The Micro and Macro Implications of Multidimensional Skill Uncertainty*

Piero De Dominicis†

September 8, 2026

Click here for the latest version.

Abstract

What are the consequences of multidimensional skill uncertainty for workers' wages and aggregate output? I develop and estimate a general equilibrium dynamic Roy model in which workers have imperfect information about their multidimensional skills and accumulate task-specific human capital. Estimated on Portuguese administrative data, the model rationalizes key patterns of occupational mobility, with learning about comparative advantage playing a central role in occupational reallocation among young and poorly matched workers. Removing information frictions raises aggregate output by 5.2%, primarily through better skill allocation across occupations, while generating the largest wage gains early in workers' careers and among the most mismatched. A feasible information treatment about comparative advantage across tasks recovers about 18% of this output loss. Finally, job transformation induced by Large Language Model adoption raises output while widening the output gap between the imperfect- and full-information economies along the transition path and in the new long run.

Keywords: Human Capital, Occupational Mobility, Multidimensional Skills, Learning, Wage Inequality

*I am indebted to Basile Grassi and Thomas Le Barbanchon for constant guidance and invaluable support, and to Jerome Adda and Luigi Bocola for many insightful conversations. I also thank Francesco Armillei, Adrien Bilal, Alexander Boca, Vasco Carvalho, Geraud Desazars, Rafael Dix-Carneiro, Jan Eckhout, Pamela Giustinelli, Cristina Gualdani, John Grigsby, Simon Jäger, Francis Kramarz, Arnaud Maurel, Romuald Meango, Andreas I. Mueller, Elena Pastorino, Nicola Pavoni, Richard Rogerson, Aysegül Sahin, Karthik Sastry, Isaac Sorkin, Wilbur Townsend and Gianluca Violante, as well as seminar participants at Bocconi, Princeton Junior Macroeconomics Workshop, Stanford Macro Lunch, Stanford Labor/Public Economics Workshop, 1st Aarhus Workshop on Labor Markets, EWMES 2025, CEPR Paris Symposium 2025, BeNA Labor Economics Workshop 2025, 2nd Boston College PhD Conference, Theories and Methods in Macro 2026, NASMES 2026, XXIX Workshop on Dynamic Macroeconomics, CEF 2026 and the NBER Summer Institute 2026 The Micro and Macro Perspectives of the Aggregate Labor Market for helpful comments. All errors are my own.

†Department of Economics, Bocconi University, piero.dedominicis@phd.unibocconi.it

1 Introduction

Workers do not possess a single one-dimensional skill but instead hold a portfolio of skills whose payoffs depend on the job that employs them. These latent multidimensional skills help explain a wide range of labor-market outcomes, including earnings trajectories and occupational choice (Heckman et al. 2006; Adda and Dustmann 2023). However, especially early in their careers, workers may have imperfect information about their skills and may therefore direct their search towards jobs based on their perceived comparative advantage (Kiss et al. 2023; Goller et al. 2026). This generates mismatch, as workers may sort into jobs that do not align with their true comparative advantage. A large literature proposes a process of experimentation and learning-by-doing, in which workers gradually discover their abilities in different jobs while accumulating imperfectly transferable experience (Miller 1984; James 2012; Antonovics and Golan 2012; Papageorgiou 2014; Pastorino 2024). Mismatch along distinct skill dimensions predicts distinct occupational moves with distinct wage and human-capital implications (Guvenen et al. 2020), and experience itself accumulates along these dimensions according to the task content of jobs (Gathmann and Schönberg 2010). Learning, imperfectly transferable human capital, and switching costs interact to create strong path dependence that prevents a substantial fraction of workers from ever sorting into their ideal career path. Quantifying these costs is crucial for understanding the importance of information frictions in generating early-career mismatch and for designing effective policies to mitigate it.

In this paper, I develop and estimate a general equilibrium dynamic Roy model to assess and quantify the consequences of imperfect information about skills for individual labor market outcomes and aggregate output. In the model, workers begin their careers with incomplete information about their multidimensional skills and gradually refine their beliefs through noisy wage signals. They also accumulate task-specific experience over time depending on which tasks they perform on the job. The productivity of skills and experience in performing a given task differs across occupations. Because of imperfect information about skills, mismatch arises endogenously, as workers sort across occupations based on beliefs rather than true comparative advantage. Two key forces drive occupational mobility in the model: learning about unobserved multidimensional skills and the accumulation of task-specific human capital. These forces, together with utility switching costs and amenities, shape both the timing and the direction of occupational transitions over the life cycle.

I estimate the model with eight occupations and two task-specific skills[1], cognitive and manual. The estimated model replicates key empirical patterns of occupational mobility, including the facts first documented by Groes et al. (2015) using Danish administrative data. Specifically, occupational mobility in Portugal is also U-shaped and directional: within an occupation, the low- and high-wage workers are more likely to switch occupation in the next period; high-wage workers move to occupations that pay more on average, while low-wage workers move down to occupations that pay less. I enrich these facts by focusing on the task content of these moves, and I show that those moving up the occupational ladder move to more cognitive occupations, while those at the bottom move to more manual occupations. The model provides a natural tool to rationalize these facts. While upward mobility of high match quality workers is

[1]Occupations are clustered based on their task content and skill requirements, mainly for identification and interpretation. More details are provided in Section 3.

mainly due to experience accumulation and non-wage related factors (amenities and idiosyncratic shocks), downward mobility of poorly matched workers is mainly driven by learning about comparative advantage.

I use the estimated model to quantify the aggregate and individual costs of information frictions about skills. I simulate a counterfactual stationary equilibrium in which workers know their true skills from entry into the labor market. I find that removing information frictions about skills would raise aggregate output by 5.2%. The individual wage gains are largest at labor-market entry, where they reach about 11.6%, and fall to roughly zero after thirty years. Almost all of the early aggregate gain comes from better sorting on skills, not from additional experience accumulation. The gains are concentrated at the bottom of the match quality distribution within occupation, as these workers are the ones most heavily mismatched. They are more likely to be specialists, that is, workers with a strong comparative advantage in one skill dimension, and therefore gain the most from sorting correctly, while generalists lose little from uncertainty.

I then study an information intervention that targets the friction directly and asks how much of the output lost to information frictions it can recover. I calibrate the intervention so that the share of workers whose beliefs about their comparative advantage match their true comparative advantage rises by 13.7 percentage points, matching the treatment effect in Kiss et al. (2023). Delivered at labor-market entry, the intervention raises aggregate output by 0.91% and lifts average wages by about 3.3% at entry, with the gains concentrated in the first decade and among the workers the baseline economy has mismatched. The output and wage gains are highest when the assessment is delivered before labor-market entry, when workers have not yet accumulated task-specific experience: the same assessment delivered two years after entering the labor market produces gains less than half of those obtained when the same signal is provided before starting their career. Furthermore, I show that reducing switching costs has non-monotonic effects on output: on the one hand, small reductions in utility switching costs increase output, as this allows workers to reduce their mismatch as they discover their skills; on the other hand, large reductions create excessive churning that destroys accumulated experience in tasks, lowering workers' productivity.

Finally, I use the model to study how information frictions about skills affect occupational reallocation to technology shocks that change the importance of specific tasks within occupations. I use the model to simulate the effects of job transformation (Freund and Mann 2025) induced by Large Language Model (LLM) adoption on aggregate output and welfare. The AI shock lowers time spent on cognitive tasks in proportion to each occupation's exposure to large language models. At the same time, it raises the importance and, in turn, the time spent on manual tasks. The shock is modeled as an unanticipated change with perfect foresight over a transition path that can be instantaneous or gradual. The shock raises aggregate output (+4.3%) at every adoption speed, with faster adoption front-loading the gains. Crucially, the shock makes skill uncertainty more costly: the output gap between the imperfect- and full-information economies widens along the transition, from about 5.2% before the shock to roughly 5.9% in the new long run.

2 Related Literature

This paper contributes to four strands of the literature. First, this paper speaks to the literature on the sources of wage growth and the misallocation of human capital (Hsieh et al. 2019). A large empirical literature documents the importance of job and occupational mobility for wage growth (Topel and Ward 1992; Neal 1999; Kambourov and Manovskii 2009; Pavan 2011; Adda and Dustmann 2023). I focus on learning about multidimensional skills and task-specific human capital accumulation as the two forces that drive reallocation to more productive matches and the wage dynamics it generates. Two recent papers document the role of information frictions in generating occupational mismatches. Goller et al. (2026) link standardized aptitude tests to administrative records in Switzerland and document that imperfect self-knowledge at labor-market entry predicts misaligned occupational aspirations, higher realized skill mismatch, and a higher probability of dropout. Kiss et al. (2023) show, in two field experiments with young jobseekers, that giving workers their assessment results redirects search toward jobs that reward the skills they score well on and substantially raises earnings, consistent with mismatch driven by imperfect information. I simulate the short-run effects of the same intervention in my estimated model to trace its general-equilibrium and life-cycle consequences.

