The Worker-Job Surplus*

Ilse Lindenlaub[†] Fabien Postel-Vinay[‡]

July 2020

Abstract

The worker-job surplus — the sum of the worker's and the employer's values of an employment relationship — is a key object in most matching models of the labor market. It drives workers' employment transitions and wages, as well as equilibrium sorting patterns. In this paper, we develop a theory-based empirical method to determine which of the *observable* worker and job characteristics impact the worker-job surplus in the data, where we exploit the mobility choices of employed workers. Our method further indicates whether workers *sort* along those surplus-relevant attributes when searching for jobs. Finally, it provides a test of the commonly used *single-index assumption*, according to which multi-dimensional worker and job heterogeneity can be collapsed into scalars. We implement our method on US data using the Survey of Income and Program Participation and the O*NET. The results suggest that a relatively sparse model underlies the data. On the job side, a cognitive and an interpersonal skill requirement impact the surplus along with the (dis)amenity of work duration as well as the workplace size. On the worker side, we find that most of the relevant characteristics are symmetric to the selected job requirements. We reject the existence of a single-index representation of these relevant multi-dimensional worker and job attributes. Finally, we use our results to shed light on multi-dimensional sorting along the relevant dimensions in the data, where we document the differential collapse of distinct job ladders during the Great Recession and its recovery.

Keywords. Worker-Job Surplus, Random Search, Multi-Dimensional Sorting, Single Index, Worker Employment-to-Employment Mobility.

*We thank various seminar audiences 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

The worker-job surplus (or match surplus) — the sum of the worker's and the employer's values of an employment relationship — is a key object in most matching models of the labor market. It is the key determinant of workers' mobility choices and their wages. The surplus is also essential for understanding the extent of mismatch and the associated welfare loss. Measuring the surplus correctly is therefore central for designing policies that improve match quality between workers and jobs; or for creating training programs that educate workers in the relevant skills. Yet our understanding of which worker and job characteristics impact the surplus in the data is limited. As a result, the standard simplifying assumption in the macro-labor literature is that the surplus is a function of a single (either observable or unobservable) worker trait and a single job trait.

In this paper, we aim to make progress on the empirical characterization of the determinants of the worker-job surplus. Our primary contribution is methodological. We develop a theory-based empirical protocol to determine how many and which of the *observable* worker and job characteristics affect the worker-job surplus in the data, where we exploit workers' mobility choices. Our method also indicates whether agents sort on those surplus-relevant attributes. As a by-product, our method reveals whether the *single index assumption* is valid, by which all relevant worker characteristics can be summarized by scalar heterogeneity without loss of information, and similarly for job characteristics. Our second contribution is applied. We demonstrate how to implement our empirical protocol on US data and show that multiple (but few) worker and job attributes are surplus-relevant. We reject the widely used assumption that the relevant multi-dimensional heterogeneity can be condensed into single indices and our application — which suggests the differential collapse of distinct job ladders after the Great Recession — highlights why accounting for multi-dimensional heterogeneity is important.

The intuitive underpinning of our method is simple. It builds upon the premise that most direct job-to-job transitions are voluntary: workers tend to move from job to job in order to improve their situation, which means moving to higher-surplus jobs. Under this assumption, observing individual job mobility decisions — and observing a set of job and worker characteristics that are the determinants of those choices — should inform us about which job and worker characteristics are indeed surplus-relevant.

For illustration, we focus on jobs and workers with two-dimensional heterogeneity. Consider two workers with the same realization of the first skill dimension x_1 (e.g. 'cognitive' skills), currently in jobs with the same (cognitive) job attribute y_1. But say they differ regarding the second skill, x_2 (e.g. 'manual' skills), and their jobs differ regarding the manual job attribute y_2. If the cognitive dimension fully captures the heterogeneity that is relevant for surplus maximization, then both workers should have identical job acceptance sets, i.e. climb the *same job ladder*. If however the manual characteristics on either side of the market also matter for the surplus, then the job acceptance decisions of the two workers should differ, meaning they climb different job ladders. Thus, testing for the heterogeneity of job acceptance sets *within* 1-D worker types (who are currently in *the same* 1-D job) allows us to either validate the assumption of scalar heterogeneity or to reject it. If we reject it, 'adding' heterogeneity is necessary to account for differences in mobility choices and therefore the surplus depends on multi-dimensional attributes. In the general case (beyond this two-dimensional example), we systematically add heterogeneity until workers' mobility decisions are fully accounted for, meaning that workers with the same attributes who are currently in jobs with the same characteristics make the same mobility choices.

We formally prove that, based on this logic, we can identify several important properties of the surplus function when the underlying framework is a general, random search job-ladder model in which workers and jobs are joint surplus maximizers. Mobility choices are driven by surplus comparisons between current job and incoming job offer. In this setting, we show that the dependence of workers' job-to-job transition rate (our proxy for job acceptance sets) on worker and current job attributes reveals that those attributes are surplus-relevant. By implication, any characteristic that the worker (or job) may have but does not enter the surplus function will *not* impact the workers' transition rates, even if it is correlated with the attributes that truly affect surplus. We also show that any surplus-relevant worker and job attribute will be detected by the fact that the job-to-job transition rate depends on it. Checking the dependence of the job-to-job transition rate on worker and job traits is the essence of our empirical protocol.

Aside from being able to detect which of the agents' characteristics are surplus-relevant, our theory indicates that the dependence of job-to-job transition rates on worker attributes conveys information about several other important properties of the surplus function. It indicates that the surplus function satisfies a single-crossing property, with two fundamental implications: First, we show that single-crossing is necessary and sufficient for sorting between workers and jobs. Second, single-crossing implies that the relevant multi-dimensional heterogeneity of workers and jobs cannot be collapsed into single indices. Combining these insights, we prove that whenever there is worker-job sorting involving multi-dimensional characteristics, then there exists no single-index representation. Owing to our theory, our empirical method thus also allows us to either validate or reject the single-index assumption, which is a pillar of this literature.

We can thus identify several properties of the surplus function from observing the dependence of the worker transition rate on worker and job attributes in the data — without estimating the model structurally. In particular, our method allows us to learn about something that is of great interest but unobserved (properties of the surplus function) based on something which we can observe in the data (dependences of the employment-to-employment — or EE — transition rate on worker and job attributes). An important advantage of our framework is its generality. Our theory-based protocol and its implications do not hinge on overly restrictive assumptions. Indeed, we do not make any functional form assumptions and we do not need a specific wage-setting protocol, which often amounts to an arbitrary choice. But, proving our results in this generality is technically challenging. We rely on tools from Differential Topology that have been used in General Equilibrium analysis but are not at all standard in the macro-labor literature.

Based on our theoretical results, we build an empirical model of the job-to-job transition rate, which allows for its dependence on all *potential* worker and job attributes that are observed in the data. We then use model selection methods to determine which worker and job attributes impact the EE transition rate.

Since we propose a new methodology to learn about properties of the worker-job surplus in the data, we want to first test it in Monte Carlo simulations. Our theory makes this exercise possible. In settings where we know the underlying data generating process, we can check whether our method successfully separates the 'true' worker and job characteristics that explain EE transitions from characteristics that do not explain them. Moreover, we want to assess which of the many available model selection tools performs best in our context. We find that our method performs well, especially so in settings where the number of potential worker and job attributes is not too large and where the data features a significant amount of worker-job sorting. Moreover, the Bayesian Information Criterion (BIC) is most effective in selecting the true worker and job heterogeneity in our context, compared to the Akaike Information Criterion (AIC), the Lasso and several others. These results then guide our choices when implementing the method on real-world data.

We apply our method to US data using the Survey of Income and Program Participation (SIPP) as our main source, supplemented with the O*NET. We first impute multi-dimensional worker characteristics (both skills and preferences for job amenities or tolerance for dis-amenities) in the SIPP using its special Education and Training module in combination with the O*NET — a method we developed in Lindenlaub and Postel-Vinay (2020).[1] In turn, we obtain the multi-dimensional job attributes (both skill requirements and job amenities) for each occupation (our empirical counterpart of the model's jobs) from the O*NET. Our baseline sample consists of the 2008 panel of the SIPP (2009-2013) but we also implement the test on the earlier panels going back to 1996, when the special Education and Training module was first launched.

We obtain the following insights about the worker-job surplus: first, our protocol indicates how many and which worker and job attributes are surplus and sorting relevant. We find that a relatively *sparse model* gets selected. On the job side, a routine cognitive and an interpersonal skill requirement get selected along with the (dis)amenity of work duration as well as workplace size. On the worker side, we find that most of the relevant worker characteristics are symmetric to the selected job attributes (cognitive skill, interpersonal skill and the workers' 'tolerance' for long work hours). In addition, two distinct manual skills predict worker EE mobility. Based on our theory, we conclude that these attributes are not only surplus but also *sorting* relevant.

Second, our results — by highlighting worker attributes as significant predictors of the EE transition rates — imply that surplus satisfies a single-crossing property, so we can *reject* that there exists a *single index* representation of the relevant multi-dimensional worker and job attributes. Thus, heterogeneity is truly multi-dimensional in the data, casting doubt on the validity of the widely-used single index assumption.

Third, given that sorting in the data is based on multi-dimensional attributes on both sides of the market that cannot be collapsed into a single index, it must be the case that workers with different attribute bundles rank jobs in different ways. Consequently, there is *no single economy-wide job ladder* that all workers agree on. Instead, workers with different skill bundles rank jobs in different ways and climb different job ladders.

We also implement the method on earlier panels of the SIPP (1996, 2001, 2004). We find that the types of attributes that impact surplus are quite consistent over time, especially for jobs.

To highlight the economic relevance of our findings, we pursue an application that investigates multi-dimensional sorting along the *relevant* dimensions of job and worker attributes, both in the cross-section and over time. We find salient changes in multi-dimensional sorting over time, showing a decline in positive sorting between worker and job attributes — or increase

[1] Compared to Lindenlaub and Postel-Vinay (2020), here we focus on a considerably larger sample of the SIPP, including not only those with college degree but also those with apprenticeship degree or occupational training.

in mismatch — between 2009 and 2013. All three main dimensions (cognitive, interpersonal and work duration) were affected but there are large quantitative differences between them. The interpersonal dimension experienced by far the most severe drop in sorting. We find that both sorting changes along the EE margin (collapse of the 'interpersonal job ladder') as well as along the employment-to-unemployment (EU) and unemployment-to-employment (UE) margins contributed to this increase in interpersonal mismatch. Our exploratory data analysis suggests that neither changes in search frictions nor in the marginal distributions of worker and job attributes were responsible for this trend. Instead, we show that the returns from sorting into jobs with high interpersonal requirements are *negative* during 2009-2013 and, moreover, being employed in those jobs, even if well matched, make separations of workers into unemployment more likely. This suggests that there are economic incentives why workers do not find it worthwhile to seek jobs whose interpersonal skill requirements fit their interpersonal skills, leading to a significant drop in interpersonal sorting. These results may be suggestive of asymmetric changes in productivity, depressing worker-job complementarities in the interpersonal dimension most severely during that time period. Our results are surprising in light of recent work that highlights the growing importance of interpersonal skills in the labor market.[2] And the differential collapse of distinct job ladders stresses the importance of accounting for multi-dimensional heterogeneity.

The Literature. A growing literature seeks to identify sorting between workers and jobs/firms in the data and quantify the degree of mismatch as well as the role of sorting in wage inequality. With very few exceptions, the focus is on identifying *unobserved scalar heterogeneity* of both workers and firms.[3] The correlation of those indices across worker-firm matches can then be used as a measure of sorting. Broadly speaking, two different approaches have been used to achieve this objective: a statistical (i.e. model-free) approach and a structural one.

Following the model-free approach, Abowd, Kramarz, and Margolis (1999) (AKM) propose in an influential paper a two-way fixed effects regression to estimate the contribution of unobserved worker and firm heterogeneity (captured by a worker and firm fixed effects) to earnings and assess sorting as the correlation between the estimated fixed effects.[4] More recently, Bon-

[2] See e.g. Deming (2017).

[3] Examples of papers that do not rely on unobserved scalar heterogeneity are Lindenlaub (2017), Lise and Postel-Vinay (2019) and Lindenlaub and Postel-Vinay (2020) who all treat heterogeneity as *observed and multi-dimensional* but fix the number and type of dimensions on both worker and job side by assumption.

[4] Borovickova and Shimer (2018) point out that the AKM measure of sorting suffers from a limited mobility bias. They propose a similar approach as AKM in the sense that wages are used to identify one-dimensional unobserved types, but they differ in the measurement of these types. They identify a worker's type by her expected log wage (across different jobs) and a firm's type by the expected log wage it pays across workers.

homme, Lamadon, and Manresa (2019) propose a two-stage approach to estimate unobserved worker and job heterogeneity and worker-job complementarities in wages, relaxing several assumptions from the fixed effects approach. In the first stage, they identify the unobserved firm types based on K-means clustering of firms' wage distributions. In the second stage, they identify the distribution of unobserved worker types over firm types as well as type-conditional earnings distributions based on the observed earnings distribution of on-the-job movers and stayers.[5]

Following the structural approach, Hagedorn, Law, and Manovskii (2017) identify unobserved scalar heterogeneity of workers and firms based on a version of the random search Shimer-Smith model where matching is one-to-one and the surplus is split by Nash-bargaining. Under the assumption that output is increasing in worker and in firm types, they can find observable statistics that are monotone in types, allowing them to recover a ranking of (unobserved) worker and firm types. In turn, Taber and Vejlin (2016), Bagger and Lentz (2018) and Sorkin (2018) all use revealed preference job-to-job mobility patterns to rank firms and thereby recover the economy-wide firm ladder. In Bagger and Lentz (2018), the firms' 'poaching rank' (fraction of hires that is poached from other firms) moves one-to-one with their unobserved scalar productivity. In Taber and Vejlin (2016) and Sorkin (2018) firms have two characteristics, productivity and amenity/residual. Taber and Vejlin (2016) assume that workers are initially heterogeneous in 1D productivity while Sorkin (2018) assumes they are homogenous, precluding sorting.

Similar to these approaches, our method also exploits the rich information provided by workers' mobility choices in order to pin down worker and job heterogeneity. Nevertheless, our approach differs significantly in its nature and objectives. In contrast to this literature, we focus on identifying what observable worker and job characteristics impact their joint surplus, and we prove that workers' EE mobility choices reveal this information. Instead of assuming that firms (and also workers) can either be ranked on a 1D scale or have 2D characteristics, we allow for the surplus to depend on *any number* of observable characteristics. We then let the data speak on what traits matter for surplus. Our approach of pinning down the surplus-relevant characteristics is thus more flexible. Our results suggest that the 1-factor assumption on either demand or supply side is too restrictive as agents have multiple relevant attributes, which cannot be collapsed into a single-index. To the best of our knowledge, this is the first paper to

[5] Compared to the two-way fixed effects regression by Abowd, Kramarz, and Margolis (1999), Bonhomme, Lamadon, and Manresa (2019) can relax three assumptions: first, the assumption that current wages do not depend on past wages and past firm types given current firm types; second, the assumption that worker mobility is independent of past wages; third the assumptions that log wages are linear in worker and firm heterogeneity.

develop and implement a micro-founded empirical protocol to uncover the determinants of the worker-job surplus in the data.

Our work also relates to the friction*less* Labor literature. Card and Lemieux (1996) propose a single-index model that has been widely used to explain the evolution of wages over time. A key implication of that model is that workers with a similar productivity index should have similar wage growth over time. Otherwise the single index assumption is not valid. Instead of wages we focus on worker mobility. But our idea is related: We are able to test the single index assumption of worker (and job) heterogeneity by analyzing whether workers with similar single index (currently in jobs with a similar single index) make similar job-to-job mobility choices.

2 Theory

We develop a theory-based empirical method to determine *how many* and *which* worker and job attributes are determinants of the worker-job surplus and matter for sorting. We first describe the theoretical framework. We then explain the intuitive underpinning of our method, and formalize it. Finally, we offer an interpretation of our method's output in light of our theory.

2.1 Framework

The theoretical framework is based on Lindenlaub and Postel-Vinay (2020) who analyze multi-dimensional sorting under random search from a theoretical point of view. Here we give a brief description of the model environment.

Time is continuous and the economy is at an aggregate steady state. There is a fixed unit mass of infinitely lived workers, each characterized by a time-invariant bundle $\mathbf{x} = (x_1, \cdots, x_X) \in \mathcal{X}$, where X denotes the number of different worker attributes and $\mathcal{X}$ is a compact and connected subset of $\mathbb{R}^X$. In what follows, we will refer to $\mathbf{x}$ and its elements as a worker's attributes, characteristics, traits, or skills interchangeably. Worker attributes are distributed with cdf L and density ℓ, strictly positive over $\mathcal{X}$. Firms are collections of perfectly substitutable (possibly vacant) jobs and face no capacity constraint. Jobs are characterized by a vector of time-invariant productive attributes, or "skill requirements" $\mathbf{y} = (y_1, \cdots, y_Y) \in \mathcal{Y}$, where Y denotes the number of different job attributes and $\mathcal{Y}$ is a compact and connected subset of $\mathbb{R}^Y$. Jobs are distributed with cdf Γ, and strictly positive and continuously differentiable density γ.

Attributes $\mathbf{x}$ and $\mathbf{y}$ are observed to both the agents in the market and to us, the analysts.

Workers can either be employed or unemployed. In both states, they face search frictions where they sample job offers randomly and sequentially. If matched, they lose their job at Poisson rate δ, and sample alternate job offers from the exogenous 'sampling distribution' of jobs Γ at rate λ_1. Unemployed workers sample job offers from the same sampling distribution at rate λ_0. Note that job contact rates and the sampling distribution are the same for all workers (they are independent of worker type $\mathbf{x}$).

We assume that utility is transferable between firms and workers, and workers are risk-neutral, ensuring a well-defined notion of the *joint match surplus*, generically a function $\sigma(\mathbf{x}, \mathbf{y})$ of all *potentially relevant* job and worker attributes, $(\mathbf{x}, \mathbf{y})$. However, not all worker or job characteristics are necessarily relevant determinants of match surplus in practice. Specifically, we assume that the surplus function σ really only depends on the first X_R (resp. the first Y_R) elements of $\mathbf{x}$ (resp. of $\mathbf{y}$). Irrelevant attributes have no impact on the surplus, meaning that for all $(\mathbf{x}, \mathbf{y})$:

$$\forall k \in \{X_R + 1, \cdots, X\} : \frac{\partial \sigma}{\partial x_k}(\mathbf{x}, \mathbf{y}) \equiv 0 \quad \text{and} \quad \forall j \in \{Y_R + 1, \cdots, Y\} : \frac{\partial \sigma}{\partial y_j}(\mathbf{x}, \mathbf{y}) \equiv 0.$$

We will refer to the first X_R elements of $\mathbf{x}$ and to the first Y_R elements of $\mathbf{y}$ as *surplus-relevant* skills and job attributes, respectively. The special case of $X_R = Y_R = 1$ is that of scalar heterogeneity. Note that we implicitly assume that $X_R \leq X$ and $Y_R \leq Y$, i.e. the vectors $\mathbf{x}$ and $\mathbf{y}$ list *all* surplus-relevant worker and job attributes (and possibly more). In other words, there is no unobserved heterogeneity. We will discuss in Section 2.3 how we can relax this assumption.

The transferable utility assumption has the important implication that jobs and workers are joint surplus maximizers. Hence, a meeting between a type-$\mathbf{x}$ unemployed worker and a type-$\mathbf{y}$ job will result in a match if and only if $\sigma(\mathbf{x}, \mathbf{y}) \geq 0$. Similarly, a meeting between a type-$\mathbf{x}$ worker, employed in job $\mathbf{y}$, and an alternative type $\mathbf{y}'$ job will result in the worker accepting the type-$\mathbf{y}'$ job if and only if $\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})$. These mobility/acceptance decisions are the only decisions made in this economy. In particular, even though there is some surplus-sharing going on in the background, our analysis does not rely on wages and therefore does not require us to specify a specific surplus-splitting rule — something we see as an advantage.

We impose the following regularity assumptions on the surplus function.

Assumption 1 *The surplus function σ is such that, for all $\mathbf{x} \in \mathcal{X}$, $\mathbf{y} \mapsto \sigma(\mathbf{x}, \mathbf{y})$ is a quasi-concave Morse function over $\mathcal{Y}$.*

Quasi-concavity will ensure that the level sets of the surplus function are well-behaved. In turn, Morse functions are smooth functions with the key property of having only isolated critical points. This regularity assumption is needed for purely technical reasons and helps us discipline the zero set of several functions that depend on the surplus. We stress that this assumption is not overly restrictive. Quasi-concave functions form a broad class. Further, Morse functions are dense in the set of smooth functions defined over $\mathcal{Y}$.

Here are some functional forms satisfying Assumption 1:[6] A common one is the bilinear form, $\sigma(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{Q}(\mathbf{y} - \mathbf{b})$, where $\mathbf{b}$ is a Y-dimensional vector that we interpret as the production technology of the unemployed and where $\mathbf{Q}$ is an $X \times Y$ matrix, capturing the complementarity structure between all job and worker characteristics. Note that only those entries in $\mathbf{Q}$ are non-zero that are associated with the surplus-relevant characteristics of workers and jobs: denoting the generic entry of $\mathbf{Q}$ as q_{kj}, we have that $q_{kj} = 0$ for all $k > X_R$ or $j > Y_R$. Another functional form satisfying Assumption 1 is the 'bliss-point' function, $\sigma(\mathbf{x}, \mathbf{y}) = A - \sum_{k,j=k}(x_k - y_j)^2, A > 0$.

Our main results will be established under the assumptions listed in this sub-section. Arguably the most restrictive ones are the independence of the contact rates or sampling distribution on worker type $\mathbf{x}$ (we note though that this assumption is almost invariably made in the literature, except for papers with endogenous search effort), or that of only observed heterogeneity. We will revisit those assumptions in Section 2.3.

2.2 Identifying Surplus-Relevant Job and Worker Attributes

Principle Of Our Approach. The principle underlying our approach to identifying surplus-relevant job and worker characteristics is intuitive: under our assumption of joint surplus maximization, workers move from lower-surplus to higher-surplus jobs. Hence, observing individual job mobility decisions — and observing the job and worker characteristics that *determine* those decisions — should inform us about which job and worker characteristics are surplus-relevant.

Consider a simple two-dimensional example ($X = Y = 2$) for illustration: workers are characterized by two potentially relevant skills: cognitive skills x_1 and manual skills x_2. Likewise, jobs have cognitive and manual skill requirements, y_1 and y_2, as potentially relevant attributes. Then, consider two workers with the same cognitive skills x_1 but different manual skills x_2, currently in jobs with the same cognitive skill requirement y_1 (but different y_2). If that cognitive dimension is the only relevant one for match surplus, then both of those workers should have

[6] These functional forms could be imposed if we treated the match surplus as a primitive as would be justified under the Sequential Auction wage-splitting protocol, see Appendix 8.1.5.

identical job acceptance sets, i.e. they should climb the *same job ladder*. If however the manual dimension also matters for surplus (on either side of the market), then the job acceptance decisions of the two workers should differ, meaning they should climb different job ladders. Thus, comparing job acceptance sets *within* 1-D match types should allow us to either validate the assumption of one-dimensional heterogeneity or to reject it. The same logic applies when testing whether the N-dimensional assumption (for $N > 1$) is justified.

FORMALIZATION. Job acceptance sets are not easily observed in worker-level panel data sets. However, a related statistic — job-to-job (EE) transition probabilities — typically are, and so we use them as a proxy for job acceptance sets. We denote the EE transition probability of worker $\mathbf{x}$ in current job $\mathbf{y}$ by $\tau(\mathbf{x}, \mathbf{y})$. It equals the joint probability of receiving a job offer, λ_1, and accepting it, $\int \mathbb{1}\{\sigma(\mathbf{x}, \mathbf{y}') > \sigma(\mathbf{x}, \mathbf{y})\} \gamma(\mathbf{y}')d\mathbf{y}'$, which is the case if the surplus with the newly drawn job is larger than the one with the current job $\mathbf{y}$

$$\tau(\mathbf{x}, \mathbf{y}) := \lambda_1 \int \mathbb{1}\left\{\sigma(\mathbf{x}, \mathbf{y}') \geq \sigma(\mathbf{x}, \mathbf{y})\right\} \gamma(\mathbf{y}')d\mathbf{y}'.$$

We will also be interested in the *average* EE transition probability in a given job $\mathbf{y}$:

$$\overline{\tau}(\mathbf{y}) := \mathbb{E}\left[\tau(\mathbf{x}, \mathbf{y}) \mid \mathbf{y}\right] = \frac{\int \tau(\mathbf{x}, \mathbf{y})h(\mathbf{x}, \mathbf{y})d\mathbf{x}}{\int h(\mathbf{x}, \mathbf{y})d\mathbf{x}},$$

where we denote by $h(\mathbf{x}, \mathbf{y})$ the equilibrium steady-state density of $(\mathbf{x}, \mathbf{y})$-type matches.

