NBER WORKING PAPER SERIES

THE WORKER-JOB SURPLUS

Ilse Lindenlaub
Fabien Postel-Vinay

Working Paper 28402
http://www.nber.org/papers/w28402

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021

We thank various seminar audiences for insightful comments. All remaining errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Worker-Job Surplus
Ilse Lindenlaub and Fabien Postel-Vinay
NBER Working Paper No. 28402
January 2021
JEL No. E24,J0,J20

ABSTRACT

The worker-job surplus — the sum of the worker's and the employer's net values of an employment relationship — is the object that drives decisions in most matching models of the labor market. 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. To do so, we exploit the mobility choices of employed workers. Our method further indicates whether workers sort along those surplus-relevant attributes when searching for jobs. It also provides a test of the commonly used single-index assumption, according to which worker and job heterogeneity can each be summarized by scalar indices. 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. We then use our results in a new approach to defining the economy's labor submarkets, highlighting a potentially important application of our methodology.

Ilse Lindenlaub
Department of Economics
Yale University
28 Hillhouse Avenue
New Haven, CT 06511
and NBER
ilse.lindenlaub@yale.edu

Fabien Postel-Vinay
Department of Economics
University College London
Gower Street
London
WC1E 6BT
United Kingdom
f.postel-vinay@ucl.ac.uk

1 Introduction

The worker-job surplus — the sum of the worker's and the employer's values of their employment relationship, net of their outside option values — is the cornerstone of most search models of the labor market. It drives hiring and separation decisions. It is the key input into most wage determination protocols. And it is the basic building block of welfare and efficiency analysis in those models. Correctly measuring the worker-job surplus is thus a pre-requisite for the design of policies aiming to improve match quality between workers and jobs, such as targeted training programs for workers, or targeted hiring subsidies for employers. 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 worker-job surplus. Our primary contribution is methodological. We develop a theory-based empirical protocol to determine which of the observable worker and job attributes are relevant determinants of the worker-job surplus. To that end, we exploit workers' mobility choices. Our method further indicates whether the surplus-relevant attributes matter for sorting. As an important by-product, our approach reveals whether all relevant worker and job heterogeneity can be summarized by scalar indices without loss of information — something we will refer to as the *single-index assumption*.

Our second contribution is applied. We implement our empirical protocol on US data and show that multiple worker and job attributes are surplus-relevant. We reject the widely used assumption that the relevant worker and job heterogeneity can be summarized by scalar indices, thus highlighting the importance of accounting for multi-dimensional heterogeneity. In an application, we show how these results can be used to draw the boundaries of the economy's distinct labor (sub)markets.

The intuitive underpinning of our method is simple. It builds upon the premise that most 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. Observing individual job mobility decisions, and identifying the set of worker and job characteristics that affects those choices, should therefore inform us about which worker and job traits are indeed surplus-relevant.

To illustrate this intuition, we focus on jobs and workers that differ in two dimensions — say 'cognitive skills' and 'manual skills' for workers (or, cognitive and manual skill requirements for jobs). Consider two workers with equal cognitive skills x_C, currently in jobs with equal cognitive job attributes y_C. Assume these workers differ in their manual skills, x_M, and their jobs differ in their manual skill requirements, y_M. If only the cognitive dimension is relevant for surplus maximization, then both workers should accept the same jobs and change jobs at the same rate. In short, they should climb the *same job ladder*. But if manual heterogeneity also matters for the surplus, then the job mobility decisions of the two workers should differ: they climb different job ladders. Thus, testing for heterogeneity of job mobility decisions conditional on cognitive job and worker types should allow us to either validate or reject the hypothesis of (cognitive) scalar heterogeneity. If we reject it, then 'adding' heterogeneity is necessary to explain differences in mobility choices, implying that the surplus depends on multiple dimensions of heterogeneity. This logic extends beyond two dimensions.

We formalize this intuition in a general class of random search job-ladder models, in which workers and jobs are joint surplus maximizers and so mobility choices are driven by surplus comparisons between current job and outside job offers. We prove that the observed dependence of individual job-to-job transition rates on worker and job attributes allows us to infer several important properties of the surplus function.

Our main result is that if workers' job-to-job transition rate depends on some worker trait or current job attribute then those attributes are *surplus-relevant*. In addition, we show that the dependence of the job-to-job transition probability on some specific worker skill x_k 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 to arise in equilibrium — a situation where workers with different levels of skill x_k are matched to jobs that systematically differ in some job trait y_j. Second, single-crossing implies that the relevant multi-dimensional heterogeneity of workers and jobs cannot be collapsed into scalar indices without loss of information. Combining these insights implies that whenever there is worker-job sorting involving multi-dimensional characteristics, there exists no scalar representation of the relevant heterogeneity. This is an important result, considering that the overwhelming majority of the literature on sorting with frictions relies on scalar heterogeneity.

We can infer these properties of the (unobserved) surplus function from the observed dependence of individual job transition rates on worker and job attributes, without estimating a specific structural model. An important advantage of our framework is its generality. Our theory-based protocol and its implications do not rely on restrictive functional form assumptions, or on any specific wage-setting protocol. To prove our results in this generality, we rely on tools from Differential Topology that have been used in General Equilibrium analysis but are not standard in the macro-labor literature.

Based on our theory, we build a flexible statistical model of the job transition rate – or employment-to-employment (EE) transition rate — allowing it to depend on any number of worker or job attributes that are observed in the data. We then use model selection techniques to determine which of those potentially relevant attributes indeed impact job mobility choices.

We begin by trying our empirical protocol in Monte Carlo experiments, where we use our theoretical framework to generate data. Knowing the underlying data generating process, we can gauge whether our method successfully separates those worker and job characteristics that truly impact EE transitions from those that do not. 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 when the number of potential worker and job traits is not too large and when the data features a significant amount of sorting. Moreover, the Bayesian Information Criterion (BIC) is the most effective model selection technique in our context, when compared to the Akaike Information Criterion (AIC), the Lasso and several others. We use those results as guidance when implementing the method on real data.

We apply our method to US data using a combination of the Survey of Income and Program Participation (SIPP) and the O*NET. We first impute multi-dimensional worker characteristics (skills and preferences for job 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] We then construct multi-dimensional job attributes (skill requirements, job amenities and firm size) for each job based on the O*NET and the SIPP.[2]

Our main results are the following. First, our protocol selects a model with relatively few — although definitely more than one — relevant dimensions of heterogeneity. On the job side, we find that a routine-cognitive and an interpersonal skill requirement get selected along with the (dis)amenity of work duration and workplace size. On the worker side, we find that most of the relevant worker characteristics are symmetric to the selected job attributes (cognitive skill,

[1] Here we focus on a considerably larger subsample of the SIPP than we did in Lindenlaub and Postel-Vinay (2020), including not only those with college degree but also those with apprenticeships/occupational training.

[2] Our baseline sample is the 2008 SIPP panel, but we also implement our method on earlier panels going back to 1996, when the special Education and Training module was first launched.

interpersonal skill and the workers' tolerance for long work hours). In addition, two distinct manual skills are found to be relevant predictors of EE mobility. Based on our theory, we conclude that all of those attributes are relevant determinants of the worker-job surplus, and that workers sort into jobs along those dimensions of heterogeneity.

Second, our results imply that the surplus satisfies a single-crossing property, ruling out the existence of a single-index representation of the relevant multi-dimensional worker and job attributes. Thus, heterogeneity is truly multi-dimensional in the data and cannot be summarized by scalar indices. In particular, workers with different attribute bundles rank jobs in different ways: there is no single, economy-wide job ladder that all workers agree on. Instead, workers with different skill bundles climb different job ladders.

To further illustrate the usefulness of our methodology, we use our results on the surplus-relevant worker and job traits to develop a new way of defining the economy's labor (sub)markets, where workers and employers typically participate in either one or a small number of markets. A recent body of work (Schmutte, 2014; Nimzcik, 2020) applies community detection algorithms to delineate labor submarkets using observed patterns of worker mobility. We use our results to build on that literature but start from a different premise, namely that labor markets are segmented by *surplus-relevant* job and worker attributes as this is the heterogeneity determining surplus comparisons, job acceptance sets and thus worker mobility. Workers with similar relevant skills can be employed in jobs with similar relevant characteristics. We then define labor submarkets so that the *potential* — as opposed to observed — mobility of workers within markets is maximized. This is reasonable since even though two different workers with similar relevant skills may not have had the exact same jobs in the past, due for example to random frictions, they are capable of performing the same tasks, and interact on the same labor market.

We find that the economy is characterized by four distinct labor markets, each of them containing jobs that are similar in terms of surplus-relevant attributes. We show that these markets have economic content: There is substantial sorting in the sense that the 'typical' (modal) labor market differs across worker types. Also, wages are higher and separations less likely when a worker is employed in their typical labor market. Finally, job ladders differ across labor markets, where improvements in certain job attributes (upon an EE move) are rewarded in some markets but penalized in others. This highlights job ladder heterogeneity — one of the defining features of an economy where multi-dimensional worker and job traits are relevant.

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]

On the reduced-form end of that literature, the influential contribution by Abowd, Kramarz, and Margolis (1999) (AKM) uses a two-way fixed-effect wage regression to estimate unobserved worker and firm heterogeneity and assesses sorting based on the correlation between the estimated fixed effects.[4] More recently, Bonhomme, Lamadon, and Manresa (2019) propose a two-stage approach to estimate unobserved worker and job heterogeneity, relaxing several assumptions from AKM's fixed-effect approach. In the first stage, they estimate unobserved firm types using k-means clustering of firms' wage distributions. In the second stage, they identify the distribution of unobserved worker types as well as type-conditional earnings distributions based on observed earnings distributions of job movers and job stayers.[5]

Following a more structural approach, Hagedorn, Law, and Manovskii (2017) identify unobserved scalar heterogeneity of workers and firms based on a random search 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 different contexts, Taber and Vejlin (2016), Bagger and Lentz (2018) and Sorkin (2018) use observed job-to-job mobility patterns to rank firms and recover an economy-wide firm ladder.[6]

Like those contributions, our method exploits the information conveyed by workers' mobility choices to identify worker and job heterogeneity. Unlike those contributions, however, we allow for the match surplus (and thus for worker and firm allocation decisions) to depend on any number of observable characteristics. We then let the data tell us which job and worker

[3]Examples of papers that do not rely on unobserved scalar heterogeneity are Lindenlaub (2017), Lise and Postel-Vinay (2020) and Lindenlaub and Postel-Vinay (2020) who all treat heterogeneity as observed and multi-dimensional as we do here, 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.

[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 assumption that log wages are linear in worker and firm heterogeneity.

[6]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 1D 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 (scalar) productivity while Sorkin (2018) assumes they are homogenous, ruling out sorting.

attributes are relevant determinants of the surplus. We thus take a fundamentally different approach, advocating to first learn about relevant observed (possibly multi-dimensional) heterogeneity instead of promptly assuming unobserved (scalar) heterogeneity. Our results suggest that the assumption of scalar heterogeneity on either side of the labor market is too restrictive as both workers and jobs have multiple relevant attributes, which cannot be collapsed into a single-index. To the best of our knowledge, ours is the first paper to develop and implement a micro-founded empirical protocol to uncover the determinants of the worker-job surplus.

2 Theory

We develop a theory-based empirical method to determine *how many* and *which* worker and job attributes affect the worker-job surplus and matter for sorting. We begin by describing the theoretical framework. We then explain our method, first intuitively, then formally. Finally, we offer an interpretation of our method's empirical output in light of our theory.

2.1 Framework

The theoretical framework is based on Lindenlaub's and Postel-Vinay's (2020) analysis of multi-dimensional sorting under random search. We describe the model environment briefly.

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 of attributes $\mathbf{x} = (x_1, \cdots, x_X) \in \mathcal{X}$, 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 $\mathbf{x}$ 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, skill requirements or amenities, $\mathbf{y} = (y_1, \cdots, y_Y) \in \mathcal{Y}$, 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 all agents *and to the econometrician.*

Workers can either be employed or unemployed. In both states, they face search frictions and 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 that are observed in the data 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, and Morse functions are dense in the set of smooth functions defined over $\mathcal{Y}$.

Here are some functional forms satisfying Assumption 1:[7] A common one is the bilinear form, $\sigma(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{Q}(\mathbf{y} - \mathbf{b})$, 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 commonly used functional form satisfying Assumption 1 is $\sigma(\mathbf{x}, \mathbf{y}) = A - \sum_{k,j=k}(x_k - y_j)^2, A > 0$.

Our main results will be established under the assumptions listed in this sub-section. 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 determining which job and worker attributes impact those decisions — should inform us about which job and worker attributes are surplus-relevant.