Closest to this paper is a strand that treats this misallocation as multidimensional, arising from a mismatch between a worker's several skills and the skills her occupation rewards (Guvenen et al. 2020; Lise and Postel-Vinay 2020; Baley et al. 2022). Guvenen et al. (2020) derive a multidimensional mismatch measure from a model of Bayesian learning about workers' learning ability, but do not estimate such a model. Lise and Postel-Vinay (2020) estimate a structural on-the-job search model with multidimensional skills, but workers know their skills, so mismatch is mainly driven by search frictions. Baley et al. (2022) let workers learn about unobserved skills, but each job uses a single skill dimension and their focus is on business-cycle dynamics rather than life-cycle sorting. In my model, learning is à la Jovanovic (1979) and Miller (1984): workers forecast a vector of time-invariant task-specific skills, and a worker is "mismatched" when, given current beliefs, she chooses an occupation that is not the one in which she is truly most productive. The approach does not require a measure of the supply of skills, such as test scores; I instead recover the distribution of unobserved skills directly from wages, using the approach of Bunting et al. (2024). Estimating the model then lets me quantify the micro and macro costs of mismatch and how much of them policy can recover.

Second, this paper contributes to the literature on learning, job assignment, and wage growth. In the terminology of Pastorino (2024), learning affects wages both directly—pay responds to beliefs about ability because it is an equilibrium outcome of what employers and workers know[2]—and indirectly, as workers and firms learn about comparative advantage and reallocate workers to jobs where their skills are used more productively. Pastorino (2024) finds that

[2]Much of the employer-learning literature emphasizes this direct channel, among others Harris and Holmström (1982), Farber and Gibbons (1996*a*), Altonji and Pierret (2001), Lange (2007*a*), and Gibbons and Katz (1991). A closely related strand develops frameworks that incorporate both wage-setting effects and assignment dynamics, among others the integrated theories of wage and promotion dynamics inside firms in Gibbons and Waldman (1999*b*) and subsequent extensions (e.g. Gibbons and Waldman (2006), and the synthesis in Gibbons and Waldman (1999*a*)), as well as sectoral and occupational sorting frameworks in which learning affects wages both directly and through mobility (e.g. Gibbons et al. (2005*a*)).

this indirect, assignment channel can be the dominant force shaping wage growth and dispersion: the direct belief-to-wage effects are relatively small, while learning substantially improves the sorting of workers to jobs, which then drives wage dynamics[3]. Motivated by this, and by the occupational mobility facts in Groes et al. (2015), I focus on the assignment channel and study how learning about multidimensional unobserved skills shapes occupational mobility and wages over the life cycle. A related structural literature studies occupational sorting when workers learn about occupation-specific abilities (James 2012; Arcidiacono et al. 2025; Papageorgiou 2014; de Melo and Papageorgiou 2025); Papageorgiou (2014) in particular shows how search frictions amplify the mismatch that learning generates. In all of these, ability is a single latent factor or one scalar per occupation, so the models are silent about which skill dimension underlies mismatch. Antonovics and Golan (2012) shows that occupational choice and wage growth in the NLSY are consistent with workers learning about their abilities on the job; their framework, however, abstracts from on-the-job human capital accumulation and non-wage amenities, both of which are central here.

This paper departs from this literature by embedding learning in a task-based model of occupational choice. I show empirically that mobility across tasks is informative about different types of occupational switches, and that these switches have distinct implications for wages and human capital accumulation. The model also enriches occupational choice by allowing for non-compensating differentials, a utility switching cost, and preference heterogeneity, which improves its quantitative fit and helps isolate the separate roles of learning and human capital accumulation in generating job mobility.

Third, this paper connects to the literatures on task-specific human capital (Gibbons and Waldman 2004; Poletaev and Robinson 2008; Gathmann and Schönberg 2010). The model in this paper is similar in spirit to the task-based frameworks developed by Gathmann and Schönberg (2010) to study the transferability of human capital across occupations, and Autor and Handel (2013), who proposed a Roy (1951) model where workers sort based on comparative advantage across tasks. In my framework, as in these task-based approaches, tasks are bundled within occupations and workers' skills map into output through occupation-specific task loadings. However, unlike these papers, I focus on how workers' occupational allocation is shaped by imperfect information about multidimensional skills. The task-based approach is also the workhorse for studying technological change: routine-biased technical change and job polarization (Autor et al. 2003, 2006; Autor and Dorn 2013; Acemoglu and Autor 2011). More recently, Freund and Mann (2025) develop a general-equilibrium task-based model to study the effects of job transformation induced by Large Language Models (Eloundou et al. 2024), in which automating a task raises the importance of those that remain, and estimate the resulting individual wage effects from a distribution of task-specific skills. However, as in related general-equilibrium dynamic Roy models studying technological change (Dvorkin and Monge-Naranjo 2019; Althoff and Reichardt 2026), workers there know their own skills exactly. In this paper, I study how job transformation affects mismatch driven by information frictions about skills, and how skill uncertainty affects occupational reallocation to this technological shock.

[3]More recently, de Paula et al. (2025) build on Pastorino (2024) with a model of heterogeneous workers and firms, monopsony power, and symmetric learning, and find that standard static methods underestimate the importance of sorting because they ignore the dynamic effects of learning and human capital accumulation on future matching.

Fourth, this paper contributes to the literature on estimating Roy-type models (Roy 1951) to study the effects of macroeconomic shocks when there are labor mobility frictions. Building on the identification results of Heckman and Honore (1990), dynamic Roy frameworks have been applied to sectoral switching costs (Lee and Wolpin 2006), labor-market responses to trade liberalization (Dix-Carneiro 2014; Traiberman 2019), and learning about occupation-specific ability (James 2011; Hincapié 2020; Arcidiacono et al. 2025). I contribute by introducing a two-stage estimation strategy that accommodates multidimensional unobserved heterogeneity, building on the nonparametric identification results of Bunting et al. (2024) for what they call the pure learning model[4].

The rest of the paper is organized as follows. In Section 3 I describe the data used and some of the key empirical facts on occupational mobility. In Section 4 I introduce the model, while in Section 5 I discuss identification and describe the estimation procedure. Finally, I present the quantitative results in Section 6.

3 Data and Empirical Facts

3.1 Data

The main data source of this paper is *Quadros de Pessoal*, a matched employer–employee dataset created by the Portuguese Ministry of Labor in the 1980s. It includes Portuguese firms with at least one employee, excluding self-employed workers. The sample used covers the period 1995–2019. The main wage measure used in the paper is a "straight-time real hourly wage," constructed following Card et al. (2016)[5]. Nominal wages are deflated to 1995 euros using the Portuguese CPI.

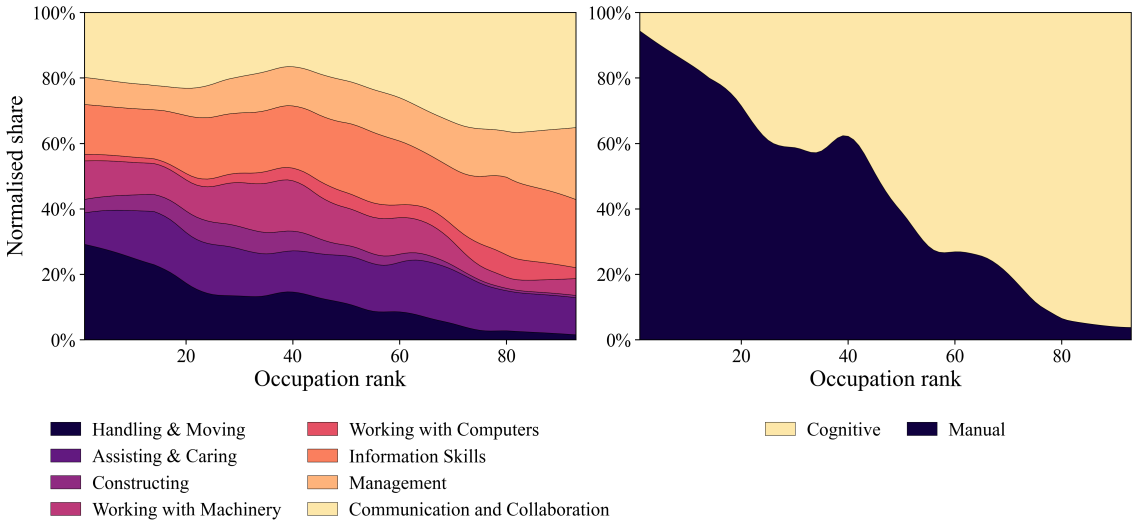
I use two related samples throughout the paper, each serving a different purpose. Detailed descriptions of the sample construction and their comparison are provided in Appendix E. The large sample is designed to maximize the number of observations and representativeness for descriptive facts and baseline comparisons. It applies standard cleaning criteria to the raw data and is used to document the overarching empirical patterns of occupational mobility at the 3-digit occupation level. The estimation sample is a cohort panel of labor market entrants, constructed by applying further restrictions to the large sample. It tracks the complete early-career trajectories of young workers entering the labor market, which is essential for proper identification and estimation of the learning and human capital parameters of the structural model.