Next, we introduce a definition of a single crossing property of the surplus function, which we will use repeatedly in the analysis that follows:

Definition 1 (Single-Crossing) *We say that the surplus function σ has the* SINGLE-CROSSING (SC) PROPERTY *for attributes (x_k, y_i, y_j) at $(\mathbf{x}, \mathbf{y})$ if and only if:*

$$\frac{\partial \sigma}{\partial y_i}(\mathbf{x}, \mathbf{y})\frac{\partial^2 \sigma}{\partial x_k \partial y_j}(\mathbf{x}, \mathbf{y}) - \frac{\partial \sigma}{\partial y_j}(\mathbf{x}, \mathbf{y})\frac{\partial^2 \sigma}{\partial x_k \partial y_i}(\mathbf{x}, \mathbf{y}) \neq 0 \tag{SC}$$

Assuming for example that $\partial\sigma/\partial y_i(\mathbf{x}, \mathbf{y}) \neq 0$, condition (SC) is equivalent to $\frac{\partial}{\partial x_k}\left(\frac{\partial\sigma/\partial y_j}{\partial\sigma/\partial y_i}\right) \neq 0$ at $(\mathbf{x}, \mathbf{y})$. Intuitively, this condition says that job attribute y_j is either more complementary (if the expression on the LHS is positive) or more substitutable (if it is negative) to worker trait x_k than job attribute y_i in the surplus function.

We are now ready to state our main result, which is the cornerstone of our empirical analysis:

Proposition 1 (Determinants of the EE Transition Probability)

Fix $\mathbf{x} \in \mathcal{X}$ and let $k \in \{1, \cdots, X_R\}$.

(i) Surplus-Relevant Job Attributes:

a. *For all $\mathbf{y} \in \mathcal{Y}$, $\frac{\partial \tau}{\partial y_j}(\mathbf{x}, \mathbf{y}) \lesseqgtr 0$ if and only if $\frac{\partial \sigma}{\partial y_j}(\mathbf{x}, \mathbf{y}) \gtreqless 0$.*

b. *For all $\mathbf{y} \in \mathcal{Y}$, $\frac{\partial \overline{\tau}}{\partial y_j}(\mathbf{y}) \neq 0$ only if there exists $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $\frac{\partial \sigma}{\partial y_j}(\tilde{\mathbf{x}}, \mathbf{y}) \neq 0$. Conversely, for all $\mathbf{y} \in \mathcal{Y}$ outside of a set of zero measure, if $\frac{\partial \sigma}{\partial y_j}(\tilde{\mathbf{x}}, \mathbf{y}) \neq 0$ on a set of $\tilde{\mathbf{x}}$ of positive measure, then $\frac{\partial \overline{\tau}}{\partial y_j}(\mathbf{y}) \neq 0$.*

(ii) Surplus-Relevant Worker Attributes:

a. *For all $\mathbf{y} \in \mathcal{Y}$, if $\frac{\partial \tau}{\partial x_k}(\mathbf{x}, \mathbf{y}) \neq 0$, then there exist job attributes $(i, j) \in \{1, \cdots, Y_R\}^2$ and a point $\tilde{\mathbf{y}} \in \mathcal{Y}$ such that the (SC) property holds for (x_k, y_i, y_j) at $(\mathbf{x}, \tilde{\mathbf{y}})$.*

b. *For all $\mathbf{y} \in \mathcal{Y}$ outside of a set of zero measure, if there exist job attributes $(i, j) \in \{1, \cdots, Y_R\}^2$ such that the (SC) property holds for (x_k, y_i, y_j) at $(\mathbf{x}, \mathbf{y})$, then $\frac{\partial \tau}{\partial x_k}(\mathbf{x}, \mathbf{y}) \neq 0$.*

(iii) Surplus-Relevant Interactions of Worker-Job Attributes:

For all $\mathbf{y} \in \mathcal{Y}$, if $\frac{\partial^2 \tau}{\partial x_k \partial y_j}(\mathbf{x}, \mathbf{y}) \neq 0$, then there exist $(\hat{\mathbf{y}}, \tilde{\mathbf{y}}) \in \mathcal{Y}^2$ such that $\frac{\partial \sigma}{\partial y_j}(\mathbf{x}, \hat{\mathbf{y}}) \neq 0$ and $\frac{\partial \sigma}{\partial x_k}(\mathbf{x}, \tilde{\mathbf{y}}) \neq 0$.

The proof is in Appendix 8.1.2. Proving this result requires some subtle arguments from Differential Topology (primarily applications of the Transversality Theorem and Preimage Theorem) to show that these derivatives are 'generically' not zero under the stated conditions. But despite technical challenges, the underlying intuition is straightforward: we can identify several properties of the surplus function σ from observing the dependency of the EE rate on worker and job attributes in the data — without estimating the model structurally. Based on this result, we will approximate the conditional expectation function of τ, $\mathbb{E}(\tau|\mathbf{x}, \mathbf{y})$, by an expansion of the following form,

$$\tau_i(\mathbf{x}, \mathbf{y}) = \alpha + \sum_{k}^{X} \beta_k x_{ik} + \sum_{j}^{Y} \gamma_j y_{ij} + \sum_{k}^{X} \sum_{j}^{Y} \delta_{kj} x_{ik} y_{ij} + \epsilon_i \tag{1}$$

which is our empirical model of the EE transition rate. Subscript i indicates an individual, α is a constant, $(\beta_k, \gamma_j, \delta_{kj})$ are parameters and ϵ_i is a mean-zero error term. We will use model selection methods to detect which worker and job attributes impact the EE transition rate. We now discuss the interpretation of such observations (that is, of the observed dependence of τ on certain worker and job attributes as well as their interactions) based on Proposition 1 and our theory more generally.

Interpreting Estimation Output. Part (i) of Proposition 1 states that from the observed dependency of the (conditional or unconditional) EE transition rate on some job attribute y_j, we can conclude that this job attribute is *surplus-relevant* (following from necessity). Moreover, any surplus-relevant job attribute will be picked up by checking the dependency of the (conditional or unconditional) EE rate on that attribute (following from sufficiency).

Part (ii) of Proposition 1 allows us to draw conclusions from the dependency of the EE transition rate τ on worker attribute x_k about the relevance of this characteristic for both surplus *and* sorting. Surplus-relevance of x_k follows directly if τ depends x_k since this dependency indicates that the (SC) property holds involving x_k, which can only be true if this attribute impacts surplus. The reason why we also learn about the sorting-relevance of worker attributes is that (SC) indicates productive complementarities and is thus tightly linked to sorting between worker and job attributes. Indeed, the statement below shows that (SC), if it holds everywhere, implies sorting in the following sense. We say that an equilibrium worker-job allocation exhibits sorting in dimension (x_k, y_j) if the equilibrium distribution of job attribute y_j conditional on worker attributes, denoted by $H_j(y|\mathbf{x}) := \int \mathbb{1}\{y_j \leq y\}\, h(\mathbf{y}|\mathbf{x})\, d\mathbf{y}$, is not invariant to worker skill x_k, i.e. if $\partial H_j/\partial x_k(y|\mathbf{x}) \neq 0$ for some y. In words, sorting between x_k and y_j means that two workers who differ in their skill x_k tend to have jobs that differ in attribute y_j. We thus do not restrict our attention to either positive or negative sorting but allow for either (including for the sign of sorting to vary across the support of y_j) since based on the dependency of τ on x_k we cannot tell them apart.[7] In our empirical application below, we address the sign of sorting directly.

Proposition 2 (Single Crossing and Sorting) *Fix $\mathbf{x} \in \mathcal{X}$, a worker attribute $k \in \{1, \cdots, X_R\}$ and a job attribute $j \in \{1, \cdots, Y_R\}$.*

a. If, for all $\mathbf{y} \in \mathcal{Y}$: $\sigma(\mathbf{x}, \mathbf{y}) > 0$, there exists a job attribute $i \in \{1, \cdots, Y_R\}$ such that (SC) holds for (x_k, y_i, y_j) at $(\mathbf{x}, \mathbf{y})$, then there is sorting in dimensions (x_k, y_j), i.e. $\partial H_j/\partial x_k(y|\mathbf{x}) \neq 0$ for some y.

b. If $\partial H_j/\partial x_k(y|\mathbf{x}) \neq 0$ for some y, there there exists a point $\mathbf{y} \in \mathcal{Y}$ and a job attribute $i \in \{1, \cdots, Y_R\}$ such that $\sigma(\mathbf{x}, \mathbf{y}) > 0$ and (SC) holds for (x_k, y_i, y_j) at $(\mathbf{x}, \mathbf{y})$.

[7] In turn, in Lindenlaub and Postel-Vinay (2020) we define a matching to be positive [negative] assortative in dimension (y_j, x_k) if and only if $\partial H_j/\partial x_k(y|\mathbf{x})$ is negative [positive] *for all* $y \in [\underline{y}_j, \overline{y}_j]$ and all $\mathbf{x} \in \mathcal{X}$.

The proof is in Appendix 8.1.3. None of the standard integral inequalities apply to the integrals that define $\partial H_j/\partial x_k(y|\mathbf{x})$. This is why part a. again uses the Transversality Theorem to show that under the stated (SC) property of surplus, $\partial H_j(y|\mathbf{x})/\partial x_k$ cannot be zero for all y, and thus there is sorting in (x_k, y_j). This result significantly extends our earlier results (Lindenlaub and Postel-Vinay, 2020): when allowing for any form of sorting (positive or negative), (SC) is sufficient to induce it.[8] Moreover, point b. of Proposition 2 shows that (SC) holding at least somewhere is necessary for sorting.

Part (iii) of Proposition 1 reinforces our conclusions from parts (i) and (ii). It says that based on the dependency of the EE transition rate on the interaction term of worker and job characteristics (x_k, y_j), we can conclude that both of these characteristics are surplus-relevant.

Besides these direct implications of Proposition 1, there are also important indirect consequences. The reason is that single crossing of the surplus function precludes a single index representation of multi-dimensional heterogeneity. We can therefore learn from the dependency of the EE transition rate on worker attributes whether the *single-index assumption* is violated in the data: if τ depends on some x_k, then there is no single index representation. This is a far-reaching insight as the macro-labor literature almost invariably assumes scalar heterogeneity of workers and jobs.

We refer to the single index representation of the surplus function as a situation where there exist three differentiable functions $\widetilde{\sigma} : \mathbb{R}^2 \to \mathbb{R}$, $I : \mathbb{R}_+^X \to \mathbb{R}_+$ and $J : \mathbb{R}_+^Y \to \mathbb{R}_+$ such that for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$, $\sigma(\mathbf{x}, \mathbf{y}) = \widetilde{\sigma}\left(I(\mathbf{x}), J(\mathbf{y})\right)$. In other words, match surplus $\sigma(\mathbf{x}, \mathbf{y})$ can be expressed as a (bivariate) function of two one-dimensional *single indices*, I and J, summarizing all of the relevant heterogeneity on the worker and on the job side, without losing any information.

Proposition 3 (Single Crossing and Single Index Representation) *If there exists $j \in \{1, \cdots, Y_R\}$ and $k \in \{1, \cdots, X_R\}$ such that (SC) holds at some $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$, then there exists no single-index representation of the surplus function.*

See Appendix 8.1.4 for the proof. The argument shows that the single index representation requires the marginal rate of substitution (MRS) between any two job attributes in the surplus function to be independent of x_k, while the (SC) property requires the opposite. A constant MRS in worker characteristics, in turn, means that during job search all workers resolve the

[8] The stronger notion of sorting in Lindenlaub and Postel-Vinay (2020) necessitates stronger conditions to guarantee sorting. We circumvent these more involved sufficient conditions here because we do not focus on sufficient conditions for positive or negative sorting *everywhere*.

trade-offs between jobs of different attributes in *the same* way. This leads to agreement on the ranking of firms across workers and thus firm heterogeneity can be described on a one-dimensional scale. This is also a situation that precludes sorting as our next result shows, which is based on Propositions 2 and 3.

Corollary 1 (Sorting and Single Index Representation.) *Suppose $X_R \geq 1$ and $Y_R > 1$. If there is sorting between any (x_k, y_j) where $j \in \{1, \cdots, Y_R\}$ and $k \in \{1, \cdots, X_R\}$, then there exists no single index representation of the surplus function.*

Corollary 1 follows from our result that the (SC) property is necessary for sorting (Proposition 2). Thus if there is sorting, (SC) of σ holds at least locally, and by Proposition 3, this precludes the single-index representation. The intuition is clear: Sorting in this multi-dimensional setting means that different worker types rank jobs in different ways and climb *different* job ladders. Jobs that differ in multiple dimensions can thus not be uniformly ranked and collapsed onto a single job ladder, preventing the single index representation.

This section shows that theory is an essential guide in interpreting the output of our empirical protocol. Said protocol allows us to detect surplus- and sorting-relevant worker and job attributes as well as violations of the single-index assumption. In particular, based on Propositions 1-3, we can go back and forth between something we would like to know but is unobserved (properties of the surplus function) and something which we can observe in the data (dependencies of the EE transition rate on worker and job attributes). Figure 1 summarizes the various potential types of results that our method can produce, and their implications.

Figure 1: Potential Output and Implications

$$
\begin{array}{lcccl}
\dfrac{\partial \tau}{\partial y_j} \neq 0 & \longrightarrow & & & y_j \text{ is surplus-relevant} \\[2ex]
\dfrac{\partial \tau}{\partial x_k} \neq 0 & \longrightarrow & \text{(SC)} & \longrightarrow & \begin{cases} x_k \text{ is surplus-relevant} \\ x_k \text{ is sorting-relevant} \\ \text{no single index representation} \end{cases} \\[2ex]
\dfrac{\partial^2 \tau}{\partial x_k \partial y_j} \neq 0 & \longrightarrow & & & x_k \text{ and } y_j \text{ are surplus-relevant}
\end{array}
$$

2.3 Discussion

Reliance on the EE Transition Rate. One may ask why we choose to infer properties of the surplus function from the EE transition rate rather than other informative labor market outcomes, such as the unemployment-to-employment (UE) transition rate or wages. We argue here that extracting relevant information from the EE rate requires fewer assumptions compared to these alternatives. Workers' UE rate is:

$$\tau_0(\mathbf{x}) := \lambda_0 \int \mathbf{1}\left\{\sigma(\mathbf{x}, \mathbf{y}') \geq 0\right\} \gamma(\mathbf{y}') d\mathbf{y}'.$$

Like in the case of the EE rate, if the UE rate depends on x_k, we can conclude that x_k is surplus-relevant. Moreover, if the UE margin is active (there exist marginally profitable matches $(\mathbf{x}, \mathbf{y})$ such that $\sigma(\mathbf{x}, \mathbf{y}) = 0$) and the surplus depends on x_k, then the UE rate will depend on that worker characteristic.[9] The UE rate therefore seems to convey similar information about the surplus-relevance of skills as the EE rate. But there are important limitations to the UE rate. First, the UE rate is only informative about surplus-relevant skills if there exist marginally profitable matches. Instead, EE mobility stays informative even if $\sigma > 0$ for all potential matches. Second, UE mobility would not allow us to learn about whether (SC) holds, or what the implications of (SC) are. Third, the UE margin conveys no information about surplus-relevant *job* attributes since it only depends on worker characteristics.

Furthermore, we deliberately chose not to base our empirical protocol on wages for several reasons. First, if we replaced the EE rate by wages, we would need to make an assumption on how wages are determined in equilibrium (e.g. sequential auctions, wage posting, Nash bargaining, etc.) in order to properly interpret the dependence of wages on agents' attributes in the data. This would be an arbitrary choice but when using the information contained in the EE transition rate we can do without it. Second, in our underlying model, wages (as opposed to surplus) do *not* drive job mobility and thus convey no information about sorting. More formally, the dependency of wages on worker or job attributes is generally no indication of single-crossing

[9] To see this, we obtain the dependence of τ_0 on worker attributes x_k by differentiating:

$$\frac{\partial \tau_0}{\partial x_k}(\mathbf{x}) = \lambda_0 \int \frac{\partial \sigma}{\partial x_k}(\mathbf{x}, \mathbf{y}') \mathbf{1}\left\{\sigma(\mathbf{x}, \mathbf{y}') = 0\right\} \gamma(\mathbf{y}') d\mathbf{y}'$$

Thus, if $\frac{\partial \tau_0}{\partial x_k}(\mathbf{x}) \neq 0$ for a given worker type $\mathbf{x}$, this implies that the integrand is not zero for all $\mathbf{y}'$, and hence that there exists a point $\mathbf{y}_0 \in \mathcal{Y}$ on the 0-level set of σ, such that $\frac{\partial \sigma}{\partial x_k}(\mathbf{x}, \mathbf{y_0}) \neq 0$, indicating that x_k is surplus-relevant. Moreover, if $\frac{\partial \sigma}{\partial x_k}(\mathbf{x}, \mathbf{y}) \neq 0$ for all $\mathbf{y}$ (and in particular for those $\mathbf{y}_0$ such that $\sigma(\mathbf{x}, \mathbf{y}_0) = 0$), then one can show — following similar steps as in the proof of Proposition 1(ii)b. — that generically $\frac{\partial \tau_0}{\partial x_k}(\mathbf{x}) \neq 0$.

properties of the surplus function and thus does not allow to draw conclusions about sorting (see Appendix 8.1.6 for the details). Third, conditional on worker and job attributes $(\mathbf{x}, \mathbf{y})$, the EE transition rate does not depend on wages in our framework. Because wages do not convey any additional information in our analysis of worker mobility, we do not include it as explanatory variables of the EE rate (1).

ASSUMPTIONS. A natural question is whether our analysis can give sensible results by revealing surplus-relevant worker and job characteristics even if our model is misspecified in some dimensions. We now discuss several assumptions of our model and their importance for the empirical analysis.

First, we assumed that EE moves happen when worker-job surplus (which is a function of observable attributes) can be improved, and there is no reallocation of employed workers for other reasons. However, allowing for *reallocation shocks* — that is, for the arrival of job offers that cannot be refused — is feasible, as long as they are in line with the typical assumption in this literature that they are neither worker nor job type specific. The EE transition rate can be modified to account for these shocks, denoted by λ_2,

$$\tau(\mathbf{x}, \mathbf{y}) = \lambda_1 \int \mathbb{1}\left\{\sigma(\mathbf{x}, \mathbf{y}') \geq \sigma(\mathbf{x}, \mathbf{y})\right\} \gamma(\mathbf{y}') d\mathbf{y}' + \lambda_2.$$

This introduction of unobserved heterogeneity would leave Propositions 1-3 unchanged.

Second, in our model we assume that search is random as opposed to directed. Suppose the data was generated by a model of *directed search* where workers move in direction of higher surplus (as in our model). Also assume that the worker and job type spaces are such that there exist a pure job ladder with one-to-one mapping between current surplus and the surplus that workers are targeting in their search (e.g. as in Garibaldi, Moen, and Sommervoll (2016) when they focus on a continuum of firm types). Different search targets are associated with different levels of market tightness as high surplus jobs are harder to get. Then any two workers, who are of the same type and are currently in the same job, have the same current surplus and target the same type of job via on-the-job search. Thus, they have the same EE transition rates. But different workers who currently do not have the same surplus (either because their types differ or because their jobs differ) have different EE transition rates. As in our random search model, the dependence of the EE transition rate on worker and firm characteristics would reveal which attributes are surplus-relevant.

Third, to obtain analytical results that have a clean interpretation, we assumed that the surplus function satisfies certain assumptions, namely that it is Morse and quasi-concave. These are regularity assumptions and not overly restrictive. For instance, if wages were bargained via sequential auctions without worker bargaining power, in which case σ is a primitive, typically used production functions yield a surplus function that satisfies these assumptions (e.g. the multiplicative technology, the bilinear one, or a bliss-point technology). Importantly, we do not impose that σ be monotonic in any of the surplus-relevant job attributes. As a result, even if the data is characterized by *one*-dimensional worker and job heterogeneity, our method will detect this. In this case, on the firm side, the EE transition rate τ would depend on a single job attribute y_j; and on the worker side, τ would depend on a single characteristic x_k, indicating horizontal sorting: workers with different one-dimensional skills have different job acceptance sets and thus climb different job ladders — something an assumption of monotonicity of σ in y would preclude.

Fourth, we assumed that the job arrival rate of employed workers λ_1 is independent of worker or job type. This assumption is essential for our method to reveal surplus-relevant characteristics of workers and jobs. If it is violated, then λ_1 varies across workers and jobs, and the EE transition rate reads:

$$\tau(\mathbf{x}, \mathbf{y}) = \lambda_1(\mathbf{x}, \mathbf{y}) \int \mathbb{1} \left\{ \sigma(\mathbf{x}, \mathbf{y}') \geq \sigma(\mathbf{x}, \mathbf{y}) \right\} \gamma(\mathbf{y}') d\mathbf{y}'.$$

Thus, the dependence of τ on $(\mathbf{x}, \mathbf{y})$ may not provide any information about the surplus but could entirely be driven by heterogeneity in job arrival rates. Now, the main reason why the offer arrival rate might depend on job or worker attributes would be endogenous search effort: in that case, workers determine their search effort based on their expected returns from search, which in turn depends on the value of their current job, $\sigma(\mathbf{x}, \mathbf{y})$. That is, the job and worker attributes that determine the arrival rate in this case are precisely the ones that are surplus-relevant.

3 Simulations

We first test this new methodology in Monte Carlo simulations. We want to check whether our procedure can separate the relevant worker and job characteristics that explain EE transitions from surplus-irrelevant attributes that do not explain transitions. Simulations have the advantage over real data that, based on our theory above, we know which job and worker attribute are truly relevant and thus which coefficients should turn out significant in the EE regressions.

Moreover, we want to assess which of the many available model selection tools performs best in our context. The results will guide our choices when implementing the procedure on real data.

3.1 Data Generating Process

Many commonly used search models with on-the-job search fit the assumptions underlying our theoretical framework. To fix ideas, we base the simulations on the sequential auction model (a special case of Lindenlaub and Postel-Vinay, 2020; see also Postel-Vinay and Robin, 2002 for the one-dimensional case). In that model, the match surplus σ is a primitive, $\sigma(\mathbf{x}, \mathbf{y}) \propto p(\mathbf{x}, \mathbf{y}) - p_0(\mathbf{x})$, where $p(\mathbf{x}, \mathbf{y})$ is the production function and $p_0(\mathbf{x})$ non-employment income.[10] We assume σ is bilinear, $\sigma(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{Q}(\mathbf{y} - \mathbf{b})$, where $\mathbf{b}$ is a Y-dimensional vector that we interpret as the production technology of the unemployed and where $\mathbf{Q}$ is an $X \times Y$ matrix. The technology matrix $\mathbf{Q}$ captures the complementarity structure between all job and worker characteristics. Note that only those entries in $\mathbf{Q}$ are non-zero that are associated with the surplus-relevant characteristics of workers and jobs: denoting the generic entry of $\mathbf{Q}$ as q_{kj}, we have that $q_{kj} = 0$ for all $k > X_R$ or $j > Y_R$.

We first need to choose different multi-dimensional models (parameterization of technology, and distributions of workers ℓ and jobs γ) underlying the data generating process of our simulations. Ideally, we would like to simulate models that look similar to the data. However, there is little guidance by the literature on what the empirical multi-dimensional technology and distributions look like. This is in fact something we want to understand better with our test. We therefore iterated between implementing the test on simulated data and real data to come up with what we believe is a reasonable set of multi-dimensional models from which we generate data to test our methodology.[11]

This process led us to consider the following models: they are sparse, meaning that the 'true' worker and job attributes impacting surplus can be described by low-dimensional vectors (of length 2 or 3), even though we allow the number of potential worker and job attributes to

[10] Treatment of the match surplus as a primitive in this model is justified in Lindenlaub and Postel-Vinay (2020) and also in Appendix 8.1.5.

[11] In more detail, here is how we iterated: we started with a broad set of models, including those that feature (i) 1D heterogeneity, (ii) multi-D heterogeneity but no sorting, (iii) multi-D heterogeneity and sorting. We ran our test on data simulated from these models and picked the selection method that performed best (BIC, see Tables 5 and 6, Appendix 8.2.2). We then applied our test to the data, focussing on BIC, to see how many dimensions turn up significant/whether there is sorting. Here is what we found empirically: The model that our test suggests generated the real data is (i) sparse, (ii) multi-dimensional, (iii) features significant sorting both in within and between the relevant worker-job dimensions. We then went back to the simulations and focussed on models that have those features. The final test results are reported in Section 3.5, Tables 1 and 2.

be large ($X = Y = \{10, 20, 40\}$, where the number of r.h.s. variables in the EE regressions is given by $m = X + Y + XY$). Further, they feature sorting between the relevant worker and job attributes. Our simulations are thus based on models whose bilinear technology matrix $\mathbf{Q}$ triggers sorting (i.e. technology satisfy at least one single crossing condition). In particular, we allow for different $\mathbf{Q}$, featuring strong, intermediate and weak sorting among the relevant attributes. Finally, we vary the correlation of the potential worker (job) attributes, ranging from very low to high correlations, where we specify both the distribution of worker and of job attributes as Gaussian copulas. Our main exercise contains 24 different models ({2D vs 3D relevant attributes}×{low vs. high correlation}×{strong vs intermediate vs weak sorting}×{two different technologies}) from which we simulate data, each of them run multiple times with different random seeds. See Appendix 8.2.1, Examples A.1.1-A.2.2, for a detailed description of the models and the 'true' vector of coefficients in the corresponding EE regressions.

3.2 Model Selection Methods

We apply model selection tools to our empirical model of the EE transition rate with the objective of identifying the subset of relevant predictors and separate them from those worker and job attributes that do not predict the transition rate. There is a large number of model selection methods, ranging from the classical information criteria (Akaike Information Criterion, AIC, after Akaike (1973) and Akaike (1974), and Bayesian Information Criterion, BIC, after Schwarz (1978)) to more recent ones (e.g. Lasso by Tibshirani (1996), or the Robust Lasso by Belloni, Chen, Chernozhukov, and Hansen (2012)) that were developed in the age of big data. These methods share the goal of choosing the subset of regressors that optimally trades off the model's goodness-of-fit (often via the likelihood function) versus overfitting (via a penalty function that is increasing in the number of estimated parameters). The unified approach to model selection is to choose a parameter vector θ that maximizes the penalized likelihood:

$$n^{-1} l_n(\theta) - \sum_{j=1}^{m} r_\lambda(|\theta_j|)$$

where $l_n(\theta)$ is the log-likelihood function of the model, θ_j indicates a regression coefficient and thus one element of θ, m is the dimensionality of the covariates vector, n is the sample size, and r_λ is a penalty function indexed by the regularization parameter $\lambda \geq 0$, which differs across model selection methods.

