Human Capital Transferability and Employer Monopsony Power*

Qingyang Han†

April 11, 2024

Latest Version

Abstract

In this paper, I study the sources of employer monopsony power from the perspective of imperfect human capital transferability, defined as the portability of skills across occupations. Utilizing a task decomposition approach, I construct a measure of human capital transferability across occupations, and integrate this into a dynamic two-sided model of the labor market. Workers in this model make job-switch decisions over their life-cycle, cognizant of the depreciation in human capital value upon changing occupations. Imperfect transferability of skills gives firms some market power and allows them to reduce the wage, but it also makes it harder to replace skilled workers, reducing that power. Employers, in turn, post wage profiles that maximize their lifetime profits. I estimate the model using the Sample of Integrated Employer-Employee Data and Task Operationalization data from German Institute of Employment Research. I have three main findings. First, occupational switches are associated with significant wage penalties, suggesting that human capital gained in one occupation will be penalized in other occupations. Second, I find the life-cycle profile of wage markdown exhibits a U shape, where middle-aged workers suffer the smallest markdown. Third, I show that restoring perfect skill transferability lowers wage markdown for senior workers but increases that for younger workers. Further policy analyses show that a set of Active Labor Market Policies and education policies that feature both general education and vocational training have the potential to reduce labor market power.

Keywords: Human Capital, Mobility, Monopsony, Elasticity of Employer Specific Labor Supply

JEL Codes: C51, J23, J24, J31, J42, J62

*I am indebted to my advisors Robert Moffitt, Yingyao Hu, and Yujung Hwang for their guidance and support. I would also like to thank Manuel Arellano, Arie Beresteanu, Greg Duffee, Jonathan Elliott, Martin Friedrich, Matthew Gentry, Michael Keane, Lixiong Li, Joan Llull, Alexandros Loukas, Nick Papageorge, Stephen Ryan, Todd Sorensen, and Matthew Wiswall for their insightful feedback. The Institute of Employment Research (IAB) of Germany generously provided data that facilitate this research. I am grateful to Philip vom Berge for his explanations about the data and institutional background, Heiner Frank who hosted my on-site visit at IAB, and Ricardo Arcos who hosted my visit at the Bank of Spain (Madrid). I acknowledge helpful comments from seminar participants at the Johns Hopkins University, the 9th Konferenz für Sozial- und Wirtschaftsdaten (KSWD) at Berlin, the International Workshop on Establishment Panel Analyses at Nürnberg, the 2024 Eastern Economic Association Annual Conference, and the 2024 Midwest Economic Association Annual Conference. All errors are my own.

†Johns Hopkins University. Email: h.q@jhu.edu.

1 Introduction

"Monopsony" was initially introduced by Joan Robinson in her 1933 book, *The Economics of Imperfect Competition*. (Robinson (1933)) This term finds significant application in the labor market, where employers possess wage-setting powers and pay workers less than their marginal product of labor (MPL). In situations of imperfect competition, employers face an imperfectly elastic labor supply curve. When the wage rate falls below the MPL, a worker does not immediately switch to a better paying position unimpeded. In the literature, two primary models are invoked to explain the sources of wage markdown: search models and compensating wage differential models. In search models, search frictions hinder workers from transitioning to better-paying outside firms. On the other hand, the compensating wage differential model posits that workers accept wages below their MPL because they receive compensation in the form of non-pecuniary amenities.

In this paper, I propose a third factor that may affect employer monopsony power, namely, the imperfect transferability of human capital across occupations. Becker (1964) makes a distinction between general and specific job training, where the latter can only boost a worker's productivity at their current job, because some acquired skills are useful only at that job. I extend the Beckerian dichotomy by explicitly measuring the degree to which human capital can be transferred upon an occupation change. The degree of transferability depends on the similarity of tasks performed in the origin and the destination occupation. For instance, an experienced physician may find their skills not valued if they become an accountant. It would be different if they become a pharmacist, in which case most of their human capital can be transferred to the new job.

This paper is, to the best of my knowledge, the first to empirically quantify the effect of imperfect human capital transferability on employers' monopsony power. I develop a dynamic model where both workers and firms have knowledge of, and rational expectations for, imperfect transferability. Workers make job-switch decisions along their lifecycle subject to search friction, moving cost, and depreciation of human capital upon occupation change, taking into account the potential values associated with different firms and occupations. Employers post and commit to a lifetime wage profile so as to maximize their profit through hiring workers.

In this model, imperfect human capital transferability affects a worker's labor supply function specific to an employer through two opposite channels. On the worker side, a lower degree of transferability suggests that human capital will likely depreciate by a larger margin upon an occupation change, thus reducing the potential wage rates at outside occupations and impeding workers' desires to seek outside jobs, although the worker will make the switch if the growth of wages in the future is sufficiently large. Consequently, at a given wage policy, the worker is less likely to separate from their current employer, giving the employer some market power. However, on the employer side, a lower degree of transferability makes external hiring more difficult, because, depending on the skills of workers at other firms, retraining or new skill acquisition will be required. Current employees who possess unsubstitutable skills become more valuable to the employer, which pushes the employer to set a higher wage rate so as to retain current workers. Summing over the two channels, there is no prior regarding the sign of the net effect, which is the empirical objective of this paper.

The model is estimated using two datasets from the Institute of Employment Research (IAB) of Germany: the Sample of Integrated Employer Employee Data (SIEED), which is an extensive panel from Social Security databases that chronicles the employment histories of a large German employees sample matched to individual establishments; and an auxiliary dataset on task operationalization that facilitates the construction of a similarity measure in terms of task composition for a pair of occupations. In particular, linked firm side information is crucial to identify firm-level wage variations and thus firm-level labor supply elasticity. To rule out the confounding factors that affect labor market transitions in unexpected ways, this paper uses the cohort of male workers from the former West German states that entered the labor market between 1996 and 1999.

I estimate the effect of transferability on wage markdown in three steps. First, I estimate the effect of transferability on the wage profile for a worker in different occupations. The wage coefficients indicate how firms price worker experiences, where experiences are specific to occupation and translates into workers' learning-by-doing human capital. When workers transition between occupations, a defined transferability equation delineates the proportion of human capital that can be ported to the subsequent new occupation. The transferability equation and the wage equation are jointly estimated, giving information on the life-cycle wage trajectories across heterogeneous firms and occupations.

Second, I estimate a model of worker mobility conditional on the wage profile in different occupations in order to recover the elasticity of the job-specific labor supply function. In a given period, subject to stochastic arrivals of external offers, a worker can either choose to stay in their current job or switch to a new job. Switching to a new job positions the worker onto a new wage trajectory, albeit at the expense of incurring a moving cost. Further, when such a transition entails a change in occupation, the worker's human capital stock specific to the incumbent occupation undergoes a depreciation in accordance with the previously described transferability equation. This results in a wage loss, positioning the transitioning worker at a wage disadvantage relative to peers with comparable tenure in the destination occupation. The worker makes a sequence of job-switch choices through the career to maximize the discounted sum of lifetime utility.

Third, the elasticity of job-specific labor supply is derived from the estimated dynamic worker choice model.[1] From the point of firm profit maximization, the recovered elasticity exhibits a life-cycle profile. This shows how the job-specific labor supply for workers of varying ages responds to a one-percent increase in anticipated lifetime earnings.

The estimated model recovers the elasticity of individual job-specific labor supply function. Contrary to the literature, this study characterizes elasticity as the percentage increase in labor supply consequent to a one-percent increase in lifetime earnings, instead of in static wage rates. The estimated elasticity is small, with values ranging from 0.50 to 0.68 for male workers, contingent on worker characteristics. This finding suggests an employer-imposed wage markdown of approximately 139% to 198%, where the wage markdown is defined as the gap between the MPL and the wage as a percentage of the wage. These estimates, which fall into the general range of existing estimates of wage markdowns in the German labor market[2], are marginally smaller in magnitude, potentially attributed to workers' attenuated responsiveness to prospective wage growth. The life-cycle elasticity exhibits a hump shape, suggesting that workers tolerate a larger markdown imposed by their employers in the early stages of their career.[3] High school graduates also have higher elasticity than their peers without high school completion, due to their better choice set along the life-cycle.

The derived estimates allow multiple counterfactual exercises and policy analyses. The first counterfactual exercise removes imperfect human capital transferability, after which the overall elasticity of labor supply increases by 1.4% to 1.7%. This shows that, overall, imperfect human capital transferability depresses the elasticity of labor supply, thereby amplifying employer monopsony power. The effect of imperfect transferability is, however, heterogeneous across the life-cycle. Workers below the age of 30 experience a decrease in their labor supply elasticity under the counterfactual scenario, because they are willing to tolerate a larger markdown imposed by their employers, anticipating better career opportunities in the subsequent phases of their career. In contrast, for worker over 30, the assumption of perfect transferability yields benefits

[1]Throughout this paper, a job is defined as an (employer, occupation) pairing. When estimating the worker-side model, I use employer types (as detailed in Section 3.4) in lieu of exact firm identifiers. Consequently, the (employer, occupation) pairing is recast as an (employer type, occupation) pairing.

[2]See Bachmann et al. (2022).

[3]Agarwal (2015) uses the term "implicit tuition" to describe such phenomenon, where medical school graduates are willing to forgo a substantial amount of salary in exchange for a position at a renowned institution.

in the form of augmented elasticity. The second counterfactual exercise assumes a 100% job arrival rate in every given period, thus removing search frictions.[4] From a life-cycle perspective, the effect of this counterfactual exercise is similar to the first exercise. Younger workers become more inelastic in response to firm wage policies. In contrast, their older counterparts exhibit increased elasticity, triggered by a richer set of employment options.

These two counterfactual exercises simulate the potential effects of a set of Active Labor Market Policies (ALMPs) originally designed to increase labor market participation. This paper shows that such ALMPs have potential in curtailing employers' monopsony power. For example, promoting vocational training alleviates the impediments of skill non-transferability, thereby diminishing monopsony power. Government provision on job counseling and placement lowers the search friction and therefore also reduces monopsony power.

Finally, this paper compares lifetime welfare contingent upon a worker's apprenticeship training status, aiming to address the policy question on the relative merits of general education vis-à-vis vocational training. This paper finds that general education (e.g. high school) and apprenticeship training are complementary in augmenting lifetime earnings. Workers possessing apprenticeship training experience demonstrate an enhanced elasticity of labor supply, an indicator that firms place a premium on skills gained through apprenticeships and have larger incentives to retain such skilled workers.

1.1 Related Literature

The empirical literature studying monopsony base their estimations on the workhorse approaches in Manning (2003) and Burdett and Mortensen (1998), where the steady-state firm-specific labor supply elasticity equals the weighted average of separation elasticity and recruitment elasticity. There are two main approaches employed to estimate firm-specific labor supply elasticity, which translates into wage markdown and thus employer monopsony power. The first approach is to estimate a reduced form relationship between firm-specific labor supply (hirings minus separations) and wage rates, thus recovering firm specific labor supply elasticities.[5] The second approach is borrowed from empirical industrial organization, by building a discrete occupation choice model following Nevo (2001).[6] Taking insights from these two approaches, this paper

[4]In this secondary counterfactual scenario, workers are not granted full discretion to select from any available position in the market. Rather, the specific firm extending a poaching offer is randomly drawn from the entire labor market.

[5]For a comprehensive review on this strand of studies, see Sokolova and Sorensen (2021). In particular, Bachmann et al. (2022) estimates the elasticity based on labor market data from Germany, and Bassier et al. (2022) states that in estimating the elasticity, the variation of the wage variable should exclusively come from firms, instead of workers.

[6]See Ashenfelter et al. (2022) for an introduction of this approach. Examples include Berger et al. (2019) and Azar et al. (2022b).

augments the estimation by building a dynamic structural model, thereby capable of analyzing the life-cycle profiles of elasticity. The existing literature focuses exclusively on the static relationship between job-specific labor supply (hiring minus separation) and wage levels, without accounting for the labor supply responses to the growth of wage in the future. In contrast, this paper considers the forward-looking behaviors of both workers and employers, and therefore the estimated elasticity has life-cycle implications.

The second literature related to this paper studies the welfare consequences of labor market concentration.[7] These papers typically find that a higher degree of labor market concentration has negative effects on workers' wage profiles. This literature confirms the nature of non-competitive labor markets. From the perspective of worker welfare, this paper analyses the potential effect of a set of ALMPs and educational policies on employers' monopsony power.

The third related literature is on task specific human capital. Becker (1964) firstly makes a distinction between general human capital and firm-specific human capital, where the latter can only raise the worker's productivity at the current job. More recently, literature focuses more on industry-specific and occupation-specific human capital.[8] Occupation-specific human capital boils down to task-specific human capital, as discussed in Autor et al. (2003).[9] Yamaguchi (2012) applies Principal Component Analysis (PCA) and models an occupation as a bundle of cognitive and motor tasks based on the US occupation requirement data O*NET.[10] This paper takes the task-based approach while proposing a novel measure on human capital transferability, so that occupation-specific experiences are still one-dimensional, facilitating tractable computation on a dynamic job choice model.

The fourth literature is on occupational choice and occupational mobility.[11] Considering occupation or task specific human capital, there are a number of structural analyses on life-cycle job mobility and wage growth. A seminal work is Keane and Wolpin (1997). More recent analyses include Lee and Wolpin (2006), Johnson and Keane (2013), Roys and Taber (2019), Adda and Dustmann (2023). This paper is new to this literature by utilizing matched employer-employee data to extract firm-level variation in wage policies, which

[7]For example, see Azar et al. (2020), Azar et al. (2022a), Benmelech et al. (2022), Qiu and Sojourner (2023), Marinescu et al. (2021), Arnold (2020), Schubert et al. (2022), Prager and Schmitt (2021), Marinescu et al. (2021).

[8]For example, see Neal (1995) on industry-specific human capital, Kambourov and Manovskii (2009) and Sullivan (2010) on occupation-specific human capital.

[9]See Gibbons and Waldman (2004) and Gathmann and Schönberg (2010).

[10]Similar PCA approaches are also used in Guvenen et al. (2020) and Lise and Postel-Vinay (2020).

[11]Some well-known styled facts can be seen in Jovanovic and Moffitt (1990), which shows that employer-employee mismatch is a major source of sectoral mobility, and in Topel and Ward (1992), which finds that worker mobility gradually decline as workers gain more experience. Other facts on occupational mobility can be seen in Kambourov and Manovskii (2008), Groes et al. (2015), and Robinson (2018).

is key to identify firm-specific labor supply elasticity and wage markdown. This paper also takes into account potential search frictions in the labor market.[12] In modeling search frictions, this paper combines insights from Lise and Postel-Vinay (2020) and Roys and Taber (2019) by integrating frictional job arrival, random draws of employer productivity, and worker selection into different occupations.

The paper proceeds as follows. Section 2 describes the datasets used in this paper and some reduced form evidence. Section 3 presents the empirical model. Section 4 discusses the identification and estimation strategies. Section 5 presents the results. Section 6 shows the counterfactual exercises and policy analyses. Section 7 concludes.

2 Data Description

There are two primary data sources for this paper, namely, the Sample of Integrated Employer Employee Data (SIEED) and the Task Operationalization data of BERUFENET[13], both of which are provided by the German Institute of Employment Research (IAB)[14]. This section describes the two datasets, the sample selection process of this paper, and the descriptive evidence.

2.1 Sample of Integrated Employer Employee Data (SIEED)

The SIEED is a large scale administrative linked employer-employee dataset managed by the IAB. (Berge et al. (2020)) The worker side data are sample of the Integrated Employment Biographies, covering all employment subject to social security. Each line of data is an employment spell, where the identity of and the information about the employer of the individual worker is merged from the Establishment History Panel (BHP). The sampling procedure of the SIEED data is described in the following three steps.