To measure the task content of occupations, I use two classifications. The first focuses on broad

[4]As I describe later in Section 5, identification in this class of models requires fewer restrictions than a standard Roy model. Moreover, parts of the functional form can accommodate heterogeneity observed by workers but unobserved to the econometrician; while this complicates identification and estimation in general, the estimation routine can incorporate such heterogeneity by extending the EM algorithm for finite mixtures. To retain tractability for the dynamic programming problem and to ease interpretation, I focus on the simpler pure learning model.

[5]This measure is computed as the sum of the worker's base wage and regular salary supplements (i.e. monetary meal allowances) and is then divided by the worker's normal hours, defined as the contractual monthly hours fixed by the individual contract. The measure is defined as "straight-time" in the sense that both wage and hours exclude overtime.

task categories, cognitive and manual skills[6]; the second classification is ESCO, the European Skills, Competences, Qualifications and Occupations dataset, which links occupations to specific skills across European labor markets. In both classifications, each occupation j receives weights ω_j^k over tasks/skills k that sum to one, $\sum_k \omega_j^k = 1$. Leveraging these measures allows for the comparison of workers within occupations and the analysis of how task content changes during occupational transitions.



(a) Task Composition by Occupation Average Wage Rank

Figure 1: Occupations are ranked by mean hourly wage within each year and mapped to percentiles (horizontal axis: 1=lowest, 100=highest; vertical axis: skill share in percent). *Left panel*: Each occupation's ESCO skill vector is normalized to shares that sum to 100%. *Right panel*: Cognitive and Manual task shares based on the cognitive/manual task classification. Both panels show smoothed average skill shares at each wage percentile. The figure is produced using the Large Sample data.

To understand how skills are distributed across the wage distribution, Figure 1 reports the average normalized share of tasks using two complementary classifications. The left panel uses the eight ESCO skills, while the right panel employs the cognitive/manual task classification. In both panels, shares sum to 100% at every point on the x-axis, and occupations are ordered by their mean wage within a year: low-paying occupations on the left, high-paying on the right.

Two facts stand out. First, although some occupations require workers to possess only one skill, on average, occupations bundle multiple skills rather than specializing in a single one across the entire wage distribution. Second, there is a clear shift toward more cognitive skills as wages rise. In the left panel, the shares of lighter-colored layers—communication & collaboration, management skills, information skills, and working with computers—increase with occupational wage rank. In contrast, the darker layers associated with more manual content—handling and moving, working with machinery, constructing, and assisting and caring—decline with occupational wage rank. The right panel confirms this pattern more starkly: cognitive tasks rise with occupational wage rank while manual tasks decline.

Taken together, the figure highlights both (i) horizontal diversity: a broad skill mix at every wage rank, and (ii) a vertical task ladder: higher-paying occupations tilt increasingly toward

[6]Detailed construction of these categories and the classification mapping over time is provided in Appendix E.

cognitive skills.

3.2 The Task Content of Occupational Mobility

Occupational transitions are not random. Following Gathmann and Schönberg (2010), I define skills distance as

$$d(j,j') = 1 - \text{Sim}(j,j'), \qquad \text{Sim}(j,j') = \frac{\boldsymbol{\omega}_j^\top \boldsymbol{\omega}_{j'}}{\|\boldsymbol{\omega}_j\|_2 \, \|\boldsymbol{\omega}_{j'}\|_2}, \tag{1}$$

where $\boldsymbol{\omega}_j$ collects the skill weights of occupation j, and $\text{Sim}(j,j') \in [0,1]$ equals 1 for identical and 0 for orthogonal skill mixes.[7] The mean skills distance among actual occupation switchers is 0.52, with a standard deviation of 0.22. The closest common move is between laborers and building frame workers ($d = 0.21$), while the most distant observed switch is between mathematicians and domestic helpers ($d = 0.99$). The most popular switch for low-educated workers is between laborers and building frame workers, while for the high-educated it is between accounting clerks and bookkeepers ($d = 0.36$) or between physical science technicians and engineers ($d = 0.43$).[8] Figure 2 compares the observed distribution of skills distance against a random mobility counterfactual, where each switcher leaving occupation j in year t is assigned a destination drawn proportional to employment shares, following Gathmann and Schönberg (2010). Observed movers switch to occupations that are substantially closer in task space than random reallocation would predict (Panel (a); the mean distance between old and new occupation is 0.52 vs. 0.68). The same holds in the coarser Cognitive–Manual task space of the eight occupation clusters used in the structural model (Panel (b)): observed moves stay closer in cognitive–manual content than the random benchmark. Workers thus sort across occupations in a way that reflects the task content of jobs, not just their availability.

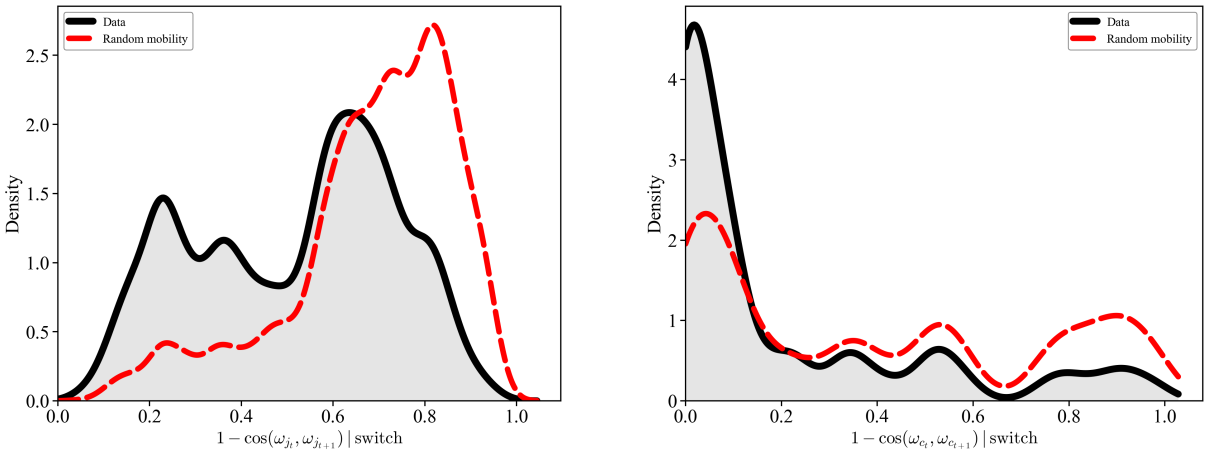
The administrative data, which covers nearly the universe of workers within each occupation, allows me to go further and characterize how mobility patterns vary across the within-occupation wage distribution. As Groes et al. (2015) show, this coverage is essential to uncover the heterogeneous nature of occupational transitions. To isolate the component of within-occupation wage variation driven by unobserved fit, I residualize wages from observable characteristics. Specifically, I estimate:

$$\log w_{it} = \alpha_j + \phi_t + f(\text{tenure}_{it}) + g(\text{experience}_{it}) + \gamma \text{Female}_i + \delta \text{Educ}_i + \mu_{ij} + \nu_{it} \tag{2}$$

where α_j are occupation fixed effects, ϕ_t are time fixed effects, $f(\cdot)$ and $g(\cdot)$ are cubic polynomials in occupational tenure and total labor market experience, μ_{ij} is the unobserved worker-occupation match quality, and ν_{it} is mean-zero noise. The estimated residual is $\widehat{\epsilon}_{it} = \widehat{\mu}_{ij} + \widehat{\nu}_{it}$.

[7] Skills distance is computed using the ESCO classification at different levels of aggregation: 8 broad skill dimensions (Matrix 1.3), 74 intermediate dimensions (Matrix 2.3), and the full 296-skill classification (Matrix 3.3). The figure in the main text uses the 296-skill classification. For interpretability, when displaying changes in skill content upon switching (Figure 4), I use the 8 aggregated ESCO skill dimensions. In addition, I use the Cognitive–Manual classification of Mihaylov and Tijdens (2019), which will serve as the task space for the structural estimation. Appendix Figure 26 shows that the patterns are robust across all granularity levels and hold for both mean and median task distance.

[8] See Appendix Table 5 for detailed summary statistics on the skills distance distribution, including the most common switches by education group.