In the case of AIC and BIC, the penalty function is given by the L_0-norm of the vector of coefficients θ, indicating the total number of nonzero elements in that vector, $\sum_{j=1}^{m} r_\lambda(|\theta_j|) = \lambda \sum_{j=1}^{m} \mathbb{1}\{\hat{\theta}_j \neq 0)\}$, where $\lambda = 1$ for AIC and $\lambda = \log(n)/2$ for BIC. In turn, the penalty function of the Lasso is the L_1-norm of the vector of coefficients, $\sum_{j=1}^{m} r_\lambda(|\theta_j|) = \frac{\lambda}{n} \sum_{j=1}^{m} |\theta_j|$, and thus penalizes through the sum of absolute parameter values. Contrary to the AIC and BIC the regularization parameter of the Lasso is not pre-specified and needs to be optimally chosen, for instance through cross-validation, or estimated. The Robust Lasso is based on the Lasso, but allows for data-driven penalty weights to deal with non-Gaussian and heteroscedastic disturbances.

A priori it is unclear which method performs best in our context. The AIC and BIC have solid micro-foundations in information theory. Moreover, the BIC is consistent in that it will select the true model with probability one as $n \to \infty$ (while the AIC tends to choose too many variables even in large samples).[12] But in finite samples, the consistency argument does not hold. Also, when using the proper *best subset algorithm* to find the best model, both methods are computationally expensive and quickly become infeasible.[13] In turn, the (Robust) Lasso is computationally much more efficient since it requires solving a convex optimization problem and it has attractive asymptotic oracle-like properties. Moreover, it can explicitly deal with non-Gaussian and heteroscedastic errors and can be used for very large m (even $m > n$). On the downside, just like with the information criteria, the Lasso's performance in finite samples is unproven.

We are most interested in finding the model selection method that performs best in selecting the true determinants of workers' EE mobility choices. As noted by Belloni, Chen, Chernozhukov, and Hansen (2012), perfect model selection is unlikely to occur in practice — independently of which method is chosen. We therefore use our simulations to compare the performance of the different methods. We focus our main analysis on the model selection methods that perform best: the AIC and the BIC as well the Robust Lasso (which in our case is the appropriate Lasso estimator due to non-Gaussian and heteroscedastic disturbances). But we also tested other methods, namely the standard Lasso, the Adaptive Lasso (Zou (2006)), the SCAD (Fan and Li (2001)) as well as simple OLS where we selected variables based on statistical significance. We make those results available on request.

[12] The consistency statement for the BIC is true if the true model is among those that are investigated.

[13] The solution to the optimization problem under the L_0-penalty function requires solving $\sum_{k \leq m} \binom{m}{k}$ least square problems, which is generically NP-hard.

3.3 Evaluating the Performance of Model Selection Methods

To be able to compare the performance of the different model selection methods, we need to summarize their accuracy in a compact way. We adopt three standard performance metrics for classification problems in machine learning, 'accuracy', 'recall', 'precision',[14] and construct a fourth one, a 'loss' function, which combines the model's 'recall' and 'precision'. These statistics summarize how well the model selection method is doing in terms of true positives (TP) and true negatives (TN), or similarly, how poorly it performs in terms of false positives (FP) and false negatives (FN). In more detail, we consider the following measures:

A. Accuracy measure: $$Accuracy = 1 - \frac{\text{FP} + \text{FN}}{\text{total \# rhs variables}}$$

The denominator refers to the total number of independent variables in the regression under consideration. The closer to one is this measure, the better the model selection works. An advantage of this measure is that it takes overall model performance into consideration, accounting for both *false positives* and *false negatives.* On the downside, this measure is not scale invariant, meaning it can increase by increasing the number of predictors in the regression.

B. Recall measure: $$Recall = 1 - \frac{\text{FN}}{\text{TP}}$$

This is a useful measure since it indicates how many of the true predictors the selection method misses. If *recall* is close to one, then among the selected variables there are those that truly matter. Another advantage is that this measure is scale invariant.

C. Precision measure: $$Precision = 1 - \frac{\text{FP}}{\text{FP+TP}}$$

This measure indicates how severe the problem of selecting *false positives* is. If *precision* is close to one, then this problem is negligible. This measure can vary with scale. Note that in many cases, a trade-off arises between *recall* and *precision.* For instance, if the model selection method sets all coefficients (including those of the true predictors) to zero, then *recall* is zero (poor performance in terms of false negatives) but *precision* is one (good performance in terms of false positives).

D. Loss function: $$Loss = (1 - Recall)^2 + (1 - Precision)^2$$

[14]See, for instance, Japkowicz and Shah (2011).

For compactness, we combine *recall* and *precision* into a *loss function*, which equals the (squared) Euclidean distance between actual model performance in the $(recall, precision)$ space and perfect performance in that same space, $(1, 1)$. This measure is low if the model selection method performs well on both counts, false negatives and false positives (i.e. if the distance between perfect and actual model performance is small). The loss function is our preferred measure of performance, which we complement with *accuracy*.

3.4 Implementation

Based on our theory, and in particular, based on Proposition 1, there are two ways of implementing our test, via a 1-Step or 2-Step procedure. To assess which one performs better, we will run both on our simulated data.

1-Step Procedure. Proposition 1(i)-(iii) suggests that we can approximate the EE transition rate of a worker *conditional* on his characteristics $\mathbf{x}$ and his job's attributes $\mathbf{y}$, $\tau(\mathbf{x}, \mathbf{y})$, using the following expansion

$$\tau_i(\mathbf{x}, \mathbf{y}) = \alpha + \sum_k^X \beta_k x_{ik} + \sum_j^Y \gamma_j y_{ij} + \sum_k^X \sum_j^Y \delta_{kj} x_{ik} y_{ij} + \epsilon_i \tag{2}$$

where subscript i indicates an employed individual, α is a constant and ϵ_i is a mean-zero error term. We can then use any of the discussed model selection methods on (2) to test *which* x_k's and y_j's are surplus/sorting-relevant. Here this selection is done in a *single step*, meaning the null model includes a constant while the full model includes *all* potential worker and job attributes and their interactions, so that the number of rhs variables is $m = X + Y + XY$.

2-Step Procedure. Based on Proposition 1(i)b., which states that the *unconditional* or *mean* EE rate, $\overline{\tau}(\mathbf{y})$, depends on a certain job attribute if and only if it is surplus-relevant, we can also split the model selection into two steps.

In a *first step*, we can select the job attributes $\mathbf{y}$ based on the unconditional EE rate $\overline{\tau}(\mathbf{y})$:

$$\overline{\tau}_i(\mathbf{y}) = \alpha + \sum_j^Y \gamma_{1j} y_{ij} + \epsilon_i. \tag{3}$$

Given the vector of selected job attributes, which we denote by $\mathbf{y}_R = (y_1, ...y_{Y_R})$, we can then in a *second step* use the model selection tools to select among worker attributes $\mathbf{x}$ and interactions $\mathbf{x}\mathbf{y}_R$, where the pre-selected $\mathbf{y}_R$ enter as fixed regressors that are no longer up for selection or penalized. Thus the relevant model in the second step resembles our model for the conditional EE rate from the 1-Step procedure but takes the vector of pre-selected job attributes as given:

$$\tau_i(\mathbf{x}, \mathbf{y}_R) = \alpha + \sum_k^X \beta_k x_{ik} + \sum_j^{Y_R} \gamma_{2j} y_{ij} + \sum_k^X \sum_j^{Y_R} \delta_{kj} x_{ik} y_{ij} + \epsilon_i \tag{4}$$

where we distinguish the coefficient on the y_j's in the first and second stage through the notation γ_{1j} and γ_{2j}. We use our simulations to inform us, which procedure — 1-Step or 2-Step — performs best. All the discussed model selection methods (AIC, BIC, Robust Lasso) can be used with either approach.

3.5 Results

The process of iterating between simulations and data induced us to focus in our main Monte Carlo exercise on simulating data from models with the following features: they are sparse, multi-dimensional, and feature significant sorting among the relevant worker and job attributes. We report the performance of the AIC, BIC and the Robust Lasso in selecting the correct surplus/sorting-relevant worker and job attributes in data generated by those models, both for the 1-Step and 2-Step procedure.[15] We split the models into two categories depending on the number of potential worker and job characteristics, those with 'small m' ($X = Y = 10$) and those with 'large m' (where we average across $X = Y \in \{20, 40\}$).

Small m. We start with the performance of models that feature a relatively small number of potential worker and job attributes. We report in Table 1 the results from the 1-Step procedure and in Table 2 the results from the 2-Step procedure. Each table has three panels, one for BIC, one for AIC and one for the Robust Lasso. In the four main columns we report the performance measures introduced above. Our preferred measures of performance is the loss function as it summarizes how the selection method performs along both dimensions, false positives and false negatives. The smaller is the loss, the better is the performance. We additionally check the Accuracy measure (the higher the accuracy the better) and we will see that Accuracy and Loss generally agree with each other. We compute those performance measures both for the full model and broken down by variable category (worker attributes X, job attributes Y and interactions XY), so each panel has four rows.

[15] To make the AIC and BIC computationally feasible, we implement them through the 'forward stepwise' procedure as opposed the 'best subset' selection. In the 2-Step procedure, the best-subset selection is computationally feasible in the first step but we found that it produces *identical* results to the forward-stepwise approach.

We start by comparing performance *within* each implementation method (1-Step and 2-Step) and across model selection methods. This reveals that the BIC performs well in identifying the relevant worker and job attributes (according to both Loss and Accuracy) and outperforms both the AIC and the Robust Lasso with lower Loss and higher Accuracy. For instance, in the 2-Step approach, the BIC does better in predicting all three types of variables: While it does slightly better than the AIC regarding the worker characteristics X and the interactions XY, it does considerably better in selecting the job characteristics Y, which in fact it predicts perfectly (loss of 0). In turn, the Robust Lasso does similarly well as the BIC in terms of predicting the correct job characteristics (loss of 0.03 versus the BIC's loss of 0) but shows a much weaker performance in selecting the correct worker characteristics and interactions. Going forward, we therefore focus on the BIC as our main model selection criterion.

Further comparing the performance of the BIC *across* implementation methods — 1-Step in Table 1 and 2-Step in Table 2 — reveals that they do equally well when considering the Full Model: they have the same high level of Accuracy (0.98) and essentially the same low level of Loss (0.04 and 0.05). Zooming into the different types of predictors — worker characteristics X, job characteristics Y and interactions XY — shows though that the 2-Step procedure does *significantly* better in terms predicting the true worker characteristics while performance in selecting the correct job characteristics and interaction terms is similar across the two approaches. Indeed, the Loss from selecting the correct worker attributes in the 2-Step procedure is less than half the Loss of the 1-Step approach (and Accuracy is higher in the 2-Step approach as well). We therefore favor the 2-Step over the 1-Step approach.

One interesting insight underlying these average statistics (i.e. averaged across simulated examples) is that model selection improves significantly when the data features stronger sorting. This is true especially when it comes to detecting the relevant *worker* attributes. As mentioned, we simulated each model for different choices of technology $\mathbf{Q}$, triggering 'strong', 'intermediate' and 'small' degrees of sorting among the surplus-relevant attributes. Figure 2 plots the Loss against the average sorting within dimensions of surplus-relevant attributes (x_k, y_j) (proxied by their correlation) in the simulated data. It shows that for all model selection methods, the Loss is *decreasing* in the strength of sorting. Our theory explains why this is the case: We show in Proposition 1 that $\partial\tau/\partial x_k \neq 0$ only if the single crossing property of technology holds. But by Proposition 2, (SC) implies sorting, meaning that $\partial\tau/\partial x_k \neq 0$ only if there is sorting. Stronger sorting makes it therefore easier for our method to detect surplus-relevant worker attributes.[16]

Large m. When considering a larger number of potential worker and job attributes, $m = X + Y + XY$ with $X = Y \in \{20, 40\}$, the performance of all our model selection methods in picking the correct worker and job attributes underlying the EE transitions worsens, see Table 8 in the Appendix 8.2.2. Similar to the case of small m, the BIC still outperforms the other methods. But compared to the case of small m, the Loss is larger, increasing from 0.05 to 0.11 (while Accuracy is similar). This is mainly due to problems of predicting the correct worker attributes and worker-job interactions. That model selection becomes more difficult in higher dimensions is well-known.

Table 1: Performance of the AIC, BIC and Robust Lasso (1 Step)

Method		Accuracy	Recall	Precision	Loss
BIC	**Full Model**	**0.98**	**0.83**	**1.00**	**0.04**
	X	0.89	0.52	1.00	0.36
	Y	0.99	0.96	1.00	0.01
	XY	0.99	0.91	1.00	0.02
AIC	**Full Model**	**0.96**	**0.95**	**0.75**	**0.08**
	X	0.95	0.83	0.97	0.09
	Y	0.98	0.97	0.96	0.02
	XY	0.96	0.98	0.64	0.15
Rlasso	**Full Model**	**0.94**	**0.52**	**0.90**	**0.30**
	X	0.81	0.27	1.00	0.58
	Y	0.87	0.56	1.00	0.34
	XY	0.96	0.60	0.84	0.28

Table 2: Performance of the AIC, BIC and Robust Lasso (2 Step)

Method		Accuracy	Recall	Precision	Loss
BIC	**Full Model**	**0.98**	**0.94**	**0.84**	**0.05**
	X	0.92	0.81	0.89	0.14
	Y	1.00	1.00	1.00	0.00
	XY	0.98	0.97	0.79	0.07
AIC	**Full Model**	**0.95**	**0.94**	**0.71**	**0.12**
	X	0.92	0.81	0.90	0.14
	Y	0.85	1.00	0.67	0.15
	XY	0.97	0.97	0.69	0.13
RLasso	**Full Model**	**0.96**	**0.61**	**1.00**	**0.19**
	X	0.75	0.00	1.00	1.00
	Y	0.97	0.92	0.99	0.03
	XY	0.98	0.75	1.00	0.16

[16] We further confirm that model selection performance deteriorates as sorting declines by running the tests on several models that feature zero sorting, see B.1.1.-B.2.2. in Appendix 8.2.1 and Table 7 in Appendix 8.2.2.

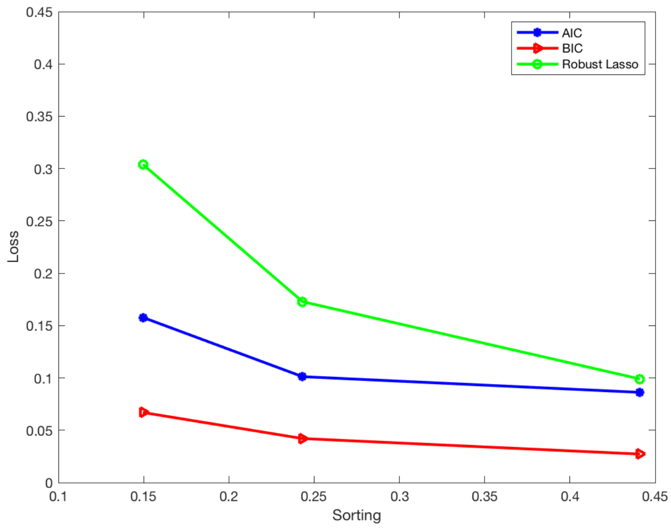

Figure 2: Test Performance as a Function of Sorting (2 Step)

We take away several insights from these Monte Carlo simulations that will guide the implementation of our test on real data: (1.) The test works well with relatively few potential job and worker attributes but performance weakens with large m. (2.) Within each approach (1-Step or 2-Step), the BIC outperforms both the AIC and the Robust Lasso, both when considering the entire model and also when focusing on each group of variables individually. (3.) Focussing on the BIC, the 2-Step procedure dominates the 1-Step procedure. We will therefore apply our procedure to data with a limited number of potential worker and job attributes and implement it using the BIC 2-Step method. Our simulation exercise reveals considerable heterogeneity in the performance of different model selection methods and implementation approaches. We stress the importance of our theory and results in enabling these simulations, allowing us to focus on the most effective model selection and implementation methods in our context.

4 The Data

We first describe the data and the construction of multi-dimensional job attributes and skills. We then discuss the implementation of our empirical protocol, and the results it produces.

4.1 Data Sources

We use two data sources: the Survey of Income and Program Participation (SIPP) and O*NET. We describe each of them in turn.

SIPP. The SIPP is our main dataset. It is a nationally representative survey with the main objective to provide information about the income and program participation of individuals and households in the United States. But the SIPP also collects extensive data on labor market outcomes like employment and, importantly, worker mobility from one job to another (EE transitions) as well as transitions from and into unemployment (UE and EU transitions).

The SIPP is administered in different panels of 4-5 year duration (with one new sample of about 40,000 individuals per panel) and conducted in waves and rotation groups. Within a SIPP panel, the entire sample is interviewed at 4-months intervals. These groups of interviews are called waves. The first time an interviewer contacts a household, for example, is called 'wave 1', the second time is 'wave 2', and so forth. Sample members within each panel are divided into four subsamples of roughly equal size, where each subsample is referred to as a rotation group. One rotation group is interviewed each month. During each interview, information is collected about the previous 4 months (the 'reference' months), so we have in fact monthly observations.

There are several advantages of the SIPP over other large labor force surveys in the US. First, the SIPP has a significant time and longitudinal dimension. So we have a relatively large sample with a large amount of labor market transitions, which is important for our method to work. Further, transitions are recorded by date (so there is no time aggregation bias, avoiding a common problem with transition data). Second, and crucial for our exercise, apart from its core dataset the SIPP features a special *Education and Training Module* with detailed information on individuals' degrees and occupational training. The Education and Training Module is available for the panels of 1996, 2001, 2004 and 2008. For each of those panels, the topical module files contain one record for each person who was a SIPP sample member during month four of wave 2. We use this special module to impute individuals' multi-dimensional skill bundles. To our knowledge, this is the first attempt at imputing multi-dimensional skills in the SIPP (which we also implemented in our companion paper Lindenlaub and Postel-Vinay (2020) but there we focussed on a considerably smaller sub-sample of only college educated individuals).

The only other US survey that provides comparably rich information on education, training (and, in addition, test scores) is the National Longitudinal Survey of Youth (NLSY). But the problem with the NLSY is that for our purposes the sample size and thus number of transitions is way too small to implement our empirical protocol in a reliable manner. Moreover, interviews are only conducted once a year, meaning at most one job-to-job transition per year is captured.

O*NET. We supplement the SIPP with the O*NET database. This database describes occupations in terms of skill and knowledge requirements, work practices, and work settings.[17] It comes as a list of around 300 descriptors, with ratings of almost 1,000 different occupations. Those descriptors are organized into ten broad categories: Abilities, Skills, Knowledge, Work Activities, Work Context, Education Requirements, Job Interests, Work Values, Work Styles, and Tasks. We retain descriptors from the O*NET files Abilities, Skills, Knowledge, Work Activities, and Work Context, as descriptors contained in the other files are less related to the skill requirements or job amenities we are interested in and are less comparable in terms of their scale. We use this dataset to learn about the multi-dimensional job attributes of the occupations that are present in the SIPP; and also to get clues into the skills of individuals who are qualified for certain occupations (as per their degree or occupational training, see below for the details).

4.2 Constructing the Variables of Interest

There are three types of variables whose construction requires additional explanations: multi-dimensional job attributes, multi-dimensional worker attributes as well as worker transition rates.

Multi-Dimensional Job Attributes. We use 'occupations' as the empirical counterpart of our model's 'jobs'. The data sets we combine (Abilities, Skills, Knowledge, Work Activities, and Work Context) contain 130 potential job attributes, corresponding to the vector $\mathbf{y}$ of potentially relevant job characteristics in our model. They range from detailed measures of cognitive, manual, interpersonal or creative skills to a wide array of job amenities such as whether there is time pressure or hazardous work conditions associated with an occupation.

Multi-Dimensional Worker Attributes. To impute the agents' bundles of potential multi-dimensional skills, we primarily rely on the special Education and Training Module in SIPP. It provides information on their college degrees, apprenticeships and vocational degrees, as well as occupational training on-the-job. We then make use of the fact that (a) college, apprenticeships and vocational programs as well as (b) occupational training qualify workers for particular occupations, where we use standard cross-walks to find out about those occupations. The final step is to combine this information with the O*NET data from above: we add the O*NET attributes of a certain occupation to a worker's skill bundle if a degree or training qualifies him for that particular occupation. For instance, if an individual holds a college degree

[17] O*NET (a.k.a. *Occupational Information Network*) is developed by the North Carolina Department of Commerce and sponsored by the US Department of Labor. Its initial purpose was to replace the old Dictionary of Occupational Titles. More information is available on `https://www.onetcenter.org`, or on the related Department of Labor site `https://www.doleta.gov/programs/onet/eta_default.cfm`.

in 'economics', we assume he is qualified for the occupation 'economist' and attach the skills required for this occupation to his skill bundle. If an individual is qualified for more than one occupation as per his training/degree we add the descriptors of multiple occupations to his skill bundle and take an average. This way we construct individuals' bundles of potentially relevant characteristics, containing 130 skills as well as attributes indicating the tolerance for occupational (dis)amenities, corresponding to vector $\mathbf{x}$ in our model. See Appendix 8.3.1 for details.

Labor Market Transitions. The key dependent variable in our empirical procedure is an indicator of EE transition, i.e. a binary variable equal to one if the individual made (at least) one EE transition in the past month. The SIPP provides start and end dates for each recorded job spell, as well as an employer number (which is individual- and panel-specific, it is not an employer identifier). We define an EE transition as a change of employers with less than a week's non-employment between the two consecutive job spells. We focus on spells of salaried employment and drop workers after their first observed spell of self-employment, if there is one.

4.3 Sample Selection and Summary Statistics

Our method of imputing the multi-dimensional skills of workers requires information on at least one of the following: degree (vocational, associate, advanced, or BA) and/or training experience in the occupation held during wave 2 (when the Education and Training module was assessed). We are thus forced to exclude those individuals from our sample for which this information is missing or who do not have such educational experiences.[18] Our sample, which we denote by 'Degree and Training Sample', is thus a subsample of the full sample featured in the Education and Training Module. Our baseline sample is based on the SIPP 2008 but below we also report the empirical results for the previous panels 1996, 2001 and 2004 for which the Education and Training Module is available. In the 2008 panel, we have 895,747 person-months observations (compared to 2,069,943 in the full sample), where we include individuals in the age range 20-60 and exclude self-employed workers.

Table 9 in Appendix 8.3 reports the summary statistics, comparing our baseline sample with the entire (representative) Education and Training Module. This shows that our sample is biased towards skilled workers (56% have a BA degree compared to 29% in the entire sample), features a lower unemployment rate (4% vs. 6%), a higher marriage rate (64% vs. 55%) and

[18]We could assign them a baseline skill based on their highest educational degree achieved but this would add considerable noise to our skill measures.

contains individuals that are on average two years older (average age of 43 vs. 41). However, the monthly labor market transitions rates (EE, UE and EU) do not differ across the two samples.

5 Empirical Implementation and Results

5.1 Implementation

To ensure that our procedure gives reliable results when implementing it on real data, we use the insights of the Monte Carlo simulations for guidance, both regarding the number of total r.h.s. variables in our empirical model of the EE transition rate and the choice of the model selection and implementation methods.

The procedure performs well when m is relatively small but performance declines as m grows. Furthermore, our initial data set contains 130 potential job attributes, many of which are highly correlated since many of the O*NET descriptors capture similar features of a job. The same is true for the 130 potential worker traits in our data set. Those high pairwise correlations between potential explanatory variables impair the performance of all model selection methods, implying that some transformation of the selectable variables is necessary.

For those reasons, we *pre-select* both worker and job attributes using principal component analysis (PCA) in order to bundle together sets of highly correlated variables before entering the model selection stage.[19] We run a standard PCA with a twist that makes the principal components interpretable, following the method from Lise and Postel-Vinay (2019): First, we run a PCA on the set of all potential job attributes $\mathbf{y}$ (constructed above) in the distribution of jobs in the SIPP and keep the first ten principal components, which explain more than 90% of the variance of the underlying data. We then impose ten exclusion restrictions (one per component) selected from the underlying 130 O*NET descriptors, each standing for one distinct and salient job attribute. An exclusion restriction is defined as exclusively loading on one of the ten components but not on any other. We transform the original ten components by recombining the loadings of each O*NET descriptor such that the resulting transformed factors satisfy these exclusion restrictions. For more details, see Lise and Postel-Vinay (2019).

This way we reduce the number of potential job attributes to ten (down from 130) while capturing most of the underlying variation and ensuring interpretability via the exclusion restrictions. Since in the construction of both, worker skill and job attributes, we use the infor-

[19]Without any pre-selection, our EE regression would have as many as $2 \times 130 + 130^2 \approx 17,000$ r.h.s. variables, with many of them highly correlated — a situation in which model selection methods do not perform well.

mation of the O*NET, this pre-selection method also reduces the number of potential worker characteristics to ten (down from 130).