Consider a simple two-dimensional example ($X = Y = 2$) for illustration, where 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 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* one-dimensional match types (the cognitive dimension, in our example) should allow us to either validate the assumption of one-dimensional

[7] 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.

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 probability of receiving a job offer, λ_1, multiplied by the probability of 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 $\mathbf{y}$:

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

We will also be interested in the average EE transition probability out of 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 type-$(\mathbf{x}, \mathbf{y})$ matches.

We define a single crossing property of the surplus function, which we will use repeatedly:

Definition 1 (Single-Crossing) *The surplus function* σ *has the* SINGLE-CROSSING (SC) PROPERTY *for attributes* $(y_i, y_j; x_k)$ *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 a worker attribute* $k \in \{1, \cdots, X_R\}$.

(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 almost all* $\mathbf{y} \in \mathcal{Y}$*, 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 attributes* $(y_i, y_j; x_k)$ *at* $(\mathbf{x}, \tilde{\mathbf{y}})$*.*

b. For almost all $\mathbf{y} \in \mathcal{Y}$*, if there exist job attributes* $(i, j) \in \{1, \cdots, Y_R\}^2$ *such that the (SC) property holds for attributes* $(y_i, y_j; x_k)$ *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. It calls on some subtle arguments from Differential Topology (primarily applications of the Transversality Theorem and Preimage Theorem) to show that the relevant derivatives of the EE transition rate are 'generically' not zero under the stated conditions. Technical challenges notwithstanding, the underlying intuition is straightforward: We can characterize several properties of the surplus function σ based on the dependence of the EE rate on worker and job attributes observed in the data — without estimating the model structurally.

To implement this result empirically, we will approximate τ using the following Linear Probability Model:

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

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 ways to interpret the results from that empirical exercise, based on Proposition 1 and our theory more generally.

Interpreting estimation output: surplus and sorting. Part (i) of Proposition 1 states that if the (conditional or unconditional) EE transition rate is found to depend on some job attribute y_j (e.g., in the case of τ if either $\gamma_j \neq 0$ or $\delta_{kj} \neq 0$ for some k), 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 dependence of the (conditional or unconditional) EE rate on that attribute (following from sufficiency).

Part (ii) of Proposition 1 allows us to draw conclusions about the relevance of worker attribute x_k for both surplus and sorting from the dependence of the EE transition rate τ on x_k (i.e. if either $\beta_k \neq 0$ or $\delta_{kj} \neq 0$ for some j). Surplus-relevance of x_k follows directly if τ depends x_k since this dependency indicates that the (SC) property holds for a triple $(y_i, y_j; x_k)$ involving x_k, which can only be true if x_k matters for the 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 $H_j(y|\mathbf{x}) := \int \mathbb{1}\{y_j \leq y\} \, h(\mathbf{y}|\mathbf{x}) \, d\mathbf{y}$, varies with worker skill x_k, i.e. if $\partial H_j(y|\mathbf{x})/\partial x_k \neq 0$ 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.[8]

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 attributes $(y_i, y_j; x_k)$ at $(\mathbf{x}, \mathbf{y})$, then there is sorting in dimensions (x_k, y_j).

b. If there is sorting in dimensions (x_k, y_j), then there exists a job attribute $i \in \{1, \cdots, Y_R\}$ and a point $\mathbf{y} \in \mathcal{Y}$, such that (SC) holds for attributes $(y_i, y_j; x_k)$ at $(\mathbf{x}, \mathbf{y})$ and $\sigma(\mathbf{x}, \mathbf{y}) > 0$.

The proof is in Appendix 8.1.3. Part a. 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, implying sorting in (x_k, y_j). This result extends our earlier results (Lindenlaub and Postel-Vinay, 2020) significantly: (SC) is sufficient to induce *some* (unspecified, positive or negative) form of sorting.[9] Moreover, point b. of Proposition 2 shows that (SC) at least locally is necessary for sorting.

[8] By contrast, 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 $\partial H_j(y|\mathbf{x})/\partial x_k$ is negative [positive] *for all* $y \in [\underline{y}_j, \overline{y}_j]$ and all $\mathbf{x} \in \mathcal{X}$.

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

Interpreting estimation output: single-index representation. The macro-labor literature almost invariably assumes scalar heterogeneity of workers and jobs. In the context of our model, that assumption is valid if the surplus function admits a *single-index assumption*, defined by the existence of three differentiable functions $\widetilde{\sigma} : \mathbb{R}^2 \to \mathbb{R}$, $I : \mathbb{R}_+^{X_R} \to \mathbb{R}_+$ and $J : \mathbb{R}_+^{Y_R} \to \mathbb{R}_+$ such that for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$, $\sigma(\mathbf{x}, \mathbf{y}) = \widetilde{\sigma}(I(\mathbf{x}), J(\mathbf{y}))$. In other words, match surplus $\sigma(\mathbf{x}, \mathbf{y})$ can be expressed as a (bivariate) function of two one-dimensional scalar indices, I and J, summarizing all of the relevant heterogeneity on the worker and on the job side, without losing any information.

It turns out that single crossing of the surplus function precludes a single index representation of multi-dimensional heterogeneity:

Proposition 3 (Single Crossing and Single Index Representation) *If (SC) holds for some attributes* $(y_i, y_j; x_k)$ *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 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 in some dimensions* (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 2b.). 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.

We can thus find out from the dependence 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 section shows how theory should guide the interpretation of 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. Propositions 1-3 establish links between something we would like to know but is unobserved (properties of the surplus function) and something which we can observe in the data (dependence of the EE transition rate on worker and job attributes). Figure 1 summarizes the various potential types of results that our method can produce based on data, and their implications.

Figure 1: Potential Output and Implications

$\frac{\partial \tau}{\partial y_j} \neq 0$	$\longrightarrow$			y_j is surplus-relevant
$\frac{\partial \tau}{\partial x_k} \neq 0$	$\longrightarrow$	(SC)	$\longrightarrow$	x_k is surplus-relevant x_k is sorting-relevant no single index representation
$\frac{\partial^2 \tau}{\partial x_k \partial y_j} \neq 0$	$\longrightarrow$			x_k and y_j are surplus-relevant

2.3 Discussion

Reliance on the EE transition rate. 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 and yields more information compared to those alternatives. The UE rate is:

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

As in the case of the EE rate, if the UE rate depends on x_k, then 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 x_k.[10] The UE rate therefore conveys 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 tell us 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.

We also deliberately chose not to base our empirical protocol on wages for several reasons. First, using wages would require making additional assumptions on the wage setting protocol (e.g. sequential auctions, wage posting, Nash bargaining, etc.) in order to allow interpretation of the dependence of wages on agents' attributes in the data. Relying solely on the information contained in the EE transition rate obviates the need for such assumptions. 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 dependence of wages on worker or job attributes is generally no indication of single-crossing of the surplus function (see Appendix 8.1.6 for 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 them as explanatory variables of the EE rate (1).

Assumptions. A natural question is whether our analysis still reveals surplus-relevant worker and job characteristics even if our model is misspecified in some dimensions. We now discuss some of the assumptions underlying our model and their importance for the empirical analysis.

First, we assume that EE transitions happen only when the worker-job surplus (which is a function of observable attributes) increases, and that 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 neither worker nor job type specific, which is the typical assumption in this literature. The EE transition rate can be

[10]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 $\frac{\partial \tau_0}{\partial x_k}(\mathbf{x}) \neq 0$.

modified to account for these shocks, denoted by λ_2:

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

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

Second, 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 the direction of higher surplus (as in our model). Also assume that the worker and job type spaces are such that there exists 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 there is more competition for higher-surplus jobs. Then, any two workers, who are of the same type and are currently in the same job, have the same current surplus and target their search to the same type of job. Therefore they have the same EE transition rates. By contrast, two workers who 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 framework, the dependence of the EE transition rate on worker and firm characteristics would reveal which attributes are surplus-relevant.

Third, we assume that the surplus function has certain regularity properties, namely that it is Morse and quasi-concave. As argued above, those assumptions are not very restrictive. 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 horizontal worker and job heterogeneity, our method can detect this. In this case, the EE transition rate τ would be found to depend on a single job attribute y_j and a single characteristic x_k. 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 rule out.

Fourth, we assume that the job loss rate δ is independent of worker or job attributes. Most of our results are unaffected if we relax this assumption as δ is part of the surplus σ, and there can be objects in σ other than the production function that depend on $(\mathbf{x}, \mathbf{y})$.[11]

Fifth, we assume 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 the surplus-relevant charac-

[11]Specifically, Proposition 1(i)a.,(ii) and (iii) as well as Proposition 3 are unaffected.

teristics. If it is violated, then λ_1 varies across workers and jobs, and the EE transition rate is:

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

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 try our method in Monte Carlo simulations. We want to check whether our procedure can detect surplus-relevant worker and job characteristics when we know the DGP (including which job and worker attributes are truly surplus-relevant). Moreover, we want to assess which of the many model selection tools available 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.[12] We assume σ is bilinear, $\sigma(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{Q}(\mathbf{y} - \mathbf{b})$, where $\mathbf{b}$ is a Y-dimensional vector that we interpret as the production technology the non-employed have access to, and where $\mathbf{Q} = \{q_{kj}\}_{\substack{1 \le k \le X \\ 1 \le j \le Y}}$ is an $X \times Y$ matrix capturing the complementarity structure between all job and worker characteristics. Note that the only nonzero entries in $\mathbf{Q}$ are those associated with surplus-relevant characteristics of workers and jobs, i.e. $q_{kj} = 0$ for all $k > X_R$ or $j > Y_R$.

We then need to choose parameterizations of the technology and distributions of workers

[12]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.

and job types (ℓ and γ) to run simulations. We simulate models with the following properties, which – based on iterating between implementing our method on simulated data and real data, see Appendix 8.2 – reflect important features of the data. First, our parameterized models are sparse, meaning that the worker and job attributes truly affecting 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 be large, $X = Y \in \{10, 20, 40\}$ (remember that the number of RHS variables in the EE regressions is $X + Y + XY$). Second, they feature sorting between the relevant worker and job attributes. Our simulations are thus based on models whose technology matrix $\mathbf{Q}$ causes sorting (i.e. satisfies at least one single crossing condition). We try different parameterizations of $\mathbf{Q}$, causing 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 spans 24 different models ({2D vs 3D relevant attributes} $\times$ {low vs. high correlation} $\times$ {strong vs intermediate vs weak sorting} $\times$ {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 Techniques

We try a number of possible model selection tools in order to determine the subset of observed worker and job attributes that are relevant predictors of the EE transition rate. Standard model selection methods range from the classical information criteria — the Akaike Information Criterion, AIC, (Akaike, 1973, 1974), and the Bayesian Information Criterion, BIC (Schwarz, 1978) — to the more recent Lasso (Tibshirani, 1996) and Robust Lasso (Belloni, Chen, Chernozhukov, and Hansen, 2012) techniques 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 and parsimony (using a penalty function that is increasing in the number of estimated parameters). The common approach to model selection is to choose a parameter vector $\boldsymbol{\theta} = \{\theta_j\}_{1 \leq j \leq m}$ that maximizes the penalized likelihood:

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

where $l_n(\boldsymbol{\theta})$ is the log-likelihood function, 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 the L_0-norm of the coefficient vector $\boldsymbol{\theta}$, which equals the number of nonzero elements in that vector, $\sum_{j=1}^{m} r_\lambda(\theta_j) = \lambda \sum_{j=1}^{m} \mathbb{1}\{\theta_j \neq 0)\}$, where $\lambda = 1$ for AIC and $\lambda = \log(n)/2$ for BIC. For the Lasso, the penalty function is the L_1-norm of $\boldsymbol{\theta}$, $\sum_{j=1}^{m} r_\lambda(\theta_j) = \frac{\lambda}{n} \sum_{j=1}^{m} |\theta_j|$. Contrary to the AIC and BIC, the regularization parameter of the Lasso is not pre-specified and needs to be optimally chosen, for instance using 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 heteroskedastic disturbances.

As noted by Belloni, Chen, Chernozhukov, and Hansen (2012), perfect model selection is unlikely to occur in practice. Which method is likely to perform best in our context is a priori unclear, mostly because, while the asymptotic properties of those techniques are well understood, their finite-sample behavior is not. We therefore use our simulations to compare the performance of those different methods. We focus our main analysis on the AIC, BIC and Robust Lasso (which in our case is the appropriate Lasso estimator due to non-Gaussian and heteroskedastic disturbances). 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.