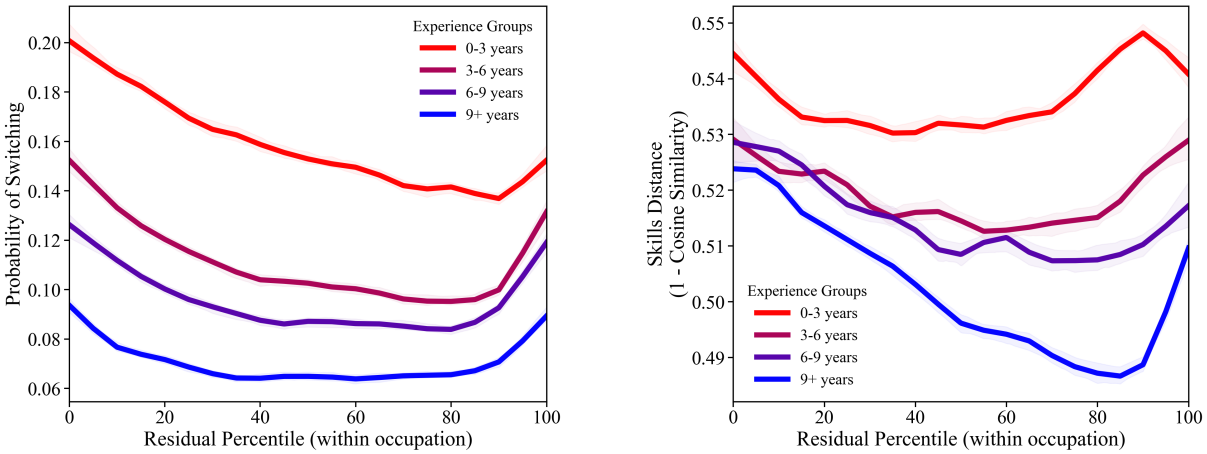


(a) ESCO skill distance (296 dimensions)

(b) Cognitive–Manual task distance (occupation clusters)

Figure 2: Task distance of occupational switches, observed (black) versus a random mobility counterfactual (red, dashed) in which each switcher's destination is drawn proportional to employment shares, following Gathmann and Schönberg (2010). Task distance is $1 - \cos(\boldsymbol{\omega}_{j_t}, \boldsymbol{\omega}_{j_{t+1}})$. *Panel (a)*: the 296-dimension ESCO skill space, between three-digit occupations. *Panel (b)*: the Cognitive–Manual task space of Mihaylov and Tijdens (2019), between the eight occupation clusters used in the structural model. In both, observed moves are closer in task space than random reallocation would predict.

Since ν_{it} is mean-zero, the percentile rank of $\widehat{\epsilon}_{it}$ within occupation j is positively correlated with μ_{ij}: workers with higher match quality tend to rank higher in the residual distribution. The "within-occupation residual percentile" used throughout this section is therefore a noisy but informative proxy for the worker's position in the match quality distribution of the current occupation. Appendix E shows that all patterns documented here are robust to using raw hourly wages instead of residualized wages.



(a) Probability of Switching by Experience

(b) Skills Distance for Occupation Switchers

Figure 3: *Panel 3(a)*: Probability of switching occupations between t and $t+1$ (vertical axis) as a function of residualized within-occupation wage percentile (horizontal axis), for different experience groups. *Panel 3(b)*: Skills distance between origin and destination occupation conditional on switching (vertical axis) by residualized within-occupation wage percentile (horizontal axis), computed using ESCO skill measures.

Figure 3(a)[9] shows two facts about occupational mobility. First, workers at both ends of the within-occupation wage distribution switch at higher rates, resembling the "U-shape" documented by Groes et al. (2015) for Denmark. Second, mobility falls sharply with experience, from roughly 15% for workers with 0–3 years of experience to below 10% after 6 years. Mobility at the bottom is mainly driven by low-educated workers, while high-educated workers tend to rank higher in the residual wage distribution. Figure 3(b) shows that the task distance between origin and destination occupations is related to the worker's match quality in the origin occupation. Workers with lower match quality make larger moves in task space, while workers with higher match quality make smaller moves. Task distance also falls with experience. This pattern is hard to reconcile with undirected search frictions, where the size of the move should be unrelated to the origin mismatch conditional on switching. As Guvenen et al. (2020) argue, if mismatch is induced by search frictions alone, the change in mismatch upon switching should be independent of the mismatch in the origin occupation. The relationship between match quality and the size of the task move is instead consistent with directed sorting, where workers learn about their comparative advantage and reallocate toward better-fitting occupations.

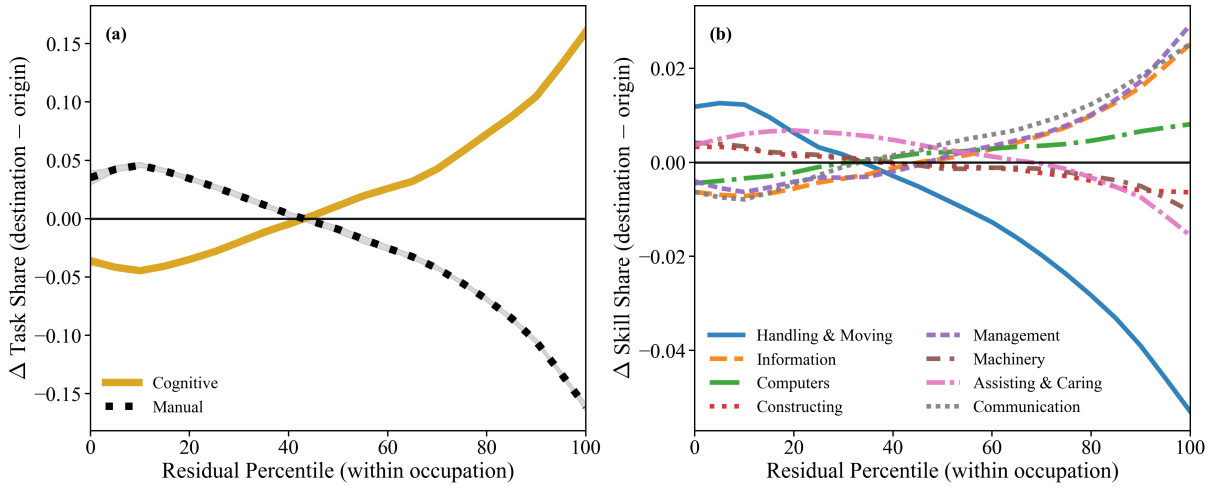


Figure 4: *Panel (a)*: Change in cognitive and manual task shares (vertical axis) by residualized within-occupation wage percentile (horizontal axis) for occupational movers. Change is $\Delta\text{Task-Share}_p^k = \text{LOWESS}_p\left(\omega^k_{j'(i)} - \omega^k_{j(i)}\right)$, the smoothed difference in task intensities between destination $j'(i)$ and origin $j(i)$ occupations. *Panel (b)*: Analogous changes based on the ESCO skill classification. By construction, given that $\sum_k \omega_j^k = 1$, the non-parametric average shares per percentile sum up to 0.

Occupational mobility in Portugal is also directional, using the terminology introduced by Groes et al. (2015). High-ability workers within an occupation are more likely to move to an occupation with a higher average wage compared to their current occupation, while low-ability workers move to an occupation with a lower average wage (Appendix Figure 27). However, when moving, low-ability workers improve their relative position in the destination occupation, while high-ability workers worsen their relative position.

Figure 4 examines the task content of occupational transitions. Workers at different positions in the match quality distribution move to occupations with different task content: high match

[9] All figures in this paper use non-parametric LOWESS regression with bandwidth 0.2. 95% confidence intervals are computed using 100 bootstrap iterations. In many figures the confidence bands are barely visible due to the precision of the estimates.

quality workers move toward more cognitive occupations, while low match quality workers move toward more manual ones. The ESCO classification (right panel) shows a similar pattern at a finer level: well-matched workers move to occupations requiring more communication, collaboration, creativity, and computer skills, while poorly matched workers move to occupations requiring more handling, moving, or caring skills. These patterns hold for both within-firm and across-firm switches (Appendix Figure 30).

Appendix E.2 documents that workers realize substantial wage gains upon switching occupations. These gains vary systematically with a worker's initial position within their occupation: those in the bottom quintile experience cumulative wage increases exceeding 30% over four years, while workers at the top initially suffer wage cuts but recover to modest gains of 10% by year four. This pattern indicates that switching is especially beneficial for workers poorly matched to their initial occupation. The delayed wage recovery for high earners suggests either backloaded compensation or non-pecuniary benefits that influence job choice (Sorkin 2018).

3.3 Discussion of the Empirical Facts

Taken together, Figures 3(b) and 4 show that workers at both ends of the match quality distribution change tasks more when they switch, with the largest moves made by workers at the bottom. The fact that, conditional on moving, the size of the task content change depends on origin match quality is a distinctive feature of the data. Workers who discover a poor fit have stronger incentives to reallocate toward occupations with different task content, and even well-matched workers who switch do so toward occupations with somewhat different task requirements. This is consistent with learning about comparative advantage. If mismatch arises from undirected search, the size of the task change upon switching should be independent of the origin mismatch (Guvenen et al. 2020). The pattern documented here points instead toward directed sorting, where occupational choices are guided by evolving beliefs about comparative advantage across tasks. These findings call for a theory that accounts for the relationship between match quality, mobility, and the task content of transitions. The model in the next section incorporates multidimensional skills, learning about comparative advantage, and task-specific human capital to quantify the consequences of this reallocation mechanism on wages over the life cycle.

4 Structural Model

The economy is populated by a continuum of workers and firms. Time is discrete and in each period t firms post piece rate wage contracts in occupations $j \in \mathcal{J}$ and workers choose an occupation j in which to work.