We report our exclusion restrictions in Table 3. The first restriction *Originality* reflects non-routine cognitive attributes, the second one *Performing Administrative Activities* reflects routine cognitive attributes, the third and fourth, *Far Vision* and *Trunk Strength*, reflect different manual characteristics, namely dexterity and strength. Restrictions six and seven, *Selling or Influencing Others* and *Assisting and Caring for Others*, stand for persuasion and empathy — two distinct interpersonal skills. Finally, the last four restrictions *Frequency of Conflict Situations*, *Responsibility for Outcomes and Results*, *Exposed to Hazardous Conditions*, and *Duration of Typical Work Week* reflect job (dis)amenities.

We selected these particular exclusion restrictions for two reasons.[20] First, restrictions 1-6 are chosen to closely resemble the broad categories that the applied labor literature has identified as important (e.g. Acemoglu and Autor (2011)):[21] (non)routine cognitive, (non)routine manual and several aspects of interpersonal skills. In turn, we chose restrictions 7-10 to additionally capture a rich set of job amenities and workers' tolerance for them, something which is usually overlooked when analyzing the worker-job surplus and sorting. Second, we selected these particular exclusion restrictions to minimize the pairwise correlations among the components, since we want each component to capture distinct information about occupations. Tables 11 and 12 in Appendix 8.3 report the resulting correlation matrices of potential worker and job attributes in their corresponding populations. Table 10 contains further summary statistics, comparing the skill and job attribute distributions in our baseline sample to the full Education and Training sample.[22] Note that we add an indicator of workplace size to the potential job attributes (y_{11}), so that $X = 10,\ Y = 11$.

In sum, our approach achieves pre-selection of worker and job attributes in a parsimonious way. It preserves interpretability as well as the covariance structure of worker and job attributes while using the entire set of underlying descriptors to construct the components.

[20] We did robustness checks varying the exclusion restrictions. Our results are not overly sensitive to this choice.

[21] Acemoglu and Autor (2011) also use the O*NET to classify skill/task attributes into those six categories: two cognitive, two manual and two interpersonal ones, `https://economics.mit.edu/faculty/dautor/data/acemoglu`. The difference is that we do not hand-pick and average a small number of O*NET descriptors for each of these categories while disregarding all the remaining O*NET descriptors. Instead, we run a PCA which uses the information from *all* descriptors but impose exclusion restrictions only for interpretability.

[22] To compute the skills in the full sample, we give each individual, who lacks both an educational degree and occupational training, a baseline skill bundle that reflects the number of years of schooling.

Principal Component	Exclusion Restriction
1	Originality
2	Performing Administrative Activities
3	Far Vision
4	Trunk Strength
5	Selling or Influencing Others
6	Assisting and Caring for Others
7	Frequency of Conflict Situations
8	Responsibility for Outcomes and Results
9	Exposed to Hazardous Conditions
10	Duration of Typical Work Week

Table 3: Exclusion Restrictions in the PCA

Regarding the technical implementation of our protocol, we choose the method that performed best in our simulations: we run the BIC 2-Step procedure (implemented via the forward-stepwise approach) to select the relevant predictors of worker mobility. The *first step* selects important job attributes among the potential job attributes $\mathbf{y}$ based on the unconditional EE rate $\overline{\tau}(\mathbf{y})$ and the *second step* selects important worker attributes and interactions among the potential skills and interactions, $\mathbf{x}$ and $\mathbf{xy}_R$, for given pre-selected $\mathbf{y}_R = (y_1, ...y_{Y_R})$:

$$\overline{\tau}_i(\mathbf{y}) = \alpha + \sum_j^{11} \gamma_{1j} y_{ij} + \varepsilon_i \tag{5}$$

$$\tau_i(\mathbf{x}, \mathbf{y}_R) = \alpha + \sum_k^{10} \beta_k x_{ik} + \sum_j^{Y_R} \gamma_{2j} y_{ij} + \sum_k^{10} \sum_j^{Y_R} \delta_{kj} x_{ik} y_{ij} + \epsilon_i \tag{6}$$

To further stabilize our results and reduce concerns about unobserved heterogeneity, we control in both stages for the monthly mean EE rate to capture time trends and we also include some fixed worker controls in the second stage (gender, age, race, marital status).

5.2 Results

SIPP 2008. Our baseline sample is the SIPP 2008. Since this panel coincides with the peak of and recovery of the US economy from the Great Recession and may be special, we also compare our results from the SIPP 2008 to those from previous SIPP panels below.

We pool observations between 2009-2013 to increase power.[23] In Table 4, we report the

[23] We start in 2009 since the Education and Training module assesses individuals' education in 2009 (wave 2).

results from running the BIC 2-Step procedure on this sample. A checkmark indicates that a certain variable was selected in one of the two stages. Checkmarks in the last row (column) refer to selected non-interaction skill (job attribute) terms, while checkmarks in the table's body refer to the selected interaction terms between worker and job attributes.

We will use our theory to interpret these results. It tells us that based on the selected variables (how many and which ones), we can learn about properties of the surplus function (see Propositions 1-3). We now discuss these implications in detail.

First, the results indicate how many and which worker and job attributes are surplus and sorting relevant. We find that a relatively sparse model gets selected. On the job side, skill requirements y_2 (routine cognitive) and y_5 (interpersonal – persuasion) get selected as significant predictors of EE mobility and thus enter the surplus (see Proposition 1(i)b. which refers to step 1 of our estimation approach); and so do (dis)amenity y_{10} (work duration) and workplace size y_{11}. To illustrate in which occupations these attributes play an important role, consider the following examples which are based on O*NET information: Cognitive routine skill requirements (y_2) are in high demand for human resources specialists, receptionists and information clerks, education administrators or accountants. In turn, interpersonal requirements (y_5) are important for advertising sales agents and sales representatives, sales managers, real estate agents and marketing managers. Finally, a high number of hours per week (y_{10}) is required in the occupations anesthesiologists, chief executives, hospitalists, or biochemists and biophysicists.

On the worker side, we find that most of the worker characteristics that get selected are symmetric to the selected job requirements (cognitive routine skill x_2, interpersonal skill x_5 and x_{10}, which we here interpret as a worker's tolerance for long work hours). In addition, the two distinct manual skills, x_3 and x_4, predict worker EE mobility. Based on Proposition 1(ii)a. and Proposition 2, these attributes are not only surplus but also *sorting* relevant. Our methodology — by highlighting several worker attributes as significant predictors of the EE transition rates — thus rejects the assumption of 'no sorting' which is common to almost all job-ladder models in this literature.

Second, based on Proposition 1(ii)a., the EE transition rate depends on some worker trait x_k only if the surplus satisfies at least one single-crossing property involving x_k and some job attribute. But then, by Propositions 3, we can *reject* that there exists a *single index* representation of the selected multi-dimensional worker and job attributes. Thus, both heterogeneity and sorting are truly multi-dimensional in the data.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Linear Terms
y_1	.	.	.	.	.	.	.	.	.	.	.
y_2	.	✓	✓	.	✓	.	.	.	.	✓	✓
y_3	.	.	.	.	.	.	.	.	.	.	.
y_4	.	.	.	.	.	.	.	.	.	.	.
y_5	.	✓	.	✓	✓	.	.	.	.	.	✓
y_6	.	.	.	.	.	.	.	.	.	.	.
y_7	.	.	.	.	.	.	.	.	.	.	.
y_8	.	.	.	.	.	.	.	.	.	.	.
y_9	.	.	.	.	.	.	.	.	.	.	.
y_{10}	.	.	✓	.	✓	.	.	.	.	✓	✓
y_{11}	.	.	.	✓	✓	.	.	.	.	.	✓
Linear Terms	.	✓	✓	✓	✓	.	.	.	.	✓	.

✓ indicates significant worker/job attribute or interaction. Variable y_{11} indicates firm size.
Sample: Degree and Training (subsample of Education Module), Pooled Across 2009-2013. Baseline controls (non-selected): male, age, married, race (all 2. stage); monthly mean EE rate (1. and 2. stage).

Table 4: Model Selection Results Based on BIC 2 Step, Years 2009-2013

Third, given that sorting in the data is based on multi-dimensional attributes on both sides of the market that cannot be collapsed into a single index, it must be the case that workers with different bundles rank jobs in different ways. Otherwise, sorting would not arise. Consequently, there is *no single economy-wide job ladder* that all workers agree on. Instead, workers with different skill bundles rank jobs in different ways and climb different job ladders.

Last, based on our theoretical result of how the conditional EE transition rate depends on job attributes in Proposition 1(i)a., we can learn from post-estimation of EE regression (6) whether the relevant job attributes have a positive or negative impact on surplus. Indeed, while cognitive skill requirements and the workplace size impact surplus positively, both high interpersonal skill requirements as well as long work duration decrease surplus during 2009-2013 in the US.

SIPP 1996, 2001 AND 2004. We chose to focus on the SIPP 2008 as our baseline sample because the earlier panels (especially 2001 and 2004) have fewer observations that are followed over a shorter time horizon. Nevertheless, we report the results produced by applying our procedure to earlier SIPP panels in Tables 13-15 in Appendix 8.4. The results are remarkably robust over time, especially as far as the model sparsity and the surplus-relevant job attributes are concerned. In all four panels, interpersonal job attribute y_5, work duration y_{10} and workplace size y_{11} are surplus-relevant as per our test. Moreover, cognitive requirement y_2 becomes surplus relevant over time, starting in the 2004 panel.

There is also a pattern in the selected surplus/sorting-relevant attributes on the worker side: cognitive routine skill x_2, manual dexterity x_3 and tolerance for long work hours x_{10} are consistently selected over time. Also, interpersonal skill x_5 turn up significant in the two panels with the largest datasets (1996 and 2008), whose results we are most confident in.

We reiterate that these results mostly *illustrate* how our new methodology can be applied, as our sample is biased towards high-skilled workers because either an educational degree or some occupational training experience is necessary to impute individuals' multi-dimensional skills in the SIPP. We believe that this is the reason why, consistently across panels, our protocol indicates that manual skill requirements do not enter the surplus function. Also note that non-routine cognitive skills or skill demands (x_1 or y_1) are never indicated to be surplus/sorting relevant. The likely reason is the relatively high correlation between those attributes with the 'duration of work week'-variables (x_{10} and y_{10}), which thus capture a considerable share of non-routine cognitive skills and skill requirements.

6 Application

We now use our empirical results to investigate multi-dimensional sorting along the relevant dimensions in the data, both in the cross-section and over time.

We first discuss how to assess the sign of sorting among those worker-job characteristics that are sorting-relevant as per our test. We then perform several checks to show that our sorting measures relate to labor market transitions in intuitive ways. The objective of our main exercise in this application is then to shed light on multi-dimensional sorting in the data: its nature, how it changed over time, and how it varies across different dimensions, i.e. along the different job ladders. We end with a discussion on the driving forces behind changes in sorting over time.

6.1 Measuring Multi-Dimensional Sorting in the Data

Our procedure informs about the specific worker and job characteristics along which agents sort but it cannot identify the *sign* or *direction* of sorting — positive or negative — in the various dimensions. However, with our results at hand, we can measure the sign of sorting of those characteristics directly in the data.

In theory (Lindenlaub and Postel-Vinay, 2020), we define positive sorting by a FOSD ordering of the marginal job distribution $H_j(y|\mathbf{x})$ w.r.t. x_k: if $H_j(y|\mathbf{x})$ is decreasing in x_k then there is PAM in dimension (x_k, y_j), since more skilled workers in x_k are matched to a better distribution of job attributes y_j. NAM is defined in an analogous way.

In the data, we measure sorting between (x_k, y_j) by the monotonicity of the conditional mean of a job attribute y_j, $\mathbb{E}(y_j|\mathbf{x})$, w.r.t. x_k. So, there is PAM in (x_k, y_j) if workers with higher x_k are on average matched with higher y_j jobs. The monotonicity of the conditional means of job attributes in worker characteristics follows from our theoretical definition of FOSD dominance and is easier to check empirically. To assess this monotonicity in the data, we run regressions of the surplus-relevant job attributes on the surplus-relevant skills. This allows us to assess the sign (and also the strength) of sorting in those dimensions that our procedure identifies as relevant.

We run the following sorting regressions in our baseline sample (SIPP 2008),

$$y_{2,i} = \alpha + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \beta_{10} x_{10,i} + \kappa^T z_i + \epsilon_i \tag{7}$$

$$y_{5,i} = \alpha + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \beta_{10} x_{10,i} + \kappa^T z_i + \epsilon_i \tag{8}$$

$$y_{10,i} = \alpha + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \beta_{10} x_{10,i} + \kappa^T z_i + \epsilon_i \tag{9}$$

where subscript i indicates an employed individual and where we include additional individual-level controls $z_i \in \{\text{gender, age, race, marital status}\}$ with associated coefficient vector κ, and where ϵ_i are mean-zero error terms. A positive (negative) $\beta_k, k \in \{2, 3, 4, 5, 10\}$ means that the conditional expectation of y_j positively (negatively) co-moves with characteristic x_k, thereby indicating positive (negative) sorting between worker trait x_k and job attribute y_j. Furthermore, the larger (in absolute values) the β coefficients are, the stronger is sorting.

In what follows, we focus on sorting in the 'within'-dimensions, that is within the symmetric dimensions on both sides of the market: sorting of routine-cognitive skills and skill demands (x_2, y_2) in (7), of interpersonal skills and skill demands (x_5, y_5) in (8) and sorting between the job requirement of long work hours and the workers' tolerance for this (x_{10}, y_{10}) in (9). 'Within'-sorting is the most natural to occur and dominates by a large margin the 'between'-sorting in the data, justifying our focus on the coefficients $(\beta_2, \beta_5, \beta_{10})$ in regressions (7)-(9), respectively.

6.2 Multi-Dimensional Sorting in the Pooled Panel

Overall Sorting. To assess overall sorting in these three within-dimensions, we first run regressions (7)-(9) in our entire sample of employed workers where we pool years 2009-2013. We here plot the within-sorting coefficients ($\beta_2, \beta_5, \beta_{10}$) along with their 95%-confidence intervals (*blue* coefficients in Figure 3, where each subfigure focuses on one dimension): Sorting in all three dimensions is positive and significant, ranging from a sorting coefficient of 0.55 in the work-duration dimension to 0.8 in the interpersonal dimension.

Sorting and Transitions. To analyze the relationship between sorting and labor market transitions, we now zoom into subsamples of workers, focussing on those who are about to make an EE move (*yellow* coefficients in Figure 3) or a UE move (*green* coefficients) or those who are about to separate from their job into unemployment through an EU transition (*red* coefficients).

In both productivity-related dimensions (cognitive routine and interpersonal), workers who are about to make an EE or EU move are more mismatched than the average worker, indicated by lower red and yellow sorting coefficients compared to the blue ones, see subfigures 1 and 2. Moreover, mismatch in jobs of newly employed workers after unemployment (after a UE transition) is larger compared to mismatch in the average job, indicated by lower green than blue coefficients. This evidence complies with our intuition about the relation between sorting and transitions: higher mismatch makes a transition (either to another job or to unemployment) more likely and mismatch is higher after unemployment since unemployed workers are less selective and start at low rungs of the job ladder they aim to climb.

In the amenity dimension (work duration), these patterns are not as clear. In particular, relatively well-matched workers separate into unemployment (the red coefficient higher than the blue one in subfigure 3, but the difference is not statistically significant). But our prior on how sorting on *amenities* is related to transitions is also less clear due to the lack of existing work on this issue. The detailed regression results underlying Figure 3 are in Tables 1-3 in the Online Appendix.

6.3 Multi-Dimensional Sorting Over Time

Time Trends. To assess how sorting has changed over time, we run regressions (7)-(9) for each month in our SIPP 2008 sample and then plot the monthly within-sorting coefficients ($\beta_2, \beta_5, \beta_{10}$) along with their 95%-confidence intervals in Figure 4. The first three subfigures again correspond to the dimensions 'routine-cognitive', 'interpersonal' and 'work duration'.

In all three dimensions, sorting has declined throughout our sample period. It was strongest during 2009 — at the peak of the Great Recession — and weakest in 2013. What stands out is that this decline did not affect all dimensions in the same way. While there is only a weak decline

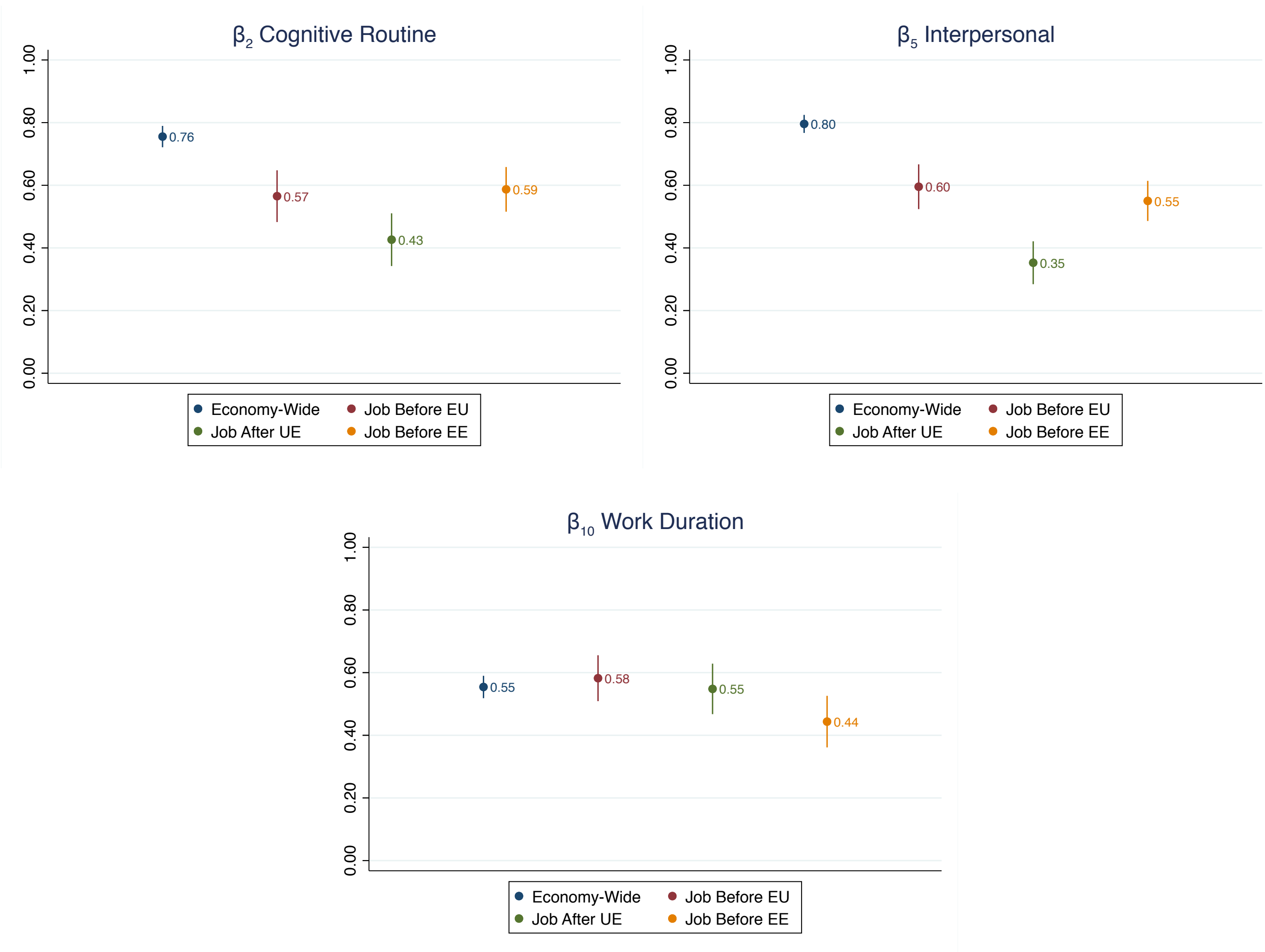

Figure 3: Average Sorting and Sorting Before and After a Labor Market Transition

— mainly concentrated in 2009/10 — in both the cognitive and work-duration dimensions (subfigures 1 and 3), there is a much larger drop in sorting in the interpersonal dimension (subfigure 2). While sorting in the interpersonal dimension was stronger compared to the other two dimensions in 2009, it then also faced the strongest decline, decreasing by more than 30% in the post-recession period. These differential sorting changes are summarized in subfigure 4, which plots the percentage decline in sorting between any two years t and $t-1$ across dimensions.

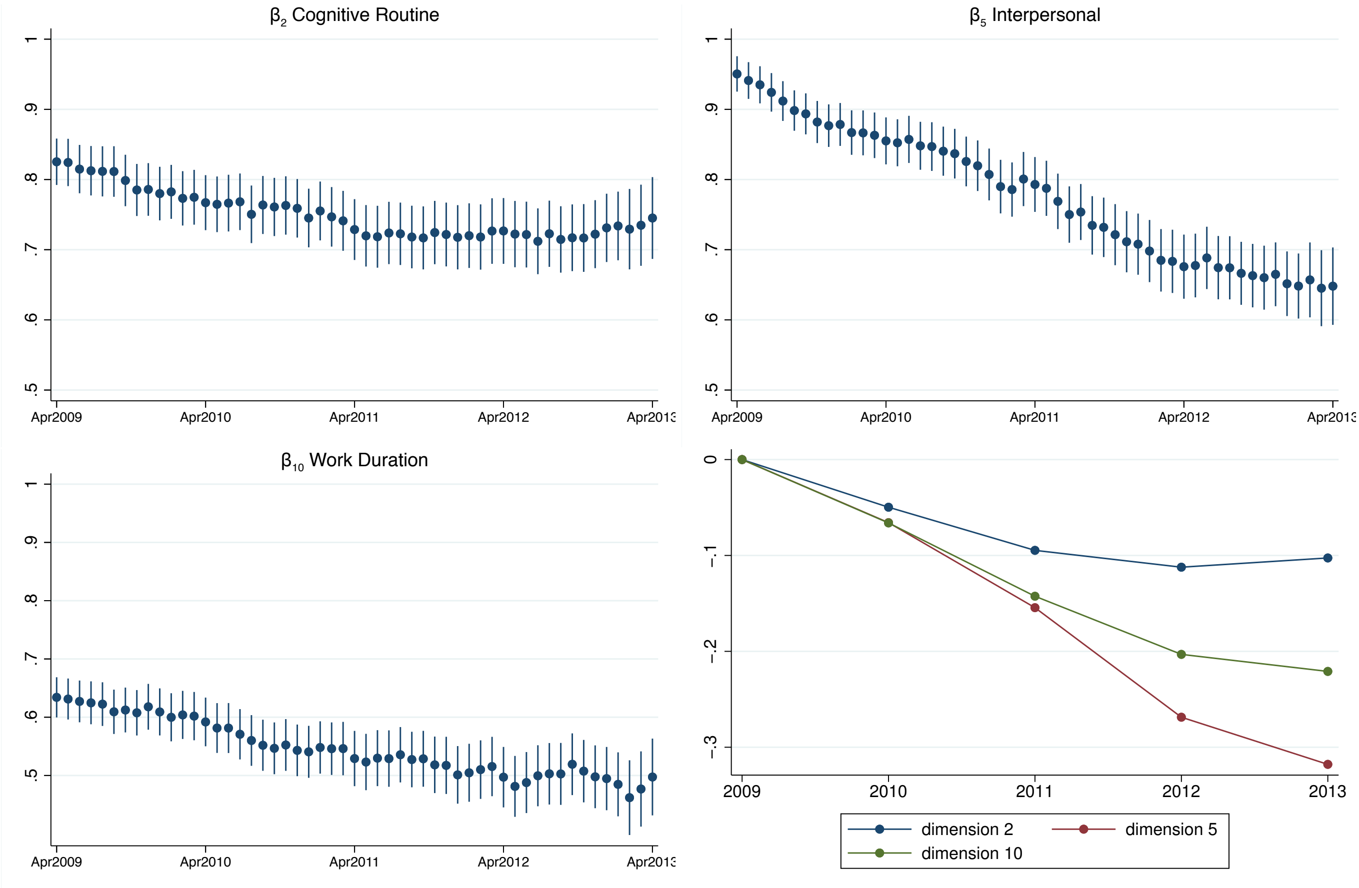

Figure 4: Sorting Over Time

Decomposition into EE, EU and UE Sorting. Changes in sorting must be driven by one of three margins: EE, UE or EU reallocation. We are interested in how much of this decline in multi-dimensional sorting is due to changes of how agents sort on the UE/EU margins and how much is due to a changing reallocation on the EE margin. Figure 5 reports the results from this decomposition. We re-run regressions (7)-(9) for each year between 2009 and 2013 and on three types of sub-samples:[24] the entire sample (blue coefficients — as before); a sample of

[24]Now we pool data by year instead of considering monthly observations since the monthly sample would become too small when breaking it down into the considered subsamples.

those that are continuously employed so we exclude those that make UE or EU transitions (red coefficients); and a sample of workers with none of them reallocating — our 'no reallocation' benchmark (green coefficients). We again focus on $(\beta_2, \beta_5, \beta_{10})$ here and the three subfigures capture the results for the cognitive, interpersonal and work duration dimension, as above. The detailed regression results underlying Figure 5 are in Tables 4-12 in the Online Appendix.