3.3 Comparing the Performance of Different Model Selection Techniques

We compare the different model selection methods along three standard performance metrics for classification problems in machine learning: Accuracy, Recall, and Precision (see, e.g., Japkowicz and Shah, 2011). We construct an additional 'loss function', which combines the Recall and Precision statistics. Those statistics summarize how well a model selection method does 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). They are defined formally as follows:

$$\text{Accuracy} = 1 - \frac{\text{FP} + \text{FN}}{\text{total \# rhs variables}} \qquad \text{Recall} = 1 - \frac{\text{FN}}{\text{TP}}$$

$$\text{Precision} = 1 - \frac{\text{FP}}{\text{FP+TP}} \qquad \text{Loss} = (1 - \text{Recall})^2 + (1 - \text{Precision})^2$$

Accuracy equals (one minus) the fraction of RHS variables that are misclassified by the method. An advantage of this measure is that it simultaneously penalizes false positives and false negatives, thus measuring the 'overall' performance of the method. On the downside, it is not scale-invariant, and can be increased artificially by increasing the number of RHS variables.

Recall equals (one minus) the fraction of true predictors that are missed by the selection method. It measures how severe the problem of false negatives is. It is scale-invariant.

Precision indicates how severe the problem of false positives is. Precision 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). We thus combine Recall and Precision into the Loss function, our preferred summary measure of performance.[13]

3.4 Implementation

Proposition 1 suggests two ways of implementing our empirical protocol, either in one or in two steps. To assess which one performs better, we run both on our simulated data.

One-step procedure. Proposition 1(i)a,(ii) and (iii) suggests that we directly apply any of the discussed model selection techniques to the Linear Probability Model (1) in order to determine which x_k's and y_j's are surplus-relevant. This is done in a single step, meaning that the null model includes a constant while the full model includes *all* potentially relevant worker and job attributes and their interactions, so that the number of RHS variables is $m = X + Y + XY$.

Two-step procedure. Based on Proposition 1(i)b., which states that the *mean* EE rate conditional on job type, $\overline{\tau}(\mathbf{y})$, 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 mean EE rate $\overline{\tau}(\mathbf{y})$ by applying model selection techniques to the following Linear Probability Model:

$$\overline{\tau}(\mathbf{y}_i) = \alpha + \sum_{j=1}^{Y} \gamma_{1j} y_{ij} + \epsilon_i. \tag{2}$$

Given the vector of selected job attributes, which we denote by $\mathbf{y}_R = (y_1, ...y_{Y_R})$, we can then select worker attributes $\mathbf{x}$ using (1), where the pre-selected $\mathbf{y}_R$ now enter as fixed regressors

[13]Loss equals the squared Euclidean distance between perfect performance $(1, 1)$ and actual model performance in the (recall, precision) space.

that are no longer up for selection or penalized. The relevant model in this second step is:

$$\tau(\mathbf{x}_i, \mathbf{y}_{R,i}) = \alpha + \sum_{k=1}^{X} \beta_k x_{ik} + \sum_{j=1}^{Y_R} \gamma_{2j} y_{ij} + \sum_{k=1}^{X} \sum_{j=1}^{Y_R} \delta_{kj} x_{ik} y_{ij} + \epsilon_i \tag{3}$$

where we use notation γ_{1j} and γ_{2j} for the coefficient on the y_j's in the first and second stage.

We try both procedures in combination with each of the model selection methods discussed above (AIC, BIC, and Robust Lasso).

3.5 Results

As explained above, we base our Monte Carlo exercise on simulated models that are sparse, multi-dimensional, and generate sorting among the relevant worker and job attributes. We report the performance of the AIC, BIC and Robust Lasso in selecting the correct surplus-relevant worker and job attributes, under both the one-step and two-step procedure.[14] We split the models into two categories based on the number of potential worker and job characteristics, those with 'small m' ($X = Y = 10$) and those with 'large m' ($X = Y \in \{20, 40\}$).

Small m. We start with the performance of models that feature a relatively small number of potential worker and job attributes. We report results from the one- and two-step procedures in in Tables 1 and 2, respectively. Each table reports the performance measures introduced above. Our preferred summary measure of performance is the loss function: the smaller the loss, the better the performance (although we note that Accuracy and Loss generally agree with each other). We compute each measure for the full model, as well as broken down by variable category (worker attributes X, job attributes Y and interactions XY).

Comparing performance across techniques and implementation procedures (one-step or two-step) suggests that, while all methods perform reasonably well, BIC tends to outperform AIC and Robust Lasso. Moreover, while the one- and two-step methods 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), the two-step procedure is measurably better at detecting worker characteristics (performance in selecting the correct job characteristics and interaction terms is similar across the two approaches). We therefore choose the two-step BIC as our preferred model selection technique.

[14] To make the AIC and BIC computationally feasible, we implement the 'forward stepwise' procedure as opposed the 'best subset' selection. In the two-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.

Table 1: Performance of the AIC, BIC and Robust Lasso (one-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
Robust Lasso	**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

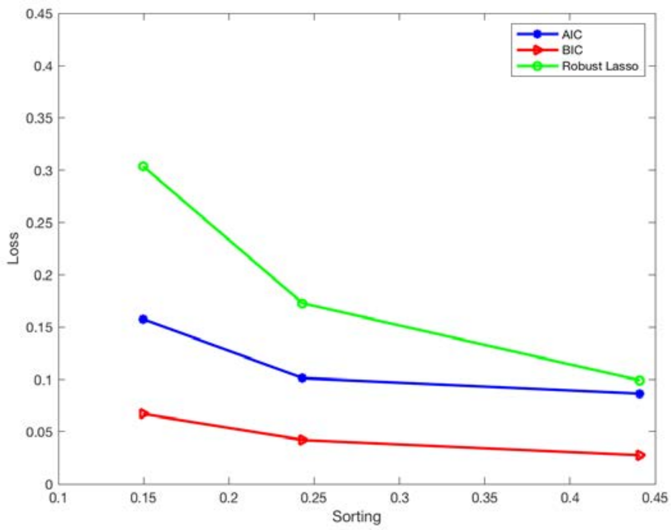
Table 2: Performance of the AIC, BIC and Robust Lasso (two-step)

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
Robust Lasso	**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

The statistics in Tables 1 and 2 are averaged across many simulated examples, concealing the fact that model selection improves significantly when the data features stronger sorting. This is especially true when it comes to detecting the relevant worker attributes. As mentioned, we simulated the model under different choices of technology **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) (measured 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 that is the case: Proposition 1 shows 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 thus makes it easier for our method to detect surplus-relevant worker attributes.[15]

[15] We further confirm that model selection performance deteriorates as sorting declines by running the tests on

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



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 worsens (see Appendix 8.2.2, Table 7). Similar to the case of small m, the BIC still outperforms other methods, although it comes out with a larger Loss value than in the small-m case (0.11 vs 0.05; Accuracy is similar in both cases). The increase in the Loss value is mainly due to difficulties predicting the correct worker attributes and worker-job interactions.

Our main takeaways from the Monte Carlo simulations are that (1) the procedure works well with relatively few potential job and worker attributes but performance weakens with large m and (2) the two-step implementation of BIC generally outperforms other methods. We will therefore apply our procedure to data with a limited number of potential worker and job attributes and implement it using the two-step BIC method.

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) provides us with information on workers, and the O*NET with information on jobs.

several models that feature zero sorting, available upon request.

SIPP. The SIPP is a nationally representative survey of the US population, which collects extensive data on labor market outcomes, including transitions from or into employment, as well as between jobs. The SIPP is administered in successive panels covering four to seven years (with a new sample of about 40,000 individuals being drawn for each panel). Data are collected in monthly interviews organized in waves and rotation groups. Within a SIPP panel, each rotation group (roughly a quarter of the sample) is interviewed every four months, so that the entire sample is interviewed over a 4-month period called a wave. During each interview, information is collected about the previous 4 months.

There are several advantages of the SIPP over other large labor force surveys in the US. First, the SIPP combines a reasonable cross-section sample size and a substantial longitudinal dimension, providing us with a large enough number of observed labor market transitions, which is important for our method to work. Second, transitions are recorded by date, in principle eliminating any time aggregation issues. Third, and crucially 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 4 of wave 2. We use this special module to impute individuals' multi-dimensional (potentially relevant) skill bundles. To our knowledge, this is the first attempt at imputing multi-dimensional skills in the SIPP (which we also implemented in Lindenlaub and Postel-Vinay, 2020 but there we focused on a considerably smaller sub-sample of college-educated individuals).

O*NET. The O*NET describes occupations in terms of skill and knowledge requirements, work practices, and work settings.[16] It comes as a list of hundreds of 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. The 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 construct the multi-

[16] 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`.

dimensional (potentially relevant) job attributes of the occupations that are present in the SIPP, and also as measures of the skills of individuals who are qualified for certain occupations (as per their degree or occupational training, see below for details).

4.2 Constructing the Variables of Interest

There are three types of variables whose construction requires additional explanations:

Multi-dimensional job attributes. The empirical counterpart of our model's 'jobs' is characterized by occupational and workplace characteristics. The datasets we combine (Abilities, Skills, Knowledge, Work Activities, and Work Context and SIPP) contain 131 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 (these form the 130 traits from O*NET) and workplace size (from SIPP).

Multi-dimensional worker attributes. To impute individuals' bundles of potential multi-dimensional skills, we primarily rely on the Education and Training Module in SIPP. It provides information on respondents' college, apprenticeship or vocational degrees, as well as occupational training on-the-job. We then make use of the fact that college, apprenticeship and vocational programs as well as occupational training programs qualify workers for particular occupations, where we use standard crosswalks to identify 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 them for that particular occupation. For instance, if an individual holds a college degree in 'economics', we assume that they are qualified for the occupation 'economist' and attach the skills required for that occupation to their skill bundle. If an individual is qualified for more than one occupation according to their training, we add the descriptors of multiple occupations to their skill bundle and average. This way we construct individuals' bundles of potentially relevant characteristics, containing 130 potential skill dimensions 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 imputation method of worker skills 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 administered). We are thus forced to exclude those individuals from our sample for which that information is missing. 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 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-month 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 8 in Appendix 8.3 reports summary statistics, comparing our baseline sample with the entire (representative) Education and Training Module. It shows that our sample is skewed 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 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

We use the Monte Carlo simulation results as guidance in implementing our empirical protocol on real data. Our baseline dataset contains 131 potential job attributes, a lot of which are highly correlated since many of the 130 O*NET descriptors capture similar features of a job. The same is true for the 130 potential worker traits in our data set. Our Monte Carlo study — and the model selection literature more generally — suggests that having large numbers of highly correlated potential explanatory variables will impair the performance of the procedure.

For those reasons, we transform the sets of worker and job attributes using principal component analysis (PCA) in order to bundle together sets of highly correlated descriptors before entering the model selection stage.[17] Specifically, we run a standard PCA in the distribution of jobs in the SIPP and keep the first ten principal components, which explain over 90% of the variance of the underlying data. Then, in order to obtain 'interpretable' job attributes, we follow Lise and Postel-Vinay (2020) and linearly recombine those ten principal components in such a way that they satisfy certain exclusion restrictions (see below Table 3 and see Lise and Postel-Vinay, 2020, for more details). This preliminary step reduces the number of potential job attributes to 11 (ten from O*NET plus firm size), down from our initial 131 measures, while capturing most of the underlying variation and ensuring interpretability of the job attributes. Since the underlying O*NET measures are also used in the construction of worker attributes, this preliminary step also reduces the number of potential such attributes to ten (down from 130).

We report our exclusion restrictions in Table 3. The first exclusion restriction imposes that the O*NET measure *Originality* only loads on job attribute y_1. We then interpret job attribute y_1 as reflecting non-routine cognitive skill requirements. The second restriction (measure *Performing Administrative Activities* only loads on y_2) means that we interpret y_2 as reflecting routine cognitive skill requirements. The third and fourth restrictions, *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*, reflect 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 Workweek* capture job (dis)amenities.

We select these particular exclusion restrictions for two reasons.[18] First, restrictions 1-6 closely resemble the broad categories that the applied labor literature has identified as important (e.g. Acemoglu and Autor, 2011):[19] (non)routine cognitive, (non)routine manual and several aspects of interpersonal skills. We then add restrictions 7-10 to capture a set of job (dis)amenities and workers' taste (or tolerance) for them, something which is usually overlooked when analyzing the worker-job surplus and sorting. Second, we select those particular exclusion restrictions to minimize pairwise correlations betweeen components, since we want

[17] Without any pre-selection, our EE regression would have as many as $130 + 131 + 131 \times 130 \approx 17,000$ RHS variables, with many of them highly correlated – a situation in which model selection methods do not perform well.