4.1 Occupations, Tasks and Human Capital

Workers supply multidimensional human capital that is used to perform different tasks $k \in \mathcal{K}$ within each occupation $j \in \mathcal{J}$. Worker i is endowed with a vector of task-specific skills $\boldsymbol{\theta}_i = (\theta_{ik})_{k\in\mathcal{K}}$ and accumulates task-specific experience $\boldsymbol{\tau}_{it} = (\tau_{ikt})_{k\in\mathcal{K}}$ while performing tasks on the

job. The output of a specific task k in a given occupation j at time t is given by:

$$Y_{ijkt} = \ell_{jk}\, \theta_{ik}^{a_{jk}}\, H_{jk}(\tau_{ikt}), \tag{3}$$

where ℓ_{jk} is the share of time allocated to task k in occupation j, θ_{ik} is worker i's latent skill in task k, and task-specific experience enters productivity through a reduced-form learning-by-doing function $H_{jk}(\tau)$: a worker who spends more time on task k accumulates task-specific know-how that is productive wherever the task is used, with returns that may differ across occupations. The function is concave in logs, disciplining diminishing returns to learning-by-doing.[10] The primitive parameters have clean interpretations: a_{jk} is the elasticity of task-k output with respect to skill θ_{ik}, and $H_{jk}(\tau)$ captures the task-k learning-by-doing schedule in occupation j. When employed in occupation j at time t, worker i uses this human capital to perform tasks $k \in \mathcal{K}$ and produces task output. Tasks aggregate within each occupation j according to a Cobb-Douglas technology with task weights ω_{jk}:

$$Y_{ijt} = \exp(\varepsilon_{ijt}) \prod_{k\in\mathcal{K}} \left(\frac{Y_{ijkt}}{\omega_{jk}}\right)^{\omega_{jk}}, \tag{4}$$

where ω_{jk} governs how intensively occupation j uses task k, with $\sum_{k\in\mathcal{K}} \omega_{jk} = 1$, and $\exp(\varepsilon_{ijt})$ is an idiosyncratic worker-occupation-time productivity shock with $\varepsilon_{ijt} \sim H_j$ i.i.d. across workers and time. Following Althoff and Reichardt (2026), I assume that the firm chooses the static allocation of worker time across tasks optimally under the symmetric belief $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ it shares with the worker, and that this allocation is non-contractible. Under the Cobb-Douglas aggregation, Appendix C.1 shows that the optimum is $\ell_{jk}^{*} = \omega_{jk}$. Task allocation is therefore independent of worker skills and experience, and depends only on the task importance ω_{jk} of the occupation. Substituting this into (4) yields

$$Y_{ijt} = \underbrace{\prod_{k\in\mathcal{K}} \theta_{ik}^{\lambda_{jk}}}_{\substack{\text{unobserved}\\ \text{skills bundle}}} \quad \underbrace{H_j(\boldsymbol{\tau}_{it})}_{\substack{\text{human-capital}\\ \text{bundle}}} \quad \underbrace{\exp(\varepsilon_{ijt})}_{\substack{\text{idiosyncratic}\\ \text{shock}}}, \tag{5}$$

where the effective elasticities at occupation j on task k are

$$\lambda_{jk} \;=\; \omega_{jk}\, a_{jk}, \tag{6}$$

and $\boldsymbol{\omega}_j = (\omega_{jk})_{k\in\mathcal{K}}$ collects the occupation's task weights. The mapping in (6) implies that each effective per-occupation–task elasticity is the Cobb-Douglas aggregator weight ω_{jk} times the corresponding primitive task-level elasticity. The occupation-level human-capital bundle aggregates the task-level functions via the Cobb-Douglas weights:

$$H_j(\boldsymbol{\tau}_{it}) \;\equiv\; \prod_{k\in\mathcal{K}} H_{jk}(\tau_{ikt})^{\omega_{jk}}. \tag{7}$$

[10] In the estimation, I parameterize the log returns to be a stepwise function: $h_{jk}(\tau) = \sum_{b\in\mathcal{B}} \delta_{jk}^{(b)} \mathbf{1}\{\tau \in b\}$, where $\mathcal{B}$ is a set of tenure bins. In this way, I make sure that I identify the wage progression along different task dimensions in each occupation flexibly as pinned down by the data.

Taking logs, we define

$$h_j(\boldsymbol{\tau}_{it}) \;\equiv\; \log H_j(\boldsymbol{\tau}_{it}) \;=\; \sum_{k\in\mathcal{K}} \omega_{jk}\, h_{jk}(\tau_{ikt}), \tag{8}$$

the additive log-form of the human-capital bundle. Notational convention: uppercase H denotes the multiplicative level, lowercase h its log. Taking logs of (5) then yields the log-output expression

$$\log Y_{ijt} \;=\; \boldsymbol{\lambda}_j'\boldsymbol{\xi}_i \;+\; h_j(\boldsymbol{\tau}_{it}) \;+\; \varepsilon_{ijt}, \tag{9}$$

with $\boldsymbol{\xi}_i = \log \boldsymbol{\theta}_i$. The occupation-specific parameters $\boldsymbol{\lambda}_j$, $h_{jk}(\cdot)$ and $\boldsymbol{\omega}_j$ jointly determine how worker human capital maps into occupation j output. They govern the relative importance of skills, experience, and the concavity of experience across tasks, and through the multiplicative structure they pin down the degree of complementarity across these components. This specification implies that worker skills are bundled in the sense that, once matched to an occupation, a worker must perform all tasks, unless $\omega_{jk} = 0$ or the elasticities are zero for some task k. The log-form bundle $h_j(\boldsymbol{\tau}_{it})$ enters the wage equation and the signal definition below.

Finally, task-specific experience evolves deterministically. Conditional on working in occupation j, the time spent on task k is given by ω_{jk}; therefore, the vector of task-specific tenure evolves following the linear specification

$$\boldsymbol{\tau}_{i,t+1} = \boldsymbol{\tau}_{it} + \boldsymbol{\omega}_j. \tag{10}$$

4.2 Labor Demand and Wage Equation

In each period t there is a continuum of competitive firms operating in each occupation j. Firms take the occupation output price in levels P_{jt} as given and compete for workers by posting output-contingent piece rate wage contracts.[11] Firms in occupation j choose employment across worker types to maximize profits

$$\max_{\{n_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\}} \Pi_{jt} = P_{jt}Y_{jt} - \int W_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, n_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, dG_t(\boldsymbol{\theta},\boldsymbol{\tau}_t), \tag{11}$$

subject to the occupation production function

$$Y_{jt} = \int n_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, Y(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, dG_t(\boldsymbol{\theta},\boldsymbol{\tau}_t), \tag{12}$$

[11] This output-contingent contracting environment differs from the standard employer-learning literature in which wages (or wage offers) depend directly on employers' beliefs about worker productivity, so that observed wages reflect employers' posterior expectations (e.g. Farber and Gibbons (1996*b*), Gibbons et al. (2005*b*), Lange (2007*b*), Pastorino (2024)). In my framework, competitive piece rates tie pay to realized output; learning affects outcomes primarily through occupational sorting and continuation values rather than through belief-based wage-setting. Similar output-contingent wage contracts are used in Groes et al. (2015) and are consistent with empirical approaches that combine workers' choices with realized outcomes, as in Bunting et al. (2024). This wage contracting can also be viewed as a special case of the general wage contract in Camargo et al. (2022), $w_{ijt} = c_{ijt} + b_{ijt}y_{ijt} = (1 - b_{ijt})\,\mathbb{E}\left[y_{ijt} \mid I_{ijt}\right] + b_{ijt}y_{ijt}$, with $b_{ijt} = 1$, so that wages depend one-for-one on realized output (up to the occupation-specific output price scaling in my model) rather than on posterior expectations.

where $G_t(\boldsymbol{\theta}, \boldsymbol{\tau}_t)$ is the cross-sectional distribution of worker types at time t. Here $Y(\boldsymbol{\theta}, \boldsymbol{\tau}_t)$ denotes the worker-occupation-specific output implied by the task technology in equation (4), which already embeds the idiosyncratic productivity shock $\exp(\varepsilon_{ijt})$. Under piece rate contracts and perfect competition, firms post contracts until the wage paid equals realized worker-specific revenue, that is, the value of realized marginal product

$$W_{ijt} = P_{jt}\, Y_{ijt}. \tag{13}$$

Taking logs yields the log wage equation

$$w_{ijt} \equiv \log W_{ijt} = p_{jt} + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_{it}) + \varepsilon_{ijt}, \tag{14}$$

where $p_{jt} \equiv \log P_{jt}$ and $\boldsymbol{\xi}_i = (\log \theta_{ik})_{k \in \mathcal{K}}$. Log wages are determined by four components: (i) the log occupation price p_{jt}; (ii) the worker's latent skill bundle $\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i$, where the occupation-specific loadings $\boldsymbol{\lambda}_j$ determine how task-specific skills map into output in occupation j; (iii) the log human-capital bundle $h_j(\boldsymbol{\tau}_{it})$, capturing the returns from accumulated task-specific experience; and (iv) the idiosyncratic productivity shock ε_{ijt}.