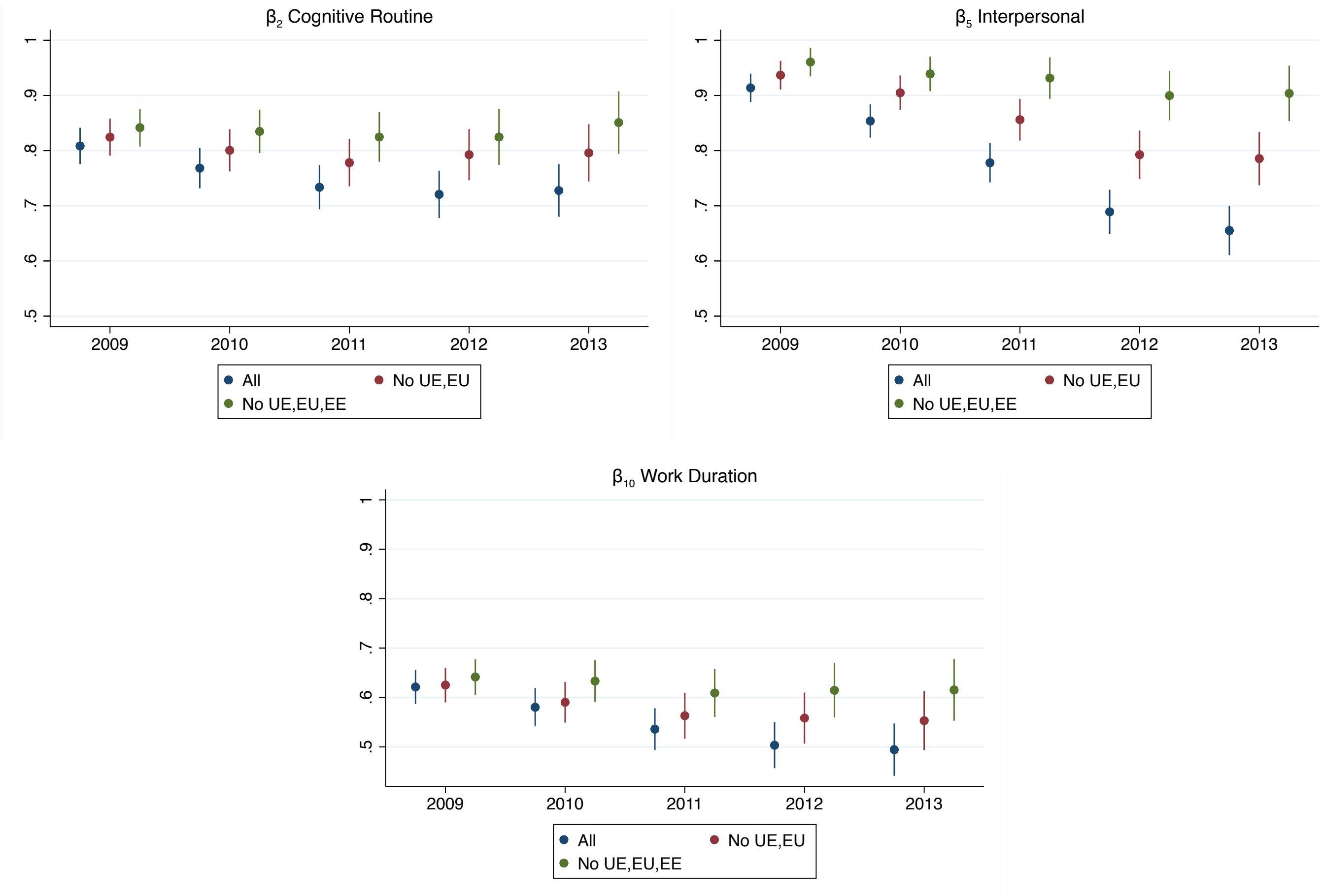

Figure 5: Decomposition of Sorting Changes into EE, UE and EU Margin

We find that along the cognitive dimension (subfigure 1), almost all of the decline in overall sorting (blue coefficients) is driven by the UE and EU margins: taking those individuals out of the sample who make a UE or EU transition nearly eliminates the drop in sorting (going from blue to red coefficients in subfigure 1, where the red coefficients almost lie on a horizontal line indicating constant sorting over time in the subsample that did not make any UE/EU transition). This is to some extent also the case in the work duration dimension (subfigure 3). But it is not true for the interpersonal dimension (subfigure 2), where both the UE/EU margin as well as the EE margin contribute about half to the decline in sorting. By definition, when focussing on the 'no reallocation' sample (green coefficients), the sorting coefficients should lie on a horizontal line, meaning sorting does not change over time in the absence of any reallocation. To better understand the changes in these various sorting margins, we zoom in further.

We first focus on the EE margin and particularly on whether EE transitions increase sorting in the three dimensions under consideration. To this end, we regress the log change in each surplus-relevant job attribute $j \in \{2, 5, 10\}$ (cognitive, interpersonal or work duration) on the surplus-relevant worker characteristics and baseline controls,

$$\log \frac{y_{j,it}}{y_{j,it-1}} = \alpha + \gamma_2 x_{2,i} + \gamma_3 x_{3,i} + \gamma_4 x_{4,i} + \gamma_5 x_{5,i} + \gamma_{10} x_{10,i} + \kappa^T z_i + \epsilon_{it},$$

where we focus on a sample of EE movers (i.e. workers i who made an EE transition between t and $t-1$). We are interested in whether $(\gamma_2, \gamma_5, \gamma_{10})$ are positive when $j \in \{2, 5, 10\}$ respectively. Positive coefficients would indicate that EE transitions lead to relative improvements in sorting in the cognitive, interpersonal and work duration dimension.

For illustration, we again report the coefficients of interest of all three regressions graphically, see Figure 6 in Appendix 8.5 (and we report the detailed regression results underlying Figure 6 in Tables 13-15 in the Online Appendix). Clearly, in the interpersonal and work duration dimension, the job ladders collapsed during and after the Great Recession, with EE transitions *not* leading to improvements in sorting. This is indicated by negative interpersonal coefficients γ_5 (subfigure 2) and insignificant coefficients in the work duration dimension, γ_{10} (subfigure 3). These results show that, in the interpersonal dimension, sorting even deteriorates through EE transitions, in line with the second subfigure of Figure 5 above. In turn, while in the cognitive dimension sorting did also not improve through EE transitions at the peak of the Great Recession and right afterwards, this changed in 2012/13. In those years after recovery, employed workers with higher cognitive skills sought jobs that are more cognitive intensive than their last one compared to workers with lower cognitive skills, reviving the 'cognitive job ladder'. These findings confirm our insights from the decomposition analysis in Figure 5. They indicated that changes in sorting on the EE margin contributed to the decline in overall sorting on the interpersonal and (to some extent) on the work duration dimension but much less so on the cognitive dimension.

Second, we take a closer look at UE and EU sorting in the data. Figure 7 in Appendix 8.5 shows the sorting coefficients $(\beta_2, \beta_5, \beta_{10})$ from regressions (7)-(9), again for the full sample (blue coefficients in each panel), for the sub-sample of workers who are about to make an EU move (red) and for the sub-sample of those who just made a UE transition (green).[25] During 2009-2013, especially in the interpersonal dimension (subfigure 2), worker-job pairs that formed after unemployment were characterized by significantly more mismatch compared to the average worker-job pair (compare green to blue coefficients). Moreover, those workers who separated from a job into unemployment were considerably better sorted on the interpersonal dimension than those who replaced them by moving from U to E (the green sorting coefficients of UE movers tend to be lower than red ones of EU movers). This again confirms our findings from the decomposition above that changes in EU/UE sorting over time have fueled mismatch during and right after the recession, especially in the interpersonal dimension.

6.4 Discussion

What caused the decline in multi-dimensional sorting and the contraction of the corresponding job ladders during the recovery of the Great Recession? We end with a discussion of potential driving forces behind the documented decline in sorting, which particularly affected the interpersonal dimension. To precisely pin down the underlying factor(s), we would need to fully estimate our model. However, analyzing the data in reduced-form already allows us to infer which explanations are more likely than others. As possible explanations we consider changes in: search frictions, worker/ job distributions, and technology. We discuss each of them in turn.

Search Frictions. Changes in search frictions could contribute to changes in sorting patterns. For instance, an increase in search frictions that slows down the process of job arrivals could induce workers to be less selective in their job choice, causing a decline in sorting. However, a look at the data reveals that changes in search frictions are unlikely the driver behind the decline in sorting. First, as indicated by Figure 3 in the Online Appendix, which plots the monthly EU, UE and EE transition rates over time, there is no apparent time trend in any of the three series. If search frictions had become more severe we would expect to see a decline in job finding rates (translating into a decline in UE and EE rates). Second, even if there was a time trend indicating that frictions had become more severe over time, it is unclear why this should have affected the interpersonal dimension most prominently.

Worker and Job Distributions. Another reason behind changes in sorting could be changes in the underlying distributions. We therefore plot in Figures 4 and 5 in the Online Appendix the marginal distributions of job and worker attributes for each year of our dataset 2009-

[25]See the detailed regression results underlying Figure 6 in Tables 16-21 in the Online Appendix.

2013, where we again focus on our three main dimensions (routine cognitive, interpersonal and work duration): Both on the worker and the job side, the CDFs of different years almost exactly lie on top of each other, indicating there were no significant changes. This finding casts doubt about the hypothesis that distributional changes are behind the documented decline in sorting.

Technology. Within our theoretical framework, a third factor underlying the observed sorting changes could have been a shift in technology $\mathbf{Q}$ (possibly related to the Great Recession) that decreased worker-job complementarities, making sorting less beneficial. In particular, given that the sorting decline is strongly biased towards the interpersonal dimension, this suggests that interpersonal complementarities took the biggest hit. Two pieces of (indirect) evidence support this hypothesis.

First, sorting into jobs with high interpersonal demands yields *negative* returns during this time period. And while the return to being in interpersonal-intensive jobs is higher for well-sorted workers with high interpersonal skills, even those workers barely earn positive returns from being in such occupations. This is revealed by a regression of log earnings on the surplus-relevant skills and job attributes, and their interactions (see Online Appendix, Section 1.3 for details). Note that the opposite is true for jobs that have high cognitive requirements: earnings increase with the cognitive requirement of the job, especially for workers with high cognitive skills. Similar findings hold for occupations that require long work hours where sorting also gets rewarded.

Second, sorting into jobs with high interpersonal demands does *not* shield from separation, even if the worker has high interpersonal skills and is thus well-matched with the job. In Section 1.2 of the Online Appendix, we report regressions that relate an individual's separation indicator to the surplus-relevant worker and job attributes as well as their interactions. The results reveal that sorting into jobs demanding high levels of interpersonal skills makes separations *more* likely. The opposite is true for cognitively demanding jobs which shield from separation.

It is worth noting that also changes in idiosyncratic firm or match-specific shocks could have contributed to the decline in sorting. We do not model such shocks explicitly. But they are unlikely the sole driver since we have documented that a significant share in the decline in sorting is due to increased mismatch after UE transitions, especially in the interpersonal dimension. Moreover, the shock process would need to be quite specific in the sense of asymmetrically affecting jobs with high interpersonal skill demands.

Our take away is that between 2009 and 2013, there were salient changes in multi-dimensional sorting and job ladders, showing a decline in sorting or increase in mismatch along the surplus-relevant worker and job characteristics. All dimensions that are relevant on both worker and job side (cognitive, interpersonal and work duration) were affected. But there are large quantitative differences in sorting changes across them, highlighting the importance of considering multi-dimensional heterogeneity. By far the most severe drop in sorting happened in the interpersonal dimension. We find that sorting changes along both the EE margin (collapse of the 'interpersonal' job ladder) as well as along the EU and UE margins contributed to this overall increase in interpersonal mismatch. Our exploratory data analysis suggests that neither changes in search frictions nor in the marginal distributions of worker and job attributes were the main drivers behind this shift. Instead we show that returns from sorting into jobs with high interpersonal requirements are negative during 2009-2013 and, moreover, those jobs see more frequent job separations. These results may be suggestive of asymmetric demand shifts. In our framework, they would stem from a decline of worker-job complementarities in technology, affecting the interpersonal dimension most severely. As a result, individuals considered it less beneficial to sort along this dimension.

7 Conclusion

The worker-job surplus is ubiquitous in any model of the labor market featuring search frictions. It is crucial for workers' employment transitions, for their wages as well as for measuring sorting and the extent of mismatch. Yet, our understanding of the determinants of the surplus is at best limited. In this paper, we try to fill this gap. Our main contribution is to develop a theory-based empirical procedure to determine how many and which of the *observable* worker and job characteristics enter the *worker-job surplus* in the data, where we exploit the mobility choices of employed workers. Our method also indicates whether these relevant attributes matter for sorting on the labor market. As an important by-product, our method also reveals whether the single-index assumption is valid, by which all surplus-relevant worker characteristics can be summarized by a scalar index without loss of information; and similarly for the job attributes.

We first test our new methodology in Monte Carlo simulations and then implement it on real data using the Survey of Income and Program Participation and the O*NET. We find that a relatively sparse model underlies the data. On the job side, a routine cognitive and an interpersonal skill requirement impact the surplus along with the (dis)amenity of work duration as well as the workplace size. On the worker side, we find that most of the relevant characteristics are symmetric to the selected job requirements. Furthermore, we reject that there exists a single index representation of these surplus-relevant multi-dimensional worker and job attributes.

We use our results in an application that aims to shed new light on multi-dimensional sorting along the relevant dimensions (as per our empirical protocol) in the data. We find salient changes in multi-dimensional sorting over time, showing a decline in positive sorting between worker and job attributes — or increase in mismatch — between 2009 and 2013, in the aftermath of the Great Recession. All three major dimensions (cognitive, interpersonal and work duration) were affected but there are large quantitative differences across them. The interpersonal dimension experienced by far the most severe drop in sorting. We find that both sorting changes along the EE margin (collapse of the 'interpersonal job ladder') as well as along the employment-to-unemployment (EU) and unemployment-to-employment (UE) margins contributed to this overall increase in interpersonal mismatch. Our results are surprising in light of recent work that highlights the growing importance of interpersonal skills in the labor market. And they underscore the importance of accounting for multi-dimensional heterogeneity since otherwise the presence and differential collapse of distinct job ladders would go unnoticed.

8 Appendix

8.1 Theory

8.1.1 An Ancillary Lemma

We begin this appendix by stating the following lemma, which will be of use in other proofs.

Lemma 1 *Fix $\mathbf{x} \in \mathcal{X}$ and let $\varphi(\mathbf{y})$ be an arbitrary smooth function (possibly depending on other variables, such as $\mathbf{x}$ or the scalar y in our application). Assume that $\sigma(\mathbf{x}, \cdot)$ is a Morse function and that the sampling density $\gamma(\cdot)$ is differentiable. Then the function:*

$$s \mapsto \int \varphi(\mathbf{y}) \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) = s\} \gamma(\mathbf{y}) \, d\mathbf{y}$$

is differentiable at all $s \in ran\, \sigma(\mathbf{x}, \cdot)$.

Proof. Consider $(\mathbf{y}, s) \in \mathcal{Y} \times \operatorname{ran} \sigma(\mathbf{x}, \cdot)$ such that $\sigma(\mathbf{x}, \mathbf{y}) = s$. Because $\sigma(\mathbf{x}, \cdot)$ is a Morse function, we can assume that $\mathbf{y}$ is not a critical point, so that there exists $i(\mathbf{y}) \in \{1, \cdots, Y_R\}$ such that $\partial\sigma/\partial y_{i(\mathbf{y})}(\mathbf{x}, \mathbf{y}) \neq 0$.

For any $j \in \{1, \cdots, Y_R\}$, let $\mathbf{y}_{-j} := (y_1, \cdots, y_{j-1}, y_{j+1}, \cdots, y_{Y_R})$. By the Implicit Function Theorem, there exists an open ball $\mathbb{B}(\mathbf{y}, s)$ of radius $r(\mathbf{y}, s) > 0$, centered at $\left(\mathbf{y}_{-i(\mathbf{y})}, s\right)$, and a continuously differentiable function $\Lambda^{(\mathbf{y},s)} : \mathbb{B}(\mathbf{y}, s) \to \mathbb{R}$ such that $\Lambda^{(\mathbf{y},s)} \left(\mathbf{y}_{-i(\mathbf{y})}, s\right) = y_{i(\mathbf{y})}$ and $\sigma\left(\mathbf{x}, \left[\mathbf{y}'_{-i(\mathbf{y})}, \Lambda^{(\mathbf{y},s)}\left(\mathbf{y}'_{-i(\mathbf{y})}, s'\right)\right]\right) = s'$ for all $\left(\mathbf{y}'_{-i(\mathbf{y})}, s'\right) \in \mathbb{B}(\mathbf{y}, s)$.

Now fix s and $\varepsilon > 0$ sufficiently small to ensure that $\left(\mathbf{y}_{-i(\mathbf{y})}, s'\right) \in \mathbb{B}(\mathbf{y}, s)$ for all $s' \in (s - \varepsilon, s + \varepsilon)$. We can then apply the Implicit Function Theorem at the point

$$\tilde{\mathbf{y}}(s') = \left[\mathbf{y}_{-i(\mathbf{y})}, \Lambda^{(\mathbf{y},s)}\left(\mathbf{y}_{-i(\mathbf{y})}, s'\right)\right]$$

and define $\mathbb{B}(\mathbf{y}, s')$ as the open ball of radius $r\left(\tilde{\mathbf{y}}(s'), s'\right)$, centered at $\left(\mathbf{y}_{-i(\mathbf{y})}, s'\right)$.

Denote the projection of $\mathbb{B}(\mathbf{y}, s')$ on the set of $\mathbf{y}_{-i(\mathbf{y})}$ (a subset of $\mathbb{R}^{Y_R - 1}$) by $\overline{\mathbb{B}}_{\mathbf{y}}(\mathbf{y}, s')$. Consider a pair $(s_1, s_2) \in (s - \varepsilon, s + \varepsilon)^2$. Let $\mathbf{y}'_{-i(\mathbf{y})} \in \overline{\mathbb{B}}_{\mathbf{y}}(\mathbf{y}, s_1) \cup \overline{\mathbb{B}}_{\mathbf{y}}(\mathbf{y}, s_2)$ (that intersection is nonempty as it contains $\mathbf{y}_{-i(\mathbf{y})}$), and $s' \in (s - \varepsilon, s + \varepsilon)$. By construction, $\Lambda^{(\tilde{\mathbf{y}}(s_1), s_1)}\left(\mathbf{y}'_{-i(\mathbf{y})}, s'\right) = \Lambda^{(\tilde{\mathbf{y}}(s_2), s_2)}\left(\mathbf{y}'_{-i(\mathbf{y})}, s'\right) = \Lambda^{(\mathbf{y},s)}\left(\mathbf{y}'_{-i(\mathbf{y})}, s'\right)$, as all three coincide with the unique solution to the equation $\sigma\left(\mathbf{x}, \left[\mathbf{y}_{-i(\mathbf{y})}, y'\right]\right) = s'$. We will denote those coinciding functions as $\Lambda^{\mathbf{y}}(\cdot)$. Note that the function $\tilde{\mathbf{y}}(s')$ now writes as $\tilde{\mathbf{y}}(s') = \left[\mathbf{y}_{-i(\mathbf{y})}, \Lambda^{\mathbf{y}}\left(\mathbf{y}_{-i(\mathbf{y})}, s'\right)\right]$.

Next, because $(\mathbf{y}', s') \mapsto \sigma(\mathbf{x}, \mathbf{y}') - s'$ is smooth, it is Lipschitz over the compact set $\mathcal{Y} \times \operatorname{ran} \sigma(\mathbf{x}, \cdot)$. Theorem 3.6 in Phien (2011) implies that $r(\mathbf{y}', s')$ is bounded below by

$$\frac{\left|\partial\sigma/\partial y_{i(\mathbf{y}')}(\mathbf{x}, \mathbf{y}')\right|}{(K+1)\sqrt{(Y_R - 1)\left[\left(\partial\sigma/\partial y_{i(\mathbf{y}')}(\mathbf{x}, \mathbf{y}')\right)^2 + K^2\right] + 1}}$$

where K is the Lipschitz constant. Applying this lower bound at the point $(\tilde{\mathbf{y}}(s'), s')$, we obtain:

$$r\left(\tilde{\mathbf{y}}(s'), s'\right) \geq \frac{\left|\partial\sigma/\partial y_{i(\mathbf{y})}(\mathbf{x}, \tilde{\mathbf{y}}(s')\right|}{(K+1)\sqrt{(Y_R - 1)\left[\left(\partial\sigma/\partial y_{i(\mathbf{y})}(\mathbf{x}, \tilde{\mathbf{y}}(s'))\right)^2 + K^2\right] + 1}}.$$

By continuity of $\tilde{\mathbf{y}}(\cdot)$ and smoothness of $\sigma(\mathbf{x}, \cdot)$, the function $s' \mapsto \partial\sigma/\partial y_{i(\mathbf{y})}(\mathbf{x}, \tilde{\mathbf{y}}(s')$ is strictly bounded away from zero when $s' \in (s-\varepsilon, s+\varepsilon)$, for ε small enough, so that for all $s' \in (s-\varepsilon, s+\varepsilon)$:

$$r\left(\tilde{\mathbf{y}}(s'), s'\right) \geq \inf_{s' \in (s-\varepsilon, s+\varepsilon)} \frac{\left|\partial\sigma/\partial y_{i(\mathbf{y})}(\mathbf{x}, \tilde{\mathbf{y}}(s')\right|}{(K+1)\sqrt{(Y_R - 1)\left[\left(\partial\sigma/\partial y_{i(\mathbf{y})}(\mathbf{x}, \tilde{\mathbf{y}}(s'))\right)^2 + K^2\right] + 1}} = \underline{r}(\mathbf{y}) > 0.$$

Crucially, this lower bound is independent of s. Let $\mathcal{B}(\mathbf{y})$ denote the open ball of radius $\underline{r}(\mathbf{y})$ centered at $\mathbf{y}_{-i(\mathbf{y})}$. The above results guarantee that $\mathcal{B}(\mathbf{y}) \subset \bigcap_{s' \in (s-\varepsilon, s+\varepsilon)} \overline{\mathbb{B}}_{\mathbf{y}}(\mathbf{y}, s')$. Moreover, by construction, for all $\mathbf{y}' \in \mathcal{B}(\mathbf{y})$ and all $(s', s'') \in (s-\varepsilon, s+\varepsilon)^2$, $\Lambda^{(\mathbf{y}, s')}\left(\mathbf{y}'_{-i(\mathbf{y})}, s''\right) = \Lambda^{\mathbf{y}}\left(\mathbf{y}'_{-i(\mathbf{y})}, s''\right)$.

Now, let $\mathcal{Y}_j$ denote the projection of $\mathcal{Y}$ on the jth coordinate, and let $\mathring{\mathcal{Y}}_j$ denote its interior. Clearly, $\bigcup_{\mathbf{y} \in \mathcal{Y}} \mathcal{B}(\mathbf{y}) \times \mathring{\mathcal{Y}}_{i(\mathbf{y})}$ is an open cover of $\mathcal{Y}$. Because $\mathcal{Y}$ is a compact set, we can extract a finite cover: $\bigcup_{n=1}^N \mathcal{B}(\mathbf{y}_n) \times \mathring{\mathcal{Y}}_{i(\mathbf{y}_n)}$. Let $\{f_n\}_{n=1}^N$ be a partition of unity subordinate to that cover, i.e. a collection of smooth functions $f_n : \mathcal{Y} \to \mathbb{R}$ such that for each n, f_n is supported in $\mathcal{B}(\mathbf{y}_n) \times \mathring{\mathcal{Y}}_{i(\mathbf{y}_n)}$, $0 \leq f_n \leq 1$, and $\sum_{n=1}^N f_n \equiv 1$. Then, for all $s' \in (s-\varepsilon, s+\varepsilon)$:

$$\int \varphi(\mathbf{y}) \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) = s'\} \gamma(\mathbf{y})\, d\mathbf{y} = \sum_{n=1}^N \int f_n(\mathbf{y}) \varphi(\mathbf{y}) \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) = s'\} \gamma(\mathbf{y})\, d\mathbf{y}$$

Now, by construction, for each n:

$$\int f_n(\mathbf{y}) \varphi(\mathbf{y}) \cdot \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) = s'\} \gamma(\mathbf{y})\, d\mathbf{y} = \int_{\mathcal{B}(\mathbf{y}_n) \times \mathring{\mathcal{Y}}_{i(\mathbf{y}_n)}} f_n(\mathbf{y}) \varphi(\mathbf{y}) \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}) = s'\} \gamma(\mathbf{y})\, d\mathbf{y}$$
$$= \int_{\mathcal{B}(\mathbf{y}_n)} f_n\left(\left[\mathbf{y}_{-i(\mathbf{y}_n)}, \Lambda^{\mathbf{y}_n}\left(\mathbf{y}_{-i(\mathbf{y}_n)}, s'\right)\right]\right) \varphi\left(\left[\mathbf{y}_{-i(\mathbf{y}_n)}, \Lambda^{\mathbf{y}_n}\left(\mathbf{y}_{-i(\mathbf{y}_n)}, s'\right)\right]\right) \gamma\left(\left[\mathbf{y}_{-i(\mathbf{y}_n)}, \Lambda^{\mathbf{y}_n}\left(\mathbf{y}_{-i(\mathbf{y}_n)}, s'\right)\right]\right) d\mathbf{y}_{-i(\mathbf{y}_n)}$$

We have thus expressed $\int \varphi(\mathbf{y})\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}) = s'\}\gamma(\mathbf{y})\, d\mathbf{y}$ as a finite sum of differentiable functions of s', proving the lemma. $\square$

8.1.2 Proof of Proposition 1

Part (i)a. To make the notation more compact, it is convenient to introduce the conditional sampling distribution $F_{\sigma|\mathbf{x}}$ of flow surplus σ, given $\mathbf{x}$ (with density $f_{\sigma|\mathbf{x}}$), defined by:

$$F_{\sigma|\mathbf{x}}(s) = \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') \leq s\right\}\gamma(\mathbf{y}')d\mathbf{y}' = \mathbb{E}_\Gamma\left[\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') \leq s\right\}\right].$$

With this notation, the job acceptance probability for an employed worker $\mathbf{x}$ is $1 - F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))$ and $\tau(\mathbf{x},\mathbf{y}) = \lambda_1\left(1 - F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right)$. The derivative of $\tau(\mathbf{x},\mathbf{y})$ w.r.t. y_k is then given by:

$$\frac{\partial \tau}{\partial y_j}(\mathbf{x},\mathbf{y}) = -\lambda_1 f_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y})$$

which is non-zero iff $\frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y}) \neq 0$, i.e. iff y_j is surplus-relevant.