1. A 1.5% sample of all establishment is drawn as of 2009.

2. All persons are drawn who worked for at least one day in one of these selected establishments between 1975 and 2018.

3. The complete employment histories are drawn for these selected people, even if they previously worked for firms that are not drawn in Step 1.

[12]Papers featuring search frictions in the labor market includes Cahuc et al. (2006), Dey and Flinn (2005), Flinn (2006), Lise et al. (2016), Lise and Robin (2017), Aizawa and Fang (2020), and Lise and Postel-Vinay (2020).

[13]https://web.arbeitsagentur.de/berufenet.

[14]Institut für Arbeitsmarkt- und Berufsforschung (https://fdz.iab.de).

On the individual side, I have information about workers' demographics, including age, gender, vocational training, residence locations, etc. On the firm side, I have information regarding firm location, firm size, occupation structures, wage structures, etc. The entire data covers the employment histories of 5,567,883 individuals in 5,248,850 establishments, and includes 172,935,233 lines of observation from 1975 to 2018.

Since the socialist German Democratic Republic (East Germany) joined the Federal Republic of Germany (West Germany) in 1990, data from the five East German states as well as East Berlin become available as well. However, in the analyses of this paper, data from the former East German states are not used for two reasons. First, there are large differences in economic conditions in East and West Germany following the reunification. If the analyses are based on a combined dataset, the economic differences may affect the estimation in unexpected ways. Second, for East German individuals that are included in the data after 1990, I do not observe their entire career histories as their jobs before 1991 are not documented.

This study uses a specified subsample of workers for its analyses:[15] male workers from the former West German states who entered the labor market between 1996 and 1999, aged 20 to 25. Subsequently, the exhaustive career records of this select cohort, ranging from 1996 to 2018, are incorporated. The rationale underpinning this sampling selection is twofold:

1. The immediate aftermath of Germany's 1990 reunification bore witness to economic and societal upheavals. Such transitions may have invariably influenced both worker and employer behavior in manners not accurately reflected by the model explored in this paper. Consequently, cohorts entering the labor market directly post-reunification are not used for analyses.

2. Concentrating the analyses on a cohort homogeneous in age facilitates the exclusion of potential distortions stemming from confounding macroeconomic determinants. Such variables might differently impact wage trajectories and mobility choices across different worker cohorts.

This study narrows its focus to job-to-job transitions, thus any recorded job spells categorized as non-full-time employment have been excluded. By the end of 2018, individuals in the chosen sample are all aged 47 or younger.[16] As such, the sample primarily encompasses the career trajectories of younger workers. Notably, within this age bracket, the incidence of non-full-time employment remains relatively low for male

[15] In Section 4, when estimating firm type assignment, the comprehensive dataset that encompasses individual career trajectories post-1991 is deployed. However, for other analyses, solely the cohort spanning 1996-1999 is employed.

[16] Wei (2022) studies the pattern that older workers take "bridge" jobs before getting fully retired. These "bridge" jobs are less demanding yet pay lower wages. The model studied in this paper does not capture the mechanisms under which older workers sort into "bridge" jobs, so only including younger workers keeps the estimation accurate.

workers. The exclusion of non-full-time job spells is, therefore, unlikely to introduce significant selection bias.

Appendix A shows the details about the definition of each variable used in this paper, and the sample construction procedures.

2.2 Task Operationalization

The task operationalization dataset is compiled by IAB based on an expert database BERUFENET. (Dengler et al. (2014)) The dataset measures the composition of tasks for different occupational classifications.

BERUFENET is an "online information portal provided by the German Federal Employment Agency for all occupations known in Germany which are mainly used in career guidance and job placement." The database contains a rich set of information regarding the required tasks in an occupational activity, required qualifications and legal regulations. From there a requirements-matrix is obtained for all occupations.

Then, the exact requirements are converted to five task types, i.e. (1) Analytical non-routine tasks, (2) interactive non-routine tasks, (3) Cognitive routine tasks, (4) manual routine tasks, and (5) manual non-routine tasks. Table 1 shows the descriptions corresponding to each of the five task dimensions. The data give a fractional weight for each of these five requirements measuring their relative intensity. For each occupation m, the task composition $s^m = (s_1^m, s_2^m, s_3^m, s_4^m, s_5^m)$ is therefore defined as the relative intensity of requirements in these five task types. For each occupation m, $\sum_{k=1}^{5} s_k^m = 1$.

Table 2 provides several examples of the task composition.

The 5-digit occupation code in the SIEED data has a hierarchical structure. The first digit identifies 9 occupation categories, while the first two digits can identify 36 occupations. I use the 2-digit code as the base classification for occupations, and then use the k-means algorithm to categorize these 36 occupations into 5 occupation-clusters based on similarity in terms of task composition. Dodini et al. (2020) uses similar approaches to cluster occupations based on task composition. Occupations that belong to the same occupation-cluster are similar in terms of their task composition. Compared to the 2-digit occupation codes, the definitions for occupations-clusters reduce the state space and accelerates the computation.

Figure 1 shows the clustering results. Each bar shows the average task compositions of an occupation cluster which includes several two-digit occupations. Occupation cluster 1 demands multiple task components relatively equally, whereas cluster 2 through 5 mainly require only one major tasks.

While occupations belonging to the same 1-digit occupation category may differ largely in terms of task composition, the occupation-clusters defined in this paper is obtained in a data-driven approach that better approximates the task requirements of each occupation. Throughout the paper I use "occupation" to refer to an occupation-cluster, where the notation $m \in \{1, \cdots, 5\}$ refers to an occupation cluster.

2.3 Task Distance

Using the task composition dataset, I construct a task distance measure between occupations m and m'.

$$d(m,m') = \sum_{k=1}^{5} (s_k^m - s_k^{m'})^2 \tag{1}$$

The task distance measures the degree of dissimilarity between two occupations; a larger $d(m,m')$ means that the two occupations m and m' require a more different set of tasks. I describe the relationship between this one-dimensional distance measure and higher-dimensional distances based on individual skills in Section 3.2 below.

2.4 Descriptive Results

Table 3 shows the descriptive statistics of the selected SIEED sample. Over 149 thousand distinct male workers are included in the selected sample, and their career paths cover over 400 thousand establishments. 34% of men have high school qualifications. Only 8% of all job spells are recorded as "non-full-time". Workers are distributed relatively evenly across the five occupations.

Figure 2 shows the wage-age profiles and the mobility-age profiles specific to high school status. On the wage profile, attending high school lowers the initial wage rates, but significantly boosts future wage growth rates: the lifetime earnings of high school graduates are notably larger than their peers without high school degree.

On the mobility profile, for all groups of workers, the probability that they stay at their current job increases as they get older, consistent with the findings of Topel and Ward (1992). High school graduates are less likely to separate from their current employers when they are above 40 years old.

2.5 Institutional Background

In the German labor market, Collective Bargaining Agreements (CBAs) play a pivotal role. As outlined by Bosch et al. (2021), these agreements predominantly occur at the sectoral level, such as trade unions and employers' organizations. The negotiations establish sector-specific minimum wages, as agreed upon in the respective CBA. It is worth noting that Germany introduced a national minimum wage in 2015, which, for the most part, is set below many pre-existing sectoral minimum wages. Consequently, this national minimum wage exerts limited influence on the majority of employers, given that their wage rates are already set above the national threshold, dictated by the higher sectoral minimum wages.

Despite the prominence of CBAs, this paper refrains from modeling collective bargaining, given three primary considerations:

1. **Employer Discretion:** Even in the presence of CBAs and the resultant sectoral minimum wages, individual employers retain a considerable discretion in determining wage growth rates.

2. **Declining CBA coverage:** Drawing from Ellguth and Kohaut (2022), there has been a consistent decline in collective bargaining coverage in Germany since 1996. The coverage rate has diminished, cascading from 66% in 1996 to a mere 40% by 2021.

3. **Data Limitations on CBA Coverage:** The firm-level dataset from SIEED does not tell the set of workers whose wage rates are subject to CBAs. This data limitation is from the nuanced within-firm variations concerning CBA affiliations.

Germany's system of collective bargaining has historically been characterized by a dual structure. While sectoral agreements, as mentioned, dominate the landscape, firm-level agreements (often augmenting the sectoral agreements) also exist. These firm-level agreements allow specific conditions tailored to individual firms, further increasing the flexibility employers may have in setting wage rates.

Exploring the effect of collective bargaining on the extent of firms' monopsony power is an interesting direction for future research on this topic.

3 Model

3.1 Overview

Each individual worker w starts their career at a specific occupation. The initial occupation and employer type[17] are endowments of each worker, reflecting unobserved premarket worker heterogeneity in skills,

[17] Types of employers will be discussed in Section 3.4. Heuristically, firms belonging to the same type share the same wage posting policy. From the perspective of the search literature, workers who are otherwise similar may get matched with a high-wage

preferences, and matching outcomes.

Individual workers gain learning-by-doing human capital through experiences accumulated while working at their job. The stock of a worker's learning-by-doing human capital determines the MPL they generate to the firm, fixing employer types. Experience and thus human capital are both occupation-specific. The MPL of a worker at a period is co-determined by the worker's human capital stock, the exact occupation, and the types of their employer. Knowing the lifetime profile of a worker's MPL, the employer posts and commits to a lifetime wage profile for the worker. The lifetime wage profile is an increasing function that maps a worker's occupation-specific experience to a wage rate, at each period.

In each period along a worker's life-cycle, an external employer may poach the worker under certain probability. The type of the poaching employer is sampled from the fixed population distribution of all firm types. Each poaching firm offers all available occupations, and the worker can decide to move to any occupation associated with the poaching employer, or to stay at their current job. If they decide to accept the new job offer at a different occupation, their existing human capital stock will be discounted by a degree that is commensurate with the dissimilarity between the new and the original occupation. As a result, they start off the new job at a lower MPL, thus suffering a wage loss, comparing to workers that work solely at the destination job for an equal amount of time. The worker trades off the lifetime earnings, the potential wage loss associated with occupation change, and the moving cost incurred from any job change, before making a moving-staying decision. The worker will move if future wage growth at the poaching job is sufficiently large.

The solution to the worker side dynamic occupation choice model recovers a labor supply function specific to a firm with jobs in a particular occupation, showing how many workers supply their labor to the firm for each wage the firm offers. This function can be used to map lifetime wage profiles to the number of newly recruited workers minus that of separated workers. Employers consequently set an optimal lifetime wage profile that maximizes their profit. The firm-side model rationalizes the estimated wage coefficient and allows the computation of worker elasticity-age profiles.

The following subsections describe the details of the model. The timing subscript t for workers denote the t^{th} year since the worker's labor market entry, instead of worker age.

firm or low-wage firm due to pure luck. In this paper I do not model the initial matching between employers and employees, but rather leave it as an endowment.

3.2 Evolution of Occupation Relevant Experiences

Let $\widetilde{x}^{m}_{wt}$ denote the occupation-m specific experience of worker w at period t. The experience variable determines the worker's human capital, which matters for the workers' productivity.

A worker w starts their career with $\widetilde{x}^{m}_{w0} = 0$. If the worker does not change occupation in period t, their experience will increase by 1 from last period, following Equation 2.

$$\widetilde{x}^{m}_{wt} = \widetilde{x}^{m}_{wt-1} + 1 \tag{2}$$

I construct a value for the relevant experience a worker has at each occupation, based upon their history of employment at other occupations. If the worker changes from occupation m to occupation m' at period t, then their *equivalent experience* relevant to the new occupation m' will be subject to a discount, which is proportional to the task distance between the two occupations m and m'. The evolution of experience in such case is parametrized by Equation 3.

$$\widetilde{x}^{m'}_{wt} = (\widetilde{x}^{m}_{wt-1} + 1)e^{-\gamma d(m,m')} \tag{3}$$

The *equivalent experience* to the new occupation m', $\widetilde{x}^{m'}_{wt}$, determines worker w's marginal productivity at the new employer. The existence of the discount term, $e^{-\gamma d(m,m')}$, suggests that human capital is occupation-specific and cannot be transferred to the new occupation seamlessly. The parameter γ governs the degree of transferability upon occupation change. If γ is greater, previous experience at a different occupation will be less relevant to the new occupation, and will thus be discounted at a larger extent. The transferability parameter γ is hypothesized to be positive so that a larger task distance between two occupations translates into a lower degree of transferability. Without knowledge about γ, the occupation-m relevant experience is a latent variable.

The model specification presented in 3 introduces a novel approach to the literature by representing occupation-specific experience through a scalar experience variable. Keane and Wolpin (1997)considers three distinct experiences, each corresponding to a separate occupation. In Yamaguchi (2012), experience is decomposed into motor-experience and cognitive-experience. These are accrued at varying rates, contingent upon the task composition inherent to a worker's profession. Moreover, Yamaguchi (2012) illustrates differential valuations for these experience dimensions across occupations. For instance, roles primarily reliant on motor skills, like truck driving, price motor-experience more than cognitive-experience. The heterogeneous pricing of distinct experiences and the heterogeneous accumulation of experience across various dimensions explains the degree of human capital transferability across different occupations. It explains why a shift from an actuary to a truck driver leads to a significant devaluation of prior human capital, whereas a transition from an actuary to an accountant largely preserves the existing human capital stock.

The task distance measure proposed in this paper, therefore, is a parsimonious simplification that preserves the transferability concept in the multidimensional task-specific experiences accumulation models, such as in Adda and Dustmann (2023). In the latter model, workers gain task-specific human capital in five dimensions. At the same time, firms in different occupations price these five experience dimensions differently. In occupations where task k ($k \in \{1, \cdots, 5\}$) is used more intensively, firms set higher prices for task-k specific human capital, compared to other task dimensions. As a result, when workers move to a different occupation, they may experience a penalty imposed by the new firm because the previously gained human capital is not perfectly compatible with the task requirements at the new term. The one-dimensional approach employed in this paper keeps the spirit of the multidimensional approach, yet allows tractable computation. A more detailed discussion and a robustness check by using transferability concept in this multidimensional approach is provided in Appendix B.

3.3 Wage Equation

A wage profile is a policy function $\phi(\widetilde{x})$ optimally chosen by the employer, which maps the occupation-m specific experience to a wage rate. In this section, I show the econometric specifications of the wage equation, which shows the observed wage posting policies of employers.

If a worker keeps working at the same job, the wage profile is a description of the lifecycle earning path of a worker. If a worker changes to a different job mid-career, they will start receiving a new wage profile for the reminder of their career. This section describes the parametric wage equations that determine the earnings of workers.

Worker w's wage in period t is approximated by Equation 4, which shows the wage as a function of worker characteristics χ, occupation m, and employer identity f.

$$\begin{aligned}\log(y_{wt}) &= \phi^{\chi m f}(\widetilde{x}^{m}_{wt}) + \psi_w + \epsilon_{wt} \\ &= \psi_0^{\chi m f} + \psi_1^{\chi m f}\ln(\widetilde{x}^{m}_{wt}) + \psi_w + \epsilon_{wt} \end{aligned} \tag{4}$$

Wage rate increases linearly in the log of experience, approximating the diminishing returns of MPL to experience. The firms choose the intercept ψ_0 and the slope ψ_1 of the wage equation optimally, taking into account: (1) the worker's background characteristics χ, (2) the occupation m, and (3) the characteristics of the firm f. (See firm side model in Section 3.6) Wage rates vary depending on these factors either because they affect the MPL of the worker, based on which firms set wage rates, or because labor supply elasticity vary with respect to these factors, based on which firms' optimal wage posting strategies differ.

Besides the part of the wage equation that is determined by the employer, a worker fixed effect ψ_w also enters the wage equation to allow for unobserved time-invariant individual heterogeneity in earnings. As a result, the specification of the wage equation 4 captures two-way unobserved heterogeneity as in Abowd et al. (1999).