[18] We did extensive robustness checks varying the exclusion restrictions, and found that our results are not overly sensitive to them.

[19] 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.

each component to capture distinct information about jobs. Tables 10 and 11 in Appendix 8.3 report the resulting correlation matrices of potential worker and job attributes in their corresponding populations. Table 9 contains further summary statistics, comparing the skill and job attribute distributions in our baseline sample to the full Education and Training sample.[20] Recall that we added an indicator of workplace size to the potential job attributes (y_{11}), so that $X = 10,\ Y = 11$, implying $m = 10 + 11 + 10 \times 11 = 131$.

Table 3: Exclusion Restrictions in the PCA

Principal Component	Exclusion Restriction
y_1	Originality
y_2	Performing Administrative Activities
y_3	Far Vision
y_4	Trunk Strength
y_5	Selling or Influencing Others
y_6	Assisting and Caring for Others
y_7	Frequency of Conflict Situations
y_8	Responsibility for Outcomes and Results
y_9	Exposed to Hazardous Conditions
y_{10}	Duration of Typical Workweek

Equipped with those transformed (potential) job and worker covariates, we implement model selection using the two-step BIC procedure (implemented using the forward-stepwise approach), which is the one that performed best in our Monte Carlo study. 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 extracted from the 2008 SIPP panel.[21] This panel begins at the trough of the Great Recession and runs through the ensuing recovery. We also produce results based on earlier SIPP panels, for comparison. Table 4 shows the results: a checkmark indicates that a certain variable was selected in one of the two steps. Checkmarks in the last row (column) refer to selected non-interacted skill (job attribute) terms, while a checkmark in

[20] 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.

[21] Specifically, the earliest observations in our sample are from 2009 since the Education and Training module assesses individuals' education in 2009 (wave 2).

the table's body indicates that the interaction between the worker attribute (in the checkmark's column) and job attribute (on the checkmark's row) was selected. Our theory tells us that, based on the selected variables (which and how many), we can infer several important properties of the surplus function (see Propositions 1-3). We now discuss these implications.

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

	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).

The results tell us how many and which worker and job attributes are relevant surplus determinants. We find that a relatively sparse model gets selected. On the job side, skill requirements y_2 (routine cognitive) and y_5 (interpersonal — persuasion) are found to be predictors of EE mobility and thus enter the surplus (see Proposition 1(i)b. which refers to step 1 of our estimation approach). So do (dis)amenity y_{10} (work duration) and workplace size y_{11}. To get a sense of the occupations in which those attributes are important, consider the following examples: cognitive routine skill requirements (y_2) are high for human resource specialists, receptionists and information clerks, education administrators or accountants. 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 'high-skill' occupations such as anesthesiologists, chief executives, biochemists and biophysicists.

On the worker side, we find that many of the worker skills that are selected mirror selected job skill requirements (cognitive routine skill x_2, interpersonal skill x_5, and x_{10}, which we here interpret as a worker's tolerance for long working hours). Two additional 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 relevant determinants of the match surplus, but are also relevant for sorting: Our results imply that sorting does occur along at least some of the relevant dimensions of worker and job heterogeneity. As a further consequence, by Propositions 3, we can reject existence of a single-index representation of the selected multi-dimensional worker and job heterogeneity. Both heterogeneity and sorting are truly multi-dimensional in the data.

Because 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.

SIPP 1996, 2001 and 2004. We chose the 2008 panel of the SIPP as our baseline sample because earlier panels (especially 2001 and 2004) have fewer observations over shorter time horizons. Yet we did apply our procedure to those earlier panels. The results (reported in Appendix 8.4, Tables 12-14) are remarkably robust over time, especially as far as model sparsity and selected job attributes are concerned. In all four panels, interpersonal job attribute y_5, work duration y_{10} and workplace size y_{11} are found to be surplus-relevant. Cognitive requirement y_2 becomes surplus relevant over time, starting in the 2004 panel.

There is also a pattern in selected worker attributes: cognitive routine skill x_2, manual dexterity x_3 and tolerance for long working hours x_{10} are consistently selected over time. Also, interpersonal skill x_5 turns up significant in the two panels with the largest datasets (1996 and 2008).

We emphasize that the results presented in this section are merely illustrative of the ways in which our new methodology can be applied, as our sample is skewed toward high-skilled workers because either an educational degree or some occupational training experience is needed to impute individuals' multi-dimensional skills in the SIPP. We believe this is the main 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 requirements (x_1 or y_1) are never found to be surplus relevant. The likely reason is the relatively high correlation between those attributes with the 'duration of workweek'-variables (x_{10} and y_{10}), which thus capture a considerable share of non-routine cognitive skills and skill requirements.

6 Application: Labor Submarkets in the US Economy

Labor economists have long since recognized that the aggregate labor market is in fact composed of a number of heterogeneous submarkets, with workers (and employers) participating in either one or a small number of them. Properly defining these markets is crucial for understanding, e.g., which segment of the labor market is most affected during a recession or by structural change, or for measuring the (local) labor market power of firms. Delineating those submarkets, however, is a thorny empirical problem. By far the most common approach has been to use observable job and worker characteristics, such as education, occupation, or industry, to segment the labor market. While very easy to implement, this approach is somewhat arbitrary: education, occupation, or industry classification systems are designed by statistical agencies to organize and describe data, which does not necessarily coincide with the definition of economically meaningful labor markets. Based on that observation, a recent body of work (Schmutte, 2014; Nimzcik, 2020) applies community detection algorithms to define labor submarkets using observed patterns of labor market mobility and thus revealed preferences.

We use our results to build on that nascent literature but start from a different premise: Labor markets are segmented by *surplus-relevant* job and worker attributes as this is the heterogeneity determining surplus comparisons, job acceptance sets and thus worker mobility. Workers with similar relevant skills $\mathbf{x}_R$ *can* be employed in jobs that are similar in terms of relevant characteristics $\mathbf{y}_R$. We then define labor submarkets so that the *potential* — as opposed to *observed* — mobility of workers within markets is maximized. This is reasonable since even though two different workers with similar relevant skills may not have had the same job in the past, due for example to random frictions, they are capable of performing the same tasks, and interact on the same labor market.

6.1 Our Approach

We proceed in two steps.

Step 1. We start by pre-clustering workers and jobs into types, based on their relevant characteristics. This step makes direct use of our previous empirical results from Section 5.2.

Our empirical protocol detected 5 relevant worker skills $\mathbf{x}_R$ and 4 relevant job attributes $\mathbf{y}_R$, based on which our sample contains 2,135 different types of workers and 1,275 different types of jobs. We want to cluster these granular job and worker types into coarser categories, grouping jobs (or workers) with similar relevant attributes together. To that end, we pre-cluster workers and jobs separately using k-means.[22]

We choose k (the number of job and worker pre-clusters in k-means) based on the following rule of thumb. If we were able to discretize each worker or job attribute down to three possible values (say low, medium and high), then we would end up with $3^5 = 243$ worker types (recall there are 5 relevant worker attributes) and $3^4 = 81$ job types (there are 4 relevant job attributes). We choose k within those bounds. We start with $k = 200$ on the job side, which after internal optimization leads to $k = 152$ (since identical prototypes are combined by the algorithm), and we set $k = 152$ on the worker side for symmetry. The number $k = 152$ strikes a balance between significantly reducing the number of worker and job categories, avoiding extreme bunching of workers or jobs, and producing clusters of non-trivial sizes.[23]

Step 2. Given the (pre-clustered) worker and job types found in Step 1, we now proceed with delineating labor submarkets in our sample. To that end, we look for clusters of job types within which mobility of worker types is highly likely, while mobility across these job clusters is less likely. We consider two job types to be connected whenever they have a worker type in common, where here worker and job *types* refer to the pre-clustered types obtained in Step 1.[24] We then use *modularity maximization* (see Blondel, Guillaume, Lambiotte, and Lefebvre (2008), and Schmutte (2014) for an application) to find clusters of highly connected job types in the resulting mobility network (we describe the technical details and intermediate results of our two-step approach in Appendix 8.5.2).

6.1.1 Results

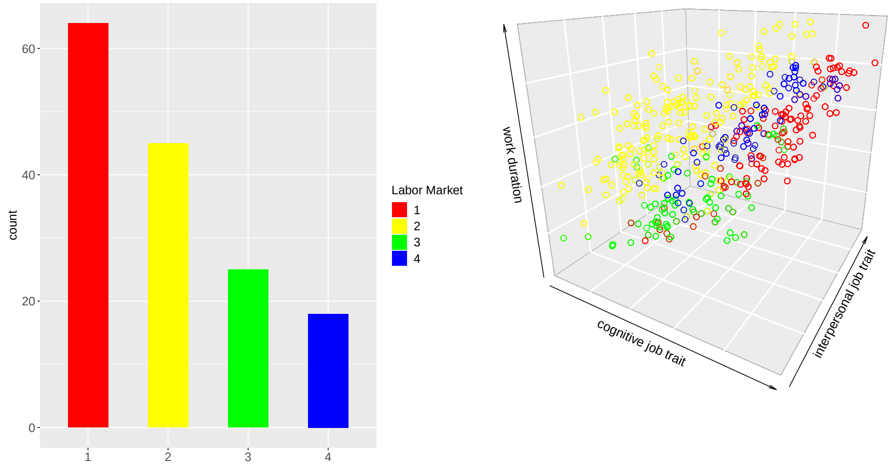
Our two-step algorithm detects *four labor markets.* This outcome is characterized by a strong community structure, with worker (type) mobility being substantially more likely within than

[22]More specifically, to pre-cluster the continuous worker attributes $\mathbf{x}_R$, k-means can be applied directly. Since $\mathbf{y}_R$ contains both continuous and categorical variables (firm size is categorical in our dataset), we use the k-prototype variant of k-means (Huang, 1998) instead. The objective function of k-prototype takes both continuous and categorical variables into account: It is the sum of the squared Euclidean distance (for continuous variables) and a similarity measure for categorical attributes, which is the number of mismatches between objects and cluster prototypes. We discuss the details in Appendix 8.5.1.

[23]We experimented with many different ways of pre-clustering: Dividing the marginal distribution of each skill into quintiles or deciles, and grouping workers together whose skills are in the same quintiles or deciles; dividing the marginal distribution of each skill into 5 or 10 equal-length intervals, and grouping workers together whose skills are in the same bins (and similarly for jobs). We also checked robustness within the k-means approach, trying $k = 100, 125, 150, 200$. All of these variants of Step 1 produced very similar numbers of final labor markets in Step 2. Specifically, we robustly obtain 4 or 5 labor markets in Step 2, showing insensitivity to how exactly we approach the pre-clustering problem.

[24]This contrasts with Schmutte (2014), where two *jobs* are connected in the mobility network whenever they share an *individual employee* in common.

Figure 3: Labor Markets based on Modularity Maximization



across markets (EE transitions are 3 to 6 times more likely to occur within than across markets). The histogram on the left of Figure 3 shows the number of job types in each market: while the four markets have unequal sizes, they are all "substantial". The histogram further introduces the color code for markets that we will adhere to for the rest of this paper.