Part (i)b. We now analyze the dependence of the EE rate on y_j, *unconditional* on $\mathbf{x}$. Let:

$$\overline{\tau}(\mathbf{y}) := \mathbb{E}\left[\tau(\mathbf{x},\mathbf{y}) \mid \mathbf{y}\right] = \frac{\int \tau(\mathbf{x},\mathbf{y})h(\mathbf{x},\mathbf{y})d\mathbf{x}}{\int h(\mathbf{x},\mathbf{y})d\mathbf{x}}$$

where $h(\mathbf{x},\mathbf{y})$ is the equilibrium density of $(\mathbf{x},\mathbf{y})$-matches. Thus:

$$\frac{\partial \overline{\tau}}{\partial y_j}(\mathbf{y}) = \mathbb{E}\left[\frac{\partial \tau}{\partial y_j}(\mathbf{x},\mathbf{y}) \mid \mathbf{y}\right] + \frac{\int \tau(\mathbf{x},\mathbf{y})\frac{\partial h}{\partial y_j}(\mathbf{x},\mathbf{y})d\mathbf{x}}{\int h(\mathbf{x},\mathbf{y})d\mathbf{x}} - \overline{\tau}(\mathbf{y}) \cdot \frac{\int \frac{\partial h}{\partial y_j}(\mathbf{x},\mathbf{y})d\mathbf{x}}{\int h(\mathbf{x},\mathbf{y})d\mathbf{x}}$$

The first term in the RHS is the average partial effect of y_j on the EE rate, while the last two terms reflect selection: due to sorting, a marginally different y_j will be matched to a different set of $\mathbf{x}$'s, impacting the average EE rate. Now note that:[26]

[26] Density h is determined by the following flow-balance equation, which embeds the optimal mobility decisions, and equates the outflow (lhs) and inflow (rhs) into matches $(\mathbf{x},\mathbf{y})$:

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

In Lindenlaub and Postel-Vinay (2020), Appendix C.1, we show how to solve this ODE for $h(\mathbf{x},\mathbf{y})$, giving us (10) above.

$$h(\mathbf{x},\mathbf{y}) = \frac{\delta\lambda_0 \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) \geq 0\right\}}{\delta + \lambda_0 \overline{F}_{\sigma|\mathbf{x}}(0)} \cdot \frac{\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2} \cdot \ell(\mathbf{x})\gamma(\mathbf{y}) \tag{10}$$

implying:

$$\begin{aligned}
\frac{\partial h}{\partial y_j}(\mathbf{x},\mathbf{y}) = {} & \frac{\delta\lambda_0\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]\ell(\mathbf{x})}{\delta + \lambda_0 \overline{F}_{\sigma|\mathbf{x}}(0)} \times \left(\frac{\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) \geq 0\right\} \cdot 2\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^3} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y})\gamma(\mathbf{y}) \right. \\
& \left. + \frac{\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) = 0\right\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y})\gamma(\mathbf{y}) + \frac{\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) \geq 0\right\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2} \frac{\partial \gamma}{\partial y_j}(\mathbf{y}) \right) \\
= {} & \frac{2\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}{\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y}) h(\mathbf{x},\mathbf{y}) \\
& + \frac{\delta\lambda_0 \ell(\mathbf{x}) \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) = 0\right\}}{\left[\delta + \lambda_0 \overline{F}_{\sigma|\mathbf{x}}(0)\right]\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)\right]} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y})\gamma(\mathbf{y}) + \frac{h(\mathbf{x},\mathbf{y})}{\gamma(\mathbf{y})} \frac{\partial \gamma}{\partial y_j}(\mathbf{y})
\end{aligned}$$

Noticing that:

$$\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}{\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y}) = -\frac{\partial \tau(\mathbf{x},\mathbf{y})/\partial y_j}{\delta + \tau(\mathbf{x},\mathbf{y})}$$

and substituting and collating terms:[27]

$$\begin{aligned}
\frac{\partial \overline{\tau}}{\partial y_j}(\mathbf{y}) = {} & \mathbb{E}\left[\frac{\partial \tau}{\partial y_j}(\mathbf{x},\mathbf{y}) \mid \mathbf{y}\right] - 2\mathbb{E}\left[\frac{\partial \tau}{\partial y_j}(\mathbf{x},\mathbf{y}) \frac{\tau(\mathbf{x},\mathbf{y}) - \overline{\tau}(\mathbf{y})}{\delta + \tau(\mathbf{x},\mathbf{y})} \mid \mathbf{y}\right] \\
& + \frac{\gamma(\mathbf{y})}{\displaystyle\int h(\mathbf{x},\mathbf{y}) d\mathbf{x}} \int \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) = 0\right\} \frac{\tau(\mathbf{x},\mathbf{y}) - \overline{\tau}(\mathbf{y})}{\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)} \frac{\delta\lambda_0 \ell(\mathbf{x})}{\delta + \lambda_0 \overline{F}_{\sigma|\mathbf{x}}(0)} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{y}) d\mathbf{x}.
\end{aligned}$$

If y_j is not output-relevant, i.e. if $\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}) = 0$ for all $\mathbf{x}$, then $\partial\overline{\tau}/\partial y_j = 0$. By contraposition, this shows that $\partial\overline{\tau}/\partial y_j \neq 0$ is sufficient for y_j being output relevant (showing the *necessity* part of the proposition). We now turn to the converse implication (*sufficiency*) and assume that y_j is output relevant.

[27] This can be written in many different ways. Notice for example that the integrand in the last term is multiplied by $\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) = 0\right\}$. Thus, for all $\mathbf{x}$ in the integration domain, $\tau(\mathbf{x},\mathbf{y}) = \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0) \geq \overline{\tau}(\mathbf{y})$. This also implies that the middle fraction in that integrand can be written as $\frac{\tau(\mathbf{x},\mathbf{y}) - \overline{\tau}(\mathbf{y})}{\delta + \tau(\mathbf{x},\mathbf{y})}$.

Spelling out $\partial\overline{\tau}/\partial y_j$ (no UE margin case):

$$\frac{\partial\overline{\tau}}{\partial y_j}(\mathbf{y}) = \frac{\displaystyle\int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}\cdot\ell(\mathbf{x})d\mathbf{x}}{\displaystyle\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}d\mathbf{x}}$$
$$- 2\frac{\displaystyle\int \frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}d\mathbf{x}}{\displaystyle\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}d\mathbf{x}} \times \frac{\displaystyle\int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^3}\cdot\ell(\mathbf{x})d\mathbf{x}}{\displaystyle\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}d\mathbf{x}}$$

implying:

$$\frac{\partial\overline{\tau}}{\partial y_j}(\mathbf{y}) \times \left(\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}d\mathbf{x}\right)^2$$
$$= \int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}\cdot\ell(\mathbf{x})d\mathbf{x}\cdot\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}d\mathbf{x}$$
$$- 2\int \frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}d\mathbf{x}\cdot\int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^3}\cdot\ell(\mathbf{x})d\mathbf{x}$$

Our objective is to show that this expression is generically not zero and we use the Transversality Theorem (Mas-Colell, Whinston, and Green (1995), Proposition 17.D.3) to do so. To this end, we first note that, by the smoothness properties of σ and γ, and by the Ancillary Lemma 1, $\partial\overline{\tau}/\partial y_j$ is itself continuously differentiable. We then apply a perturbation argument. We will perturb ℓ. Let $m(\mathbf{x})$ be an arbitrary integrable function such that $\int m(\mathbf{x})d\mathbf{x} = 1$, and let $\widetilde{\ell}(\mathbf{x};t) = (1-t)\cdot\ell(\mathbf{x}) + t\cdot m(\mathbf{x})$, for t in a neighborhood of 0. Then use the RHS of the last expressions where we let:

$$\psi(\mathbf{y};t) = \int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}\cdot\widetilde{\ell}(\mathbf{x};t)d\mathbf{x}\cdot\int \frac{\widetilde{\ell}(\mathbf{x};t)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2}d\mathbf{x}$$
$$- 2\int \frac{\widetilde{\ell}(\mathbf{x};t)}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}d\mathbf{x}\cdot\int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^3}\cdot\widetilde{\ell}(\mathbf{x};t)d\mathbf{x}$$

Finally, let $\mathbf{y}_0$ be a vector of job attributes s.t. $\partial\overline{\tau}(\mathbf{y}_0)/\partial y_j = 0$ (which by construction is equivalent to $\psi(\mathbf{y}_0;0)=0$). Then, the derivative w.r.t. to the 'perturbation parameter' is given by:

$$\begin{aligned}
\frac{\partial\psi}{\partial t}(\mathbf{y}_0;0) =& \int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2} m(\mathbf{x})d\mathbf{x}\cdot\int\frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x} \\
&-2\int\frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}\cdot\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^3}m(\mathbf{x})d\mathbf{x} \\
&+\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}\cdot\int\frac{m(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x} \\
&\quad-2\int\frac{m(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}\cdot\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^3}\ell(\mathbf{x})d\mathbf{x}
\end{aligned}$$

Further substituting the identity implied by $\partial\overline{\tau}(\mathbf{y}_0)/\partial y_j=0$ into the last expression yields:

$$\begin{aligned}
\frac{\partial\psi}{\partial t}(\mathbf{y}_0;0) =& \int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2} m(\mathbf{x})d\mathbf{x}\cdot\int\frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x} \\
&-2\int\frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}\cdot\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^3}m(\mathbf{x})d\mathbf{x} \\
&+\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}\cdot\int\frac{m(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x} \\
-\int\frac{m(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}\cdot&\frac{\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}_0)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}\cdot\int\frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x}}{\int\frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}}
\end{aligned}$$

Next, we set the function $m(\cdot)$ to be a Dirac mass at some point $\mathbf{x}_0$:

$$\begin{aligned}
\frac{\partial\psi}{\partial t}(\mathbf{y}_0;0) =& \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)\right]^2}\frac{\partial\sigma}{\partial y_j}(\mathbf{x}_0,\mathbf{y}_0)\cdot\int\frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x} \\
&-2\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)\right]^3}\frac{\partial\sigma}{\partial y_j}(\mathbf{x}_0,\mathbf{y}_0)\cdot\int\frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x} \\
&+\frac{1}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)\right]^2}\cdot\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\cdot\ell(\mathbf{x})d\mathbf{x} \\
-\frac{1}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}\cdot&\frac{\int\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}\cdot\int\frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x}}{\int\frac{\ell(\mathbf{x})}{\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)}d\mathbf{x}}
\end{aligned}$$

and thus:

$$\frac{\partial \psi}{\partial t}(\mathbf{y}_0;0) = \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)\right]^2}\frac{\partial \sigma}{\partial y_j}(\mathbf{x}_0,\mathbf{y}_0)\cdot \int \frac{\ell(\mathbf{x})}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}$$
$$\times\left(\frac{\displaystyle\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x}}{\displaystyle\int \frac{\ell(\mathbf{x})}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}} - \frac{2}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}\right)$$
$$-\frac{\displaystyle\int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}$$
$$\times\left(\frac{\displaystyle\int \frac{\ell(\mathbf{x})}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}d\mathbf{x}}{\displaystyle\int \frac{\ell(\mathbf{x})}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x}} - \frac{1}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}\right)$$

Now, choose $\mathbf{x}_0$ such that:

$$\frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)}\frac{\partial \sigma}{\partial y_j}(\mathbf{x}_0,\mathbf{y}_0)\cdot \int \frac{\ell(\mathbf{x})}{\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)}d\mathbf{x} = \int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}$$

(which is possible by the Mean Value Theorem). Then, for that choice of $\mathbf{x}_0$:

$$\frac{\partial \psi}{\partial t}(\mathbf{y}_0;0) = -\frac{\displaystyle\int \frac{\lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y})}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y}_0)\right)\right]^2}\ell(\mathbf{x})d\mathbf{x}}{\left[\delta+\lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x}_0,\mathbf{y}_0)\right)\right]^2}$$

which is nonzero under the proposition's assumption. We can then apply the Transversality Theorem (Mas-Colell, Whinston, and Green (1995), Proposition 17.D.3), and as a result ψ intersects the point 0 transversally. Since 0 is a single point, ψ being transversal to 0 implies that 0 is a regular value of ψ. Last, we use the generalization of the Preimage Theorem in Guillemin and Pollack (1974, Preimage Theorem, p.21), which states that, if y is a regular value of $f : X \to Y$, then the preimage $f^{-1}(y)$ is a submanifold in X, with dimension $\dim f^{-1}(y) = \dim X - \dim Y$. Applied to our context, the preimage $\psi^{-1}(0)$ is a submanifold in $\mathbb{R}^Y$, with dimension $\dim \psi^{-1}(0) = \mathbb{R}^Y - 1$, which has measure zero in $\mathbb{R}^Y$. It follows that if y_j is surplus-relevant, then $\partial\overline{\tau}(\mathbf{y})/\partial y_j \neq 0$ for almost all $\mathbf{y}$. $\square$

Part (ii)a. We first show that the (SC) property of σ is necessary for $\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) \neq 0$.

Note that:

$$\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) = \lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\left(\mathbb{E}_\Gamma\left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') \mid \sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\right] - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right)$$
$$= \lambda_1 \int \left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right] \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\}\gamma(\mathbf{y}')d\mathbf{y}' \tag{11}$$

The first observation is that if $\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) \neq 0$, then there exists a point $\mathbf{y}' \neq \mathbf{y}$, which satisfies the level set condition $\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})$, and such that $\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}) \neq 0$.

Next, because $\sigma(\mathbf{x},\cdot)$ is quasi-concave, its level sets are smooth-path connected.[28] Thus, there exists a differentiable function $\mathbf{f}$ that maps $[0,1]$ into the $\sigma(\mathbf{x},\mathbf{y})-$level set of $\sigma(\mathbf{x},\cdot)$ such that $\mathbf{f}(0)=\mathbf{y}$ and $\mathbf{f}(1)=\mathbf{y}'$. Moreover, because $\mathbf{f}$ maps $[0,1]$ into the $\sigma(\mathbf{x},\mathbf{y})-$level set of $\sigma(\mathbf{x},\cdot)$, $\sigma(\mathbf{x},\mathbf{f}(t))=\sigma(\mathbf{x},\mathbf{y})$ for all $t\in[0,1]$. Thus:

$$\forall\, t \in [0,1], \quad \sum_{j=1}^{Y_R} \frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{f}(t)) \cdot f_j'(t) = 0 \tag{12}$$

where $\mathbf{f}(t) = (f_1(t),\cdots,f_{Y_R}(t))$.

By contrast, $t \mapsto \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{f}(t))$ is not constant over $[0,1]$ (this is because $\mathbf{y}'$ was chosen such that $\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}) \neq 0$). Thus:

$$\exists\, t^\star \in [0,1]: \quad \sum_{j=1}^{Y_R} \frac{\partial^2 \sigma}{\partial x_k \partial y_j}(\mathbf{x},\mathbf{f}(t^\star)) \cdot f_j'(t^\star) \neq 0 \tag{13}$$

Moreover, because $\sigma(\mathbf{x},\cdot)$ is a Morse function, its critical points are isolated, so we can always choose $t^\star$ such that $f(t^\star)$ is not a critical point of $\sigma(\mathbf{x},\cdot)$, i.e. such that $\exists\, i \in \{1,\cdots,Y_R\}$: $\partial\sigma/\partial y_i(\mathbf{x},\mathbf{f}(t^\star)) \neq 0$. We can then solve for $f_i'(t^\star)$ in (12):

$$f_i'(t^\star) = -\frac{\sum_{j\neq i} \partial\sigma/\partial y_j(\mathbf{x},\mathbf{f}(t^\star)) \cdot f_j'(t^\star)}{\partial\sigma/\partial y_i(\mathbf{x},\mathbf{f}(t^\star))}$$

and substitute into (13):

$$\sum_{j\neq i} \frac{\partial^2 \sigma}{\partial x_k \partial y_j}(\mathbf{x},\mathbf{f}(t^\star)) \cdot f_j'(t^\star) - \frac{\partial^2 \sigma}{\partial x_k \partial y_i}(\mathbf{x},\mathbf{f}(t^\star)) \cdot \frac{\sum_{j\neq i} \partial\sigma/\partial y_j(\mathbf{x},\mathbf{f}(t^\star)) \cdot f_j'(t^\star)}{\partial\sigma/\partial y_i(\mathbf{x},\mathbf{f}(t^\star))} \neq 0$$

[28] The level sets of quasi-concave functions are convex and compact. Moreover, the boundary of a compact convex set in $\mathbb{R}^n$ is path-connected, since it is homeomorphic to the unit sphere $S^{n-1} = \{\mathbf{x} : ||\mathbf{x}||\}$ in $\mathbb{R}^n$ (where $||\cdot||$ is the standard norm in $\mathbb{R}^n$), and it is well-known that this is a path-connected set (e.g., Munkres (1975), Topology, p. 156). Because we are in $\mathbb{R}^{Y_R}$, path connectedness implies (indeed is equivalent to) smooth-path connectedness (see this link).

Rearranging:

$$\sum_{j\neq i} \frac{f_j'(t^\star)}{\partial\sigma/\partial y_i(\mathbf{x},\mathbf{f}(t^\star))} \cdot \left[\frac{\partial^2\sigma}{\partial x_k \partial y_j}(\mathbf{x},\mathbf{f}(t^\star))\frac{\partial\sigma}{\partial y_i}(\mathbf{x},\mathbf{f}(t^\star)) - \frac{\partial^2\sigma}{\partial x_k \partial y_i}(\mathbf{x},\mathbf{f}(t^\star))\frac{\partial\sigma}{\partial y_j}(\mathbf{x},\mathbf{f}(t^\star))\right] \neq 0$$

Because the sum above is nonzero, at least one of its terms must be nonzero, which proves the claim that the (SC) property of σ holds for $\tilde{\mathbf{y}} = \mathbf{f}(t^\star)$. □

Part (ii)b. To show sufficiency of (SC) for $\frac{\partial\tau}{\partial x_k}(\mathbf{x},\mathbf{y}) \neq 0$, we use a perturbation argument. Consider a point $(\mathbf{x},\mathbf{y})$ such that for some $j \neq i$:

$$\frac{\partial\sigma}{\partial y_i}(\mathbf{x},\mathbf{y})\frac{\partial^2\sigma}{\partial x_k \partial y_j}(\mathbf{x},\mathbf{y}) - \frac{\partial\sigma}{\partial y_j}(\mathbf{x},\mathbf{y})\frac{\partial^2\sigma}{\partial x_k \partial y_i}(\mathbf{x},\mathbf{y}) \neq 0.$$

Note that if the above condition is true, then $\frac{\partial\sigma}{\partial y_i}(\mathbf{x},\mathbf{y})$ and $\frac{\partial\sigma}{\partial y_j}(\mathbf{x},\mathbf{y})$ cannot simultaneously be zero. We focus on the case $\frac{\partial\sigma}{\partial y_i}(\mathbf{x},\mathbf{y}) \neq 0$. Moreover, in what follows we assume that $i < j$; this is merely to fix ideas, nothing except the notation depends on that assumption.

Consider the function $\Delta\sigma : \mathbb{R}^2 \to \mathbb{R}$ defined as follows:

$$\Delta\sigma : (\tilde{y}_i, \tilde{y}_j) \mapsto \sigma\left(\mathbf{x}, [y_1, \cdots, y_{i-1}, \tilde{y}_i, y_{i+1}, \cdots, y_{j-1}, \tilde{y}_j, y_{j+1}, \cdots, y_Y]\right) - \sigma(\mathbf{x},\mathbf{y}).$$

By the Implicit Function Theorem, there exists an open interval $\mathbb{I}$ around y_i and a continuously differentiable function $\Upsilon : \mathbb{I} \to \mathbb{R}$ such that $\Upsilon(y_j) = y_i$ and $\Delta\sigma\left(\Upsilon(\tilde{y}_j), \tilde{y}_j\right) = 0$ for all $\tilde{y}_j \in \mathbb{I}$. Moreover,

$$\Upsilon'(\tilde{y}_j) = -\frac{\dfrac{\partial\sigma}{\partial y_j}\left(\Upsilon(\tilde{y}_j), \tilde{y}_j\right)}{\dfrac{\partial\sigma}{\partial y_i}\left(\Upsilon(\tilde{y}_j), \tilde{y}_j\right)}$$

For later use, let $\tilde{y}_j \in \mathbb{I}$ and let:

$$\boldsymbol{\psi}(\tilde{y}_j) = [y_1, \cdots, y_{i-1}, \Upsilon(\tilde{y}_j), y_{i+1}, \cdots, y_{j-1}, \tilde{y}_j, y_{j+1}, \cdots, y_Y]$$

i.e. $\boldsymbol{\psi}(\tilde{y}_j)$ is the vector whose elements are all equal to those of $\mathbf{y}$, except for elements j and i that equal $\tilde{y}_j$ and $\Upsilon(\tilde{y}_j)$, respectively. Note that, by construction, $\sigma(\mathbf{x}, \boldsymbol{\psi}(\tilde{y}_j)) = \sigma(\mathbf{x},\mathbf{y})$ for all $\tilde{y}_j \in \mathbb{I}$. Also note that, by σ being twice continuously differentiable and by the Single Crossing condition, we can choose $\tilde{y}_j$ sufficiently close to y_j to ensure that $\frac{\partial\sigma}{\partial y_i}(\mathbf{x}, \boldsymbol{\psi}(\hat{y}_j))\frac{\partial^2\sigma}{\partial x_k \partial y_j}(\mathbf{x}, \boldsymbol{\psi}(\hat{y}_j)) - \frac{\partial\sigma}{\partial y_j}(\mathbf{x}, \boldsymbol{\psi}(\hat{y}_j))\frac{\partial^2\sigma}{\partial x_k \partial y_i}(\mathbf{x}, \boldsymbol{\psi}(\hat{y}_j)) \neq 0$ for all $\hat{y}_j \in [y_j, \tilde{y}_j]$.

Next, recall that

$$\begin{aligned}\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) &= \lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\left(\mathbb{E}_\Gamma\left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') \mid \sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\right] - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right)\\ &= \lambda_1 \int \left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right]\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\}\gamma(\mathbf{y}')d\mathbf{y}'\end{aligned}$$

We now show that the set of points $\mathbf{y}$ such that the (SC) condition holds and, simultaneously, $\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y})=0$, has measure zero as claimed. To this end, we perturb γ. Let $m(\mathbf{y})$ be an arbitrary integrable function with $\int m(\mathbf{y}')d\mathbf{y}'=1$ and let $t>0$ be a parameter. Denote by

$$\phi(\mathbf{x},\mathbf{y};t) := \lambda_1 \int \left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right]\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\}((1-t)\gamma(\mathbf{y}')+tm(\mathbf{y}'))d\mathbf{y}'$$

Take the derivative w.r.t. t and evaluate it at point $(\mathbf{x},\mathbf{y};0)$, using the premise $\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y})=0$:

$$\frac{\partial \phi(\mathbf{x},\mathbf{y};0)}{\partial t} = \lambda_1 \int \left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right]\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\}m(\mathbf{y}')d\mathbf{y}'$$

Let m be the Dirac mass at $\boldsymbol{\psi}(\tilde{y}_j)$. Then, the last expression reads:

$$\begin{aligned}\frac{\partial \phi(\mathbf{x},\mathbf{y};0)}{\partial t} &= \lambda_1 \left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j)) - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right]\\ &= \lambda_1 \left[\frac{\partial^2 \sigma\left(\Upsilon(\hat{y}_j),\hat{y}_j\right)}{\partial x_k \partial y_i}\Upsilon'(\hat{y}_j) + \frac{\partial^2 \sigma\left(\Upsilon(\hat{y}_j),\hat{y}_j\right)}{\partial x_k \partial y_j}\right](\tilde{y}_j - y_j) \neq 0\end{aligned}$$

where $\hat{y}_j \in [y_j, \widetilde{y}_j]$. The second equality follows from the Mean Value Theorem. The fact that the resulting expression is nonzero follows from the (SC) property and the expression of $\Upsilon'(\cdot)$.

It then follows from the Transversality Theorem and the Preimage Theorem that the set $\{\mathbf{y} : \partial\tau/\partial x_k(\mathbf{x},\mathbf{y}) \neq 0\}$ has measure zero. By the Transversality Theorem, the above perturbation guarantees that $\partial\tau/\partial x_k$ intersects the point 0 transversally. Since 0 is a single point, $\partial\tau/\partial x_k$ being transversal to 0 implies that 0 is a regular value of $\partial\tau/\partial x_k$. By the generalization of the Preimage Theorem in Guillemin and Pollack (1974, Preimage Theorem, p.21), for every fixed $\mathbf{x}$, the preimage $(\partial\tau/\partial x_k)^{-1}(0)$ is a submanifold in $\mathbb{R}^Y$, with dimension $\dim(\partial\tau/\partial x_k)^{-1}(0) = \mathbb{R}^Y - 1$, which has measure zero in $\mathbb{R}^Y$. □

Remark. We note that for the proof of Part(ii)b. neither quasi-concavity nor the Morse property is needed.