3.4 Firm Heterogeneity

In the wage equation 4, the intercept and slope coefficient $(\psi_0^{\chi m f}, \psi_1^{\chi m f})$ are specific to individual characteristics χ, occupation m, and firm identifier f. The firm specificity of wage coefficients captures firm-level heterogeneity in wage determination. Consistent with the profit maximization problem described below in Section 3.6, firms of different characteristics may differ in posting their optimal wage profiles. While the full discussion of the wage posting problem is postponed until the last part, the observed firm-specific wage profiles differ in starting salaries and in growth rates. A natural approach, therefore, is to estimate the wage equation allowing firm specific intercepts and firm specific slopes, so that the coefficients $(\psi_0^{\chi m f}, \psi_1^{\chi m f})$ are specific to each individual firm.

Abowd et al. (1999) (henceforth AKM1999) is the seminal work in decomposing the worker and firm heterogeneity in wage determination. There are two challenges when directly applying the AKM1999 approach. First, the identification of firm fixed effects depends on inter-employer mobility. The fixed effect of a specific firm is identified if the firm, at one time or another, has hired at least one worker who has changed jobs. That means for firms of which all employees have never changed jobs, their corresponding fixed effects are not identified. In AKM1999, over 10% of observations do not contribute to the estimation of firm fixed effects because they fail to satisfy the identification condition. Second, for firms with relatively small size of employees, their fixed effects are estimated imprecisely due to the large standard error. Given millions of firms in the data, all firm fixed effects cannot be estimated with reliable efficiency level.

To overcome the identification and efficiency problem, I employ an unobserved group heterogeneity approach that was first developed by Bonhomme and Manresa (2015) (henceforth BM2015). In the BM2015 framework, each firm belongs to a finite set of unobserved groups and heterogeneity operates at the group level; each group possesses a group fixed effect while firms within a group are assumed to be homogeneous. BM2015 estimates the group fixed effects and the group membership in an iterated algorithm, using the k-means clustering approach. This paper follows the more recent multidimensional clustering approach by Cheng et al. (2023), where a firm has two unobserved group memberships: a group for intercepts, and a group for slopes. I define these two groups as $i-$types and $s-$types, respectively. Firms within the same $i-$type share the same intercepts in 4, while firms within the same $s-$type share the same slopes. Therefore, the wage equation 4 becomes equation 5.

$$\begin{aligned}\log(y^m_{wt}) &= \phi^{\chi mis}(\widetilde{x}^m_{wt}) + \psi_w + \epsilon_{wt}\\ &= \psi_0^{\chi mi(f)} + \psi_1^{\chi ms(f)} \ln(\widetilde{x}^m_{wt}) + \psi_w + \epsilon_{wt} \qquad (5)\end{aligned}$$

where $i(f)$ and $s(f)$ denote the $i-$type and $s-$type memberships of firm f, respectively. For two firms f, f' such that $i(f) = i(f')$, their ψ_0 coefficients are same holding fixed worker characteristics χ and occupation m: $\psi_0(\chi, m, i(f)) = \psi_0(\chi, m, i(f'))$. Likewise, for two firms belonging to the same $s-$type, they share the same ψ_1 coefficients.

While the details of the estimation procedures will be discussed in Section 5, I denote θ^{wage} as the collection of all wage coefficients $(\psi_0^{\chi mi}, \psi_1^{\chi ms})$ over all (χ, m, i, s).

3.5 Worker Dynamic Choice

With the wage equation facing workers established, I now state the worker's optimal choice of occupational mobility over the lifecycle. The timing of the structural model can be summarized below.

- At the beginning of period t, Nature decides if there is an arriving outside offer.
  - With probability π, there is an outside offer. This outside offer is characterized by a specific employer (i, s) type, drawn from the population distribution of (i, s) type.
  - With probability $1 - \pi$, there is no outside offer, and the worker stays in the current job and moves on to the next period.
- Any job arrival from employer type (i, s) consists of **all** occupations $m \in \{1, \cdots, 5\}$. The worker can choose to move to any occupation associated with the (i, s) type of the arriving offer, or otherwise, to stay at the current job.
- Time moves on to the next period $t + 1$, until the worker retires at age 65.

Henceforth, a **job** is defined as an occupation, $i-$type, $s-$type combination (m, i, s).

At the beginning of every period t, with probability π, an outside offer arrives to the worker. The outside offer is a new firm (i, s) type, which is randomly sampled from the population distribution of (i, s): $F(i, s)$.[18] An arriving outside offer (i, s) entails all five occupations $m \in \{1, \cdots, 5\}$, among which the worker can choose any option. In addition, the worker can also choose to stay at their current job.[19] With probability $(1 - \pi)$, there will be no outside opportunity in this period, so the worker can only choose to stay in the current job.

Worker w whose current job is (m, i, s) makes a job-switch choice d_{wt} in each period t. They can choose to stay at their current job (d_{wt} = stay), move to a different job within the current occupation ($d_{wt} = (m, i', s')$), or move to a different occupation ($d_{wt} = (m', i', s')$), subject to the arrival of an outside job opportunity. The worker makes choice d_{wt} in order to maximize their lifetime value functions. While the outside job opportunity may offer higher wages or more promising future wage growth, the worker may suffer from depreciation to their human capital pertaining to the current occupation. In case the worker moves to a different job but within the current occupation, there will be no human capital depreciation. This job mobility decision by the worker is a function of the wage the current firm is offering to the worker, so this is the part of the model that determines whether a worker supplies their labor to the firm as a function of the

[18] This specification of the job arrival process is similar to Lise and Postel-Vinay (2020), which proposes an undirected search model where the productivity of the arrival job is randomly drawn from the population.

[19] Roys and Taber (2019) proposes a directed search model where workers can choose to search for a job in any specific occupation, and following their search decision, a new offer may arrival with certain probability. My model assumes a reversed order in the sense that the outside offer arrives first with friction, and conditional on arrival of outside offers, workers make occupation choices. This paper establishes an undirected search model with occupation choices to take into account the added firm side information compared to the typical structural work with only worker side information.

wage. This, in turn, goes into the determination of the wage elasticity of the labor supply curve.

The modeling of the job arrival process takes into account the limitations workers face in choosing a desired job, because typically very few people have the chance to choose the highest-paying jobs, while also allowing workers to choose different occupations, keeping the insights of directed search models such as Roys and Taber (2019).

State variables that determine the value functions of each choice includes:

- χ_{wt}: individual demographic characteristics, including age and high school status. Age determines the number of remaining periods, since individuals retire at 65. Unlike a canonical lifecycle model, this paper does not consider the utility gained after retirement, as these issues are not relevant to the purpose of this paper. As a result, workers essentially solve a finite-horizon dynamic discrete choice model. High school statuses affect the potential wage rates through equation 4, though they do not necessarily affect the differences in MPL.

- m_{wt-1}: the incumbent occupation of the worker.

- x_{wt-1}: the experiences that are specific to the incumbent occupation m. The incumbent occupation and experience determine the potential earnings the individual will get at an outside job, according to the transferability equation 3 and the wage equation 4.

- (i, s) type: the $i-$type and $s-$type of the worker's incumbent employer. The types of the incumbent employer determine the earning profiles and thus the value functions if the individual keeps working at the current job.

- $CS^{i,s}_{wt}$: the realized choice set at period t. This is an unobserved state variable as researchers cannot tell from the data if any outside job offer arrives at a given period. Let d_{wt} denote the choice made by worker w in period t. $CS^{i,s}$ means that a worker can move to a new job among all 5 occupations from type (i, s) employers, in addition to staying at the current job, thus totaling 6 choices:

$$d_{wt} \in CS^{i,s} = \{\text{stay}, (m = 1, i, s), \cdots, (m = 5, i, s)\}$$

  In contrast, $CS^{\varnothing}$ means that a worker can only stay in the current job, in which case the choice set is a singleton:

$$d_{wt} \in CS^{\varnothing} = \{\text{stay}\}$$

Let θ denote the set of all parameters that determine the worker's dynamic choices. It consists of three parts: the pre-estimated wage equation parameters θ^{wage} in Section 3.3, the pre-estimated transferability parameter γ, and the structural parameters specific to this section θ^{d}.

$$\theta = (\theta^{wage}, \gamma, \theta^{d}) \tag{6}$$

The structural parameters specific to the dynamic job choice model consists of five coefficients:

$$\theta^{d} = \left(\sigma^2, c_0, c_1, \pi^{nhs}, \pi^{hs}\right) \tag{7}$$

where σ^2 determines the relative weight workers' utility assigns to current period wage payoffs, (c_0, c_1) decides the moving cost incurred from any job change, and (π^{nhs}, π^{hs}) decides the arrival rates of outside job opportunities for workers without and with high school qualifications, respectively.

Let Ω_{wt} denote the observable state variables of the worker.

$$\Omega_{wt} = (\chi_{wt}, m_{wt-1}, \widetilde{x}_{wt-1}, (i,s)_{wt-1})$$

Worker w derives a choice specific value function $\widetilde{V}(\Omega_{wt}, CS_{wt}, d_{wt}; \theta)$ by choosing option d_{wt}, shown in Equation 8.

$$\begin{aligned}\widetilde{V}(\Omega_{wt}, CS_{wt}, d_{wt}; \theta) = \frac{1}{\sigma^2} y(d_{wt}, \Omega_{wt}; \theta^{wage}, \gamma) - (c_0 + c_1 \times age) 1(d_{wt} \neq \text{stay}) + \nu_{wtd} \\ + \beta \mathbb{E}_{CS_{wt+1}}[V(\Omega_{wt+1}; \theta) | \Omega_{wt}, d_{wt}]\end{aligned} \tag{8}$$

In Equation 8, σ^2 is a scale parameter that governs the relative utility contributed by current period wage rate, $y(d_{wt}, \Omega_{wt}; \theta^{wage})$. The current period wage rate is determined by the wage profile of job (m, i, s), i.e. Equation 5, which is governed by wage coefficients θ^{wage} and the transferability parameter γ if the worker experiences occupation change. A larger γ implies that human capital is less transferable, therefore lowering the starting wage rates at all outside occupations. The future wage will also be lower because of a lower starting wage. Therefore, a lower degree of skill transferability lowers the choice specific value function at all outside occupations.

Parameters (c_0, c_1) determine the moving cost incurred from any job change, which depends on the age of the worker. The discount factor β is a preset parameter. The structural error term ν_{wtd} follows type-I extreme value distribution, which allows idiosyncratic tastes with respect to different career choices. The expectation over the continuation value function $V(\Omega_{wt+1};\theta)$ is taken over possible choice sets CS_{wt+1} in the next period.

Worker w therefore chooses an option d_{wt} from the realized choice set CS to maximize the Bellman Equation 9.

$$V^{CS}(\Omega_{wt};\theta) = \max_{d_{wt}\in CS} \widetilde{V}(\Omega_{wt}, CS, d_{wt};\theta) \tag{9}$$

By integrating out the unobserved state variable CS_{it}, the value function 10 is recovered.

$$V(\Omega_{wt};\theta) = (1-\pi)V^{CS^{\emptyset}}(\Omega_{wt};\theta) + \pi \int_{i,s} V^{CS^{i,s}}(\Omega_{wt};\theta)dF(i,s) \tag{10}$$

Following the properties of the extreme value distribution, the dynamic discrete choice model described so far implies the separation function, $s_\Omega\left(\theta^{wage},\gamma,\theta^d\right)$, and the recruiting function $r_\Omega\left(\theta^{wage},\gamma,\theta^d\right)$, both of which are functions of the wage coefficients at all possible jobs, θ^{wage}. According to the job (m,i,s) specific wage profiles determined by Equation 5, the wage coefficients of all jobs θ^{wage} fully determine the potential wage rates a worker can receive at every possible job. Therefore, these wage profiles completely determine the choice-specific value functions associated with all possible job. Based on these fully determined value functions, the fractions of workers with any state variable Ω that stays at or moves into any job (m,i,s) are fully determined, too, giving Equations 11 and 12.

$$s_\Omega\left(\theta^{wage},\gamma,\theta^d\right) = \int_{i,s}\left(1 - \frac{\widetilde{V}(\Omega, CS^{i,s}, d = stay;\theta)}{\sum_{d\in CS^{i,s}} \widetilde{V}(\Omega, CS^{i,s}, d;\theta)}\right)dF(i,s) \tag{11}$$

$$r_\Omega\left(\theta^{wage},\gamma,\theta^d\right) = \sum_{\Omega'\neq\Omega}\left(n_{\Omega'}\int_{i,s}\frac{\widetilde{V}(\Omega', CS^{i,s}, d = (m,i,s);\theta)}{\sum_{d\in CS^{i,s}} \widetilde{V}(\Omega', CS^{i,s}, d;\theta)}dF(i,s)\right) \tag{12}$$

In the notations above, χ is the demographic vector that is part of the state variable Ω. Likewise, χ' is part of state variable Ω'.

Equation 11 is an integration of staying probabilities over the distribution of arriving job opportunities, for a given individual state variables Ω. Specially, when the choice set CS is a singleton, the probability inside the parenthesis is one. Equation 12 is the summation over all workers not currently be in state Ω, but moves to state Ω. Both the separation function and the recruiting function are specific at the worker state variable Ω level. $n_{\Omega'}$ is the size of all workers whose current state variable is Ω'. Note that the inner integration inside Equation 12 is 0 if the demographic variables χ' embedded in Ω' is different to χ embedded in Ω.

Both Equation 11 and 12 are decreasing in γ. An increase of the transferability parameter γ lowers the values of all outside occupations, therefore reducing the probabilities that a worker choose any outside occupations, equivalently reducing the separation probabilities. This will also lower the numerator inside the integration in 12, reducing the likelihood that any worker from outside occupation m moves into occupation m.

By taking summation over the total number of job stayers and the total number of external recruits, I construct the labor supply function of workers with state variable Ω, faced by each individual firm with type (i,s). This labor supply function $l_\Omega(\cdot,\cdot,\cdot)$ is a function of the wage coefficients θ^{wage}, the transferability parameter γ, as well as the structural parameters θ^d, as shown in Equation 13.

$$l_\Omega\left(\theta^{wage},\gamma,\theta^d\right) = n_\Omega\left(1 - s_\Omega\left(\theta^{wage},\gamma,\theta^d\right)\right) + r_\Omega\left(\theta^{wage},\gamma,\theta^d\right) \tag{13}$$

The labor supply function 13 is an input to the employer's optimal wage setting problem, which I describe in detail in the next part. With the firm-specific labor supply equation 13, it is possible now to define the main objectives of this paper, namely, the effect of human capital transferability on recruiting, separation, and labor supply elasticity.

The effect of human capital transferability on labor supply is characterized by the derivative of the labor supply function 13 with respect to the transferability parameter γ, which is a weighted summation of its effects on separations and recruits, respectively, as shown in Equation 14:

$$\frac{\partial l_\Omega\left(\theta^{wage},\gamma,\theta^d\right)}{\partial\gamma} = n_\Omega\left(1 - \frac{\partial s_\Omega\left(\theta^{wage},\gamma,\theta^d\right)}{\partial\gamma}\right) + \frac{\partial r_\Omega\left(\theta^{wage},\gamma,\theta^d\right)}{\partial\gamma} \tag{14}$$

The model does not impose a steady state, in which case Equation 14 should be restricted to zero. Indeed, since both the recruiting function and the separation function increases in γ, the overall effect of γ on employer-specific labor supply function can be either positive or negative; some jobs may experience a net inflow while other jobs may experience a net outflow depending on the change of value functions associated with each career option.