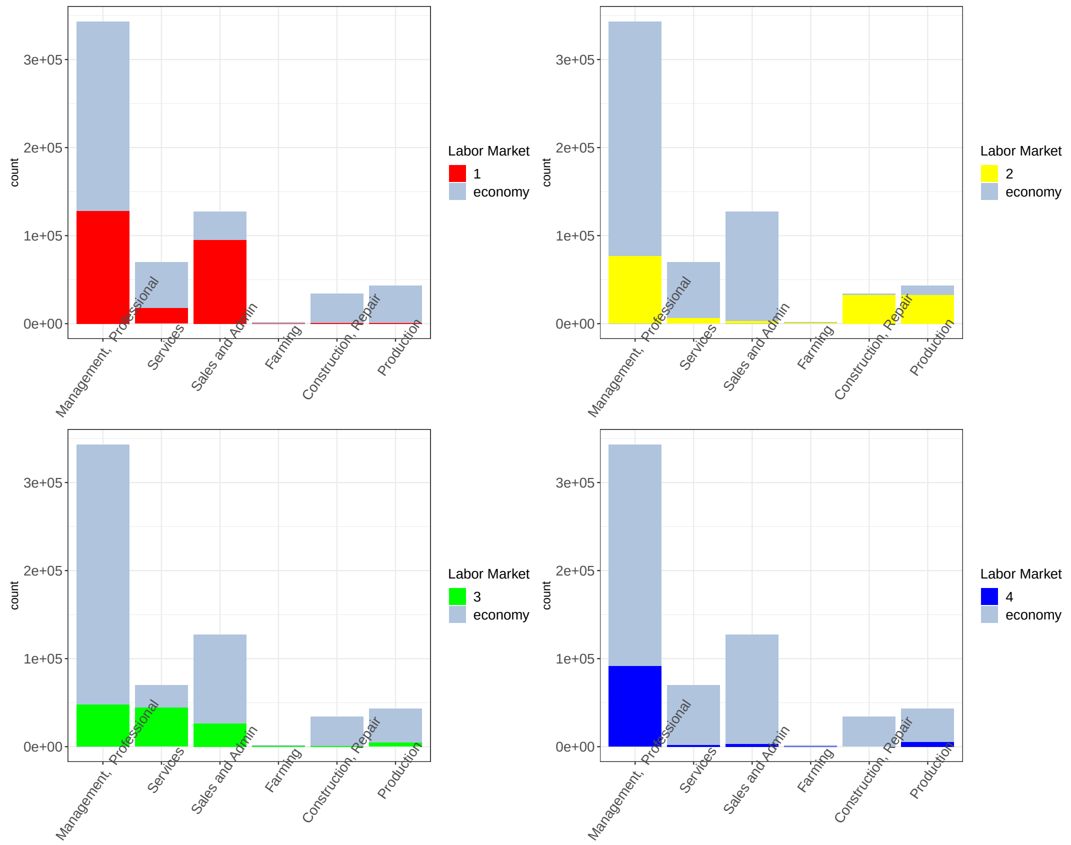
We then investigate the composition of those four markets in terms of relevant job attributes. The scatterplot on the right of Figure 3 shows how cognitive, interpersonal and hours requirements differ across the fours markets (leaving out, for plotting purposes, the workplace size indicator). The four clouds of job types clearly suggest substantial similarity in job attributes *within* markets and differences *across* markets. This is the case even though the modularity maximization clustering algorithm we used in the second step is not designed to group jobs by similarity of relevant attributes (rather, it maximizes within-cluster mobility of workers across jobs). Labor Market 1 (red) is relatively intensive in cognitive and interpersonal requirements, but relatively low on work duration. Labor Market 2 (yellow) has relatively low skill requirements across the board but relatively high hours requirements. Labor Market 3 (green) is similar in terms of cognitive and interpersonal requirements to Labor Market 2, but is also low in terms of hours. Finally, Labor Market 4 (blue) is similar to Labor Market 1 but more specialized in cognitive skill requirements and also requiring longer work hours.

6.2 Comparison to Alternative Ways of Defining Labor Markets

The literature has proposed various alternative ways of how to define labor markets or group firms: by occupation (Azar, Marinescu, and Steinbaum (2017)), by industry (Berger, Herkenhoff, and Mongey, 2019), by similarity of firms' wage distribution (Bonhomme, Lamadon, and Manresa, 2019), by observed mobility of *individual* workers across industry-occupations (Schmutte, 2014) or firms (Nimzcik, 2020). The main contribution of our application is to recognize that workers accept or reject job offers based on match surplus, which is determined by the surplus-relevant heterogeneity, which in turn should convey the information needed to map out the relevant labor market. We now show that our approach partitions the aggregate labor market quite differently from previous contributions.

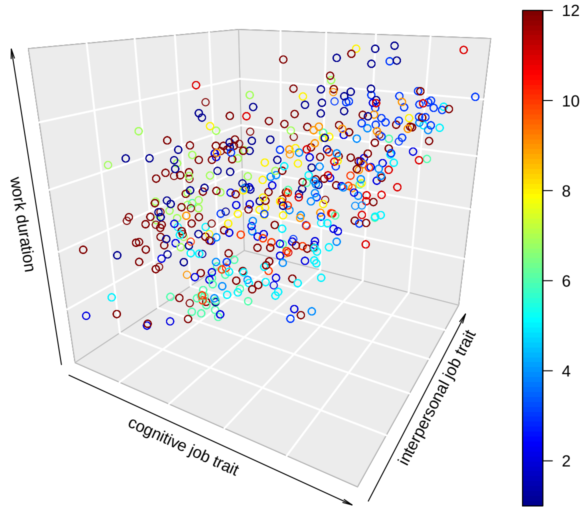
Figure 4 plots the distribution of broad occupations within each labor market, along with the economy-wide distribution for comparison. Clearly, our labor markets cannot be easily mapped to standard occupation groups. Figure 12, Appendix 8.5.4, shows a similar plot but for broad industries, suggesting that our markets cannot easily be mapped into industries either.

Figure 4: Occupation Distribution by Labor Market



Next, we try clustering labor markets based on observed individual worker mobility following Schmutte (2014). That is, we use modularity maximization on the observed mobility network, *without* pre-clustering individuals and jobs based on their relevant characteristics. In that approach, the nodes in the bipartite mobility network are the (roughly) 20,000 individual workers on one side, and the 460 occupations on the other. Modularity maximization generates 44 labor markets (the last 12 of which contain a single original occupation, and the last 24 contain only very few occupations). The large number of clusters may be due to our short panel, as the limited mobility of workers over a short time horizon causes fewer occupations to be connected through hiring the same individual. One disadvantage of this approach is that it fails to take into account that a worker's labor market should be defined by which jobs they are *willing to accept* rather than the jobs they have indeed held in the past. Our approach takes this *potential* mobility into account when defining labor markets. Figure 5 shows a scatterplot of job attributes across the markets obtained by direct modularity maximization, where we group all observations in markets above the 11th into the 12th one (mimicking Schmutte's combination of all clusters beyond those that account for around 80% of job nodes). In contrast to our approach, no community structure in terms of relevant attributes is visually evident.

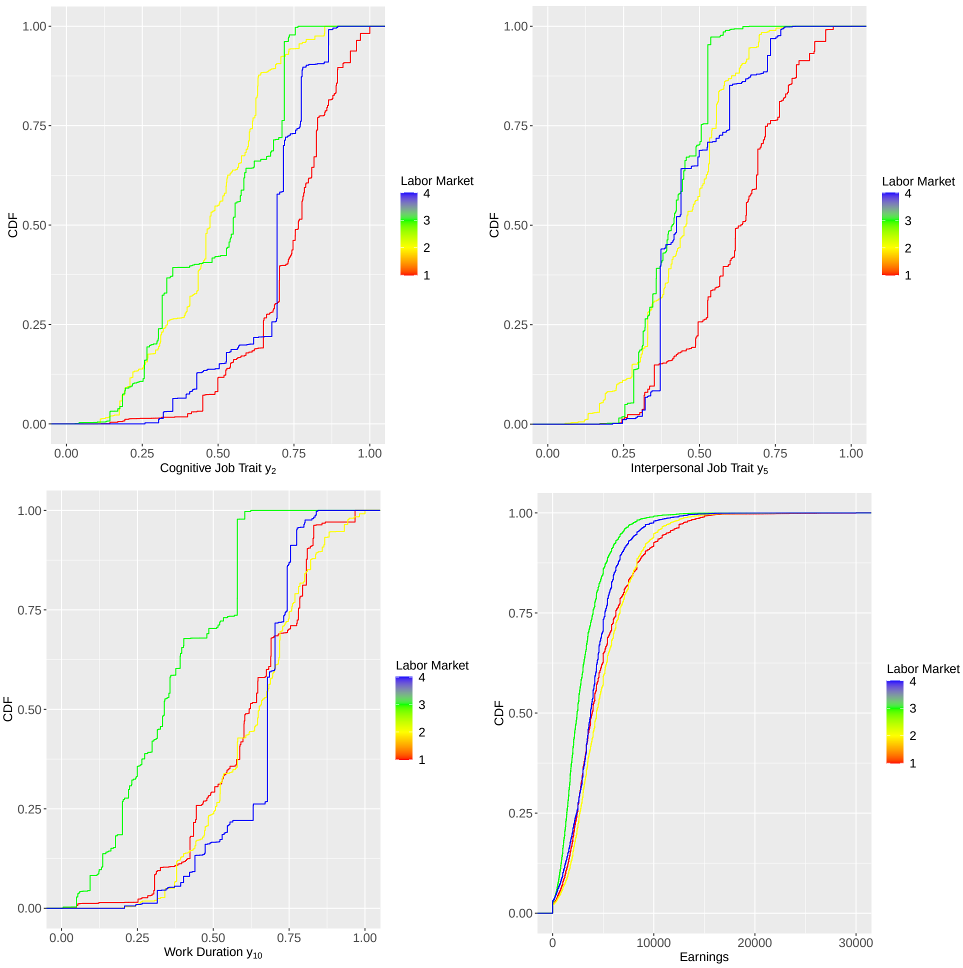
Figure 5: Relevant Job Attributes By Job Cluster, No Pre-Clustering



Finally, even though our submarkets are relatively homogeneous in terms of surplus-relevant job characteristics, they cannot simply be ranked according to any one attribute — or indeed in terms of wages. Figure 6 plots the marginal CDF of each relevant (continuous) job trait, and the CDF of wages, within each labor market, showing multiple crossings.

This section demonstrates that our approach to define labor markets is not nested by any of the existing ones. It has the unique feature that labor markets are groups of jobs with similar surplus-relevant traits that fit workers who share similar surplus-relevant skills. Importantly, our approach easily accommodates multi-dimensional worker and job heterogeneity, which cannot be collapsed into a scalar index that can be used to define labor markets.

Figure 6: CDFs of Relevant Job Attributes by Labor Market

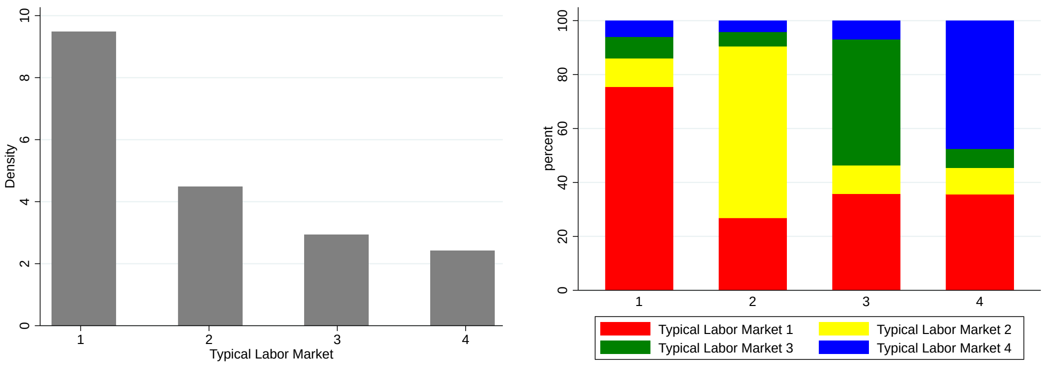


6.3 The Economic Content of the Estimated Labor Markets

We now investigate the economic contents of the labor submarkets identified in the previous section. We first show that there is substantial sorting in the sense that the 'typical' labor market differs across worker types. We then show that wages are higher and separations are lower when a worker is employed in the labor market that is typical of their type.

Typical labor markets and sorting. We start by defining a worker's 'typical' labor market as the modal labor market of their pre-clustered worker type. The idea is that each worker type (depending on the skill bundle) has a natural labor market to be employed in. Figure 7, left panel, shows the histogram of typical labor markets in our sample of workers. The right panel shows the fraction of employed workers in each of the four markets whose typical market is the first one (red), the second one (yellow), the third one (green) and the fourth one (blue). It suggests that there is substantial sorting, where most workers are employed in their typical market, even though the matching is not perfect, as expected in the presence of labor market frictions.