4.3 Learning and Information Structure

Workers belong to an observed education group $g \in \mathcal{G}$. The objects of learning are the worker's latent log skills $\boldsymbol{\xi}_i \equiv \log \boldsymbol{\theta}_i$ introduced in Section 4.1. Upon entry into the labor market,

$$\boldsymbol{\xi}_i \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{15}$$

so that $\boldsymbol{\theta}_i$ is lognormal conditional on g. Workers are Bayesian and at the beginning of period t hold a subjective belief density $\tilde{f}_{it}(\boldsymbol{\xi})$ over $\boldsymbol{\xi}_i$ given their information set $\mathcal{I}_{it}$, which contains group membership and the history of past occupational choices, wages, and experience. I assume rational initial beliefs in the sense that the initial prior coincides with the true group distribution,

$$\tilde{f}_{i0}(\boldsymbol{\xi}) = f_{\xi \mid g(i)}(\boldsymbol{\xi}) = \phi\Big(\boldsymbol{\xi}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\Big), \tag{16}$$

where $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal density.

After choosing occupation j at time t and observing the wage-implied signal

$$s_{ijt} \equiv w_{ijt} - p_{jt} - h_j(\boldsymbol{\tau}_{it}) = \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}, \tag{17}$$

beliefs update according to Bayes' rule

$$\tilde{f}_{it+1}(\boldsymbol{\xi}) = \frac{f_{s \mid \xi}\big(s_{ijt} \mid \boldsymbol{\xi}\big)\, \tilde{f}_{it}(\boldsymbol{\xi})}{\int f_{s \mid \xi}\big(s_{ijt} \mid \boldsymbol{\xi}'\big)\, \tilde{f}_{it}(\boldsymbol{\xi}')\, d\boldsymbol{\xi}'}. \tag{18}$$

Assume $\varepsilon_{ijt} \sim \mathcal{N}(0, \varsigma_{\varepsilon j}^2)$. Since the prior at $t = 0$ is multivariate normal and the signal in (17) is linear Gaussian, Bayes' rule implies that posteriors remain multivariate normal at all dates. Therefore the relevant beliefs can be summarized by a mean vector $\tilde{\boldsymbol{\xi}}_{it}$ and a variance-

covariance matrix $\tilde{\boldsymbol{\Sigma}}_{it}$ such that

$$\boldsymbol{\xi}_i \mid \mathcal{I}_{it} \sim \mathcal{N}\left(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}\right), \qquad \tilde{\boldsymbol{\xi}}_{i0} = \boldsymbol{\mu}_{g(i)}, \quad \tilde{\boldsymbol{\Sigma}}_{i0} = \boldsymbol{\Sigma}_{g(i)}. \tag{19}$$

Under normality, Bayes updating can be written in the Kalman filter form

$$\tilde{\boldsymbol{\xi}}_{it+1} = \tilde{\boldsymbol{\xi}}_{it} + \boldsymbol{K}_{ijt}\left(s_{ijt} - \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\xi}}_{it}\right), \tag{20}$$

$$\tilde{\boldsymbol{\Sigma}}_{it+1} = \left(\boldsymbol{I}_{|\mathcal{K}|} - \boldsymbol{K}_{ijt}\boldsymbol{\lambda}_j'\right)\tilde{\boldsymbol{\Sigma}}_{it}, \tag{21}$$

where $\boldsymbol{K}_{ijt}$ is the Kalman gain defined by

$$\boldsymbol{K}_{ijt} \equiv \tilde{\boldsymbol{\Sigma}}_{it}\boldsymbol{\lambda}_j\left(\boldsymbol{\lambda}_j'\tilde{\boldsymbol{\Sigma}}_{it}\boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2\right)^{-1}. \tag{22}$$

Because $\tilde{\boldsymbol{\Sigma}}_{it}$ may have nonzero off-diagonal elements, observing s_{ijt} updates beliefs about all skill dimensions even if an occupation loads heavily on only a subset of skills.

4.4 Occupational Choice

Let the worker state at the beginning of period t be

$$x_{it} \equiv \left(g(i),\, \tilde{\boldsymbol{\xi}}_{it},\, \tilde{\boldsymbol{\Sigma}}_{it},\, \boldsymbol{\tau}_{it},\, d_{it-1},\, \{p_{j\ell}\}_{j\in\mathcal{J},\ell\geq t}\right), \tag{23}$$

where $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ are beliefs about latent log skills, $\boldsymbol{\tau}_{it}$ is task experience, and d_{it-1} is the last period's occupation. The sequence $\{p_{j\ell}\}_{\ell\geq t}$ denotes log occupation prices from period t onward. I assume workers have perfect foresight over the future path of occupational prices and take $\{p_{j\ell}\}$ as given. Figure 5 below summarizes the within-period timing of the worker's choice and the resulting wage realization.

In each period t worker i chooses an occupation $j \in \mathcal{J}$ before observing the current wage shock. I assume log utility over wages, $u(W) = \log W$, and a switching cost $\mathcal{C}(j, d_{it-1})$ that is paid whenever the destination occupation j differs from the origin occupation d_{it-1}. The realized current-period utility from choosing j is

$$U_{ijt}(x_{it}) = \log W_{ijt} + \eta_{jg(i)} - \mathcal{C}\left(j, d_{it-1}\right) + \epsilon_{ijt}, \tag{24}$$

where W_{ijt} is the wage paid in levels, $\eta_{jg(i)}$ is a non-pecuniary compensating differential specific to occupation j and the worker's education group $g(i)$, and ϵ_{ijt} is i.i.d. Extreme Value Type I with scale ϱ.[12] The switching cost is incurred only when the worker changes occupation, and I

[12]The taste shocks ϵ_{ijt} serve two purposes. First, they add idiosyncratic noise to the occupational choice in every period. When entering in the labor market, all workers in a given education group enter with the same prior beliefs and no experience, so without the shocks every group would make the same choice. The shocks instead disperse entrants across occupations and, in later periods, help rationalize occupational choices and moves that the systematic part of the model cannot account for. Appendix C.4 illustrates the entry choice with and without the shocks. Second, they smooth the integrated value function and the conditional choice probabilities, which keeps the Bellman fixed point and the likelihood well-behaved for estimation.

parametrize it as a single utility constant:

$$\mathcal{C}\left(j, d_{it-1}\right) = \kappa_0 \, \mathbf{1}_{\{j \neq d_{it-1}\}}. \tag{25}$$

When switching occupation, workers have to pay a cost of κ_0 utils. The cost captures, in reduced form, all the non-wage related factors involved with changing occupation. For example, search and retraining effort, or the psychological cost of changing the type of job one does. Keeping this cost parsimonious allows us to understand the role of learning about task-specific skills and accumulating task-specific experience in generating the observed empirical patterns related to the task content of occupational mobility. Because the wage equation already makes moves costly through the task-specific human capital that a switcher leaves behind, κ_0 absorbs only the residual, non-pecuniary component of the cost of moving. Under log utility and the wage equation (14), expected utility is linear in expected log productivity:

$$\mathbb{E}_{\boldsymbol{\xi}_i, \varepsilon_{ijt}}\left[\log W_{ijt} \mid \tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}\right] = p_{jt} + \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_{it} + h_j(\boldsymbol{\tau}_{it}).$$

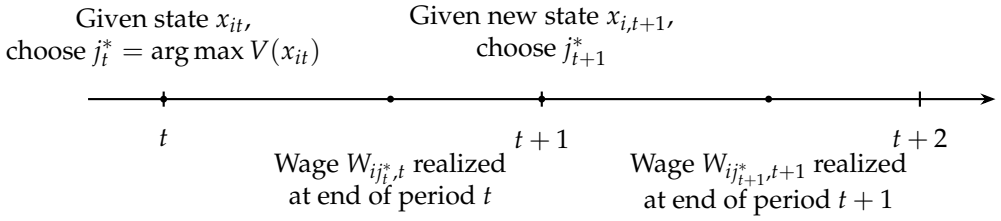


Figure 5: Timing of choices and wage realizations.

Dropping the worker index for clarity, and emphasizing dependence on the price path, the worker's dynamic program is

$$\begin{aligned}
V_t\left(g, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\right) = \max_{j \in \mathcal{J}} \Big\{ & \mathbb{E}_{\boldsymbol{\xi}, \varepsilon_{jt}}\left[\log W_{ijt} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t\right] + \eta_{jg} - \mathcal{C}\left(j, d_{t-1}\right) + \epsilon_{jt} \\
& + \beta \, \mathbb{E}_{\boldsymbol{\xi}, s_{j_t t}, \boldsymbol{\epsilon}_{t+1}}\left[V_{t+1}\left(g, \tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, j; \{p_{j\ell}\}_{\ell \geq t+1}\right)\right] \Big\}, \\
\text{subject to:} \quad \boldsymbol{\tau}_{t+1} &= \boldsymbol{\tau}_t + \boldsymbol{\omega}_j, \\
s_{jt} &= \boldsymbol{\lambda}_j' \boldsymbol{\xi} + \varepsilon_{jt}, \\
\tilde{\boldsymbol{\xi}}_{t+1} &= \tilde{\boldsymbol{\xi}}_t + \boldsymbol{K}_{jt}\left(s_{jt} - \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_t\right), \\
\tilde{\boldsymbol{\Sigma}}_{t+1} &= \left(\boldsymbol{I}_{|\mathcal{K}|} - \boldsymbol{K}_{jt} \boldsymbol{\lambda}_j'\right) \tilde{\boldsymbol{\Sigma}}_t, \\
\text{given} \quad \left(\tilde{\boldsymbol{\xi}}_0, \tilde{\boldsymbol{\Sigma}}_0, \boldsymbol{\tau}_0, d_{-1}\right) &= \left(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \mathbf{0}, \varnothing\right) \text{ for an entrant of group } g.
\end{aligned} \tag{26}$$