The firm-specific labor supply elasticity ϵ_Ω is defined as the percentage change in labor supply l_Ω following a one percent change in own wage profile θ_Ω^{wage} (relevant to workers with state variable Ω), as shown in Equation 15.

$$\epsilon_\Omega\left(\theta^{wage},\gamma,\theta^d\right)=\frac{\partial l_\Omega\left(\theta^{wage},\gamma,\theta^d\right)}{\partial\theta_\Omega^{wage}}\frac{\theta_\Omega^{wage}}{l_\Omega\left(\theta^{wage},\gamma,\theta^d\right)} \tag{15}$$

The effect of imperfect human capital transferability on labor supply elasticity is a second order derivative suggested in Equation 16:

$$\frac{\partial\epsilon_\Omega\left(\theta^{wage},\gamma,\theta^d\right)}{\partial\gamma}=\frac{\partial^2\ln l_\Omega\left(\theta^{wage},\gamma,\theta^d\right)}{\partial\ln\theta_\Omega^{wage}\partial\gamma} \tag{16}$$

Without estimating the model, it is theoretically unclear whether Equation 16 is positive or negative. The exact sign and magnitude of this effect will be recovered from the estimation.

3.6 Firm Optimal Wage Posting

The previous section 3.5 recovers firm-specific labor supply elasticity with respect to wage coefficient θ^{wage}. This section proposes a firm-side wage posting model that will formally define the elasticity from a lifecycle perspective. The firm-side model is not to be estimated, rather, it rationalizes the estimated elasticity from the perspective of firm profit maximization.

Firms earn profits from the output of workers. Workers with different background χ and different human capital stock $\widetilde{x}$ may differ in either their MPLs or their labor supply functions. In turn, firms will set different wage profiles (ψ_0,ψ_1) for workers with different characteristics in order to maximize profits.

Each worker type Ω is associated with a path for the marginal product of labor:

$$(q_a,q_{a+1},\cdots,q_{65})$$

where a is the current age of the worker, and the MPL path assumes that the worker never changes job until retirement. Because Ω contains the firm side characteristics, the (i,s) type, the MPL path is interpreted as a path for the match surplus between a worker and a firm. Given this path, the present discounted value (PDV) of the total MPL contributed by a worker with state variable Ω can be derived in Equation 17.

$$Q_\Omega = \sum_{t=a}^{65} \beta^t q_t \tag{17}$$

Fixing the intercept and slope parameters (ψ_0, ψ_1) and the worker fixed effect ψ_w, the present discounted value of the lifetime earnings for a worker with state variable Ω can be derived in Equation 18.

$$Y(\psi_0, \psi_1) = \sum_{t=a}^{65} \beta^t \left(\psi_0 + \psi_1 \ln(\widetilde{x}_a + t) + \psi_w + \epsilon_{wt}\right) \tag{18}$$

where a refers to the current age of the worker, m is the current occupation, and $\widetilde{x}_a$ is the current experience. Y is the presented discounted value of lifetime earnings if the worker with state variable Ω stays with the firm forever until retirement.

Suppose that the employer takes the wage profiles of all other jobs as exogenously fixed, its own wage profile posted in Equation 18 will affect the staying probability for existing workers and the size for external recruits. A higher wage profile will increase both the number of staying workers and the size of new recruits, at the cost of lower profits earned by the firm. Let n_Ω be the current size of employees of type Ω within the firm f, $s_\Omega(Y_\Omega)$ be the probability of separation given wage profile Y_Ω, and r_Ω be the size of outside recruits of type Ω. Firm f's profit function is defined in Equation 19.

$$\begin{aligned}
\Pi_f &= \sum_\Omega \underbrace{\mathbb{E}[Q_\Omega - Y_\Omega]}_{\text{taken over possible future separations}} \left[n_\Omega(1 - s_\Omega(Y_\Omega, \gamma, \theta^d)) + r_\Omega(Y_\Omega, \gamma, \theta^d)\right] \\
&= \sum_\Omega \mathbb{E}[Q_\Omega - Y_\Omega] l_\Omega(Y_\Omega, \gamma, \theta^d)
\end{aligned} \tag{19}$$

Note that compared to Equations 14, 15 and 16 in the previous section, in Equation 19 the lifetime PDV of earnings posted by the firm, Y_Ω, replaces the wage parameter θ_Ω^{wage}. This is because that the intercept and the slope parameters (ψ_0, ψ_1) in the wage equation affects the workers' dynamic job choices only through their effects on the lifetime earnings Y_Ω, by Equation 18.

From the point of view of an employer, it takes as given the current size of employees with state variable Ω, n_Ω. Both its recruiting function $r_\Omega(Y_\Omega, \gamma, \theta^d)$ and its separation function $s_\Omega(Y_\Omega, \gamma, \theta^d)$ are solved from the worker side model: the dynamic discrete choice model supplies the fraction of workers with state variable Ω that separates from any job (m,i,s), and at the same time also provides the fraction of moving workers from (m,i,s) to (m',i',s'). The summation of the recruiting function and the separation function is the labor supply function $l(Y_\Omega, \gamma, \theta^d)$ to a specific firm, as suggested in Equation 13.

An employer, by changing the promised wage profile Y_Ω, changes the workers' value functions, therefore also changes workers' likelihood to leave or join the firm. As the model assumes that other employers always commit to the observed wage profiles, the recruiting and separation functions only vary in the own wage profile Y_Ω.

There are a few remarks about the profit function 19.

Remark 1: There is no complementarity between different types of workers, so that the profit function is additively separable in worker types.

Remark 2: Equation 19 is the projected profit function supposing that the future separation probability of the worker is fixed and thus unaffected by a change in wage policy today. This is an simplifying condition to avoid the nested relationship between the wage profile and the future separation probability. By simply assuming known separation probability in the remaining periods, firms' wage posting strategies do not depend on the future separation probabilities, which are in turn functions of future wage PDVs.

Remark 3: When maximizing the profit function, the firm excludes the worker fixed effects ψ_w and the error term ϵ_{wt}, which are unknown to the firm at the time of recruitment.

Remark 4: Instead of maximizing profit by choosing an intercept-slope combination (ψ_0, ψ_1), the firm simply chooses the PDV of lifetime wages Y_Ω to maximize 19. This assumption guarantees the existence of a unique maximizer Y_Ω, avoiding the possible multiple optima of (ψ_0, ψ_1), as different combinations of intercepts and slopes may yield the same PDV of wages.

Remark 5: The model does not consider separations into unemployment or hiring from unemployment. This is because this paper does not study unemployment decisions, and this is reasonable simplification because the sample covers prime age male workers only, whose unemployment rate is sufficiently low.

By taking the first order condition with respect to Y_Ω, the overall wage markdown can be solved in Equation

21.[20]

$$
\begin{aligned}
(Q_\Omega - Y_\Omega)(r_{\Omega 1}(Y_\Omega,\gamma,\theta^d) - n_\Omega s_{\Omega 1}(Y_\Omega,\gamma,\theta^d)) &= (1 - s_\Omega(Y_\Omega,\gamma,\theta^d))n_\Omega + r_\Omega(Y_\Omega,\gamma,\theta^d) \\
\implies \frac{Q_\Omega - Y_\Omega}{Y_\Omega} &= \frac{(1 - s_\Omega(Y_\Omega,\gamma,\theta^d))n_\Omega + r_\Omega(Y_\Omega,\gamma,\theta^d)}{(r_{\Omega 1}(Y_\Omega,\gamma,\theta^d) - n_\Omega s_{\Omega 1}(Y_\Omega,\gamma,\theta^d))Y_\Omega} \\
&= \frac{l(Y_\Omega,\gamma,\theta^d)}{l_{\Omega 1}(Y_\Omega,\gamma,\theta^d)Y_\Omega} \qquad (20) \\
&= \frac{1}{\epsilon_\Omega(Y_\Omega,\gamma,\theta^d)} \qquad (21)
\end{aligned}
$$

Line 20 suggests that the wage markdown equals the inverse of the labor supply elasticity, where the labor supply equals new hires minus separations. The overall markdown 21 can be recovered from the estimated elasticity of firm-specific labor supply. The relationship between equation 20 and 21 follows the classic results on market power: if the labor supply is more inelastic, firms will exercise a greater market power.

The separation function 11, the recruiting function 12, the firm-specific labor supply function 13, the firm-specific labor supply elasticity 20, and consequently the wage markdown 21 are all affected by the structural parameters and the transferability parameter γ. If any of these parameters changes, workers' choice-specific value functions 8 change accordingly. The next section talks about the details of the whole estimation scheme, where the eventual goal is to recover the wage markdown 21 as well as its relationship to the structural and transferability parameters. The estimated results indicate how imperfect transferability affects hiring, separation, and market power of employers, as suggested by Equations 14, 15 and 16 in the previous section.

4 Estimation and Identification

This section describes the estimation approaches for the unobserved firm type assignment, the transferability equation 3 and the wage equation 4, as well as the worker side dynamic discrete choice model. The identification argument for the dynamic choice model is also discussed in this section, with a more detailed proof shown in the appendix. After obtaining the estimation results for the three model elements mentioned above, I recover the elasticity of separation, the elasticity of recruiting, and the overall firm-specific labor supply elasticity. Employers' wage markdown can subsequently be recovered.

[20] $s_{\Omega 1}, r_{\Omega 1}, l_{\Omega 1}$ denotes the derivative of each function with respect to the first argument.

Estimation approaches

Model Elements	Estimation approach
Firm type assignment	Iterated K-means clustering (Cheng et al. (2023))
Transferability and wage equation	EM algorithm
Dynamic worker job choice model	Maximum likelihood via Nested Fixed Point Algorithm

4.1 Unobserved Firm Type Assignment

A real-world firm can employ individuals in various occupations, whereas in my conceptual framework, a firm is limited to hiring in just one specific occupation. A conceptual firm in the model only hires in one occupation. In this paper, a real firm is subdivided into several distinct entities, each hiring in a single occupation. For the sake of clarity in my subsequent discussion, I will use the term "firm" to refer to this conceptual pairing of a firm with a particular occupation.

Firm level heterogeneity operates through the intercept parameter ψ_0 and the slope parameter ψ_1 in the wage equation 4, at the level of $i-$types and $s-$types, respectively. Different firms that belong to the same $i-$type offer the same intercept ψ_0 to workers, fixing worker side characteristics χ. Likewise, different firms in the same $s-$type offer the same slope ψ_1. Let there be a total number of $\mathcal{I}$ $i-$types and $\mathcal{S}$ $s-$types, and with abuse of notation, we denote $i(f) \in \{1, \cdots, \mathcal{I}\}$ the i type for firm f, and $s(f) \in \{1, \cdots, \mathcal{S}\}$ the s type for firm f.

The object to be estimated is therefore the assignment of i-types and s-types for all firms:

$$\{i(f), s(f)\}_{f \in \mathcal{F}} \tag{22}$$

Compared to the one-dimensional clustering approach, the multi-dimensional clustering approach proposed by Cheng et al. (2023) has a few advantages: It is a parsimonious approach to efficiently capture heterogeneity in multiple dimensions without the need to categorize the observations into too many clusters. Using the one-dimensional approach, firms should be clustered into $\mathcal{I} \times \mathcal{S}$ clusters, yet the multidimensional approach only requires the determination of $\mathcal{I} + \mathcal{S}$ memberships. Second, multi-dimensional clustering accommodates sparse interactions. The one-dimensional approach may generate clusters that only includes few firms, because it requires both slopes and intercepts to be the same within each cluster. In contrast, multi-dimensional clustering does not cut the data too fine, thus yielding more robust estimates.

The Cheng et al. (2023) approach employs a recursive Lloyd approach to simultaneously estimate type assignments and coefficients, denoted as (ψ_0, ψ_1). This approach updates firms' i-types and s-types using a k-means clustering methodology.

In each iteration, the following steps are taken:

- Initially, the wage equation is estimated, incorporating i-type-specific intercepts and s-type-specific slopes based on the current type assignment.
- Next, the i-type of each firm is updated to maximize the within-firm likelihood.
- The third and fourth steps repeat the first two but focus on updating the s-type.
- The recursion continues until the pre-set convergence criterion is met.

The estimation for type assignments carries over to the next two sections, serving as inputs to the wage equation and state variables in the worker dynamic choice problem.

The detailed description of the algorithm is shown in Appendix C.

4.2 Transferability and Wage Equation

Sections 3.2 and 3.3 establishes a two-equation system that determines the wage rates, namely, the transferability equation 3 and the wage equation 4. The main obstacle in estimating the wage equation 4 is the unobserved occupation-relevant experience, $\widetilde{x}$. I use the expectation-maximization (EM) algorithm initially developed in Dempster et al. (1977) to sequentially update the expectations for $\widetilde{x}$ and maximize the likelihood for the wage equation. Given a value for the transferability parameter γ, the expectation for $\widetilde{x}$ can be constructed according to Equation 3. The constructed experience variable $\widehat{x}$ joins the wage equation and produces an estimate for the wage coefficients. The recursive estimation approach is continued until the pre-set convergence criterion is met.

Intuitively, the estimated value of the transferability parameter $\widehat{\gamma}$ maximizes the likelihood of the wage equation. The estimation results in this section carried over to the next section, serving as inputs to the worker dynamic choice problem.

The detailed description of the EM algorithm used in this paper is shown in Appendix C.

4.3 Worker Dynamic Choice

After obtaining the firm type assignments, the transferability parameter γ, and the wage coefficients, the worker side model becomes a canonical dynamic discrete choice model with only one unobserved state variable, i.e., the choice set $CS^{i,s}$ following the realization of job arrivals. For each realized $CS^{i,s}$, workers can choose to move to any of the 5 occupations belonging to firm type (i,s). The distribution of $CS^{i,s}$ follows the observed joint distribution of employees working for type (i,s) firms: $F(i,s)$, therefore, the latent state variable $CS^{i,s}$ can be integrated out. The model can thus be estimated using the Nested FiXed Point Algorithm (NFXP) Rust (1987). The NFXP algorithm is a full solution approach which allows the evaluation of the value functions given any values of state variables, and is thus able to implement relevant counterfactual analyses.

There are in total five structural parameters to be estimated in this step:

$$\theta = (\sigma^2, c_0, c_1, \pi^{nhs}, \pi^{hs}) \tag{23}$$

where the job arrival rates can vary depending on high school statuses.

4.3.1 Identification

This section states the general identification argument for the moving cost and the job arrival rate. Although the moving cost depends on age profiles, and the job arrival rate depends on high school statuses, these two sets of parameters co-determine the mobility patterns. Since the model is estimated to fit the patterns of the actual data, it is noted that these two forces affect mobility in different ways.

The moving cost parameters affect the relative values associated with all alternatives except for "stay". An increase in moving cost decreases the attractiveness of all options in the choice set other than "stay". It also affects the ratio of values associated with any two choices other than "stay". This could be illustrated by the extreme case: suppose a worker can choose to move to outside jobs A or B. A offers a value of 100, while B offers a value of 1. Without moving cost, the value of A is 100 times greater than B, suggesting that very few people would move to B compared to those who move to A. However, a moving cost equaling 10,000 will almost eliminate the differences of choices A and B as both become unattractive.

The job arrival rate affects the likelihood of workers choosing "stay", but unlike the effects of moving cost, it does not affect the ratio of choice probabilities into two alternative external options. An increase in job arrival rate raises the choice probabilities of all external options without changing the relative choice probabilities. Therefore, the moving cost parameters can be separately identified from the job arrival rates from the relative choice probabilities among all external options.

The detailed proof of the identification arguments is deferred to Appendix D.

The following table lists the set of pre-set parameters, including the time discount factor, the numbers of $i-$types and $s-$types.

