) Z \geq \frac{q_{kj} q_Y(\mathbf{x}) - q_{kY} q_j(\mathbf{x})}{q_Y(\mathbf{x})}, \; j = 1, \cdots, Y-1$$

$$Z \geq 0$$

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:

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

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

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

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

Sufficiency. 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 nonpositive for $Y = 2$ which collapses to $\underline{y}_2 \geq b_2$ (Assumption 3 in Theorem 3).

Necessity. To show that single crossing condition (SC-2d) is also necessary for PAM when considering *all* possible sampling distributions γ, recall from Corollary 2 that there is PAM in dimension (x_k, y_j) on the NE margin if and only if

$$\Bigg[\Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = 0 \right\} \times \left\{ \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = 0, y'_j \leq y \right] - \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = 0 \right] \right\}$$
$$+ \mathbb{E}_\Gamma \left[\frac{\partial \sigma(\mathbf{x}, \mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x}, \mathbf{y}') = 0 \right]$$
$$\times \left[\int_0^{+\infty} G'_{\sigma|\mathbf{x}}(s) \left[\Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = 0 \right\} - \Pr_\Gamma \left\{ y'_j \leq y | \sigma(\mathbf{x}, \mathbf{y}') = s \right\} \right] ds \right] \Bigg] \times G'_{\sigma|\mathbf{x}}(0) \tag{23}$$

is negative. It thus suffices to show that there exists *a* sampling distribution γ, under which (SC-2d) is necessary for (23) to be negative.

We focus on bilinear technology and $Y = 2$. First, note the first line in (23) is negative if and only if (SC-2d) holds (by an analogous argument as in Theorem 2). Second, note that if γ is log-supermodular with log-concave marginals, then $\left[\Pr_\Gamma\left\{y_j' \leq y|\sigma(\mathbf{x},\mathbf{y}') = 0\right\} - \Pr_\Gamma\left\{y_j' \leq y|\sigma(\mathbf{x},\mathbf{y}') = s\right\}\right] ds \geq 0$, see the proof of Theorem 5 and the Sufficiency-part above. Hence, e.g., if γ is uniform with independent marginals then $\frac{\partial^2 \ln \gamma}{\partial y_1 \partial y_2} = 0$ and $\frac{\partial^2 \ln \gamma}{\partial y_2^2} = 0$ so that $\left[\Pr_\Gamma\left\{y_j' \leq y|\sigma(\mathbf{x},\mathbf{y}') = 0\right\} - \Pr_\Gamma\left\{y_j' \leq y|\sigma(\mathbf{x},\mathbf{y}') = s\right\}\right] ds = 0$. Only the first part in (23) remains, which is negative only if (SC-2d) holds. □

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

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

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

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:

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

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 Interrelation of Sorting Patterns Across Dimensions

A.6.1 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:

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

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 $\partial_{x_k}\mathbb{E}(\mathbf{y}|\mathbf{x}) = \left(\partial\mathbb{E}(y_1|\mathbf{x})/\partial x_k,\cdots,\partial\mathbb{E}(y_Y|\mathbf{x})/\partial x_k\right)^\top$. 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 (10), we have that

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

$$= \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i + a_i)q_{ij}\frac{\delta\left[\delta + \lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty} \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta + \lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3}$$

$$\times \left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = s\right\} y'_j \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}'\right.$$

$$\left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = s\right\} y'_j \gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}') = s\right]\right\} ds$$

$$= \frac{\delta\left[\delta + \lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty} \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta + \lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3}$$

$$\times \left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = s\right\} \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i + a_i)q_{ij}y'_j \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}'\right.$$

$$\left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = s\right\} \sum_{i=1}^{X}\sum_{j=1}^{Y}(x_i + a_i)q_{ij}y'_j \gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}') = s\right]\right\} ds$$

$$= \frac{\delta\left[\delta + \lambda_1\overline{F}_{\sigma|\mathbf{x}}(0)\right]}{\overline{F}_{\sigma|\mathbf{x}}(0)} \times \int_0^{+\infty} \frac{2\lambda_1 F'_{\sigma|\mathbf{x}}(s)}{\left[\delta + \lambda_1\overline{F}_{\sigma|\mathbf{x}}(s)\right]^3} \times \left(s + (\mathbf{x}+\mathbf{a})^\top \mathbf{Q}\,\mathbf{b}\right)$$

$$\times \left\{\int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = s\right\} \frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\gamma(\mathbf{y}')d\mathbf{y}'\right.$$

$$\left. - \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = s\right\} \gamma(\mathbf{y}')d\mathbf{y}' \times \mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k} \mid \sigma(\mathbf{x},\mathbf{y}') = s\right]\right\} ds = 0,$$

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

A.6.2 Proof of Corollary 6.