Part (iii). To analyze the dependence of $\tau(\mathbf{x},\mathbf{y})$ on the interaction $x_k y_j$, recall that:

$$\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) = \lambda_1 \int \left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right] \mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') = \sigma(\mathbf{x},\mathbf{y})\}\gamma(\mathbf{y}')d\mathbf{y}'$$

Fix $\mathbf{x} \in \mathcal{X}$. Clearly, if $\forall\, \mathbf{y}' \in \mathcal{Y}$, $\partial\sigma/\partial x_k(\mathbf{x},\mathbf{y}') = 0$ (i.e. if worker attribute x_k is not surplus-relevant), then $\partial\tau/\partial x_k(\mathbf{x},\mathbf{y}) = 0$, and so is $\partial^2\tau/\partial x_k \partial y_j(\mathbf{x},\mathbf{y})$.

Next, following the exact same steps as in Lemma 1, we can rewrite $\partial\tau/\partial x_k(\mathbf{x},\mathbf{y})$ as:

$$\begin{aligned}\frac{1}{\lambda_1}\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) = \sum_{n=1}^{N} \int_{\mathcal{B}(\mathbf{y}_n)} & f_n\left(\left[\mathbf{y}'_{-i(\mathbf{y}_n)}, \Lambda^{\mathbf{y}_n}\left(\mathbf{y}'_{-i(\mathbf{y}_n)}, \sigma(\mathbf{x},\mathbf{y})\right)\right]\right) \\ & \times \left[\frac{\partial \sigma}{\partial x_k}\left(\mathbf{x}, \left[\mathbf{y}'_{-i(\mathbf{y}_n)}, \Lambda^{\mathbf{y}_n}\left(\mathbf{y}'_{-i(\mathbf{y}_n)}, \sigma(\mathbf{x},\mathbf{y})\right)\right]\right) - \frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right] \\ & \qquad \times \gamma\left(\left[\mathbf{y}'_{-i(\mathbf{y}_n)}, \Lambda^{\mathbf{y}_n}\left(\mathbf{y}'_{-i(\mathbf{y}_n)}, \sigma(\mathbf{x},\mathbf{y})\right)\right]\right) d\mathbf{y}'_{-i(\mathbf{y}_n)}\end{aligned}$$

where $\{\mathbf{y}_n\}_{n=1}^N$ is a collection of points of $\mathbb{R}^{Y_R-1}$, $i(\mathbf{y}_n)$ indicates the "missing coordinate" of $\mathbf{y}_n$ as explained in the proof of Lemma 1, $\mathcal{B}(\mathbf{y}_n)$ is an open ball around $\mathbf{y}_n$, $\Lambda^{\mathbf{y}_n}\left(\mathbf{y}'_{-i(\mathbf{y}_n)}, \sigma(\mathbf{x},\mathbf{y})\right)$ is the implicit function defined by $\sigma(\mathbf{x},\mathbf{y}') = \sigma(\mathbf{x},\mathbf{y})$ solved for $y'_{i(\mathbf{y}_n)}$ as a function of $\mathbf{y}'_{-i(\mathbf{y}_n)}$ and $\sigma(\mathbf{x},\mathbf{y})$ over $\mathcal{B}(\mathbf{y}_n)$, and $\{f_n\}_{n=1}^N$ is a partition of unity, again as constructed in the proof of Lemma 1.

As is clear by simple inspection of this last expression of $\partial\tau/\partial x_k(\mathbf{x},\mathbf{y})$, $\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}') = 0$ for all $\mathbf{y}' \in \mathcal{Y}$ implies $\partial^2\tau/\partial x_k\partial y_j(\mathbf{x},\mathbf{y}) = 0$. $\square$

8.1.3 Proof of Proposition 2

Sufficiency. We focus on the EE margin to prove this statement about sorting. The "sorting on the EE margin" term is:

$$\frac{\partial H_j}{\partial x_k}(y \mid \mathbf{x}) = \int_{\underline{y_j}}^{y} \int_{\mathbf{y}_{-j}} \frac{2\delta(\delta+\lambda_1)\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}') \geq 0\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}'))\right]^3} \frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}')\gamma(\mathbf{y}')d\mathbf{y}'_{-j}dy'_j \tag{14}$$

where we define $\overline{F}_{\sigma|\mathbf{x}}(s) := 1 - F_{\sigma|\mathbf{x}}(s)$, and $\mathbf{y}_{-j} := (y_1, \cdots, y_{j-1}, y_{j+1}, \cdots, y_{Y_R})$. See Lindenlaub and Postel-Vinay (2020, Appendix C.2) for a derivation of (14). The objective is to show that (14) is nonzero over a nontrivial set of y.

For ease of exposition, we first express (14) as

$$\frac{\partial H_j}{\partial x_k}(y \mid \mathbf{x}) = \int_{\underline{y_j}}^{y} \int_{\mathbf{y}_{-j}} g\left(\mathbf{x}, \left[y'_j, \mathbf{y}'_{-j}\right]\right) \frac{\partial \tau}{\partial x_k}\left(\mathbf{x}, \left[y'_j, \mathbf{y}'_{-j}\right]\right) d\mathbf{y}'_{-j}dy'_j$$

where

$$g\left(\mathbf{x},[y_j,\mathbf{y}_{-j}]\right) := \frac{2\delta(\delta+\lambda_1)\mathbf{1}\{\sigma\left(\mathbf{x},[y_j,\mathbf{y}_{-j}]\right)\geq 0\}}{\left[\delta+\lambda_1\overline{F}_{\sigma|\mathbf{x}}\left(\sigma\left(\mathbf{x},[y_j,\mathbf{y}_{-j}]\right)\right)\right]^3}\gamma\left([y_j,\mathbf{y}_{-j}]\right)$$

is a non-negative function.

We now proceed to proving the claim by contradiction. Assume that $\partial H_j/\partial x_k(y \mid \mathbf{x}) = 0$ for all y. Then $\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)\frac{\partial\tau}{\partial x_k}\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)d\mathbf{y}'_{-j} = 0$ for almost all y. By the Mean Value Theorem for integrals, there exists a vector $\tilde{\mathbf{y}}_{-j}(y)$ such that

$$\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)\frac{\partial\tau}{\partial x_k}\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)d\mathbf{y}'_{-j} = \frac{\partial\tau}{\partial x_k}\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)d\mathbf{y}'_{-j}.$$

Because $\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)d\mathbf{y}'_{-j} > 0$, $\partial\tau/\partial x_k\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)$ must then equal 0 for almost all y. We now show that, under the Proposition's assumption, that is generically not the case, causing a contradiction.

Recalling that

$$\frac{\partial\tau}{\partial x_k}(\mathbf{x},\mathbf{y}) = \lambda_1\int\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}-\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k}\right]\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = \sigma(\mathbf{x},\mathbf{y})\right\}\gamma(\mathbf{y}')d\mathbf{y}'$$

and substituting into (14), we first note that the function

$$\begin{aligned} y \mapsto\ & \frac{\partial\tau}{\partial x_k}\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right) \\ =\ & \frac{\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)\lambda_1\int\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}'')}{\partial x_k}-\frac{\partial\sigma\left(\mathbf{x},\left[y,\mathbf{y}''_{-j}\right]\right)}{\partial x_k}\right]\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}'') = \sigma\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)\right\}\gamma(\mathbf{y}'')d\mathbf{y}''d\mathbf{y}'_{-j}.}{\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)d\mathbf{y}'_{-j}.} \end{aligned}$$

is a continuously differentiable function of the scalar variable y. (Continuous differentiability follows from the properties of σ and γ, the definition of g, and application of Lemma 1.)

Now consider a point y such that $\partial\tau/\partial x_k\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right) = 0$. To show that this is an isolated zero, we apply the Transversality Theorem by perturbing γ. Let $m(\mathbf{y})$ be an arbitrary integrable function with $\int m(\mathbf{y}')d\mathbf{y}' = 1$ and let $t > 0$ be a parameter. Define

$$\begin{aligned} \phi(y,t) := \int\left[\frac{\partial\sigma}{\partial x_k}(\mathbf{x},\mathbf{y}')-\frac{\partial\sigma}{\partial x_k}\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)\right]&\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = \sigma\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)\right\} \\ &\times\left((1-t)\gamma(\mathbf{y}')+tm(\mathbf{y}')\right)d\mathbf{y}'. \end{aligned}$$

Take the derivative w.r.t. t and evaluate it at point y, using the premise $\partial\tau/\partial x_k\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right) = 0$:

$$\frac{\partial\phi}{\partial t}(y,0) = \int\left[\frac{\partial\sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') - \frac{\partial\sigma}{\partial x_k}\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)\right]\mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}') = \sigma\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)\right\}m(\mathbf{y}')d\mathbf{y}'.$$

To streamline the notation, let $\mathbf{y}^* := [y,\tilde{\mathbf{y}}_{-j}(y)]$. Now, by assumption, there exists $i \in \{1,\cdots,Y_R\}$ such that (SC) holds for (x_k,y_i,y_j) at $(\mathbf{x},\mathbf{y}^*)$, i.e.

$$\frac{\partial\sigma}{\partial y_i}(\mathbf{x},\mathbf{y}^*)\frac{\partial^2\sigma}{\partial y_j\partial x_k}(\mathbf{x},\mathbf{y}^*) - \frac{\partial\sigma}{\partial y_j}(\mathbf{x},\mathbf{y}^*)\frac{\partial^2\sigma}{\partial y_i\partial x_k}(\mathbf{x},\mathbf{y}^*) \neq 0.$$

This implies, in particular, that $\partial\sigma/\partial y_i$ and $\partial\sigma/\partial y_j$ cannot simultaneously equal zero at $(\mathbf{x},\mathbf{y}^*)$. In the rest of this proof, we assume that $\partial\sigma/\partial y_i(\mathbf{x},\mathbf{y}^*) \neq 0$. Adjusting the proof to the case $\partial\sigma/\partial y_j(\mathbf{x},\mathbf{y}^*) \neq 0$ is straightforward. We also assume that $i < j$, only to fix the notation.

Next, define the function $\Delta\sigma : \mathbb{R}^2 \to \mathbb{R}$ as:

$$\Delta\sigma(\tilde{y}_i,\tilde{y}_j) := \sigma\left(\mathbf{x},\left[y_1^*,\cdots,y_{i-1}^*,\tilde{y}_i,y_{i+1}^*,\cdots,\cdots,y_{j-1}^*,\tilde{y}_j,y_{j+1}^*,\cdots,y_{Y_R}^*\right]\right) - \sigma(\mathbf{x},\mathbf{y}^*).$$

By the Implicit Function Theorem, there exists an open interval $\mathbb{I}$ around $y_j^* = y$ and a continuously differentiable function $\Upsilon : \mathbb{I} \mapsto \mathbb{R}$ such that $\Upsilon(y_i^*) = y_j^* = y$ and $\Delta\sigma(\Upsilon(\tilde{y}_j),\tilde{y}_j) = 0$ for all $\tilde{y}_j \in \mathbb{I}$. Moreover,

$$\Upsilon'(\tilde{y}_j) = -\frac{\dfrac{\partial\sigma}{\partial y_j}(\Upsilon(\tilde{y}_j),\tilde{y}_j)}{\dfrac{\partial\sigma}{\partial y_i}(\Upsilon(\tilde{y}_j),\tilde{y}_j)}$$

Below, we consider the point $\boldsymbol{\psi}(\tilde{y}_j) = \left[y_1^*,\cdots,y_{i-1}^*,\Upsilon(\tilde{y}_j),y_{i+1}^*,\cdots,y_{j-1}^*,\tilde{y}_j,y_{j+1}^*,\cdots,y_Y^*\right]$, i.e. the vector whose elements are all equal to those of $\mathbf{y}^* = [y,\tilde{\mathbf{y}}_{-j}(y)]$, except for elements j and i that equal $\tilde{y}_j$ and $\Upsilon(\tilde{y}_j)$, respectively. Note that, by construction, $\sigma(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j)) = \sigma(\mathbf{x},\mathbf{y}^*)$ for all $\tilde{y}_j \in \mathbb{I}$ and that, by the smoothness of σ, we can choose $\mathbb{I}$ narrow enough that $\frac{\partial\sigma}{\partial y_i}(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j))\frac{\partial^2\sigma}{\partial y_j\partial x_k}(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j)) - \frac{\partial\sigma}{\partial y_j}(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j))\frac{\partial^2\sigma}{\partial y_i\partial x_k}(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j)) \neq 0$ for all $\tilde{y}_j \in \mathbb{I}$.

Let m be the Dirac mass at $\boldsymbol{\psi}(\tilde{y}_j)$. Then:

$$\begin{aligned}\frac{\partial\phi}{\partial t}(y,0) &= \lambda_1\left[\frac{\partial\sigma}{\partial x_k}(\mathbf{x},\boldsymbol{\psi}(\tilde{y}_j)) - \frac{\partial\sigma}{\partial x_k}\left(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)]\right)\right]\\ &= \lambda_1\left[\frac{\partial^2\sigma(\Upsilon(\hat{y}_j),\hat{y}_j)}{\partial x_k\partial y_i}\Upsilon'(\hat{y}_j) + \frac{\partial^2\sigma(\Upsilon(\hat{y}_j),\hat{y}_j)}{\partial x_k\partial y_j}\right](\tilde{y}_j - y)\end{aligned}$$

where $\hat{y}_j \in [y,\widetilde{y}_j]$. The second equality follows from the Mean Value Theorem. By the

(SC) property, the second line is not zero (to see this plug in the expression for $\Upsilon'(\hat{y}_j)$). But then, by the Transversality Theorem, point y_j must be an isolated zero (recall that $\partial\tau/\partial x_k\,(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)])=0$ is one equation in one variable y: the Transversality Theorem implies that it is a regular system — see Definition 17.D.3 in Mas-Colell, Whinston, and Green (1995) — and regular systems with the same number of equations as unknowns have isolated solutions). As a result, $\partial\tau/\partial x_k\,(\mathbf{x},[y,\tilde{\mathbf{y}}_{-j}(y)])$ is cannot be zero for almost all y and so there exists a set of positive measure of y for which it is not zero. $\square$

Necessity. If $\partial H_j/\partial x_k(y\mid\mathbf{x})\neq 0$, the integrand in (14) cannot be identically zero. Hence, the part of the integrand in curly brackets cannot be identically zero. In particular, $\partial\tau/\partial x_k(\mathbf{x},\mathbf{y}')$ cannot be identically zero. Proposition 1(ii)a. states that the (SC) condition needs to hold at some point $(\mathbf{x},\tilde{\mathbf{y}})$ (such that $\sigma(\mathbf{x},\tilde{\mathbf{y}})=\sigma(\mathbf{x},\mathbf{y}')$) for $\partial\tau(\mathbf{x},\mathbf{y}')/\partial x_k$ to be nonzero. $\square$

8.1.4 Proof of Proposition 3

Suppose there exist three differentiable functions $\widetilde{\sigma}:\mathbb{R}^2\to\mathbb{R}$, $I:\mathbb{R}_+^X\to\mathbb{R}_+$ and $J:\mathbb{R}_+^Y\to\mathbb{R}_+$ such that for all $(\mathbf{x},\mathbf{y})\in\mathcal{X}\times\mathcal{Y}$, $\sigma(\mathbf{x},\mathbf{y})=\widetilde{\sigma}\left(I(\mathbf{x}),J(\mathbf{y})\right)$. Then, using the identity $\sigma(\mathbf{x},\mathbf{y})=\widetilde{\sigma}\left(I(\mathbf{x}),J(\mathbf{y})\right)$ we can form the ratio,

$$\frac{\frac{\partial\sigma}{\partial y_j}}{\frac{\partial\sigma}{\partial y_1}}=\frac{\frac{\partial\tilde{\sigma}}{\partial J}\frac{\partial J}{\partial y_j}}{\frac{\partial\tilde{\sigma}}{\partial J}\frac{\partial J}{\partial y_1}}=\frac{\frac{\partial J}{\partial y_j}}{\frac{\partial J}{\partial y_1}}$$

which shows that under the SI representation (on the r.h.s.), this ratio does not depend on $\mathbf{x}$ (and thus not on x_k) and therefore the (SC) condition cannot hold strictly at any $(\mathbf{x},\mathbf{y})$, contradicting the premise. $\square$

8.1.5 Surplus under the Sequential Auctions Framework

Workers and firms are risk-neutral and have equal time discounting rates $\rho>0$. Under those assumptions, the total present discounted value of a type-$(\mathbf{x},\mathbf{y})$ match is independent of the way in which it is shared, and 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 being employed under 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, the total surplus generated by a type-$(\mathbf{x},\mathbf{y})$ match is $P(\mathbf{x},\mathbf{y})-U(\mathbf{x})$.

We here discuss the case when wage contracts are set as in the sequential auction model without worker bargaining power of Postel-Vinay and Robin (2002). In the sequential auction model, firms offer take-it-or-leave-it wage contracts to workers. Wage contracts are long-term contracts specifying a fixed wage that can be renegotiated by mutual agreement only. In particular, when an employed worker receives an outside offer, the current and outside employers Bertrand-compete for the worker. Consider a type-$\mathbf{x}$ worker employed at a type-$\mathbf{y}$ firm and receiving an outside offer from a firm of type $\mathbf{y}'$. Bertrand competition between the type-$\mathbf{y}$ and type-$\mathbf{y}'$ employers results in the worker matching with the employer where the total match value is higher, while extracting the full surplus from the lower-surplus match. This implies that he stays in his initial job if $P(\mathbf{x},\mathbf{y}) \geq P(\mathbf{x},\mathbf{y}')$, moves to the type-$\mathbf{y}'$ job otherwise, and ends up with a new wage contract worth $W' = \min\{P(\mathbf{x},\mathbf{y}), P(\mathbf{x},\mathbf{y}')\}$ (provided that W' exceeds the value of the worker's initial contract, W, as otherwise the worker would not have initiated the contract renegotiation in the first place).

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

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

The annuity value of the match, $\rho P(\mathbf{x},\mathbf{y})$, equals the output flow $p(\mathbf{x},\mathbf{y})$ plus the expected capital loss $\delta[U(\mathbf{x}) - P(\mathbf{x},\mathbf{y})]$ of the firm-worker pair from job destruction.[29]

Given that $U(\mathbf{x})$ is independent of firm type, the optimal mobility choices of workers hinge on the comparison of match surplus $\sigma(\mathbf{x},\mathbf{y}) := P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})$ across jobs. It solves $(\rho + \delta)\left[P(\mathbf{x},\mathbf{y}) - U(\mathbf{x})\right] = p(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x})$. Denote the *flow surplus* of a match by:

$$s(\mathbf{x},\mathbf{y}) := p(\mathbf{x},\mathbf{y}) - \rho U(\mathbf{x}).$$

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 $s(\mathbf{x},\mathbf{y}) = p(\mathbf{x},\mathbf{y}) - b(\mathbf{x})$, i.e. the surplus $\sigma(\mathbf{x},\mathbf{y}) = s(\mathbf{x},\mathbf{y})/(\rho+\delta)$ is pinned down by technology. Thus, optimal mobility decisions are entirely determined by technology, where in the case of bilinear surplus: $p(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{Q}\mathbf{y}$ and $b(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{b}$.

[29]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 $P(\mathbf{x},\mathbf{y})$.

8.1.6 Wages

Can the dependence of wages on worker characteristics $\mathbf{x}$, job characteristics $\mathbf{y}$ and their interactions give clues into surplus *and* sorting-relevant attributes?

In order to investigate this, we need to take a stance on the bargaining protocol. For concreteness, we here assume sequential auctions without worker bargaining power (see Appendix 8.1.5), in which case the wage for a worker $\mathbf{x}$ in current job $\mathbf{y}$ and previous job $\mathbf{z}$ reads:

$$w(\mathbf{x},\mathbf{y},\mathbf{z}) = p(\mathbf{x},\mathbf{z}) - \frac{\lambda_1}{\rho+\delta}\int_{\sigma(\mathbf{x},\mathbf{z})}^{\sigma(\mathbf{x},\mathbf{y})}(1-F_{\sigma|\mathbf{x}}(s'))ds'$$

We first investigate the dependence of wages on y_j conditional on $(\mathbf{x},\mathbf{z})$:

$$\frac{\partial w(\mathbf{x},\mathbf{y},\mathbf{z})}{\partial y_j} = -\frac{\lambda_1}{\rho+\delta}\left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial y_j}$$

which is non-zero if the σ depends on y_j (which is the same condition under which the EE rate depends on y_j). So, wages depend on job attribute y_j if and only if it is surplus-relevant.

Next, we analyze the dependence of wages on x_k conditional on $(\mathbf{y},\mathbf{z})$:

$$\begin{aligned}
\frac{\partial w(\mathbf{x},\mathbf{y},\mathbf{z})}{\partial x_k} &= \frac{\partial p(\mathbf{x},\mathbf{z})}{\partial x_k} - \frac{\lambda_1}{\rho+\delta}\left[\left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} - \left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{z}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{z})}{\partial x_k}\right] \\
&\quad + \int_{\sigma(\mathbf{x},\mathbf{z})}^{\sigma(\mathbf{x},\mathbf{y})}\frac{\partial(1-F_{\sigma|\mathbf{x}}(s'))}{\partial x_k}ds' \\
&= \frac{\partial p(\mathbf{x},\mathbf{z})}{\partial x_k} - \frac{\lambda_1}{\rho+\delta}\left[\left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} - \left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{z}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{z})}{\partial x_k}\right] \\
&\quad + \int_{\sigma(\mathbf{x},\mathbf{z})}^{\sigma(\mathbf{x},\mathbf{y})}\mathbb{E}_\Gamma\left[\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\middle|\sigma(\mathbf{x},\mathbf{y}')=s'\right]f_{\sigma|\mathbf{x}}(s')ds' \\
&= \frac{\partial p(\mathbf{x},\mathbf{z})}{\partial x_k} - \frac{\lambda_1}{\rho+\delta}\left[\left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k} - \left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{z}))\right)\frac{\partial\sigma(\mathbf{x},\mathbf{z})}{\partial x_k}\right] \\
&\quad + \int_{\sigma(\mathbf{x},\mathbf{z})}^{\sigma(\mathbf{x},\mathbf{y})}\int\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=s'\}\gamma(\mathbf{y}')d\mathbf{y}'ds'
\end{aligned}$$

So $\frac{\partial w(\mathbf{x},\mathbf{y},\mathbf{z})}{\partial x_k}\neq 0$ indicates that x_k is surplus-(or output-) relevant.

Last, we investigate the dependence of wages on interaction $x_k y_j$ conditional on $\mathbf{z}$:

$$\begin{aligned}
\frac{\partial^2 w(\mathbf{x},\mathbf{y},\mathbf{z})}{\partial x_k\partial y_j} &= -\frac{\lambda_1}{\rho+\delta}\left[\left(1-F_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\right)\frac{\partial^2\sigma(\mathbf{x},\mathbf{y})}{\partial x_k\partial y_j} + f_{\sigma|\mathbf{x}}(\sigma(\mathbf{x},\mathbf{y}))\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial x_k}\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial y_j}\right] \\
&\quad + \int\frac{\partial\sigma(\mathbf{x},\mathbf{y}')}{\partial x_k}\frac{\partial\sigma(\mathbf{x},\mathbf{y})}{\partial y_j}\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\}\gamma(\mathbf{y}')d\mathbf{y}'
\end{aligned}$$

So $\frac{\partial^2 w(\mathbf{x},\mathbf{y},\mathbf{z})}{\partial x_k \partial y_j} \neq 0$ indicates that both x_k and y_j are surplus-relevant.

The bottom line is that the dependence of wages on worker and job characteristics can only be used to learn about surplus-relevant characteristics but gives no insights into whether σ features a (SC) property. It thus gives no insights into sorting or the single index representation of multi-dimensional heterogeneity. Our approach, which is based on EE mobility, is therefore more informative.

8.2 Simulations

8.2.1 Simulated Examples

In our main analysis, we generate data from the following examples (A):

A.1.1 *2D is surplus-relevant with varying degrees of sorting, and with zero correlation among the irrelevant y's and among the irrelevant x's.*

Q is an $X \times Y$ non-zero matrix along the 2D surplus-relevant dimensions and has zeros in all other entries. The Q matrix is chosen to reflect (i) two cases of 'strong' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; (ii) two cases of 'intermediate sorting' between relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; and (iii) two cases of 'weak' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$. The vector of true coefficients in the EE regression has zeros everywhere except for the following: (i) the coefficients of the 2D surplus-relevant y's, (ii) the coefficients of the 2D surplus-relevant x's (since they are involved in sorting), and (iii) the cross-partials of the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$.

A.1.2 *2D is surplus-relevant with varying degrees of sorting, and with non-zero correlation among the irrelevant y's and among the irrelevant x's.*

Q is an $X \times Y$ non-zero matrix along the 2D surplus-relevant dimensions and has zeros in all other entries. The Q matrix is chosen to reflect (i) two cases of 'strong' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; (ii) two cases of 'intermediate sorting' between relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; and (iii) two cases of 'weak' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$. The vector of true coefficients in the EE regression has zeros everywhere except for the following: (i) the coefficients of the 2D surplus-relevant y's, (ii) the coefficients of the 2D surplus-relevant x's (since they are involved in sorting), and (iii) the cross-partials of the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$.

A.2.1 *3D is surplus-relevant with varying degrees of sorting, and with zero correlation among the irrelevant y's and among the irrelevant x's.*

Q is an $X \times Y$ non-zero matrix along the 3D surplus-relevant dimensions and has zeros in all other entries. The Q matrix is chosen to reflect (i) two cases of 'strong' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; (ii) two cases of 'intermediate sorting' between relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; and (iii) two cases of 'weak' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$. The vector of true coefficients in the EE regression has zeros everywhere except for the following: (i) the coefficients of the 3D surplus-relevant y's, (ii) the coefficients of the 3D surplus-relevant x's (since they are involved in sorting), and (iii) the cross-partials of the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$.