Figure 7: Typical Labor Markets and Sorting



Separations and earnings. We now show that our estimated labor markets have predictive power for separations and earnings. We are interested in whether 'well-matched' individuals — those employed in their typical labor market — have a lower probability of separation and higher earnings. Our empirical specification is given by:

$$z_{it} = \alpha_0 + \sum_{r=1}^{4} \alpha_{1,r} \mathbb{1}\{\text{Labor Market}_{it} = r\} \\ + \sum_{r=1}^{4} \alpha_{2,r} \mathbb{1}\{\text{Labor Market}_{it} = r\} \times \mathbb{1}\{\text{Typical Labor Market}_{it}\} + \kappa^T q_{it} + \epsilon_{it}$$

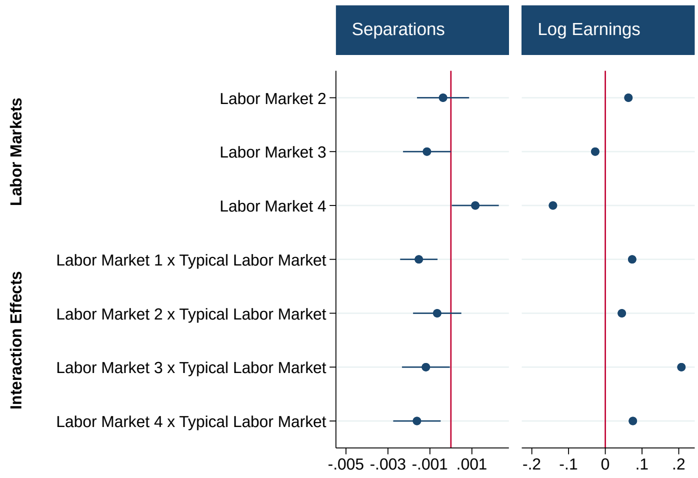
where z_{it} is the outcome of interest (a separation indicator or log-earnings of individual i at time t), $\mathbb{1}\{\text{Labor Market}_{it} = r\}$, $r \in \{1, 2, 3, 4\}$ is an indicator equal to one if individual i is employed in labor market r in period t, and $\mathbb{1}\{\text{Typical Labor Market}_{it}\}$ is an indicator equal to one if i works in their typical market in t. Further, q_{it} is a vector of additional controls such as the worker's surplus-relevant skills, the surplus-relevant attributes of the job they are employed in, their interactions, as well as some basic demographic traits.

Our main coefficients of interest are $\alpha_{2,r}$, measuring the effect of being employed in one's typical labor market on our outcomes. We show those, along with the main-effect coefficients $\alpha_{1,r}$, on Figure 8. There we see that being employed in the typical labor market is associated with a reduction in separation rate of 0.1-0.2 percentage points (roughly 10 to 20%), and with an increase in log earnings by 5 to 20% depending on the labor market.[25]

Job ladder heterogeneity. In a world with multi-dimensional heterogeneity, workers with different skill bundles rank jobs differently and climb different job ladders. In most models of the labor market, EE transitions yield (at least on average) individual earnings gains. We use

[25] More detailed results are reported in Table 16, Appendix 8.5.5.

Figure 8: Separations and Log Earnings: Effect of Employment in One's Typical Labor Market



earnings gains associated with an EE transition as a proxy for the gain in match quality and thus for the functionality of a certain ladder. And we assess how improvements in the various dimensions of surplus-relevant job attributes are 'rewarded' through wage gains on different submarkets. We are interested in whether those rewards are different for workers who move jobs within their typical market. Specifically, we run the following regression:

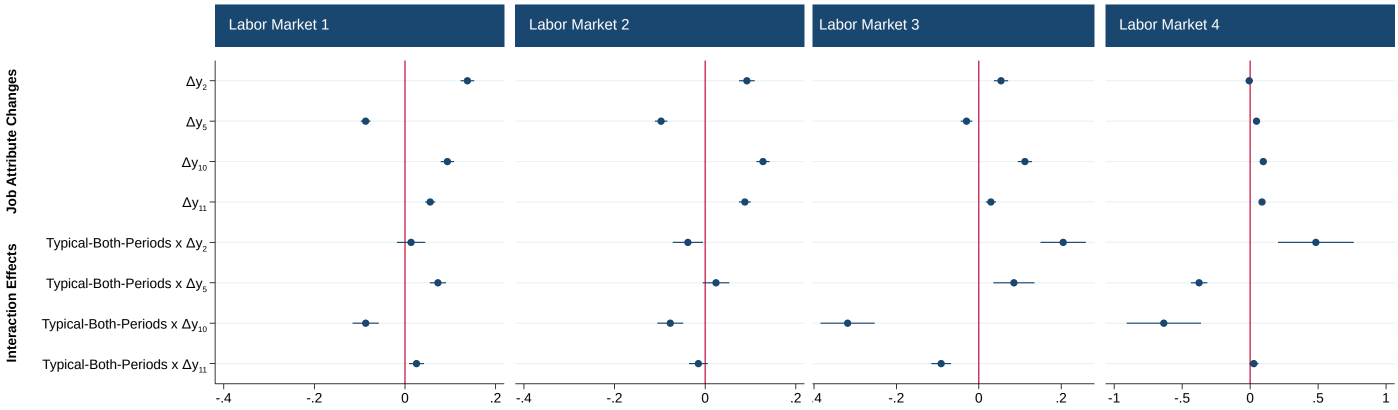
$$\Delta \log w_{it} = \beta_0 + \beta_1 \mathbb{1}\{\text{eetrans}_{it}\} + \beta_2 \mathbb{1}\{\text{eetrans}_{it}\} \times \mathbb{1}\{\text{Typical Both Periods}_{it}\}$$
$$+ \sum_{j \in \{2,5,10,11\}} \beta_{3,j} \Delta y_{j,it} + \sum_{j \in \{2,5,10,11\}} \beta_{4,j} \mathbb{1}\{\text{Typical Both Periods}_{it}\} \times \Delta y_{j,it} + \kappa^T q_{it} + \epsilon_{it} \quad (4)$$

where $\Delta \log w_{it}$ is log earnings growth of individual i between time $t-1$ and t, $\mathbb{1}\{\text{eetrans}_{it}\}$ is an indicator of individual i making an EE move between $t-1$ and t, $\mathbb{1}\{\text{Typical Both Periods}_{it}\}$ is an indicator of individual i being employed in their typical market in both periods $t-1$ and in t, and $\Delta y_{j,it}$ is the difference in surplus-relevant job attribute j between individual i's jobs in t and $t-1$ (note that $\Delta y_{j,it} = 0$ for all observations such that $\mathbb{1}\{\text{eetrans}_{it}\} = 0$). We also include other individual controls, such as surplus-relevant skills $\mathbf{x}_{R,it}$ and basic demographic covariates. Our main coefficients of interest are $\beta_{4,j}$, $j \in \{2, 5, 10, 11\}$, the effect of improvements in surplus-relevant job attributes on earnings growth when experiencing an EE transition within one's *typical* labor market. We implement (4) for each labor market $r = \{1, 2, 3, 4\}$ separately.

For each market, Figure 9 summarizes these coefficients of interest, $\beta_{4,j}$, $j \in \{2, 5, 10, 11\}$, in the bottom rows, along with the effects of improving job attributes when switching labor markets or making an EE transition in an atypical market, $\beta_{3,j}$, $j \in \{2, 5, 10, 11\}$, in the top rows.[26]

[26]Table 17, Appendix 8.5.6, reports the results for all four labor markets in detail.

Figure 9: Marginal Effects of Changes in Job Attributes



It reveals that climbing the job ladder (i.e. improving the various job attributes through EE moves) is beneficial in all labor markets, indicated by mostly positive estimates of $\beta_{3,j}$. Moreover, climbing the *typical* job ladder tends to be even more beneficial, indicated by several positive estimates of $\beta_{4,j}$, but there is significant heterogeneity as to which job attribute improvements are rewarded in the different markets. For example, an improvement in cognitive skill requirements, $\Delta y_{2,it} > 0$, is rewarded in all labor markets, but much more so in markets 3 and 4 for workers who consider these their typical markets. In turn, increases in firm size (often used as a proxy for firm productivity), $\Delta y_{11,it} > 0$, is generally rewarded and especially so in market 1 if that is workers' typical market. However, improvements in the interpersonal skill requirement, $\Delta y_{5,it} > 0$, are rewarded only in market 3 and only if that is workers' typical market, while they are penalized in the other markets. This analysis suggests that there is no unique job ladder in the economy, along which improvements in job requirements benefit all workers in the same way. Instead, different labor markets feature different job ladders, and workers with different skill bundles sort into different labor markets, climb different job ladders, and they get rewarded for climbing the ladder in their typical market.

7 Conclusion

The worker-job surplus is ubiquitous in models of (labor) markets with frictions. It is the driver of worker reallocation and wages, and it is the object used to measure the extent of sorting and mismatch. Yet, our understanding of the determinants of the surplus is limited. In this paper, we develop a theory-based empirical procedure to determine how many and which observable worker and job characteristics are relevant determinants of the worker-job surplus. To do so, we exploit workers' mobility choices. Our method also indicates whether these relevant attributes matter for sorting. As an important by-product, our method 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 job attributes).

We first test our methodology in Monte Carlo simulations before implementing it on real data from the SIPP and the O*NET. We find that a relatively sparse model underlies the data, where only a handful of job and worker characteristics matter for the match surplus. Furthermore, we reject the existence of a single index representation of these surplus-relevant, multi-dimensional worker and job attributes.

To highlight the economic relevance of our methodology, we use our results to develop a new way of delineating labor submarkets in the US economy. Our approach builds on the premise that the economy's labor markets are segmented by surplus-relevant worker and job attributes. We find that (our sample of) the US economy has four distinct labor markets. Workers who are employed in their 'typical' market enjoy lower separation rates, higher earnings and often higher wage gains when climbing that market's job ladder.

Another important application of our methodology is to provide structural analysis with guidelines on the types and dimensionality of agents' heterogeneity in models of the labor market. Mis-specifying surplus-relevant heterogeneity leads to sizable mistakes in the quantification of sorting and mismatch, as we show in Lindenlaub and Postel-Vinay (2020).

References

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

Acemoglu, D., and D. Autor (2011): *Skills, Tasks and Technologies: Implications for Employment and Earnings*vol. 4 of *Handbook of Labor Economics*, chap. 12, pp. 1043–1171. Elsevier.

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.

Azar, J., I. Marinescu, and M. I. Steinbaum (2017): "Labor Market Concentration," Working Paper 24147, National Bureau of Economic Research.

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.

Berger, D., K. Herkenhoff, and S. Mongey (2019): "Local Labor Market Power," Working Paper.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008): "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.

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.

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.

Huang, Z. (1997): “Clustering large data sets with mixed numeric and categorical values,” in *In The First Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 21–34.

——— (1998): “Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values,” *Data Mining and Knowledge Discovery*, 2(3), 283–304.

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 (2020): “Multidimensional Skills, Sorting, and Human Capital Accumulation,” *American Economic Review*, 110(8), 2328–76.

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

Nimzcik, J. (2020): “Job Mobility Networks and Data-driven Labor Job Mobility Networks and Data-Driven Labor Markets,” Working Paper.

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.

Schmutte, I. (2014): "Free to Move? A Network Analytic Approach for Learning the Limits to Job Mobility," *Labour Economics*, 29(C), 49–61.

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

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

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

Szepannek, G. (2018): "clustMixType: User-Friendly Clustering of Mixed-Type Data in R," *The R Journal*, 10(2), 200–208.

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.

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:[27]

[27] 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 (5) above.

$$h(\mathbf{x},\mathbf{y}) = \frac{\delta\lambda_0 \mathbf{1}\left\{\sigma(\mathbf{x},\mathbf{y}) \geq 0\right\}}{\delta + \lambda_0 \overline{F}_{\sigma|\mathbf{x}}(0)} \cdot \frac{\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(0)}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\right]^2} \cdot \ell(\mathbf{x})\gamma(\mathbf{y}) \tag{5}$$

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:[28]

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

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.

[28] 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:

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

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:

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

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:

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

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:

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

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 $\mathbf{y}$. □

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:

$$\begin{aligned}\frac{\partial \tau}{\partial x_k}(\mathbf{x},\mathbf{y}) &= \lambda_1 f_{\sigma|\mathbf{x}}\left(\sigma(\mathbf{x},\mathbf{y})\right)\left(\mathbb{E}_\Gamma\left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}') \mid \sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\right]-\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right)\\ &= \lambda_1\int\left[\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y}')-\frac{\partial \sigma}{\partial x_k}(\mathbf{x},\mathbf{y})\right]\mathbf{1}\{\sigma(\mathbf{x},\mathbf{y}')=\sigma(\mathbf{x},\mathbf{y})\}\gamma(\mathbf{y}')d\mathbf{y}' \end{aligned}\tag{6}$$

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.[29] 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:

$$\forall\, t\in[0,1],\quad \sum_{j=1}^{Y_R}\frac{\partial \sigma}{\partial y_j}(\mathbf{x},\mathbf{f}(t))\cdot f_j'(t)=0 \tag{7}$$

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:

$$\exists\, t^\star\in[0,1]:\quad \sum_{j=1}^{Y_R}\frac{\partial^2 \sigma}{\partial x_k \partial y_j}(\mathbf{x},\mathbf{f}(t^\star))\cdot f_j'(t^\star)\neq 0 \tag{8}$$

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 (7):

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

and substitute into (8):

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

[29] 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 $\partial^2 \tau / \partial x_k \partial y_j(\mathbf{x}, \mathbf{y}) = 0$. □

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:

$$\frac{\partial H_j}{\partial x_k}(y \mid \mathbf{x}) = \int_{\underline{y_j}}^{y} \int_{\mathbf{y}_{-j}} \frac{2\delta(\delta + \lambda_1) \mathbf{1}\{\sigma(\mathbf{x}, \mathbf{y}') \geq 0\}}{\left[\delta + \lambda_1 \overline{F}_{\sigma|\mathbf{x}}(\sigma(\mathbf{x}, \mathbf{y}')) \right]^3} \frac{\partial \tau}{\partial x_k}(\mathbf{x}, \mathbf{y}') \gamma(\mathbf{y}') d\mathbf{y}'_{-j} dy'_j \tag{9}$$

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 (9). The objective is to show that (9) is nonzero over a nontrivial set of y.