Pre-set parameters

Parameter	Explannation	Preset value
β	discount factor	0.975
$\mathcal{I}$	number of i-types	5
$\mathcal{S}$	number of s-types	5

5 Results

5.1 Reduced Form Gravity Equation

Cortes and Gallipoli (2018) constructed a similar "task distance" measure using the O*NET data from the US. To compare my distance measure with theirs, I estimate a gravity equation of occupational mobility with the same specification to Cortes and Gallipoli (2018):

$$ln(\frac{sw_{mm'}}{sw_{mm}}) = \alpha_m + \alpha_{m'} - \beta_0 d(m,m') - \sum_{s=1}^{5} \beta_s \lambda^s_{mm'} + \epsilon_{mm'} \tag{24}$$

The $sw_{mm'}$ term is the number of workers who switch from occupation m to m', and sw_{mm} is the number of workers who stayed in occupation m. The dependent variable is the fraction of people from occupation m who switch to m'. The α_m and $\alpha_{m'}$ terms are origin and destination fixed effects, capturing the average wage and amenities within an occupation. The key explanatory variable is the task distance between the two occupations, $d(m,m')$. If the distance measure is valid, the coefficient β_0 should be positive since a larger task distance means it is harder to transfer between the two occupations. $\{\lambda^s_{mm'}\}^5_{s=1}$ is a set of dummies which equal 1 if and only if m and m' are in different major task categories and the major task category of destination occupation m' is s.

Table 4 shows regression results of the gravity equation estimation from 1996 to 2018. The coefficients in front of the task distance is significantly negative across all specifications, therefore confirming the validity of the task distance measure. The magnitude of the estimates is also comparable to the results obtained by Cortes and Gallipoli (2018).

5.2 Wage Equation

Table 6 shows the estimation results for all wage coefficients. The intercept coefficient is specific to high school, occupation, and employer $i-$type, while the slope coefficient is specific to high school, occupation, and employer $s-$type. With 2 education categories, 5 occupations, 5 $i-$types, and 5 $s-$types, there are a total of 100 wage coefficients.[21] The first column of table 6 shows the estimation result by using the latent experience as inputs, while the second column shows the results for the alternative specification with observable occupation specific experiences (as in Keane and Wolpin (1997)) as inputs. The second specification yields much lower R squared, showing evidence that the estimated latent experience does a better job in fitting the wage profile.

Table 7 shows the average wage coefficients by different firm types, education categories, and occupations, taken over the entire selected sample. The first panel shows there are large across-firm variations in both initial wage levels (as reflected by intercepts) and growth rates (as reflected by slopes). This result shows the value-added by incorporating firm types as state variables in the structural model: this added firm-level variation helps identify worker mobility in response to firm-specific wage posting policies, which cannot be completed in previous structural work using only worker-side panel.

Table 7 also shows that high school graduates, compared to workers without high school qualification, have lower initial wage rates but much faster wage growth rate. It also shows across-occupation wage differentials. Occupation 4 and 5, which mainly requires analytical non-routine tasks and cognitive routine tasks, respectively, observe lower initial wage rates but much faster wage growth rates than the other 3 occupations.

[21] Worker fixed effects are controlled.

5.3 Transferability Parameter

Table 5 shows the estimation for the transferability parameter γ. The estimated $\widehat{\gamma} = 0.282$, implying a human capital depreciation of 3.14% to 21.00%. The exact magnitude of the depreciation depends on the task distance between the current occupation and the new occupation. For example, a worker who moves across two least related occupations will lose 21% of their existing human capital stock specific to the incumbent occupation.

5.4 Firm Type Distribution

The upper panel of Figure 3 is a heat-map showing the distribution of workers among different types of firms, by education status. It is worth noting that the $i-$type and $s-$type of employers are negatively correlated: workers are more densely populated in (high-i, low-s) and (low-i, high-s) firms, while they are sparsely populated in (high-i, high-s) and (low-i, low-s) firms. This shows that workers will need to trade off between the starting wage levels and the future wage growth rates. For workers with and without high school education, most of them work for firms with low intercepts and high slopes, suggesting that workers are willing to forgo some salary at the beginning of their career in exchange for a more rapid wage growth in the future.

Comparing with high school graduated workers, workers without high school are less likely to work in firms with either high $i-$type or high $s-$type. This evidence shows that workers without high school education are more likely to be matched with employers of lower profitability.

The lower panel of Figure 3 shows the (i, s) type distribution of destination employers among all workers who have changed jobs. This shows the same pattern as the overall employer type distribution, and shows evidence that high school educated workers have better choice set along their lifecycle.

5.5 Structural Parameters

Table 8 presents the estimated values for the five structural parameters. The moving cost (slope) parameter c_1, is positive, indicating that the costs associated with job changes increase with age. High school graduates are characterized by higher job arrival rates, implying that individuals with high school degrees are likely to encounter better external job opportunities throughout their careers compared to those without a high school education. This result is consistent with the findings in Section 5.4.

5.6 Goodness of Model Fit

The model's estimation produces simulated job separation probabilities across all potential combinations of state variables, denoted as Ω. This section evaluates the model's goodness of fit by comparing these simulated staying probabilities with the probabilities observed from the dataset.

Figure 4 shows both in-sample and out-of-sample model fit. The x-axis shows the 20 ventiles of the predicted staying probabilities, while the y-axis displays the corresponding staying probabilities observed from data. A 45-degree line facilitates the comparison between the simulated and actual data. A majority of the data points closely adhere to this 45-degree line, indicating a solid model fit. The binscatter graph on the left hand side represents the in-sample fit, and the one on the right hand side represents the out-of-sample fit. The congruence observed between the in-sample and out-of-sample binscatter plots suggests robustness of the model estimation and absence of over-fitting.

5.7 Elasticity and Wage Markdown

The job-specific labor supply elasticity is recovered from estimating and solving the structural model. Column (1) in Table 9 shows the actual elasticity. There are a few findings worth pointing out.

First, the overall magnitude of the elasticity estimates is small, implying a large wage markdown imposed by employers. Bachmann et al. (2022) adopts a reduced-form approach to estimate employer-specific labor supply elasticity using a dataset from Germany and reports estimates ranging between 0.9 and 1.6. Comparing to their estimates, my estimates are smaller in magnitude. This disparity is likely due to the consideration of workers' forward-looking behaviors in my analysis. If workers are less responsive to future wage growth compared to current wage levels, it can account for the smaller elasticity. When dynamic considerations are incorporated, the data suggests that firms impose a more pronounced wage markdown than what earlier estimations indicated.

Second, workers who have completed high school exhibit a greater elasticity. This suggests that firms levy a smaller markdown on them, likely because high school graduates encounter better external job opportunities, as shown from section 5.5.

Figure 5 shows the elasticity-age profile. For both education groups, the elasticity-age profile exhibits a hump shape. At the beginning of their careers, the elasticity is low, which gradually increases as they accumulate more experiences before eventually decreases after they reach the mid-30s. The low elasticity at initial periods is again evidence showing young workers' willingness to accommodate larger markdown imposed by their employers. Older workers have smaller benefits and larger opportunity costs associated with job changes, thereby displaying low elasticity.

6 Counterfactual Exercises and Policy Analyses

The estimation of the structural model allows a range of counterfactual analyses. These simulations provide insights into the underlying drivers of employer monopsony power. Moreover, they illuminate the potential influence of Active Labor Market Policies (ALMPs) in moderating this monopsony power.

6.1 Imposing Perfect Transferability

The first counterfactual exercise assumes perfect transferability, namely, by setting the transferability parameter γ to 0. Column (2) of table 9 shows the effect of the counterfactual exercise on the elasticity. Figure 6 further illustrates the effect on the elasticity-age profile.

First, assuming perfect human capital transferability increases the elasticity for all education groups. The increase in elasticity is more pronounced for high school graduated workers than their peers without high school degree. An explanation is that workers with high school degree tend to experience more rapid wage growth rate, and by removing imperfect skill transferability, they will benefit more than the non-high school group. When switching to an occupation with steeper wage growth rate, they do not need to worry about the loss of human capital upon switching. For non-high school group, such benefit is much smaller due to a flatter wage-age profile.

In terms of the elasticity-age profile, the elasticity increases more for workers at an older age. The elasticity slightly decreases at the early years of their career following a resumption of perfect transferability, likely due to their increased tolerance over wage markdown. They become more patient, expecting that moving opportunities in the future will be better.

This counterfactual exercise first answers the main research question of this study: overall, imperfect human capital transferability lowers workers' elasticity of labor supply, and therefore adds to the employers' monopsony power. However, the effect is heterogeneous with respect to different stages in the life-cycle.

This simulation mimics the potential effects of ALMPs that promote vocational training. By subsidizing vocational training, the government lowers the skill barrier for workers to change occupation, therefore increases their human capital transferability. Such policy should mitigate employers' monopsony power.

6.2 Imposing 100% Job Arrival Rate

The second counterfactual exercise restores a 100% job arrival rate by setting the π parameter to 1. Column (3) of table 9 shows the effect of the counterfactual exercise on the elasticity. Figure 7 illustrates the effect on the elasticity-age profile.

It is worth noting that younger workers' elasticity decreases following the counterfactual change, while older workers' elasticity increases. The intuition for this result is the same to the previous counterfactual exercise in the sense that younger workers become more patient as they expect a better choice set in the future.

This simulation resembles the effects of ALMPs that provide job counseling and placement service, which reduces search friction in the labor market. By doing so, it is not guaranteed that firm monopsony power will decrease for everyone, but older workers can benefit from a lower wage markdown.

6.3 Comparing General Education and Vocational Training

A key feature of the education system in Germany is the "dual apprentice" system where students spend time in both attending high school (general education) and participating in vocational training. In order to compare the effects of education types on worker elasticity and lifetime welfare, I resolve the structural model using the same parameters, while adding apprenticeship training as an additional state variable. Before labor market entry, students can choose an apprenticeship training status ranging from six options: no training, or training at one of the five occupations. Working in an occupation where the worker has received vocational training will increase the wage profiles received by the worker.

Table 10 shows the comparison of elasticity based on apprenticeship training statuses. Workers who have gone through vocational training have larger elasticity. This result shows that as trained workers gain more experiences, their skills become more unsubstitutable from the views of employers. That explains why vocationally trained workers will have larger elasticity.

Table 11 compares the value functions of workers at the beginning of their labor market entry, conditional on different apprenticeship training options. It is shown that not receiving vocational training yields the least lifetime utility. Vocational training combined with high school completion yields the largest lifetime utility, suggesting that high school education and vocational training are complementary. Attending high school also helps young workers getting better informed about the conditions of the labor market, so that they are more likely to make the right apprenticeship choice, as shown from the last row in Table 11. 74% of high school graduates make the apprenticeship choice that maximizes their lifetime utility, while only 59% of non-high school graduates make the "right" decision.

7 Conclusion

This paper argues that imperfect transferability of human capital is a source of employer monopsony power. Using matched employer-employee panel data from Germany, I build and estimate a finite-horizon dynamic model capturing worker job-switching behaviors. This model integrates elements of search friction, moving costs, and imperfect skill transferability, aiming to recover the elasticity of job-specific labor supply. Empirical results reveal that imperfect skill transferability reduces workers' labor supply elasticity overall, increasing employers' market power. Notably, the effects of imperfect transferability manifest diversely across age groups. Younger workers, when faced with a hypothetical scenario without imperfect transferability, display a greater willingness to bear wage markdowns—an observation aligning with the concept of "implicit tuition" in Agarwal (2015).

Furthermore, the findings suggest that a set of Active Labor Market Policies, encompassing initiatives like job placement counseling and vocational training, can potentially curb employers' market power. These policies are especially influential in augmenting elasticity for senior workers. Contrary to a possible "lock-in" effect, vocational training seems to fortify workers' position; those with such training exhibit greater elasticity, resulting from employers' heightened valuation of their skills and the consequent motivation to retain them.

It's important to acknowledge the study's focus on job-to-job transitions exclusively for male workers, thereby ignoring transitions involving unemployment—transitions which might bear more significance for female workers. The potential influence of firm-specific experience, compared with the occupation-specific experience discussed herein, could further shape worker labor supply. Delving into these areas, accompanied by relevant policy implications, presents intriguing avenues for future research.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Adda, J. and Dustmann, C. (2023). Sources of wage growth. *Journal of Political Economy*, 131(2):456–503.

Agarwal, N. (2015). An empirical model of the medical match. *American Economic Review*, 105(7):1939–78.

Aizawa, N. and Fang, H. (2020). Equilibrium labor market search and health insurance reform. *Journal of Political Economy*, 128(11):4258–4336.

Arnold, D. (2020). Mergers and acquisitions. *Local Labor Market Concentration, and Worker Outcomes*.

Ashenfelter, O., Card, D., Farber, H., and Ransom, M. R. (2022). Monopsony in the labor market: new empirical results and new public policies. *Journal of Human Resources*, 57(3):S1–S10.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333.

Azar, J., Marinescu, I., and Steinbaum, M. (2022a). Labor market concentration. *Journal of Human Resources*, 57(S):S167–S199.

Azar, J., Marinescu, I., Steinbaum, M., and Taska, B. (2020). Concentration in us labor markets: Evidence from online vacancy data. *Labour Economics*, 66:101886.

Azar, J. A., Berry, S. T., and Marinescu, I. (2022b). Estimating labor market power. Technical report, National Bureau of Economic Research.

Bachmann, R., Demir, G., and Frings, H. (2022). Labor market polarization, job tasks, and monopsony power. *Journal of Human Resources*, 57(S):S11–S49.

Bassier, I., Dube, A., and Naidu, S. (2022). Monopsony in movers: The elasticity of labor supply to firm wage policies. *Journal of Human Resources*, 57(S):S50–s86.

Becker, G. S. (1964). *Human capital theory*, volume 1964.

Benmelech, E., Bergman, N. K., and Kim, H. (2022). Strong employers and weak employees: How does employer concentration affect wages? *Journal of Human Resources*, 57(S):S200–S250.

Berge, P. v. et al. (2020). The sample of integrated employer-employee data (sieed): Sieed 7518 version 1. *Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB)*, 56(6):635–662.

Berger, D., Herkenhoff, K., and Mongey, S. (2019). Labor market power (no. w25719).

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bosch, G., Kupets, O., DWT, I., et al. (2021). Collective bargaining in germany and ukraine: lessons learned and recommendations for ukraine.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273.

Cahuc, P., Postel-Vinay, F., and Robin, J.-M. (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, 74(2):323–364.

Cheng, X., Schorfheide, F., and Shao, P. (2023). Clustering for multi-dimensional heterogeneity. Technical report, Working paper.

Cortes, G. M. and Gallipoli, G. (2018). The costs of occupational mobility: An aggregate analysis. *Journal of the European Economic Association*, 16(2):275–315.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22.

Dengler, K., Matthes, B., and Paulus, W. (2014). Occupational tasks in the german labour market. *FDZ Methodenreport*, 12.

Dey, M. S. and Flinn, C. J. (2005). An equilibrium model of health insurance provision and wage determination. *Econometrica*, 73(2):571–627.

Dodini, S., Lovenheim, M., Salvanes, K. G., and Willén, A. (2020). Monopsony, skills, and labor market concentration.

Ellguth, P. and Kohaut, S. (2022). Tarifbindung und betriebliche interessenvertretung: Ergebnisse aus dem iab-betriebspanel 2021. *WSI-Mitteilungen*, 75(4):328–336.

Flinn, C. J. (2006). Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, 74(4):1013–1062.

Gathmann, C. and Schönberg, U. (2010). How general is human capital? a task-based approach. *Journal of Labor Economics*, 28(1):1–49.

Gibbons, R. and Waldman, M. (2004). Task-specific human capital. *American Economic Review*, 94(2):203–207.

Groes, F., Kircher, P., and Manovskii, I. (2015). The u-shapes of occupational mobility. *The Review of Economic Studies*, 82(2):659–692.

Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. (2020). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics*, 12(1):210–244.

Johnson, M. and Keane, M. P. (2013). A dynamic equilibrium model of the us wage structure, 1968–1996. *Journal of Labor Economics*, 31(1):1–49.