A.2.2 *3D is surplus-relevant with varying degrees of sorting, and with non-zero correlation among the irrelevant y's and among the irrelevant x's.*

Q is an $X \times Y$ non-zero matrix along the 3D surplus-relevant dimensions and has zeros in all other entries. The Q matrix is chosen to reflect (i) two cases of 'strong' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; (ii) two cases of 'intermediate sorting' between relevant $\mathbf{x}_R$ and $\mathbf{y}_R$; and (iii) two cases of 'weak' sorting between the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$. The vector of true coefficients in the EE regression has zeros everywhere except for the following: (i) the coefficients of the 3D surplus-relevant y's, (ii) the coefficients of the 3D surplus-relevant x's (since they are involved in sorting), and (iii) the cross-partials of the relevant $\mathbf{x}_R$ and $\mathbf{y}_R$.

We also report the results for the following examples (B):

B.1.1. *1D is surplus-relevant, no sorting, and with zero correlation among irrelevant y's and among the irrelevant x's.* Q is an $X \times Y$ matrix of zeros, except for the first entry. The vector of true coefficients in the EE regression has zeros everywhere except for the coefficient of the single output-relevant y, which should be significantly different from zero.

B.1.2. *1D is surplus-relevant, no sorting, and with non-zero correlation among the irrelevant y's and among the irrelevant x's.* Q is an $X \times Y$ matrix of zeros, except for the first entry. The vector of true coefficients in the EE regression has zeros everywhere except for the coefficient of the single output-relevant y, which should be significantly different from zero.

B.2.1 *2D is surplus-relevant, no sorting, and with zero correlation among the irrelevant y's and among the x's.* Q is a matrix of ones for the relevant attributes and has zero entries otherwise. The vector of true coefficients in the EE regression has zeros everywhere except for the coefficients of 2D surplus-relevant y's, which should be significantly different from zero.

B.2.1 *2D is surplus-relevant, no sorting, and with non-zero correlation among the irrelevant y's and among the x's.* Q is a matrix of ones for the relevant attributes and has zero entries otherwise. The vector of true coefficients in the EE regression has zeros everywhere except for the coefficients of 2D surplus-relevant y's, which should be significantly different from zero.

B.3.1 *3D is surplus-relevant, no sorting, and with zero correlation among the irrelevant y's and among the x's.* Q is a matrix of ones for the relevant attributes and has zero entries otherwise. The vector of true coefficients in the EE regression has zeros everywhere except for the coefficients of 3D surplus-relevant y's, which should be significantly different from zero.

B.3.2 *3D is surplus-relevant, no sorting, and with non-zero correlation among the irrelevant y's and among the x's.* Q is a matrix of ones for the relevant attributes and has zero entries otherwise. The vector of true coefficients in the EE regression has zeros everywhere except for the coefficients of 3D surplus-relevant y's, which should be significantly different from zero.

8.2.2 Additional Results Based on Simulated Data

Table 5: Performance of the BIC, AIC and Robust Lasso Across *All* Models (1 Step, small p)

Method		Accuracy	Recall	Precision	Loss
BIC	**Full Model**	**0.99**	**0.87**	**1.00**	**0.03**
	X	0.92	0.52	1.00	0.36
	Y	0.99	0.97	1.00	0.01
	XY	0.99	0.91	1.00	0.02
AIC	**Full Model**	**0.96**	**0.96**	**0.64**	**0.18**
	X	0.96	0.83	0.97	0.09
	Y	0.98	0.98	0.95	0.02
	XY	0.96	0.98	0.64	0.15
Rlasso	**Full Model**	**0.95**	**0.64**	**0.90**	**0.23**
	X	0.86	0.27	1.00	0.58
	Y	0.91	0.67	1.00	0.25
	XY	0.97	0.60	0.63	0.28

The table reports averages across all models A.1.1.-B.3.2.

Table 6: Performance of the BIC, AIC and Robust Lasso Across *All* Models (2 Step, small p)

Method		Accuracy	Recall	Precision	Loss
BIC	**Full Model**	**0.98**	**0.95**	**0.74**	**0.12**
	X	0.92	0.81	0.89	0.14
	Y	1.00	1.00	1.00	0.00
	XY	0.98	0.97	0.79	0.07
AIC	**Full Model**	**0.95**	**0.95**	**0.61**	**0.21**
	X	0.92	0.81	0.90	0.14
	Y	0.85	1.00	0.67	0.15
	XY	0.97	0.97	0.69	0.13
Rlasso	**Full Model**	**0.97**	**0.71**	**1.00**	**0.14**
	X	0.81	0.00	1.00	1.00
	Y	0.98	0.94	1.00	0.02
	XY	0.98	0.75	1.00	0.16

The table reports averages across all models A.1.1.-B.3.2.

Table 7: Performance of the BIC, AIC and Robust Lasso Across Models With No Sorting (2 Step, small p)

Method		Accuracy	Recall	Precision	Loss
BIC	**Full Model**	**0.98**	**1.00**	**0.43**	**0.33**
	X	0.94	–	–	–
	Y	1.00	1.00	1.00	0.00
	XY	0.98	–	–	–
AIC	**Full Model**	**0.95**	**1.00**	**0.32**	**0.48**
	X	0.94	–	–	–
	Y	0.83	1.00	0.67	0.17
	XY	0.96	–	–	–
Rlasso	**Full Model**	**1.00**	**1.00**	**1.00**	**0.00**
	X	1.00	–	–	–
	Y	1.00	1.00	1.00	0.00
	XY	1.00	–	–	–

The table reports averages across all models B.1.1.-B.3.2.

Table 8: Performance of the AIC, BIC and Robust Lasso (2 Step, large p)

Method		Accuracy	Recall	Precision	Loss
BIC	**Full Model**	**0.99**	**0.93**	**0.72**	**0.11**
	X	0.94	0.77	0.82	0.22
	Y	1.00	1.00	1.00	0.00
	XY	0.99	0.97	0.63	0.16
AIC	**Full Model**	**0.97**	**0.93**	**0.53**	**0.25**
	X	0.95	0.76	0.88	0.20
	Y	0.86	1.00	0.51	0.26
	XY	0.98	0.97	0.48	0.30
RLasso	**Full Model**	**0.99**	**0.59**	**1.00**	**0.21**
	X	0.88	0.00	1.00	1.00
	Y	0.98	0.90	1.00	0.03
	XY	0.99	0.73	1.00	0.18

The table reports averages across models A.1.1.-A.2.2. and $X = Y = \{20, 40\}$.

8.3 Data

8.3.1 Construction of Multi-Dimensional Worker and Job Attributes

Multi-Dimensional Job Attributes. The O*NET has many different subsamples, each of them describing certain attributes of occupations. We keep those subsamples whose variables have a comparable scale (namely the 'importance' of certain tasks/skills/abilities/amenities for an occupation) and drop the remaining ones. In particular, the O*NET datasets we keep are those called 'Work Activities', 'Skills' and 'Abilities'. Combined they contain 130 distinct job attributes. However, many of these job attributes are similar in nature and highly correlated. To perform some variable pre-selection that combines similar job attributes while preserving interpretability we use PCA with exclusion restrictions (see Lise and Postel-Vinay (2019) for details).

Multi-Dimensional Worker Attributes. We here focus on the construction of worker attributes in the SIPP 2008 — our baseline sample. To impute the agents' multi-dimensional attributes, in a first step we extract from the SIPP's Education and Training Module information on individuals' *college degrees, apprenticeships and vocational degrees*, as well as occupational training on-the-job (variables `evocfld`, `eassocfd`, `ebachfld`, `eadvncfd`). We first construct a crosswalk between between the SIPP fields of study (which are quite coarse) and the more standard CIP codes, using a fuzzy merge in a first stage and correcting the poor matches by hand in a second stage. The CIP codes come from the crosswalk `FINALCIPtoSOCcrosswalk_022811.dta` downloaded from the National Center for Education Statistics (see `https://nces.ed.gov/ipeds/cipcode/resources.aspx?y=55` using the link 'CIP 2010 to SOC 2010 Crosswalk'). Now our fields of study associated with degrees in the SIPP are at CIP level. Crosswalk `FINALCIPtoSOCcrosswalk_022811.dta` then links CIP codes (and thus field of studies in the SIPP) to the occupation codes SOC 2010. We drop the observations whose field of study cannot be matched to an occupation. We then merge the the fields of study to the ONET occupational codes by using a crosswalk that links occupational codes SOC 2010 to occupational codes ONET-SOC. This crosswalk is called `2010_to_SOC_Crosswalk.dta`, downloaded here: `https://www.onetcenter.org/taxonomy/2010/soc.html/2010_to_SOC_Crosswalk.xls?fmt=xls`. Finally, we merged in the Census 2002 codes (denoted by variable 'censcode' in the datasets) using crosswalk `Crosswalk-census2002-soc2000.dta` (downloaded here: `http://data.widcenter.org/download/xwalks/` with original file name: `cen02soc.txt`). This step is needed since the constructed job attributes above are created based on Census 2002 occupational codes. Finally, we merge the job attributes constructed above (saved to `job_attributes.dta`) to the fields of study, meaning every field of study is now associated with certain skill requirements as well as job amenities.

In a second step, we add skills that are based on completed *training* experiences in the current occupation (where we only take those training experiences into account *before* our baseline sample starts, i.e. before April 2009, to avoid a mechanically high correlation between skills and job attributes). We focus on individuals who got trained to improve skills in the current job (according to variable `ercvtrn2`). For 'current occupation', we focus on the main current occupation as employee `tjbocc1` (coded using the 2002 census occupational classification); if this occupation is not in the universe, we focus on main current occupation as business owner `tbsocc1` (also coded using the 2002 census occupational classification). Finally, we merge the job attributes into those occupations that the individuals have been trained for (using `job_attributes.dta`) as above.

In a third step, to arrive at a vector of attributes for each worker, we take each individual and average each skill in the vector across all education and training experiences. Thus every worker ends up with $X = 10$ worker attributes.

For the previous SIPP panels (1996, 2001, 2004) we use essentially the same method for multi-d skill imputation but need to adjust the occupational codes and crosswalks since occupational codes in the SIPP have changed across panels.

8.3.2 Summary Statistics SIPP 2008

	Degree and Training Sample (Baseline)				Education and Training Module			
	Mean	Std Dev.	Min	Max	Mean	Std Dev.	Min	Max
HSD	.006	.077	0	1	.096	.295	0	1
HS	.051	.221	0	1	.257	.437	0	1
Other-Degree	.38	.485	0	1	.361	.48	0	1
>=BA	.563	.496	0	1	.286	.452	0	1
Unemployed	.044	.206	0	1	.064	.244	0	1
Married	.637	.481	0	1	.552	.497	0	1
Male	.438	.496	0	1	.456	.498	0	1
Black	.198	.398	0	1	.211	.408	0	1
Age	42.769	10.572	20	60	40.862	11.794	20	60
EE (monthly)	.014	.003	.001	.02	.014	.003	.001	.02
EU (monthly)	.011	.004	.005	.026	.011	.004	.005	.026
UE (monthly)	.125	.027	.069	.197	.124	.027	.069	.197
N (person-month)	895,747				2,069,943			

Table 9: Summary Statistics Across Different Samples

	Degree and Training Sample (Baseline)				Education and Training Module			
	Mean	Std Dev.	Min	Max	Mean	Std Dev.	Min	Max
x1	.636	.124	.036	1	.444	.202	.036	1
x2	.675	.144	.029	1	.5	.193	.029	1
x3	.509	.095	0	1	.441	.096	0	1
x4	.397	.147	0	1	.53	.161	0	1
x5	.563	.1	.057	1	.465	.121	.057	1
x6	.401	.133	.054	1	.389	.092	.054	1
x7	.566	.108	.056	1	.498	.103	.056	1
x8	.57	.125	0	1	.492	.11	0	1
x9	.349	.132	0	1	.371	.092	0	1
x10	.677	.15	.005	1	.467	.223	.005	1
y1	.552	.181	0	1	.477	.197	0	1
y2	.612	.206	0	1	.533	.228	0	1
y3	.48	.143	0	1	.455	.151	0	1
y4	.448	.217	0	1	.515	.228	0	1
y5	.509	.172	0	1	.482	.17	0	1
y6	.428	.219	0	1	.401	.195	0	1
y7	.575	.162	0	1	.537	.161	0	1
y8	.544	.174	0	1	.529	.175	0	1
y9	.339	.183	0	1	.371	.2	0	1
y10	.57	.204	0	1	.503	.225	0	1
y11	2.19	.85	1	3	2.079	.866	1	3

Principal components are re-scaled to be between 0 and 1. Firm size y_{11} is a categorical variable, taking values $\{1, 2, 3\}$.

Table 10: Summary Statistics of Potential Skills and Job Attributes

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1.000									
x_2	0.689	1.000								
x_3	0.332	0.113	1.000							
x_4	-0.674	-0.775	0.186	1.000						
x_5	0.612	0.639	-0.005	-0.491	1.000					
x_6	-0.127	0.081	0.115	0.370	-0.130	1.000				
x_7	0.149	0.549	0.333	0.013	0.222	0.628	1.000			
x_8	0.236	0.338	0.439	-0.118	0.387	-0.106	0.327	1.000		
x_9	-0.278	-0.517	0.568	0.615	-0.308	-0.131	-0.254	0.234	1.000	
x_{10}	0.810	0.663	0.579	-0.535	0.525	-0.314	0.199	0.498	0.127	1.000

Table 11: Correlations Worker Attributes

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
y_1	1.000										
y_2	0.655	1.000									
y_3	0.438	0.137	1.000								
y_4	-0.481	-0.714	0.310	1.000							
y_5	0.589	0.577	0.042	-0.342	1.000						
y_6	0.092	0.127	0.371	0.412	-0.053	1.000					
y_7	0.374	0.551	0.499	0.070	0.271	0.630	1.000				
y_8	0.391	0.428	0.358	-0.054	0.444	0.107	0.415	1.000			
y_9	-0.282	-0.436	0.482	0.660	-0.247	0.102	-0.131	0.258	1.000		
y_{10}	0.792	0.667	0.559	-0.421	0.506	-0.133	0.349	0.616	0.105	1.000	
y_{11}	0.103	0.081	0.092	-0.076	-0.069	0.052	0.011	0.019	0.011	0.105	1.000

Table 12: Correlation of Job Attributes

8.4 Results

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Linear Terms
y_1	.	.	.	.	.	.	.	.	.	.	.
y_2	.	.	.	.	.	.	.	.	.	.	.
y_3	.	.	.	.	.	.	.	.	.	.	.
y_4	.	.	.	.	.	.	.	.	.	.	.
y_5	.	✓	✓	✓	✓	.	.	.	✓	.	✓
y_6	.	.	.	.	.	.	.	.	.	.	.
y_7	.	✓	✓	✓	✓	.	.	.	✓	.	✓
y_8	.	.	.	.	.	.	.	.	.	.	.
y_9	.	✓	.	✓	✓	.	.	.	✓	.	✓
y_{10}	.	✓	✓	✓	.	.	.	.	.	✓	✓
y_{11}	.	.	.	✓	✓	.	.	.	✓	.	✓
Linear Terms	.	✓	✓	✓	✓	.	.	.	✓	✓	.

✓ indicates significant worker/job attribute or interaction. Variable y_{11} indicates firm size.
Sample: Degree and Training (subsample of Education Module), Pooled Across 1996-1999. Baseline controls (non-selected): male, age, married, race (all second stage); monthly mean EE rate (1. and 2. stage).

Table 13: Model Selection Results Based on BIC 2 Step, Years 1996-1999

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Linear Terms
y_1	.	.	.	.	.	.	.	.	.	.	.
y_2	.	.	.	.	.	.	.	.	.	.	.
y_3	.	.	✓	.	.	.	.	.	.	✓	✓
y_4	.	.	.	.	.	.	.	.	.	.	✓
y_5	.	.	.	.	.	.	.	.	.	✓	✓
y_6	.	.	.	.	.	.	.	.	.	.	.
y_7	.	.	.	.	.	.	.	.	.	.	.
y_8	.	.	.	.	.	.	.	.	.	.	.
y_9	.	.	.	.	.	.	.	.	.	.	.
y_{10}	.	.	.	.	.	.	.	.	.	.	✓
y_{11}	.	.	✓	.	.	.	.	.	.	✓	✓
Linear Terms	.	.	✓	.	.	.	.	.	.	✓	.

✓ indicates significant worker/job attribute or interaction. Variable y_{11} indicates firm size.
Sample: Degree and Training (subsample of Education Module), Pooled Across 2001-2003. Baseline controls (non-selected): male, age, married, race (all second stage); monthly mean EE rate (1. and 2. stage).

Table 14: Model Selection Results Based on BIC 2 Step, Years 2001-2003

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Linear Terms
y_1	.	.	.	.	.	.	.	.	.	.	.
y_2	.	✓	.	.	.	.	.	.	.	✓	✓
y_3	.	.	.	.	.	.	.	.	.	.	.
y_4	.	.	.	.	.	.	.	.	.	.	.
y_5	.	✓	✓	.	.	.	.	.	.	✓	✓
y_6	.	.	.	.	.	.	.	.	.	.	.
y_7	.	.	.	.	.	.	.	.	.	.	.
y_8	.	.	.	.	.	.	.	.	.	.	.
y_9	.	.	.	.	.	.	.	.	.	.	.
y_{10}	.	.	✓	.	.	.	.	.	.	✓	✓
y_{11}	.	✓	.	.	.	✓	.	.	.	.	✓
Linear Terms	.	✓	✓	.	.	✓	.	.	.	✓	.

✓ indicates significant worker/job attribute or interaction. Variable y_{11} indicates firm size.
Sample: Degree and Training (subsample of Education Module), Pooled Across 2004-2007. Baseline controls (non-selected): male, age, married, race (all second stage); monthly mean EE rate (1. and 2. stage).

Table 15: Model Selection Results Based on BIC 2 Step, Years 2004-2007

8.5 Application: Additional Material

8.5.1 EE, UE and UE Sorting

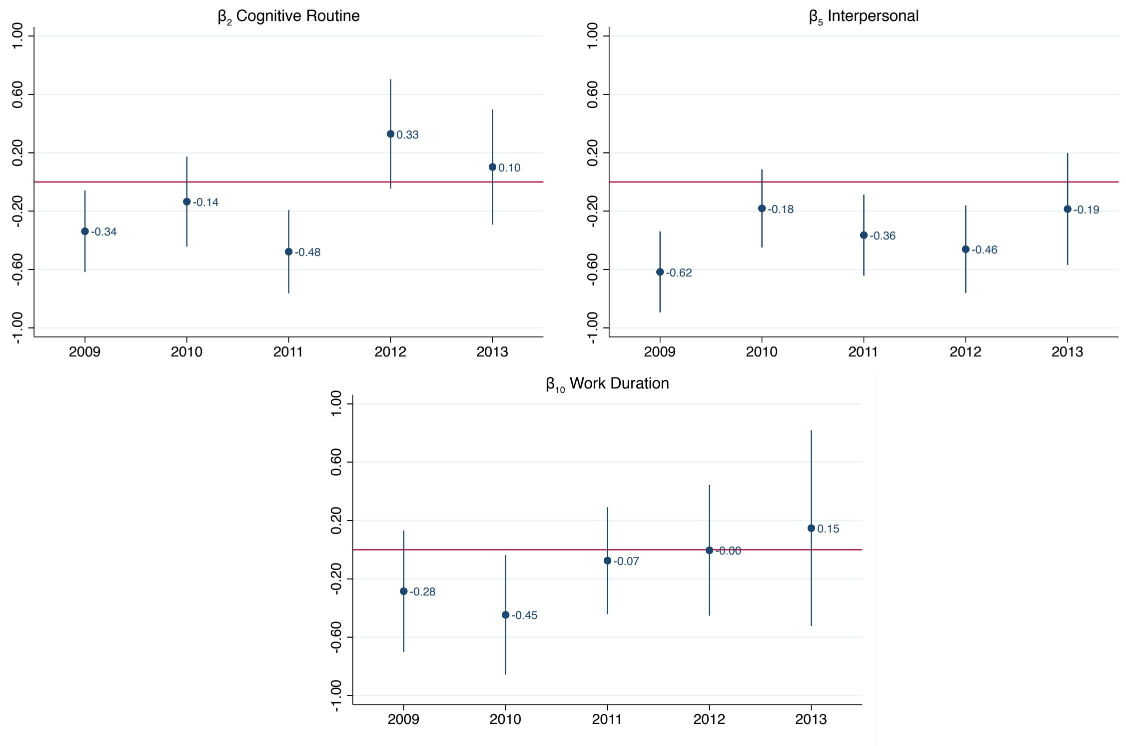

Figure 6: EE Sorting Over Time

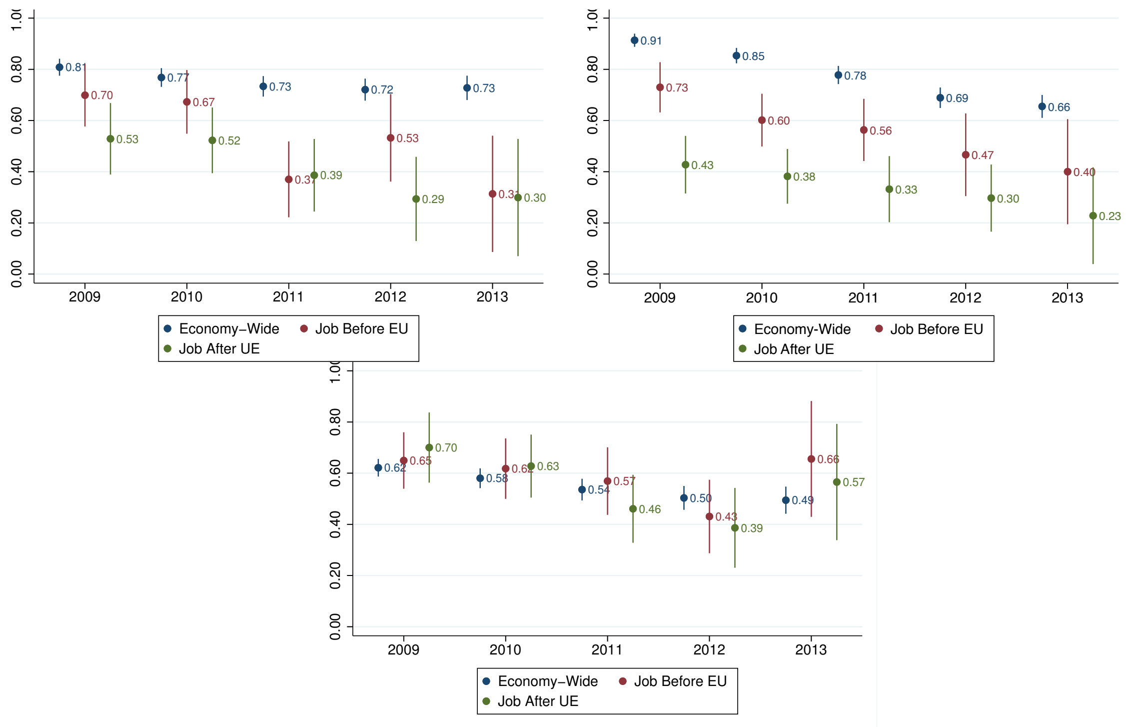

Figure 7: EU and UE Sorting Over Time

References

Abowd, J. M., F. Kramarz, and D. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Akaike, H. (1973): *Information Theory and an Extension of the Maximum Likelihood Principle*pp. 199–213. Springer New York, New York, NY.

Akaike, H. (1974): "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Bagger, J., and R. Lentz (2018): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86(1), 153–190.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80(6), 2369–2429.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87(3), 699–739.

Borovickova, K., and R. Shimer (2018): "High Wage Workers Work for High Wage Firms," Working Paper.

Card, D., and T. Lemieux (1996): "Wage dispersion, returns to skill, and black-white wage differentials," *Journal of Econometrics*, 74(2), 319–361.

Deming, D. J. (2017): "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, 132(4), 1593–1640.

Fan, J., and R. Li (2001): "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *Journal of the American Statistical Association*, 96(456), 1348–1360.

Garibaldi, P., E. R. Moen, and D. E. Sommervoll (2016): "Competitive on-the-job search," *Review of Economic Dynamics*, 19, 88 – 107, Special Issue in Honor of Dale Mortensen.

Guillemin, V., and A. Pollack (1974): *Differential Topology.* AMS Chelsea Publishing.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85(1), 29–65.

Japkowicz, N., and M. Shah (2011): *Evaluating Learning Algorithms: A Classification Perspective.* Cambridge University Press.

Lindenlaub, I. (2017): "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 84(2), 718–789.

Lindenlaub, I., and F. Postel-Vinay (2020): "Multidimensional Sorting Under Random Search," *Working Paper.*

Lise, J., and F. Postel-Vinay (2019): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, forthcoming.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*, no. 9780195102680 in OUP Catalogue. Oxford University Press.

Phien, P. (2011): "Some Quantitative Results on Lipschitz Inverse and Implicit Functions Theorems," *East West Journal of Mathematics*, 13(1).

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Schwarz, G. (1978): "Estimating the dimension of a model," *The Annals of Statistics*, 6, 461–464.

Sorkin, I. (2018): "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 133(3), 1331–1393.

Taber, C., and R. Vejlin (2016): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Working Paper.*

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.

Zou, H. (2006): "The Adaptive Lasso and Its Oracle Properties," *Journal of the American Statistical Association*, 101(476), 1418–1429.