For ease of exposition, we first express (9) as

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

where

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

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

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

Because $\int_{\mathbf{y}_{-j}} g\left(\mathbf{x},\left[y,\mathbf{y}'_{-j}\right]\right)d\mathbf{y}'_{-j}>0$, $\partial\tau/\partial x_k\left(\mathbf{x},\left[y,\tilde{\mathbf{y}}_{-j}(y)\right]\right)$ must under the premise 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 (9), we first note that the function

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

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 $\partial\tau/\partial x_k\left(\mathbf{x},\left[y,\tilde{\mathbf{y}}_{-j}(y)\right]\right)=0$. 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

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

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 $(y_i,y_j;x_k)$ 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:

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

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 $\Delta\sigma\left(\Upsilon(\tilde{y}_j),\tilde{y}_j\right)=0$ for all $\tilde{y}_j\in\mathbb{I}$. Moreover,

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

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:

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

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. Therefore, it cannot be that $\partial H_j/\partial x_k(y \mid \mathbf{x}) = 0$ for all y. □

Necessity. If $\partial H_j/\partial x_k(y \mid \mathbf{x}) \neq 0$, the integrand in (9) 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 holding at some point $(\mathbf{x}, \tilde{\mathbf{y}})$ (such that $\sigma(\mathbf{x}, \tilde{\mathbf{y}}) = \sigma(\mathbf{x}, \mathbf{y}')$) is necessary for $\partial\tau(\mathbf{x}, \mathbf{y}')/\partial x_k$ to be nonzero. □

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 $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$, $\sigma(\mathbf{x}, \mathbf{y}) = \widetilde{\sigma}\,(I(\mathbf{x}), J(\mathbf{y}))$. Then, using the identity $\sigma(\mathbf{x}, \mathbf{y}) = \widetilde{\sigma}\,(I(\mathbf{x}), J(\mathbf{y}))$ we can form the ratio,

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

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. □

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.[30]

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

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

technology, where in the case of bilinear surplus: $p(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{Q}\mathbf{y}$ and $b(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{b}$.

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 $\mathbf{z}$:

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

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

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.

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.

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.*

$\mathbf{Q}$ is an $X \times Y$ non-zero matrix along the 2D surplus-relevant dimensions and has zeros in all other entries. The $\mathbf{Q}$ 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.*

$\mathbf{Q}$ is an $X \times Y$ non-zero matrix along the 2D surplus-relevant dimensions and has zeros in all other entries. The $\mathbf{Q}$ 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.*

$\mathbf{Q}$ is an $X \times Y$ non-zero matrix along the 3D surplus-relevant dimensions and has zeros in all other entries. The $\mathbf{Q}$ 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.*

$\mathbf{Q}$ is an $X \times Y$ non-zero matrix along the 3D surplus-relevant dimensions and has zeros in all other entries. The $\mathbf{Q}$ 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 $\mathbf{y}_R$.

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.

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.

Table 7: 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 (2020) 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 8: 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 9: 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 10: 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 11: 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 12: 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	.	.	$\checkmark$	.	.	.	.	.	.	$\checkmark$	$\checkmark$
y_4	.	.	.	.	.	.	.	.	.	.	$\checkmark$
y_5	.	.	.	.	.	.	.	.	.	$\checkmark$	$\checkmark$
y_6	.	.	.	.	.	.	.	.	.	.	.
y_7	.	.	.	.	.	.	.	.	.	.	.
y_8	.	.	.	.	.	.	.	.	.	.	.
y_9	.	.	.	.	.	.	.	.	.	.	.
y_{10}	.	.	.	.	.	.	.	.	.	.	$\checkmark$
y_{11}	.	.	$\checkmark$	.	.	.	.	.	.	$\checkmark$	$\checkmark$
Linear Terms	.	.	$\checkmark$	.	.	.	.	.	.	$\checkmark$	.

$\checkmark$ 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 13: 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	.	$\checkmark$	.	.	.	.	.	.	.	$\checkmark$	$\checkmark$
y_3	.	.	.	.	.	.	.	.	.	.	.
y_4	.	.	.	.	.	.	.	.	.	.	.
y_5	.	$\checkmark$	$\checkmark$	.	.	.	.	.	.	$\checkmark$	$\checkmark$
y_6	.	.	.	.	.	.	.	.	.	.	.
y_7	.	.	.	.	.	.	.	.	.	.	.
y_8	.	.	.	.	.	.	.	.	.	.	.
y_9	.	.	.	.	.	.	.	.	.	.	.
y_{10}	.	.	$\checkmark$	.	.	.	.	.	.	$\checkmark$	$\checkmark$
y_{11}	.	$\checkmark$	.	.	.	$\checkmark$	.	.	.	.	$\checkmark$
Linear Terms	.	$\checkmark$	$\checkmark$	.	.	$\checkmark$	.	.	.	$\checkmark$	.

$\checkmark$ 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 14: Model Selection Results Based on BIC 2 Step, Years 2004-2007

8.5 Application

8.5.1 K-Prototype Clustering

K-prototype clustering is the appropriate variant of k-means clustering if the variables are a mix of categorical and continuous variables. The objective function to be minimized is given by:

$$\sum_{i=1}^{n}\sum_{j=1}^{k} u_{ij} d\left(z_i, \mu_j\right)$$

where $z_i, i = 1, \ldots, n$ are the observations in the sample (in our case: the realizations of the surplus-relevant characteristics $\mathbf{y}_R$ of each filled job), $\mu_j, j = 1, \ldots, k$ are the cluster prototype observations, and u_{ij} are the elements of the binary partition matrix $U_{n\times k}$ satisfying $\sum_{j=1}^{k} u_{ij} = 1, \forall i$. And the distance function d is given by

$$d\left(z_i, \mu_j\right) = \sum_{m=1}^{q}\left(z_i^m - \mu_j^m\right)^2 + \lambda \sum_{m=q+1}^{p} \delta\left(z_i^m, \mu_j^m\right)$$

where m is an index over all variables in the data set, and where the first q variables are numeric and the remaining $p-q$ variables are categorical. Note that mismatch measure $\delta(a,b) = 0$ for $a = b$ and $\delta(a,b) = 1$ for $a \neq b$ and d corresponds to weighted sum of Euclidean distance between two points in the metric space and simple matching distance for categorical variables (i.e. the count of mismatches).

The categorical similarity measure enters with a positive weight λ in the objective function, which needs to be chosen by the analyst. We choose $\lambda = 0.25$, which is in the recommended range but we also experimented with several different values for λ without affecting the results in significant ways.[31]

For implementation, we use `R` package `clustMixType`, see here.

8.5.2 Details on 2-Step Clustering Algorithm

Step 1: Pre-Clustering Worker and Jobs into Types Based on Surplus-Relevant Traits

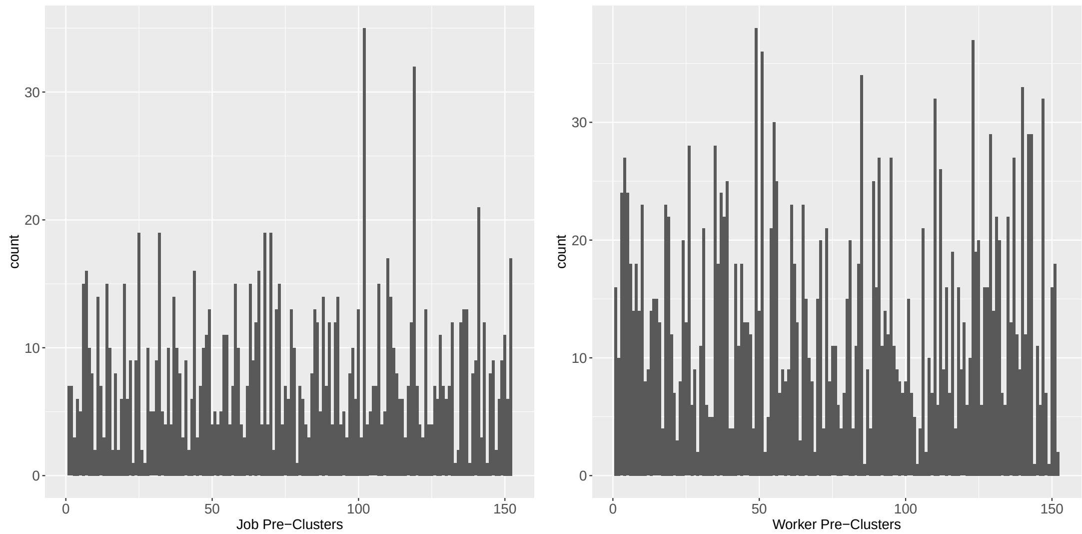
Jobs. We construct pre-clustered job types using the k-prototype algorithm with 25 randomly generated starting values, see Section 8.5.1. As discussed in the text, we choose $k = 200$ as our baseline, which yields $k = 152$ pre-clustered job types, as the `R` algorithm automatically combines any groups with equal prototypes in the process. The resulting ratio of between sum of squares over total sum of squares is 0.996. Figure 10, left panel, gives a histogram reporting the count of original job types (in the raw data, there are 1275 of them) that fall into each of the 152 pre-clustered job types.

Workers. We construct pre-clustered worker types using k-means clustering with 25 randomly generated starting values. We choose $k = 152$ for symmetry with pre-clustered job types. Figure 10, right panel,

[31] Huang (1997) – who invented the k-prototype clustering – suggests based on simulations for various applications to set $1/3\sigma \leq \lambda \leq 2/3\sigma$, where σ is the average standard deviation of the numeric variables. In turn, Szepannek (2018) suggests either $\lambda = \sigma/h$ or $\lambda = \sigma^2/h$ where $h = 1 - \sum p_c^2$ and p_c is the probability of certain category c of the categorical variable (so h measures the dispersion of the categorical variable). These considerations suggests that in our case $\lambda \in [0.06, 0.31]$.

gives a histogram reporting the count of original worker types (in the raw data, there are 2135 of them) that fall into each of the 152 pre-clustered worker types. The resulting ratio of between sum of squares over total sum of squares is 0.974.

Figure 10: Pre-Clustered Job and Worker Types



Step 2: Clustering Job Types into Labor Markets

We apply modularity maximization to find the economy's *labor markets*, such that worker type mobility within the estimated markets is likely relative to across-market mobility. We proceed as follows after grouping individuals into worker types and jobs into job types (as per Step 1), where now the relevant identifier is at the *type* level for both sides of the market:

1. Form the *bipartite adjacency matrix* of worker types and job types, indicating the observed mobility/matches over the years 2009-2013 in a single matrix. Denote this matrix by $\mathbf{M}$. The dimension of this matrix is 152×152 (worker type in rows and job types in columns), reflecting the number of pre-clustered job and worker types.

2. Create the *one-mode job projection network*: this is a network of *indirect* links of length 2 across jobs, a 152×152 matrix (because of 152 pre-clustered job types) $\mathbf{J} := \mathbf{M}^T\mathbf{M}$. An entry of this matrix is positive if the two job types under consideration are connected, i.e. have a worker *type* in common. The exact entry indicates how many indirect links there are between the two job types based on common worker types. A large entry means a strong connection between two job types, while a small entry means that the two jobs are loosely connected through worker type mobility.

3. We apply *modularity maximization* to the one-mode job projection matrix $\mathbf{J}$ (see, for instance, Schmutte (2014) for details on the algorithm). We obtain 4 job clusters (or labor markets), with modularity score is to ~ 0.1.