The flow expectation integrates over the worker's posterior uncertainty about her latent skills $\boldsymbol{\theta}_i$ (and the zero-mean wage shock ε_{ijt}, which drops out under log utility); the continuation expectation integrates over the predictive distribution of the wage signal $s_{ij_t t}$, which drives the Bayesian update of beliefs, and the next-period taste shock vector $\boldsymbol{\epsilon}_{i,t+1}$. The next-period covariance is deterministic given the chosen occupation, so the only stochastic state evolution is the posterior mean $\tilde{\boldsymbol{\xi}}_{t+1}$. Integrating out ϵ yields the multinomial logit conditional choice

probabilities:

$$\mathbb{P}_t\left(d_t = j \mid g, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\right) = \frac{\exp\left(v_{jt}\left(g, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\right)/\varrho\right)}{\sum_{m \in \mathcal{J}} \exp\left(v_{mt}\left(g, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\right)/\varrho\right)}. \quad (27)$$

With a continuum of workers, these conditional choice probabilities represent the mass of workers in each state choosing each occupation. The education group g is a permanent characteristic of each worker and it shifts the payoff of every occupation through the occupation-specific amenities η_{jg}. In the empirical implementation, rather than treating past occupational choice and education group as discrete states, I solve for a set of value functions, where each value function is occupation-education specific. Accordingly, the choice probabilities are also computed by education group. More details are provided in Appendix D.

4.5 Closing the Model

The model is closed by introducing a representative final goods producer who aggregates occupation outputs $\{Y_{jt}\}_{j \in \mathcal{J}}$ into final output Y_t using a CES technology with elasticity of substitution $\sigma > 0$:

$$Y_t = \left(\sum_{j \in \mathcal{J}} \alpha_j Y_{jt}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \qquad \alpha_j > 0. \quad (28)$$

The aggregator weights $\{\alpha_j\}_{j \in \mathcal{J}}$ govern the relative importance of each occupation's output in final production. I do not impose the simplex constraint $\sum_j \alpha_j = 1$; instead the overall scale of $\{\alpha_j\}$ serves as the price-level normalization, as detailed below. Let P_t denote the price of the final good and P_{jt} the price of occupation j output. Profit maximization by the final goods producer yields the inverse demand function for occupation j output:

$$P_{jt} = P_t \, \alpha_j \left(\frac{Y_t}{Y_{jt}}\right)^{\frac{1}{\sigma}}. \quad (29)$$

This expression determines occupation prices as a function of the final-good price and relative outputs. Taking the final good as the numeraire ($P_t = 1$, Section 4.6) gives occupation prices directly.

4.6 Stationary Competitive Equilibrium

I define a stationary competitive equilibrium in which occupation prices are constant over time. With a continuum of workers, let $\pi(x)$ denote the mass of workers in state $x = (g, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$ (excluding prices), which augments the decision state (23) with the worker's true latent skill $\boldsymbol{\xi}$[13]. Entrants are characterized by group masses π_g (with $\sum_{g \in \mathcal{G}} \pi_g = 1$) and initial distributions

[13]The worker's problem depends on the state only through the decision-relevant sub-vector $(g, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$: the value function and conditional choice probabilities condition on g through the amenity η_{jg} and on beliefs through expected payoffs, and they do not depend on $\boldsymbol{\xi}$ directly. True skill enters on the aggregation side, determining realized output $Y_j(x)$ and, through the realized wage signal $s_j = \boldsymbol{\lambda}_j' \boldsymbol{\xi} + \varepsilon_j$, the Bayesian belief update.

$\pi_g^0(\cdot)$ that draw true skill from the group prior, $\boldsymbol{\xi} \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, set rational initial beliefs $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}) = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, and start from $\boldsymbol{\tau} = \mathbf{0}$. A stationary competitive equilibrium[14] consists of constant occupation prices $\{P_j\}_{j \in \mathcal{J}}$, a stationary mass distribution $\pi^*(\cdot)$ over worker states, aggregate output Y, occupation outputs $\{Y_j\}_{j \in \mathcal{J}}$, a value function $V(\cdot)$, and conditional choice probabilities $\mathbb{P}(d = \cdot \mid \cdot)$ satisfying the following conditions:

1. **Labor supply.** Taking the constant prices $\{P_j\}_{j \in \mathcal{J}}$ as given, workers solve the Bellman equation (26) with constant discount factor β, generating the value function $V(\cdot)$ and conditional choice probabilities (27). Experience accumulates according to (10), and beliefs update via Bayes' rule (18). In every period, workers face a constant and exogenous probability δ of leaving the labor market (this acts as a death shock). This exit risk is not internalized by workers and serves only to generate a well-defined stationary distribution: workers who exit are replaced by an entrant with skills drawn from the same education group-g skill distribution $\boldsymbol{\xi} \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$.

2. **Labor demand and wages.** For each occupation $j \in \mathcal{J}$, aggregate output equals the integral of worker output contributions over the mass of workers:

$$Y_j = \int \left(\int \exp(\varepsilon)\, Y_j(x)\, dH_j(\varepsilon) \right) \mathbb{P}(d = j \mid x)\, \pi^*(x)\, dx, \tag{30}$$

   where $Y_j(x)$ is the task technology (5) evaluated at the worker's true skill $\boldsymbol{\xi}$ and task experience $\boldsymbol{\tau}$, H_j is the distribution of the wage shock ε, and $\mathbb{P}(d = j \mid x)\pi^*(x)$ gives the mass of workers in state x choosing occupation j. Wages satisfy the piece-rate adjusted revenue condition (14) in each occupation.

3. **Price determination.** Occupation prices satisfy the inverse demand system (29) given aggregate output Y and occupation outputs $\{Y_j\}_{j \in \mathcal{J}}$. Because CES inverse demand pins down only relative prices, I take the final good as the numeraire and normalize its price to one. The price of the final good equals the dual CES unit-cost index of the occupation outputs, so the numeraire is

$$P = \left(\sum_{j \in \mathcal{J}} \alpha_j^{\sigma}\, P_j^{1-\sigma} \right)^{\frac{1}{1-\sigma}} = 1, \tag{31}$$

   the cost function dual to the aggregator (28).[15] Rather than rescaling the weights to the simplex, I choose the scale of $\{\alpha_j\}$ so that (31) holds exactly at the Stage-1 estimated prices (Section 5), which makes the estimated price vector itself the stationary equilibrium under the $P = 1$ numeraire.

4. **Stationary distribution.** The mass distribution π^* is stationary (invariant), satisfying

$$\pi^*(x') = \sum_{j \in \mathcal{J}} \int \left[(1 - \delta)\, Q(x' \mid x, j) + \delta \sum_{g \in \mathcal{G}} \pi_g\, \pi_g^0(x') \right] \mathbb{P}(d = j \mid x)\, \pi^*(x)\, dx, \tag{32}$$

[14] I provide the algorithm used to compute the stationary competitive equilibrium in Appendix C.2.

[15] The dual index carries α_j^{σ}, not α_j; this is the unique scale-one unit cost consistent with (28) and its first-order conditions (29).

for all states $x' \in \mathcal{X}$, where $Q(\cdot \mid x, j)$ is the transition density over worker states induced by choosing occupation j: true skill $\boldsymbol{\xi}$ and group g are carried forward unchanged, task experience accumulates via (10), and the belief moments update by Bayes' rule (18) from the realized wage signal $s_j = \boldsymbol{\lambda}_j' \boldsymbol{\xi} + \varepsilon_j$, whose distribution is governed by the worker's true skill; and $\pi_g^0(\cdot)$ is the initial mass distribution for entrants from group g[16].

In practice the equilibrium is computed as a damped fixed point in the vector of occupation log prices: given a price guess I solve the worker's dynamic program, simulate a stationary panel of workers (with exogenous exit and entrant replacement) to obtain occupation outputs $\{Y_j\}$, update prices from the CES inverse demand (29) subject to the numeraire normalization (31), and iterate to convergence. Appendix C.2 details the algorithm.

5 Identification and Estimation

To ensure a computationally tractable state space dimension, I estimate the model with two task specific skills: cognitive and manual. Additionally, I aggregate the 103 three-digit ISCO occupations into $K = 8$ clusters. Clusters are designed by minimizing the within-cluster dispersion of the cognitive and manual task content across occupations within a given cluster, while being penalized for distorting the original empirical occupational mobility graph. Details for the clustering algorithm and the resulting groups are provided in Appendix E.1. While this aggregation mechanically lowers measured mobility levels compared to the 3-digit ISCO analysis, the empirical facts documented in Section 3 remain qualitatively consistent within this clustered state space, as shown in the robustness checks in Appendix E.