Jovanovic, B. and Moffitt, R. (1990). An estimate of a sectoral model of labor mobility. *Journal of Political Economy*, 98(4):827–852.

Kambourov, G. and Manovskii, I. (2008). Rising occupational and industry mobility in the united states: 1968–97. *International Economic Review*, 49(1):41–79.

Kambourov, G. and Manovskii, I. (2009). Occupational specificity of human capital. *International Economic Review*, 50(1):63–115.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, 105(3):473–522.

Lee, D. and Wolpin, K. I. (2006). Intersectoral labor mobility and the growth of the service sector. *Econometrica*, 74(1):1–46.

Lise, J., Meghir, C., and Robin, J.-M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, 19:63–87.

Lise, J. and Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–2376.

Lise, J. and Robin, J.-M. (2017). The macrodynamics of sorting between workers and firms. *American Economic Review*, 107(4):1104–1135.

Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.

Marinescu, I., Ouss, I., and Pape, L.-D. (2021). Wages, hires, and labor market concentration. *Journal of Economic Behavior & Organization*, 184:506–605.

Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of labor Economics*, 13(4):653–677.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

Prager, E. and Schmitt, M. (2021). Employer consolidation and wages: Evidence from hospitals. *American Economic Review*, 111(2):397–427.

Qiu, Y. and Sojourner, A. (2023). Labor-market concentration and labor compensation. *ILR Review*, 76(3):475–503.

Robinson, C. (2018). Occupational mobility, occupation distance, and specific human capital. *Journal of Human Resources*, 53(2):513–551.

Robinson, J. (1933). *The economics of imperfect competition*. London: MacMillan Press.

Roys, N. A. and Taber, C. R. (2019). Skill prices, occupations, and changes in the wage structure for low skilled men. Technical report, National Bureau of Economic Research.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033.

Schubert, G., Stansbury, A., and Taska, B. (2022). Employer concentration and outside options. *Available at SSRN 3599454*.

Sokolova, A. and Sorensen, T. (2021). Monopsony in labor markets: A meta-analysis. *ILR Review*, 74(1):27–55.

Sullivan, P. (2010). Empirical evidence on occupation and industry specific human capital. *Labour economics*, 17(3):567–580.

Topel, R. H. and Ward, M. P. (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics*, 107(2):439–479.

Wei, S. (2022). Income, employment and health risks of older workers. *Centro de estudios monetarios y financieros*.

Yamaguchi, S. (2012). Tasks and heterogeneous human capital. *Journal of Labor Economics*, 30(1):1–53.

Tables

Table 1: Description of the five task dimensions

Task Dimensions	Contents
Analytical non-routine	Research, analyse, evaluate, plan, construct, design, create, work out rules/regulations, apply and interpret rules
Interactive non-routine	Negotiate, represent interests, coordinate, organise, teach or train, sell, purchase, acquire customers, advertise, entertain, present, employ or manage staff
Cognitive routine	Calculate, accounting, correct texts/data, measure length/height/temperature
Manual routine	Operate or control machines, equip machines
Manual non-routine	Repair or refurbish houses/flats/machines/vehicles, renovate paintings/monuments, serve or accommodate guests

Source. Dengler et al. (2014)

Table 2: Task compositions of some example occupations

Occupation	Major Task Type	s1	s2	s3	s4	s5
Professions in law and administration	1: Analytical Non-routine	**0.56**	0.10	0.34	0.00	0.00
Performing, entertaining professions	1: Analytical Non-routine	**0.41**	0.16	0.26	0.01	0.16
Advertising, marketing, commercial, editorial media jobs	2: Interactive Non-routine	0.37	**0.47**	0.15	0.00	0.00
Mechanical and automotive engineering jobs	3: Cognitive Routine	0.13	0.01	**0.35**	0.30	0.20
Plastic and wood manufacturing and processing	4: Manual Routine	0.06	0.01	0.07	**0.65**	0.21
Cleaning professions	5: Manual Non-routine	0.04	0.00	0.09	0.13	**0.73**
Tourism, hotel and catering professions	5: Manual Non-routine	0.13	0.29	0.18	0.01	**0.38**

Notes. This table shows the task decomposition of several 2-digit occupations. Each occupation is associated with 5 task dimensions, namely, analytical non-routine (s_1), interactive non-routine (s_2), cognitive routine (s_3), manual routine (s_4), and manual non-routine (s_5). The five dimensions add up to one.

Table 3: Descriptive statistics of the selected sample

# Individuals		149,547	
# Establishments		433,998	
Total observations		1,917,458	
		Mean	s.d.
High School		0.34	(0.47)
Non-full-time employment		0.08	(0.27)
Occupation cluster	Major task dimension	Total spells	Percentage
Cluster 1	Versatile	423,852	19.5%
Cluster 2	Manual non-routine	367,558	16.9%
Cluster 3	Manual routine	256,554	11.8%
Cluster 4	Analytical non-routine	733,199	33.8%
Cluster 5	Cognitive routine	390,853	18.0%

Notes. The selected sample are individual workers from West German states entering the labor market between 1996 and 1999 aged 20 to 25.

Table 4: Gravity Equation

	(1)	(2)	(3)
Task Distance	-1.780	-2.045	-2.536
	(0.039)	(0.026)	(0.036)
origin occ. f.e.		controlled	controlled
destination occ. f.e.		controlled	controlled
year f. e. (1996-2018)		controlled	controlled
dest. occ. task category f.e.			controlled
R squared	0.0711	0.6935	0.6995
Observations	27,321	27,321	27,321

Note. An observation is at year-origin occupation-destination occupation level. The outcome variable is the log of the ratio of occupation movers from the origin occupation to the destination occupation over the the stayers at the origin occupation, measured on a yearly basis. The dependent variable Task Distance is defined in Task Distance. Standard errors are shown in parenthesis.

Table 5: Estimates for the transferability parameter

	Transferability Equation
Estimates for γ	0.282 (0.003)
Range of depreciation	[3.14%, 21.00%]
Observations	2,172,016
Log-likelihood	-0.891

Notes. The table shows results of the estimates for the transferability parameter, γ. This estimation is in joint with the estimation of the wage equation, using the EM algorithm. The range of depreciation refers to the percentage of previous human capital stock that will be forgone following an occupation change, where the exact magnitude depends on the task distance between the origin and the destination occupation. For example, if a worker moves between the two least related occupation, his human capital stock will depreciate by 21% as indicated from this result table.

Table 6: Wage equation

			(1) Latent experience				(2) Observable experience			
Occupation	High school	Firm i/s type	Intercept	(s.e.)	Slope	(s.e.)	Intercept	(s.e.)	Slope	(s.e.)
1	0	1			0.119	(0.001)			-0.020	(0.002)
1	1	1	-0.964	(0.007)	0.308	(0.002)	-1.342	(0.006)	0.065	(0.003)
1	0	2	0.284	(0.002)	0.198	(0.001)	0.278	(0.003)	0.074	(0.001)
1	1	2	-0.673	(0.007)	0.386	(0.002)	-1.021	(0.006)	0.164	(0.003)
1	0	3	0.484	(0.002)	0.268	(0.001)	0.466	(0.003)	0.157	(0.001)
1	1	3	-0.472	(0.007)	0.457	(0.002)	-0.770	(0.006)	0.265	(0.002)
1	0	4	0.679	(0.003)	0.354	(0.001)	0.643	(0.003)	0.253	(0.001)
1	1	4	-0.284	(0.007)	0.553	(0.002)	-0.551	(0.006)	0.387	(0.002)
1	0	5	0.896	(0.003)	0.431	(0.001)	0.857	(0.003)	0.346	(0.001)
1	1	5	-0.078	(0.007)	0.633	(0.002)	-0.271	(0.007)	0.487	(0.002)
2	0	1	0.071	(0.004)	0.094	(0.001)	0.006	(0.004)	-0.037	(0.001)
2	1	1	-0.807	(0.008)	0.261	(0.004)	-1.126	(0.008)	-0.005	(0.004)
2	0	2	0.414	(0.004)	0.165	(0.001)	0.315	(0.003)	0.046	(0.001)
2	1	2	-0.503	(0.009)	0.327	(0.003)	-0.840	(0.008)	0.081	(0.004)
2	0	3	0.631	(0.004)	0.230	(0.001)	0.518	(0.003)	0.115	(0.001)
2	1	3	-0.291	(0.009)	0.393	(0.003)	-0.632	(0.008)	0.160	(0.004)
2	0	4	0.790	(0.004)	0.305	(0.001)	0.683	(0.004)	0.192	(0.001)
2	1	4	-0.156	(0.009)	0.461	(0.003)	-0.481	(0.008)	0.246	(0.004)
2	0	5	1.014	(0.004)	0.382	(0.001)	0.917	(0.004)	0.276	(0.001)
2	1	5	0.059	(0.009)	0.534	(0.003)	-0.185	(0.009)	0.347	(0.004)
3	0	1	0.066	(0.004)	0.144	(0.002)	0.085	(0.004)	-0.073	(0.002)
3	1	1	-0.762	(0.008)	0.304	(0.004)	-1.121	(0.007)	-0.025	(0.004)
3	0	2	0.437	(0.004)	0.226	(0.001)	0.454	(0.004)	0.038	(0.002)
3	1	2	-0.434	(0.009)	0.384	(0.003)	-0.749	(0.008)	0.073	(0.004)
3	0	3	0.605	(0.004)	0.294	(0.001)	0.622	(0.004)	0.120	(0.002)
3	1	3	-0.268	(0.009)	0.446	(0.003)	-0.532	(0.008)	0.162	(0.004)
3	0	4	0.783	(0.004)	0.362	(0.001)	0.786	(0.004)	0.204	(0.002)
3	1	4	-0.105	(0.009)	0.528	(0.003)	-0.365	(0.008)	0.270	(0.004)
3	0	5	0.948	(0.005)	0.432	(0.001)	0.977	(0.004)	0.287	(0.002)
3	1	5	0.045	(0.009)	0.586	(0.003)	-0.132	(0.008)	0.356	(0.003)
4	0	1	-0.101	(0.004)	0.264	(0.001)	0.181	(0.004)	0.046	(0.002)
4	1	1	-0.868	(0.006)	0.366	(0.001)	-1.151	(0.005)	0.189	(0.002)
4	0	2	0.119	(0.004)	0.350	(0.001)	0.378	(0.003)	0.156	(0.001)
4	1	2	-0.574	(0.006)	0.468	(0.001)	-0.840	(0.005)	0.311	(0.001)
4	0	3	0.279	(0.004)	0.427	(0.001)	0.549	(0.003)	0.250	(0.001)
4	1	3	-0.407	(0.006)	0.545	(0.001)	-0.648	(0.005)	0.396	(0.001)
4	0	4	0.442	(0.004)	0.500	(0.001)	0.697	(0.004)	0.339	(0.001)
4	1	4	-0.252	(0.006)	0.613	(0.001)	-0.493	(0.005)	0.472	(0.001)
4	0	5	0.602	(0.005)	0.565	(0.001)	0.882	(0.004)	0.416	(0.001)
4	1	5	-0.083	(0.006)	0.687	(0.001)	-0.314	(0.005)	0.554	(0.001)
5	0	1	-0.257	(0.004)	0.289	(0.001)	-0.106	(0.003)	0.011	(0.001)
5	1	1	-0.838	(0.004)	0.346	(0.003)	-1.098	(0.005)	-0.080	(0.003)
5	0	2	0.046	(0.004)	0.377	(0.001)	0.223	(0.004)	0.126	(0.001)
5	1	2	-0.538	(0.004)	0.445	(0.002)	-0.716	(0.005)	0.058	(0.003)
5	0	3	0.244	(0.004)	0.470	(0.001)	0.456	(0.004)	0.239	(0.001)
5	1	3	-0.363	(0.004)	0.550	(0.002)	-0.437	(0.005)	0.192	(0.002)
5	0	4	0.417	(0.004)	0.538	(0.001)	0.640	(0.004)	0.317	(0.001)
5	1	4	-0.171	(0.004)	0.616	(0.002)	-0.219	(0.005)	0.291	(0.002)
5	0	5	0.590	(0.005)	0.606	(0.001)	0.839	(0.004)	0.397	(0.001)
5	1	5			0.675	(0.002)			0.405	(0.002)
Constant			3.342	(0.003)			3.851	(0.002)		
R squared			0.793				0.716			
Observations			1,917,458				1,917,458			

Notes. This table shows the estimates for the wage equation. The intercept and slope parameters are specific to corresponding firm $i-$type and $s-$type, respectively, as shown in column "Firm i/s type". The first specification uses the log of latent occupation specific experience as regressor, where the second specification uses the observable occupational experience. Individual fixed effects are controlled in both regressions.

Table 7: Wage equation summary

	Intercept	(s.e.)	Slope	(s.e.)
i/s-type				
1	-1.37	(0.14)	0.22	(0.10)
2	-1.07	(0.16)	0.30	(0.11)
3	-0.88	(0.17)	0.40	(0.12)
4	-0.70	(0.17)	0.49	(0.12)
5	-0.52	(0.19)	0.57	(0.12)
w/o high school				
	-0.86	(0.33)	0.34	(0.13)
w/ high school				
	-1.07	(0.27)	0.56	(0.11)
occupation				
1	-0.88	(0.33)	0.33	(0.15)
2	-0.71	(0.31)	0.25	(0.11)
3	-0.76	(0.31)	0.32	(0.12)
4	-1.04	(0.25)	0.54	(0.12)
5	-1.09	(0.29)	0.48	(0.12)

Notes. This table summarizes the wage coefficients over the selected sample. In particular, the first panel shows the intercept coefficients by firm $i-$type, and the slope coefficient by firm $s-$type.

Table 8: Estimates for the structural parameters

Parameters	Meaning	Men
c_0	Moving cost (intercept)	0.246
		(0.039)
c_1	Moving cost (slope)	0.045
		(0.001)
σ^2	Scale parameter	30.46
		(0.091)
π^{nhs}	Job arrival rate (no high school)	0.410
		(0.004)
π^{hs}	Job arrival rate (high school)	0.455
		(0.004)
	Observations	1,082,901
	Log-likelihood	-0.841

Notes. This table shows the estimates for the five structural parameters. Standard errors (shown in parenthesis) are computed from the inverse Hessian matrix following the maximum likelihood estimation.

Table 9: Overall elasticity

	(1) Actual	(2) $\gamma = 0$	(3) $\pi = 0$
Without high school	0.504	0.511	0.497
Markdown	198%	196%	201%
With high school	0.675	0.687	0.720
Markdown	148%	146%	139%

Notes. This table shows the overall elasticity of job-specific labor supply from the estimated structural model. Column (1) and (4) shows the actual elasticity recovered from the estimated model. Column (2) and (5) shows the counterfactual elasticity solved from the model with perfect human capital transferability. Column (3) and (6) shows the counterfactual elasticity solved from the model with 100% job arrival rate. The row "Markdown" translates the elasticity into the employer's wage markdown.

Table 10: Elasticity by apprenticeship training status

	w/o high school		w/ high school	
Age range	No training	Training	No training	Training
20-25	0.424	0.478	0.593	0.581
26-31	0.441	0.631	0.625	0.824
32-37	0.421	0.662	0.589	0.869
38-43	0.377	0.653	0.511	0.815
44-49	0.273	0.592	0.39	0.899

Notes. This table shows the elasticity of job-specific labor supply by apprenticeship training status, high school status, and age group. "No training" means the corresponding worker does not go through apprenticeship training at their current occupation (but they may have received training in other occupations), while "Training" means the worker has gone through apprenticeship training at their current occupation.