8.5.3 Results

Figure 11: Relevant Job Attributes by Labor Markets

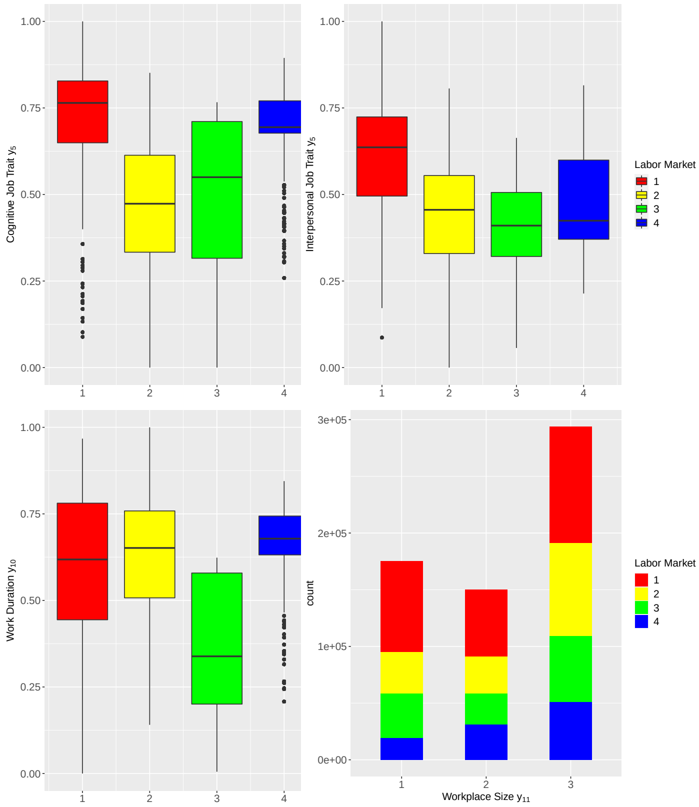


Table 15: EE Transition Matrix

	labor market 1	labor market 2	labor market 3	labor market 4
lag labor market 1	.63699156	.12125863	.15732924	.08442057
lag labor market 2	.196	.62971429	.104	.07028571
lag labor market 3	.23081201	.11067853	.5756396	.08286986
lag labor market 4	.2528626	.12977099	.13072519	.48664122

Examples of occupations that fall into each of the labor markets:

Labor Market 1: Management and Professional Services, Services, Sales and Admin. More specifically: Accountants, Data Entry Keyers, Sales Workers, Social Workers, Secretaries, Marketing and Sales Managers, Human Resource Officers, Lawyers, Real Estate Agents, Financial Service Workers; Hotel, Motel, and Resort Desk Clerks.

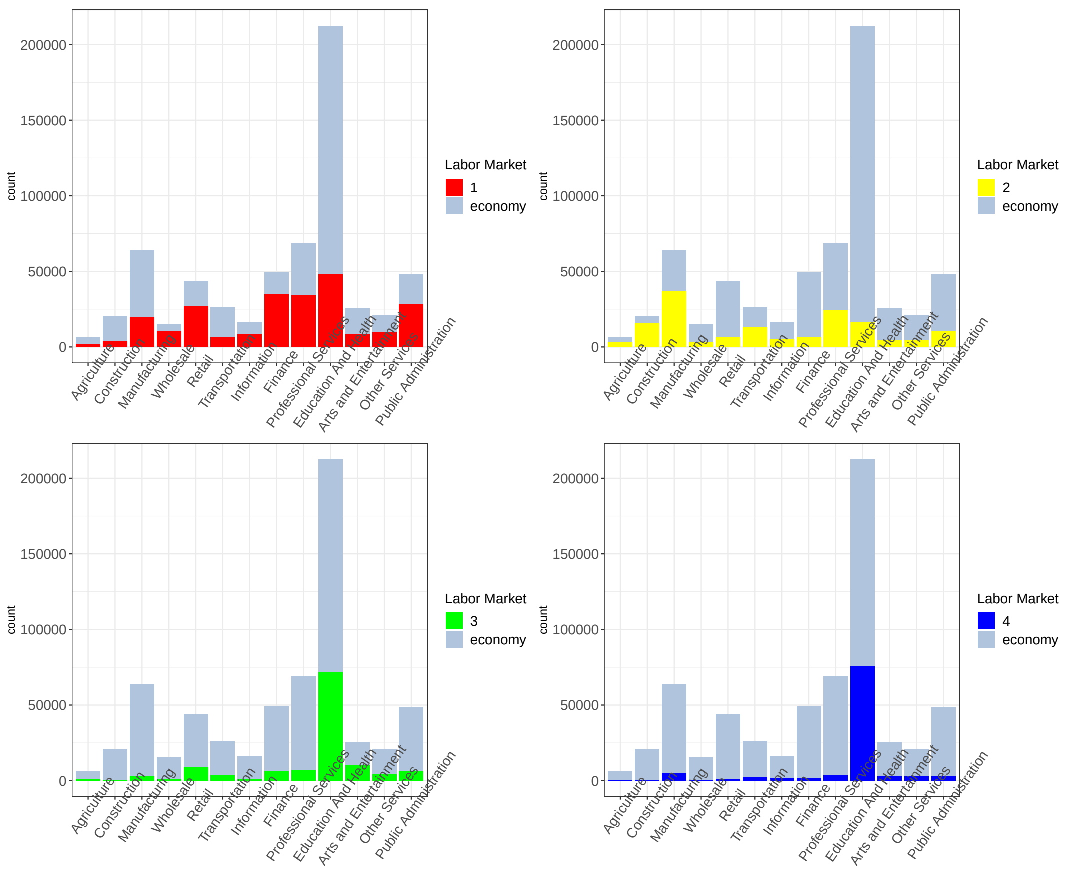
Labor Market 2: Management and Professional, Construction and Repair, Production. More specifically: Printing Machine Operators, Hazardous Materials Removal Workers, Radio and Telecommunications Equipment Installers and Repairers, Automotive Body and Related Repairers, Chemical Engineers, Sales Engineers, Helpers–Extraction Workers, Astronomers and Physicists, Producers and Directors, Dancers and Choreographers.

Labor Market 3: Management and Professional Services, but relatively more Services, Sales and Admin compared to Cluster 1. More specifically: Dental Assistants, Nurses, Counselors, Physical Therapist Assistants and Aides, Waiters and Waitresses, Maids and Housekeeping Cleaners, Personal and Home Care Aides, Medical Assistants and Other Healthcare Support Occupations, Child Care Workers, Library Technicians, Procurement Clerks Radiation Therapists.

Labor Market 4: Management and Professional Services. More specifically: Elementary and Middle School Teachers, Special Education Teachers, Postsecondary Teachers, Secondary School Teachers, Other Teachers and Instructors, Teacher Assistants, Architects, Education Administrators, Librarians, Aircraft Pilots and Flight Engineers, Computer Support Specialists, Shipping, Receiving, and Traffic Clerks, Writers and Authors, Psychologists, Desktop Publishers, Dietitians and Nutritionists, Economist.

8.5.4 Alternative Approaches of Defining Local Labor Markets

Figure 12: Industry Distribution by Labor Market



8.5.5 Separations and Earnings

Table 16: Job Ladders, Separations and Earnings

	Separations	Log Earnings
Labor Market 2	-0.000374	0.0630***
	(0.000633)	(0.00498)
Labor Market 3	-0.00114*	-0.0275***
	(0.000582)	(0.00448)
Labor Market 4	0.00116*	-0.142***
	(0.000571)	(0.00451)
Labor Market 1 × Typical Labor Market	-0.00153***	0.0732***
	(0.000454)	(0.00378)
Labor Market 2 × Typical Labor Market	-0.000654	0.0451***
	(0.000587)	(0.00470)
Labor Market 3 × Typical Labor Market	-0.00119*	0.207***
	(0.000581)	(0.00450)
Labor Market 4 × Typical Labor Market	-0.00162**	0.0751***
	(0.000577)	(0.00477)
y_2	-0.00246***	0.118***
	(0.000289)	(0.00203)
y_5	0.000811***	-0.00952***
	(0.000189)	(0.00140)
y_{10}	-0.000409	0.235***
	(0.000286)	(0.00204)
y_{11}	-0.00166***	0.146***
	(0.000137)	(0.00110)
$y_2 \times x_2$	0.00000312	0.0260***
	(0.000236)	(0.00160)
$y_2 \times x_3$	0.0000827	0.0170***
	(0.000226)	(0.00175)
$y_2 \times x_5$	0.000807**	-0.0194***
	(0.000284)	(0.00186)
$y_2 \times x_{10}$	-0.000726*	-0.00799**
	(0.000360)	(0.00244)
$y_5 \times x_2$	-0.000429	-0.000496
	(0.000260)	(0.00200)
$y_5 \times x_4$	-0.000354	0.0000348
	(0.000218)	(0.00174)
$y_5 \times x_5$	-0.000710***	0.0236***
	(0.000158)	(0.00120)
$y_{10} \times x_3$	-0.000674**	-0.0263***
	(0.000215)	(0.00165)
$y_{10} \times x_5$	-0.000248	-0.0241***
	(0.000225)	(0.00160)
$y_{10} \times x_{10}$	0.000808*	0.0270***
	(0.000315)	(0.00215)
Observations	611715	600343
R^2	0.002	0.301
Adjusted R^2	0.002	0.301
Year FE	X	X
Demographic Controls	X	X
Surplus-Relevant Skills	X	X

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. OLS regression with separation indicator as dependent variable (column 1) and log earnings as dependent variable (column 2), based on a pooled sample 2009-2013 of all employed workers in $t-1$ (column 1) and all employed workers in t (column 2). Time period t is one month. Separation is an indicator variable equal to 1 if the individual made a EU or EN transition between $t-1$ and t. Demographic controls include married, age, gender and black. We also control for the surplus relevant skills ($x_2, x_3, x_4, x_5, x_{10}$), omitted from the table. All surplus-relevant worker and job attributes are standardized for interpretability.

8.5.6 Earnings Growth

Table 17: Job Ladders and Earnings Growth

	Log Earnings Growth			
	Labor Market 1	Labor Market 2	Labor Market 3	Labor Market 4
eetrans	0.0194***	0.0730***	0.0746***	-0.00374
	(0.00713)	(0.00802)	(0.00741)	(0.00833)
eetrans × Typical-Both-Periods	-0.00436	-0.0610***	-0.0787***	-0.0269*
	(0.00926)	(0.0112)	(0.0120)	(0.0160)
$\Delta y_{2,t}$	0.138***	0.0920***	0.0539***	-0.00654
	(0.00770)	(0.00885)	(0.00881)	(0.0136)
$\Delta y_{5,t}$	-0.0869***	-0.0973***	-0.0298***	0.0471***
	(0.00550)	(0.00699)	(0.00718)	(0.00919)
$\Delta y_{10,t}$	0.0935***	0.128***	0.112***	0.0969***
	(0.00749)	(0.00743)	(0.00881)	(0.0123)
$\Delta y_{11,t}$	0.0556***	0.0875***	0.0294***	0.0878***
	(0.00563)	(0.00673)	(0.00642)	(0.00778)
$\Delta y_{2,t} \times$ Typical-Both-Periods	0.0134	-0.0382**	0.205***	0.484***
	(0.0159)	(0.0171)	(0.0281)	(0.142)
$\Delta y_{5,t} \times$ Typical-Both-Periods	0.0727***	0.0238	0.0852***	-0.375***
	(0.00909)	(0.0150)	(0.0255)	(0.0310)
$\Delta y_{10,t} \times$ Typical-Both-Periods	-0.0867***	-0.0770***	-0.318***	-0.635***
	(0.0148)	(0.0146)	(0.0335)	(0.140)
$\Delta y_{11,t} \times$ Typical-Both-Periods	0.0254***	-0.0150	-0.0914***	0.0273
	(0.00843)	(0.0106)	(0.0123)	(0.0174)
Typical-Both-Periods	-0.00107	0.000283	-0.000905	-0.000860
	(0.00116)	(0.00168)	(0.00148)	(0.00168)
Observations	232111	144804	117096	95716
R^2	0.010	0.010	0.007	0.008
Year FE	X	X	X	X
Demographic Controls	X	X	X	X
Surplus-Relevant Skills	X	X	X	X

Note: ***(**)(*) indicates significance at the 1%(5%)(10%) level. Standard errors in parentheses. OLS regression with log earnings differences between t and $t-1$ as dependent variable, based on a pooled sample 2009-2013 of all employed workers in t and $t-1$. Time period t is one month. eetrans is an indicator equal to 1 in t if individual made EE transition between t and $t-1$. $\Delta y_{j,t} = y_{j,t} - y_{j,t-1}, j \in \{2, 5, 10, 11\}$. Typical-Both-Periods is an indicator equal to 1 if individual was employed in typical job cluster and did not change job clusters between t and $t-1$. Demographic controls include married, age, gender and race. Surplus-relevant skills are $x_2, x_3, x_4, x_5, x_{10}$. Surplus-relevant worker and job attributes were standardized for interpretability (mean zero, standard deviation of one).