For the bilinear technology with $q_j(\mathbf{x}) \geq 0$ for all j, vector $(\mathbf{x}+\mathbf{a})^T\mathbf{Q}$ has only positive elements. Theorem 8 then implies that if sorting is positive in (y_j, x_k) for all $j \in \{1, \cdots, Y-1\}$, implying $\frac{\partial \mathbb{E}(y_j|\mathbf{x})}{\partial x_k} > 0$ for all $j \in \{1, \cdots, Y-1\}$, then it must be that $\frac{\partial \mathbb{E}(y_Y|\mathbf{x})}{\partial x_k} < 0$. Since $\frac{\partial \mathbb{E}(y_Y|\mathbf{x})}{\partial x_k} > 0$ is necessary for $\frac{H_Y(y|\mathbf{x})}{\partial x_k} < 0$ to hold for all values of $\mathbf{y}$ (i.e. necessary for PAM in dimension (y_Y, x_k)), it must be that $\frac{H_Y(y|\mathbf{x})}{\partial x_k} > 0$ on some set of values $\mathbf{y}$ of positive measure and thus there cannot be PAM in dimension (y_Y, x_k).

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:

$$(\rho+\delta)S(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x}) + \lambda_1 \int m(\mathbf{x},\mathbf{y},\mathbf{y}')\left[\Omega^m(\mathbf{x},\mathbf{y},\mathbf{y}') - S(\mathbf{x},\mathbf{y})\right]\gamma(\mathbf{y}')d\mathbf{y}' \tag{24}$$

$$\rho U(\mathbf{x}) = b(\mathbf{x}) + \lambda_0 \int m(\mathbf{x},\mathbf{0}_Y,\mathbf{y}')\Omega^m(\mathbf{x},\mathbf{0}_Y,\mathbf{y}')\gamma(\mathbf{y}')d\mathbf{y}' \tag{25}$$

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:

$$\begin{aligned}\rho U(\mathbf{x}) &= b(\mathbf{x}) + \lambda_0\beta\int \mathbf{1}\left\{S(\mathbf{x},\mathbf{y}') \geq 0\right\} S(\mathbf{x},\mathbf{y}')\gamma(\mathbf{y}')d\mathbf{y}' \\ &= b(\mathbf{x}) + \lambda_0\beta\int_0^{+\infty} S dF_{S|\mathbf{x}}(S) = b(\mathbf{x}) + \lambda_0\beta\int_0^{+\infty}\overline{F}_{S|\mathbf{x}}(S)dS\end{aligned} \tag{26}$$

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 (26) 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:

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

As is clear from (27), $S(\mathbf{x},\mathbf{y})$ 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 (27), is given by:

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

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 (27) that $S(\mathbf{x},\mathbf{y}) = 0 \iff \sigma(\mathbf{x},\mathbf{y}) = 0$. Hence, the job acceptance rule is equivalent to $m(\mathbf{x},\mathbf{y},\mathbf{y}') = \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') > \sigma(\mathbf{x},\mathbf{y})\right\}$ for employed workers and $m(\mathbf{x},\mathbf{0}_Y,\mathbf{y}') = \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') > 0\right\}$ 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 $m(\mathbf{x},\mathbf{y},\mathbf{y}')=\mathbf{1}\left\{S(\mathbf{x},\mathbf{y}')>S(\mathbf{x},\mathbf{y})\right\}$. 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:

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

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

$$\left[\rho+\delta+\lambda_1(1-\beta)\overline{F}_{S|\mathbf{x}}\left(S(\mathbf{x},\mathbf{y})\right)\right]S(\mathbf{x},\mathbf{y})=\sigma(\mathbf{x},\mathbf{y})-\lambda_1\beta\int_0^{S(\mathbf{x},\mathbf{y})}\overline{F}_{S|\mathbf{x}}(S)dS \tag{28}$$

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:

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

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:

$$\frac{SF'_{S|\mathbf{x}}(S)}{\overline{F}_{S|\mathbf{x}}(S)}\leq\frac{\rho+\delta+\lambda_1}{\lambda_1(1-\beta)}\quad\iff\quad\frac{S\int\mathbf{1}\left\{S(\mathbf{x},\mathbf{y}')=S\right\}\gamma(\mathbf{y}')d\mathbf{y}'}{\int\mathbf{1}\left\{S(\mathbf{x},\mathbf{y}')\geq S\right\}\gamma(\mathbf{y}')d\mathbf{y}'}\leq\frac{\rho+\delta+\lambda_1}{\lambda_1(1-\beta)}. \tag{29}$$

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 (29) 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 $F_{f|\mathbf{x}}$ 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 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.

[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 $\widetilde{b}(\mathbf{x}) = \max\{R(\mathbf{x}); \min_{\mathbf{y}' \in \mathcal{Y}} f(\mathbf{x}, \mathbf{y}')\}$. Then, the integration constant $C(\mathbf{x})$ is zero if $\widetilde{b}(\mathbf{x}) = R(\mathbf{x})$ and some positive function of $\mathbf{x}$ otherwise.