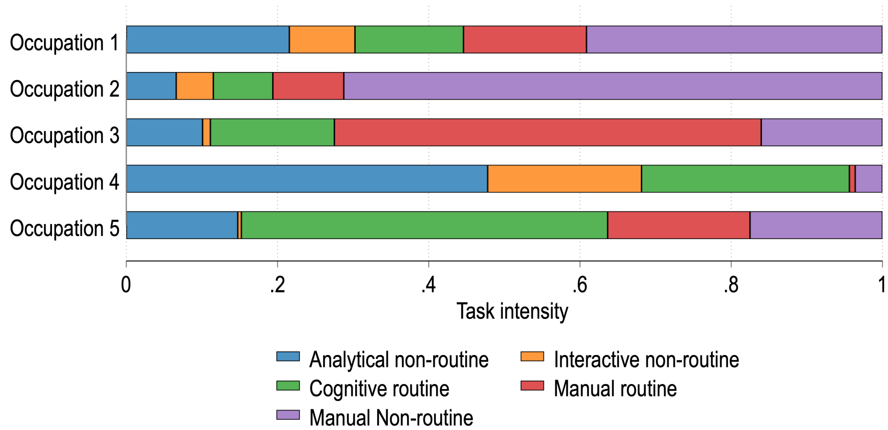
Table 11: Comparison of initial value functions depending on selection of apprenticeship training

Apprenticeship training	w/o high school	w/ high school
no training	1.000	1.122
training in occupation 1	1.016	1.132
training in occupation 2	1.007	1.126
training in occupation 3	1.019	1.136
training in occupation 4	1.018	1.158
training in occupation 5	1.027	1.150
% choosing maximum option	0.593	0.741
Number of individuals	44581	22275

Notes. This table shows the comparison of value functions at the beginning of one's career, conditional on different apprenticeship choices. Workers can choose six different apprenticeship statuses: no apprenticeship training, getting training in occupation 1, $\cdots$, getting training in occupation 5. The value associated with no apprenticeship training and no high school is normalized to one. The row "% choosing maximum option" shows the fraction of individuals that select the apprenticeship choice which maximizes their initial-period value functions.

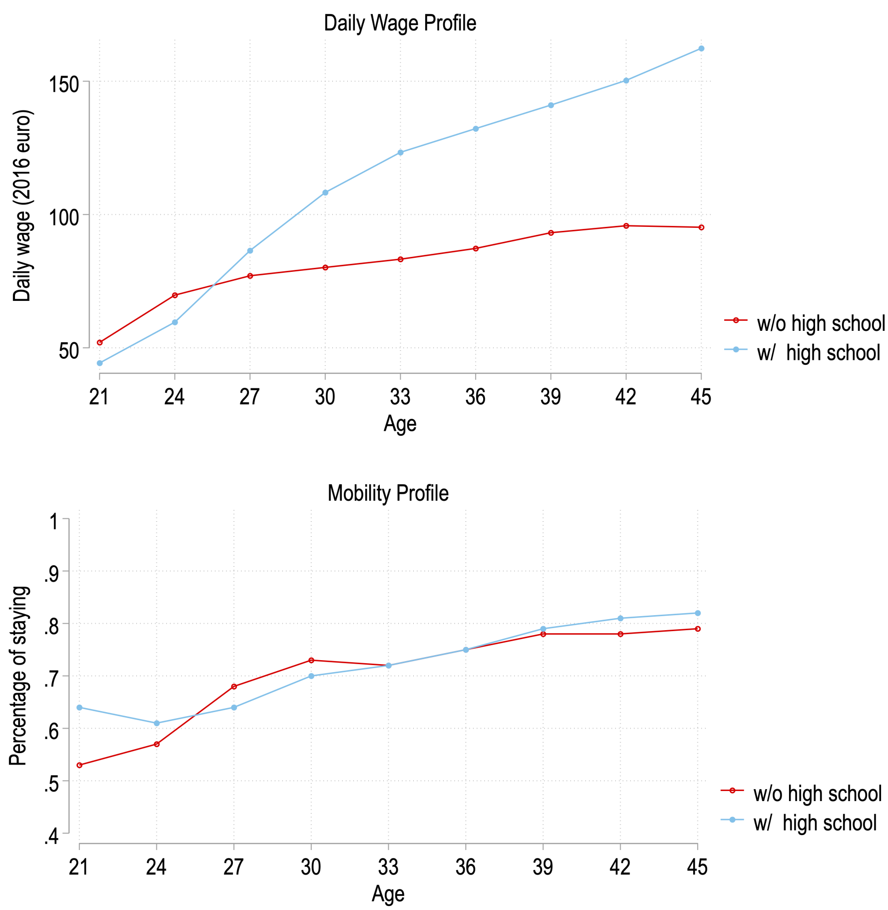
Figures

Figure 1: Task decomposition of the five occupation clusters



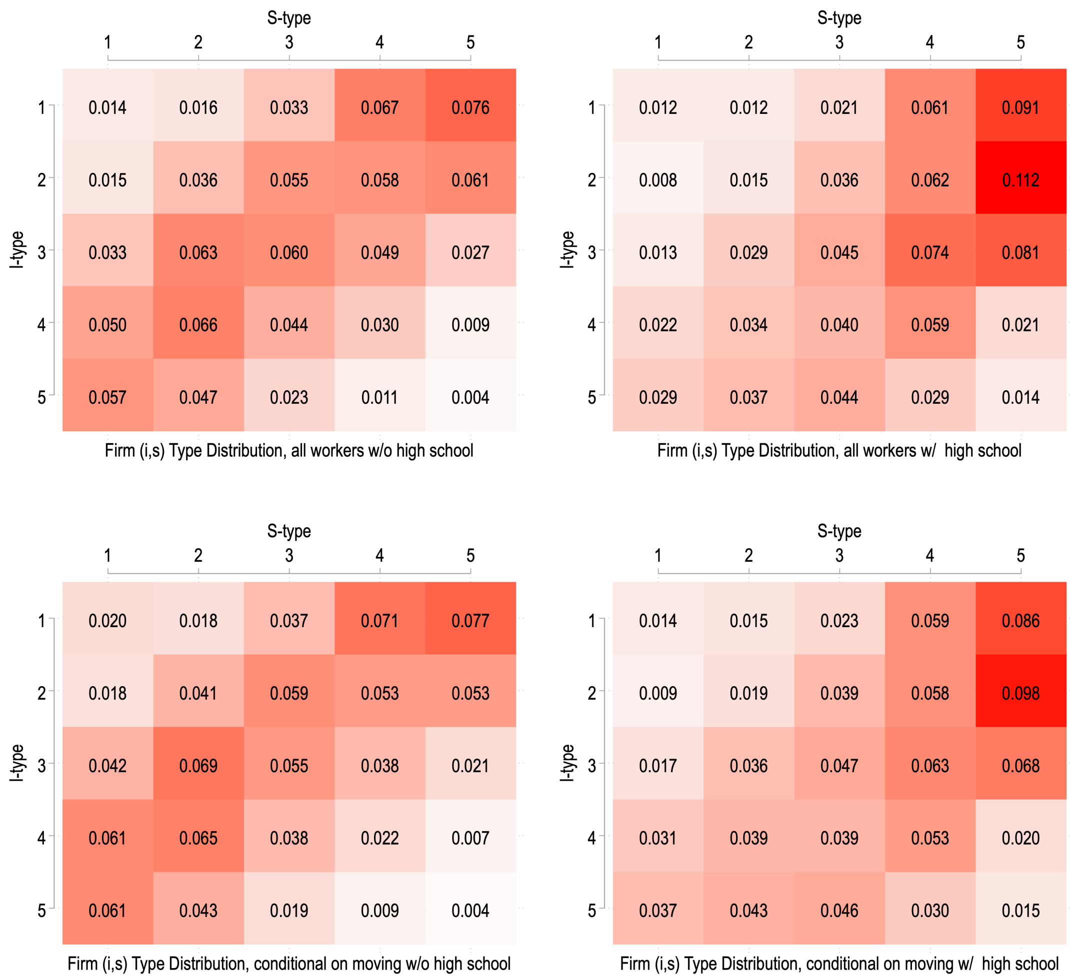
Notes. This figure shows the decomposition of the five task components of the five occupation clusters. The occupation clusters are constructed from the 36 two-digit occupations, such that the two-digit occupations that fall in the same cluster share similar task decomposition.

Figure 2: Wage and mobility profiles



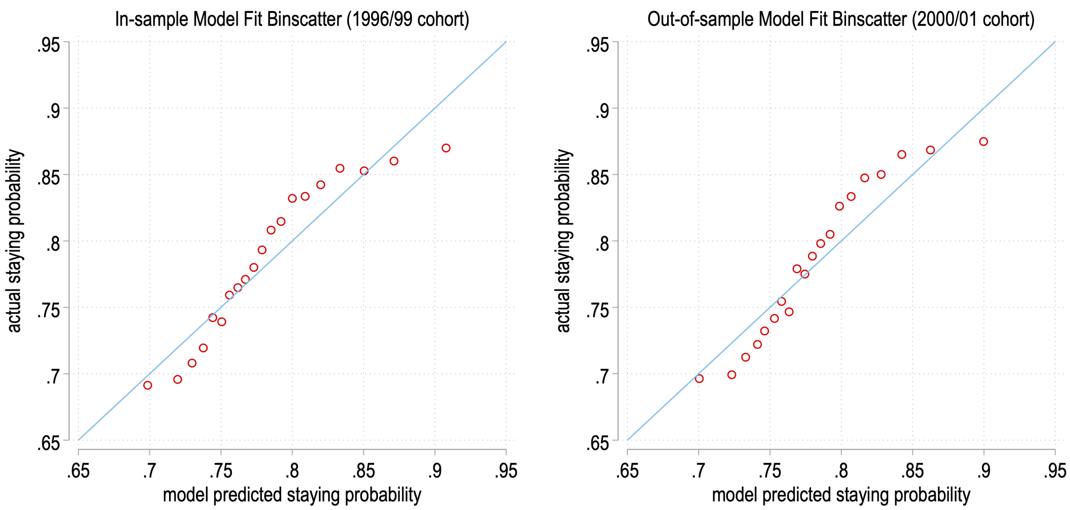
Notes. This upper figure shows the average daily-wage profiles in the life-cycle by high school status. (Unit: Euro in 2016.) The lower figure shows the average mobility profiles in the life-cycle.

Figure 3: Firm type distribution



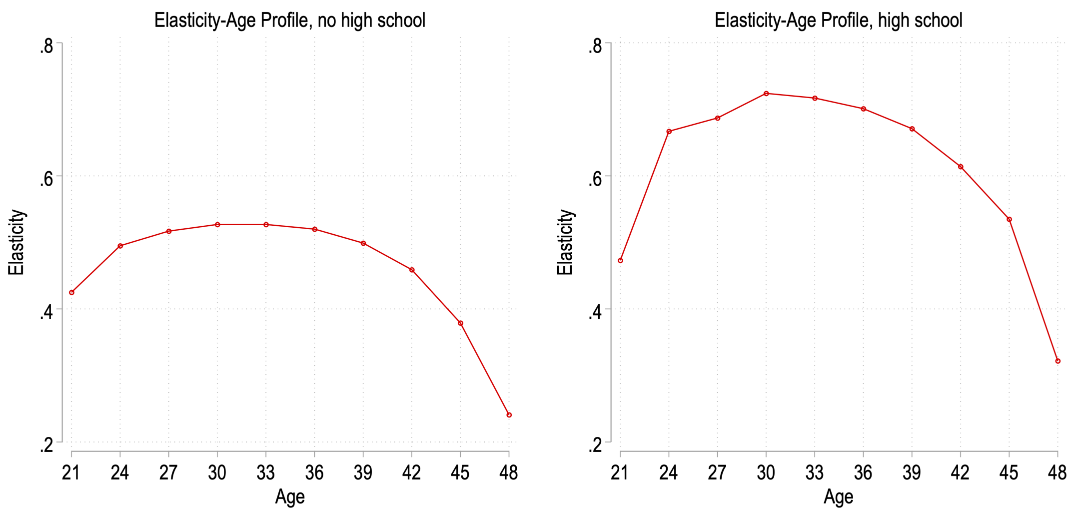
Notes. The heat maps show the distribution of workers in different types of firms. The horizontal axis refers to a firm's $i-$type, while the vertical axis refers to a firm's $s-$type. The upper panel shows the overall distribution of firm types among all workers, separated by education status. The lower panel shows the distribution of destination firm types among those who changed jobs.

Figure 4: Model fit binscatter plots



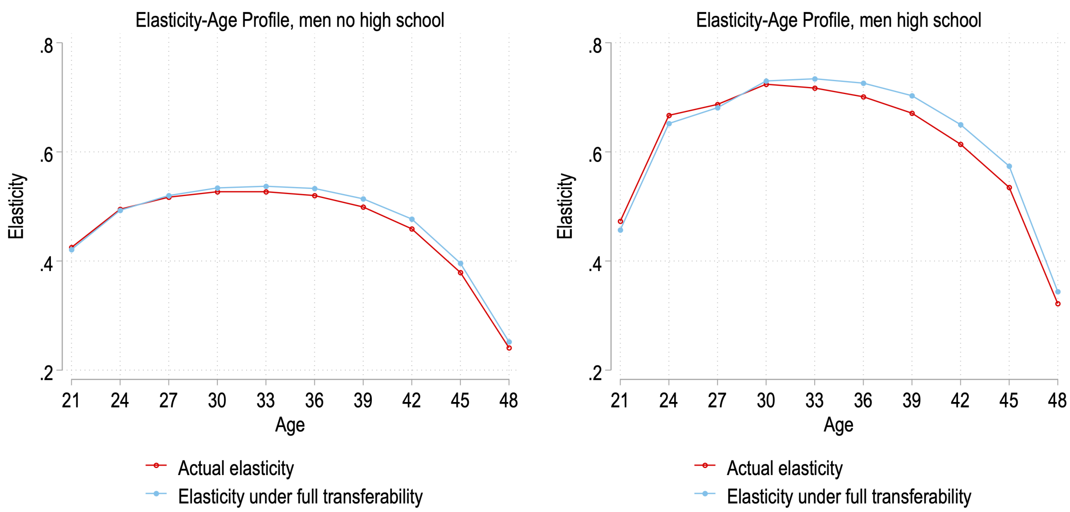
Notes. The two binscatter graphs present the in-sample model and out-of-sample fit. Along the x-axis, I clustered the predicted staying probabilities into 20 equidistant quantiles. These quantiles then align with the mean staying probabilities as indicated by the real data. The graph on the right hand side depicts the out-of-sample fit. Here, the model's forecasted staying probabilities pertain to the 2000-2001 cohort. These probabilities are compared against their corresponding actual staying probabilities.

Figure 5: Elasticity-age profile



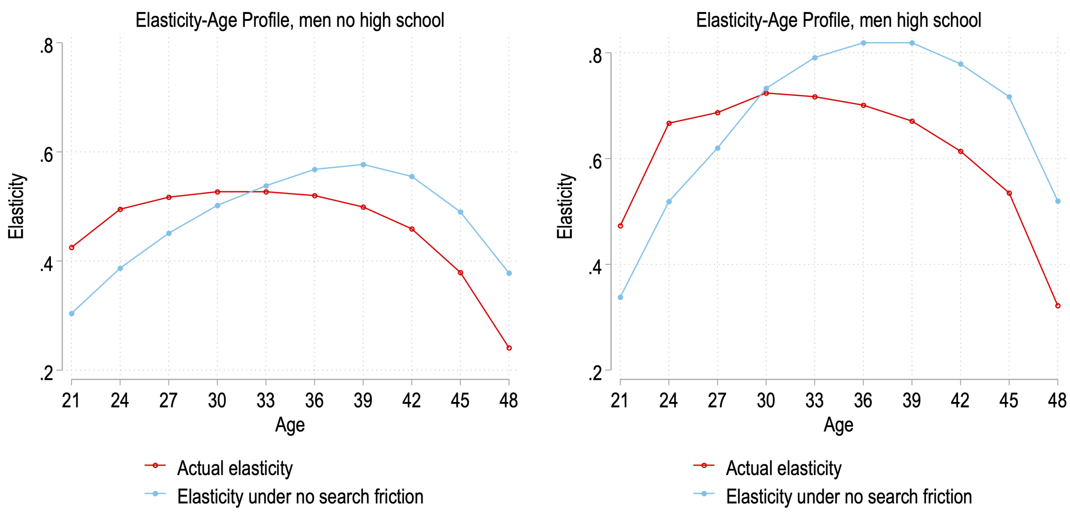
Notes. This figure shows the recovered elasticity-age profile by high school status.

Figure 6: Counterfactual elasticity under perfect human capital transferability



Notes. This figure shows the comparison between the actual elasticity and the counterfactual elasticity obtained from imposing perfect transferability.

Figure 7: Counterfactual elasticity under 100% job arrival rate



Notes. This figure shows the comparison between the actual elasticity and the counterfactual elasticity obtained from imposing a 100% job arrival rate.

Appendix A. Sample Construction and Variable Definition

This appendix describes the sample construction and variable definition. The original SIEED data covers job spells that are reported by employers to the Social Security Administration (SSA). Each spell can last between one day and one year.

Constructing Yearly Panel

In case an employment spell continues after one year, the SSA updates the spell at June.30 every year, so the dataset records a new entry representing the job spell at the current year. If the spell terminates within an one year, the SSA either records a new spell associated with the worker's new job, or leaves blank for this worker if they do not keep working. A worker, therefore, can have multiple job spells within a given year.

In case a worker has multiple job spells in a given year, I take the weighted average of their daily wage rates across all job spells, weighted by the duration of each job spell. The occupation and associated employer representing the longest job spell within a year will apply to the whole year. After this step, the cleaned data have a unique identifier: person id, year.[22]

Determining Non-full-time Employment

A worker in a given year is coded as ”non-full-time” employment if any of the following is true.

1. There is no recorded data entry for the worker in this year.
2. The duration of all job spells in this year is shorter than 180 days.
3. The average daily wage rate in this year is less than 17 euros while the worker is not undergoing vocational training.[23]

Education Status

The original school-leaving qualifications variable (schule) identifies the following categories:

1 No school leaving certificate

4 Lower secondary school certificate/ grade school certificate

6 Intermediate school leaving certificate

8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling

I recoded categories 1, 4, and 6 into "no high school" and category 8 into "high school".

[22] In the sample, the original variable names are persnr and jahr.

[23] A worker is undergoing vocational training if erwstat is either 102, 105, or 106.

Apprenticeship Status

If a person has gone through vocational training before the age 25, I record the occupation where the worker has spent the longest training time as the worker's trained occupation.

Appendix B. Model with multidimensional human capital accumulation

Setup: Suppose that workers accumulate 5 kinds of experiences in analytical non-routine tasks, interactive non-routine tasks, cognitive routine tasks, manual routine tasks, and manual non-routine tasks, respectively, denoted by

$$\overrightarrow{exp} = (exp_1, \cdots, exp_5)$$

. where the superscript m refer to occupations.

In occupation m, the wage rate is determined by the 5-dimensional experiences $\overrightarrow{exp}$:

$$w^m = p_1^m \ln(exp_1) + \cdots + p_5^m \ln(exp_5)$$

where $\vec{p}^m = (p_1^m, \cdots, p_5^m)$ is the occupation-specific prices for each of the 5 experiences. (Each occupation values each experiences differently.)

Denote the occupation task decomposition as $\vec{\lambda}^m = (\lambda_1, \cdots, \lambda_5)$, which is available from data.

When working in occupation m, every year, a worker will gain λ_k years of experience in task dimension k:

$$exp_{kt+1}^m = exp_{kt}^m + \lambda_k^m, \quad \forall k = 1,2,3,4,5$$

Discount Compare the wage rates of a worker A who has been working in occupation m for t years but transfers to occupation $\widetilde{m}$ at year t, with a worker B who has been working in occupation $\widetilde{m}$ for t years.

The wage rate of worker A is

$$w_A = \sum_{k=1}^{5} p_k^{\widetilde{m}} \ln(\lambda_k^m t) = \sum_{k=1}^{5} p_k^{\widetilde{m}} \ln(\lambda_k^m) + p_k^{\widetilde{m}} \ln(t) \tag{25}$$

whereas the wage rate of worker B is

$$w_B = \sum_{k=1}^{5} p_k^{\widetilde{m}} \ln(\lambda_k^{\widetilde{m}} t) = \sum_{k=1}^{5} p_k^{\widetilde{m}} \ln(\lambda_k^{\widetilde{m}}) + p_k^{\widetilde{m}} \ln(t) \tag{26}$$

The intuition is that as long as the pricing vector $\vec{p}^{\widetilde{m}}$ is closer to $\vec{\lambda}^{\widetilde{m}}$ than to $\vec{\lambda}^m$, Eq (2) is greater than Eq (1). The gap between (1) and (2) becomes greater when $\vec{\lambda}^{\widetilde{m}}$ and $\vec{\lambda}^m$ gets further away from each other, hence comes the task distance measure.

Special case when $\vec{\lambda} = \vec{p}$: In this case,

$$
\begin{aligned}
w_A - w_B &= \sum_{k=1}^{5} \lambda_k^{\widetilde{m}} \left(\ln(\lambda_k^m) - \ln(\lambda_k^{\widetilde{m}}) \right) \\
&\approx \sum_{k=1}^{5} \lambda_k^{\widetilde{m}} \left(\frac{\lambda_k^m - \lambda_k^{\widetilde{m}}}{\lambda_k^{\widetilde{m}}} - \frac{(\lambda_k^m - \lambda_k^{\widetilde{m}})^2}{2\lambda_k^{\widetilde{m}2}} \right) \\
&= -\sum_{k=1}^{5} \frac{(\lambda_k^m - \lambda_k^{\widetilde{m}})^2}{2\lambda_k^{\widetilde{m}}} \qquad (27)
\end{aligned}
$$

First, $w_A - w_B$ is always less than or equal to 0, because the maximum of $\sum_{k=1}^{5} \lambda_k^{\widetilde{m}} \ln(\lambda_k^m)$ is obtained when $\lambda_k^m = \lambda_k^{\widetilde{m}}$ for all $k = 1, 2, 3, 4, 5$.

Second, the wage differential $w_A - w_B$ becomes bigger when (27) is bigger. Eq (27) is a weighted distance measure between the two task decompositions λ^m and $\lambda^{\widetilde{m}}$.

In general cases, if we assume that the pricing vector $\vec{p}^m$ is closer to the task decomposition $\vec{\lambda}^m$ than to any other $\vec{\lambda}^{m'}$, the two results should still hold. But a drawback of these results is that the wage differential $w_A - w_B$ does not depend on the years of experience t.

In practice, the pricing coefficients $\vec{p}^m$ can be estimated. Based on these estimated coefficients I can test whether my current approach produces similar results to the 5-dimensional approach.

Appendix C. Estimation Details

C1. Firm Type Assignment

This paper uses a recursive Lloyd's approach to estimate the type assignment 22. Similar to the typical k-means algorithm, in each round of iteration, the I-type and S-type of each firm is updated by picking the cluster that minimizes the within-firm likelihood function.

1. **Initialization** Estimate the following equation with firm fixed effects and firm specific slopes.

$$
y_{wft} = \psi_0^f + \psi_1^f \ln(x_{wt}^m) + \psi_w + \epsilon_{wt} \qquad (28)
$$

   Obtain the distributions of the estimated coefficients: $F^{\psi_0}(\cdot)$ and $F^{\psi_1}(\cdot)$ of $\widehat{\psi}_0^f$ and $\widehat{\psi}_1^f$, respectively. Cluster the estimated $\widehat{\psi}_0^f$ into $\mathcal{I}$ quantiles and $\widehat{\psi}_1^f$ into $\mathcal{S}$ quantiles. These quantiles are the initial type assignment of the (ψ_0, ψ_1) parameters.

2. **Iterations** At step $k \geq 1$, execute the following steps until the convergence criterion is reached.

   (a) Using the estimated group assignment $\{\widehat{i}^{(k-1)}(f), \widehat{s}^{(k-1)}(f)\}_{f \in \mathcal{F}}$, change the wage equation 28 by replacing the firm specific coefficients with (i, s) specific coefficients, and obtain the parameter

$(\widehat{\psi}_{0i}, \widehat{\psi}_{1s})^{(k)}$ by estimating 29:

$$y_{wist} = \psi_0^{i^{(k-1)}} + \psi_1^{s^{(k-1)}} \ln(x_{wt}^m) + \psi_w + \epsilon_{wt} \tag{29}$$

(b) For each firm $f = 1, \cdots, \mathcal{F}$, update the I-type membership by picking the group i that maximizes the likelihood function $\widehat{L}_f$, without changing the previously assigned s-type:

$$i^{(k)}(f) = arg \max_{i \in \{1, \cdots, \mathcal{I}\}} \widehat{L}_f(\widehat{\psi}, i, s(f)^{(k-1)})$$

(c) Re-estimate the parameter ψ in:

$$y_{wist} = \psi_0^{i^{(k)}} + \psi_1^{s^{(k-1)}} \ln(x_{wt}^m) + \psi_w + \epsilon_{wt}$$

(d) For each firm $f = 1, \cdots, \mathcal{F}$, update the S-type membership by picking the group s that maximizes the likelihood function $\hat{L}_f$, without changing the previously assigned i-type:

$$s^{(k)}(f) = arg \min_{s \in \{1, \cdots, \mathcal{I}\}} \widehat{L}_f(\widehat{\psi}, i(f)^{(k)}, s)$$

(e) Assess the convergence criterion and finish the iteration if

$$\left| \widehat{L}^{(k)} - \widehat{L}^{(k-1)} \right| \leq \epsilon.$$

Given two reasons listed below, this paper estimate the wage coefficients in a separate step after type assignment is recovered.

1. The wage equation 4 involves an unobserved term $\widetilde{x}$, i.e. the occupation relevant experience. Without the knowledge of the transferability parameter γ, the wage equation cannot be estimated directly. Therefore, in order to proceed estimating type assignment without knowing γ, I will replace the latent experience $\widetilde{x}$ with the observable actual number of years worked at a specific occupation m, x_m. As a result, the $\hat{\psi}_0$ and $\hat{\psi}_1$ parameters estimated from this stage are biased due to measurement error in $\widetilde{x}$. Nevertheless, the clustering patterns of the biased parameters $(\hat{\psi}_0, \hat{\psi}_1)$ could pin down the type assignment of each firm.

2. The data used for assigning latent firm types are different to the data used to estimate the wage equations, transferability equations, and the dynamic discrete choice model. In order to maintain a constant type assignment throughout different years, and to avoid the generated regressor problem, I use the entire panel (1992 to 2018) to estimate latent firm types, while only use the cohort-wise panels to estimate the remaining parts.

C2. Transferability Equation and Wage Equation

When estimating this system of equations, there is a potential selection bias: there are unobserved factors that affect the odds of job changes, and these unobserved factors may correlate with the error terms in the wage equation. Intuitively, workers who find extra benefits associated with external jobs are more likely to change jobs, but these extra benefits may not be known by econometricians. Private connections to the new employer or unobserved skills relevant to the new job may simultaneously increase the wage rates at outside jobs and increase the likelihood of a job separation. Without controlling the selection bias, the estimation may falsely attribute the observed job switches to a high degree of human capital transferability, resulting to an underestimation of the γ parameter.

To overcome the selection problem, I add a selection equation 30.

$$\Pr(JobChange_{wt}) = \Phi(i, s, \widetilde{x}_{wt}(\gamma), \chi_w) \tag{30}$$

The unknown transferability parameter, γ, affects both the selection equation and the wage equation, therefore capturing the correlation of unobserved factors in both equations. The latent experience variable have predictive powers in both the wage rates and the job switch patterns. The EM algorithm estimates γ so that the joint likelihood of the selection equation and the wage equation is maximized.

The following paragraphs describe the procedures in estimating γ.

1. **Initialization** Set $\gamma = \gamma^0$.

2. **Iteration** Given At step $k \geq 1$, execute the following steps until the convergence criterion is reached.

   (a) Using the estimated $\widehat{\gamma}^{(k-1)}$ obtained from the previous iteration, construct the latent occupation relevant experience $\widetilde{x}_{wt}(\widehat{\gamma}^{(k-1)})$.

   (b) Use the constructed $\widetilde{x}_{wt}(\widehat{\gamma}^{(k-1)})$ to estimate the wage coefficients $(\widehat{\psi}_0^{(k)}, \widehat{\psi}_1^{(k)})$ as well as the selection equation 30. Evaluate the criterion function $\widehat{Q}^{(k)}$ (the likelihood function of the wage equation).

   (c) Iterate until the criterion function reaches optimum.

Appendix D. Identification of the Structural Model

The identification primarily focuses on two sets of parameters: those that govern the moving cost and those that govern job arrival rates. Both two parameter groups have a direct impact on the overall separation probability. However, it's crucial to note that they differ in how they influence the relative choice probabilities among external options, aside from the option to "stay". This distinction is essential as it forms the basis of the identification argument.

The following proof is based on backward induction, and the moving cost is denoted as c, whereas the job arrival rate is denoted as π.

Last Period In the final period, there is no continuation payoff; job choices therefore depends solely on current period payoff. Without loss of generosity, suppose a worker's current job is $(m, i, s) = (1, 1, 1)$. The instantaneous payoff associated with any possible choice (m, i, s) is denoted as y_{mis}. The staying probability is therefore

$$\Pr\left(\text{stay}|(1,1,1)\right) = (1-\pi) + \pi \underbrace{\sum_{i=1}^{\mathcal{I}} \sum_{s=1}^{\mathcal{S}} F_{is} \frac{exp(y_{111})}{exp(y_{111}) + \sum_m exp(y_{mis} - c)}}_{<1} \tag{31}$$

It can be seen that both c and π affect the staying probability 31 monotonically: an increase in c or a decrease in π both unambiguously increase the staying probability. Therefore, c and π cannot be separately identified solely from the staying probability.

The probability that the worker moves to job (m, i, s) is given in equation 32.

$$\Pr\left(\text{move to } (m,i,s)|(1,1,1)\right) = \pi F_{is} \frac{exp(y_{mis} - c)}{exp(y_{111}) + \sum_{\tilde{m}} exp(y_{\tilde{m}is} - c)} \tag{32}$$

Similarly, the probability that the worker moves to job (m', i', s') is given in equation 33.

$$\Pr\left(\text{move to } (m',i',s')|(1,1,1)\right) = \pi F_{i's'} \frac{exp(y_{m'i's'} - c)}{exp(y_{111}) + \sum_{\tilde{m}} exp(y_{\tilde{m}i's'} - c)} \tag{33}$$

The ratio of equation 32 to 33 is given in Equation 34.

$$\frac{\Pr\left(\text{move to } (m,i,s)|(1,1,1)\right)}{\Pr\left(\text{move to } (m',i',s')|(1,1,1)\right)} = \frac{F_{is} \dfrac{exp(y_{mis} - c)}{exp(y_{111}) + \sum_{\tilde{m}} exp(y_{\tilde{m}i's'} - c)}}{F_{i's'} \dfrac{exp(y_{m'i's'} - c)}{exp(y_{111}) + \sum_{\tilde{m}} exp(y_{\tilde{m}is} - c)}} \tag{34}$$

Equation 34 does not depend on π, and therefore is solely determined by c. The moving cost parameter c can therefore be identified from the ratios of moving probabilities towards external options.

Backward Induction In any given period $t < T$, the continuation payoffs $V' := V_{t+1}$ are already proved to identify c and π. The ratios in Equation 34 still do not depend on π, and therefore can form as an identification constraint.