5.1 Identification of the Wage Equation under Learning

This section establishes identification of the wage equation in the presence of multidimensional skills that are unobserved to the workers and the econometrician. The argument follows closely Arcidiacono et al. (2025) and is based on the formal nonparametric identification result of Bunting et al. (2024), adapted to the multidimensional-skill setting of this paper. The key identifying idea is that the model is a pure-learning environment, in the sense of Bunting et al. (2024): workers sort on beliefs about latent skills, and those beliefs are formed from the same wage and occupational histories observed by the econometrician. Therefore, conditional on the worker's history, occupational choice selects on no additional unobserved component, so that controlling for the worker history is enough to solve the selection problem.

Let $D_{it} \in \mathcal{O}$ denote worker i's occupation in period t, let $g_i \in \mathcal{G}$ denote the worker's observed group, and let $\boldsymbol{\tau}_{it}$ be predetermined at the time of the occupational choice. The potential log wage in occupation j is

$$W_{it}(j) = p_j + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}, \tag{33}$$

[16]Due to the learning structure, the dynamic problem of workers is Markovian. Hence under standard regularity conditions, the death-and-rebirth mechanism makes the worker-state dynamics regenerative and causes initial conditions to wash out over time. The process therefore admits a unique invariant distribution π^*, and simulation averages converge to expectations under π^*. Appendix C.2 describes the simulation procedure used to approximate this distribution in practice.

where $\boldsymbol{\xi}_i \in \mathbb{R}^K$ is a time-invariant latent skill vector and the observed wage is $W_{it} = W_{it}(D_{it})$. The piece-rate assumption implies that, after subtracting the occupation price and the human-capital component, both worker and econometrician observe a noisy linear measurement of $\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i$.

The result below adapts Bunting et al. (2024)'s pure-learning identification theorem to the multidimensional-skill model presented in the previous section. The key idea is that workers select on beliefs that are themselves formed from the same wage and occupational histories the econometrician observes, so that

$$D_{it} \perp\!\!\!\perp \boldsymbol{\xi}_i \mid \mathcal{H}_{it},$$

where $\mathcal{H}_{it} \equiv (g_i, D_i^{t-1}, W_i^{t-1}, \boldsymbol{\tau}_i^t)$ is the observed history conditional on the observed education level g_i. Conditional on the history, occupational choices carry no additional information about latent skills, so the endogenous-occupational-choice problem reduces to one of selection on observables. Combined with standard rank and support conditions and the anchor and weighted-zero-mean normalizations, this restriction delivers point identification of the wage-equation parameters $\{(p_j, \boldsymbol{\lambda}_j, \delta_j, \varsigma_{\varepsilon j}^2)_{j \in \mathcal{J}}, (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g \in \mathcal{G}}\}$ from the joint distribution of observed wages, occupational choices, groups, and predetermined state variables (Theorem 1 in the appendix).

The formal identification argument is deferred to Appendix A.1 and follows the same inverse-probability weighting proposed by Bunting et al. (2024). Intuitively, the "history-specific" CCPs serve as inverse-probability weights that recover the potential wage path under any supported occupational choice sequence. Under the linear-Gaussian wage equation this recovered path is multivariate normal: its means identify occupation prices plus an education-specific mean, and its within-worker covariances identify the latent skill covariance and the occupation-specific loadings. Repeated observations in the same occupation separate the permanent component from transitory noise and identify the occupation-specific task-experience returns. Occupation switchers are needed in order to identify how different occupations load on the common skill vector. As in standard factor models, the scale of loadings and factors is indeterminate, and they are not separately identifiable unless we impose some restrictions. Because we leave the variance-covariance matrix of group-specific skills unrestricted, we need to impose two restrictions on the loadings. For $K = 2$, we use an anchor occupation to fix the cognitive coordinate and one that fixes the manual coordinate; covariances of all other occupations with these two anchors recover their cognitive and manual loading vectors.

This identification argument can equivalently be viewed as a control-function argument: the observed history $\mathcal{H}_{it}$ is a sufficient control for endogenous occupational choice, so that the wage regression in expectation becomes

$$\mathbb{E}\left[W_{it}(j) \mid \mathcal{H}_{it},\, g_i,\, D_{it} = j\right] \;=\; p_j + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \, \mathbb{E}\left[\boldsymbol{\xi}_i \mid \mathcal{H}_{it},\, g_i\right],$$

where the posterior mean $\mathbb{E}[\boldsymbol{\xi}_i \mid \mathcal{H}_{it}, g_i]$ enters as a control regressor and absorbs selection based on beliefs.

5.2 Identification of the Dynamic Choice Parameters

Stage 1 identifies all the parameters from the wage equation, so for every supported state x_{it} and every occupation j the econometrician can compute the expected log wage $\tilde{w}_j(x_{it})$ and the distribution of the wage signals, hence the law of motion of beliefs. Importantly, I assume that workers are rational in the sense that their beliefs upon entry into the labor market coincide with the true distribution of skills identified in the first step. This assumption, together with the identified wage signal and the law of motion of beliefs, allows the econometrician to compute the exact sequence of beliefs for every worker in the sample. Therefore, CCPs are identified directly from the joint distribution of choices and states. Because the data are generated by a stationary equilibrium with constant occupation prices, the choice-specific value index and the conditional choice probabilities are time invariant and I write $v_j(x)$ and $P_j(x)$. The state x contains the education group g, so v_j, P_j and the ex-ante value W can be defined as group-specific. With β fixed externally and ϱ calibrated, the only unknown parameters driving the dynamic choice left in the model are the amenities η_{jg} and the single switching cost κ_0 of (25). We can identify these parameters using the level and composition of mobility across destinations observed in the data. Under the Type-I extreme value assumption, the observed log-odds recover differences in the choice-specific value function,

$$\varrho \, \log \frac{P_j(x)}{P_m(x)} = v_j(x) - v_m(x),$$

and the same logit structure that produces (27) can also be inverted, following Hotz and Miller (1993), to express the ex-ante value function W in terms of any single choice using a probability correction term, that is $W(x) = v_j(x) - \varrho \log P_j(x)$ for any j. The correction $-\varrho \log P_j(x)$ is the gap between the ex-ante value and the value of committing to j in a particular state x: it is small when j is chosen often, and hence close to optimal, and large when j is chosen rarely.

Applying this inversion recursively along the path on which the worker keeps choosing the occupation she has just chosen yields the representation of Theorem 1 in Arcidiacono and Miller (2011), which writes the conditional value function as a discounted sequence of expected wages, each corrected by the choice-probability term $-\varrho \log P_j$. Intuitively, we can express continuation value using CCPs that are directly identified from the data rather than through the unknown structural parameters. Collecting terms, the choice-specific value function is linear in the two unknowns,

$$v_j(x) = \underbrace{\tilde{w}_j(x) + \beta \, \Psi_j(x)}_{\text{known from Stage 1 and the CCPs}} + \frac{\eta_{jg}}{1-\beta} - \kappa_0 \, \mathbf{1}_{\{j \neq d\}}, \qquad d \equiv d_{t-1} \text{ in } x,$$

where $\Psi_j(x)$ is the discounted stream of expected wages net of the choice-probability correction accumulated along that path, known from Stage 1 and the observed CCPs (Lemma 1 in Appendix A.2 gives its expression). Continuation values are thus not unknown parameters, but rather they are a linear function of unknown parameters with an intercept that can be computed using wage equation parameters and observed CCPs.

Two comparisons then separate the two kinds of parameter. Comparing two destinations $j \neq d$ and $m \neq d$ reached from a common origin d, a move pays κ_0 regardless of where the worker

goes, so the switching cost cancels and the log-odds isolate the amenity difference:

$$\eta_{jg} - \eta_{mg} = (1-\beta)\left[\varrho \log \frac{P_j(x)}{P_m(x)} - \big(\tilde{w}_j(x) - \tilde{w}_m(x)\big) - \beta\big(\Psi_j(x) - \Psi_m(x)\big)\right].$$

It is thus the relative frequency with which workers leaving a common origin go to each different destination that identifies amenities, and the normalization $\eta_{0g} = 0$ within each education group turns these differences into levels. Comparing instead a move to j with staying in d, the amenities are already known and the switching cost is the only remaining unknown:

$$\kappa_0 = \big(\tilde{w}_j(x) - \tilde{w}_d(x)\big) + \beta\big(\Psi_j(x) - \Psi_d(x)\big) + \frac{\eta_{jg} - \eta_{dg}}{1-\beta} - \varrho \log \frac{P_j(x)}{P_d(x)}.$$

Every term on the right is known, so κ_0 is identified. The two parameters are separated by a single feature of the data, whether the chosen occupation coincides with the origin.

The scale ϱ is conceptually separate from amenities and switching costs. Because expected log wages are measured in utility units, ϱ could in principle be identified from variation in expected wage differences. In this learning model, however, the same state variables that shift current expected wages also shift future beliefs and continuation values, so identification of ϱ is fragile without payoff variation that moves current utility but not the continuation value. I therefore calibrate a value of ϱ that allows the estimated model to match the average occupational mobility rate in the large sample rather than estimating it jointly by maximum likelihood. Given this calibration, the normalization $\eta_{0g} = 0$ for every g, and support conditions requiring that the occupations used are chosen with positive probability in the states used, that each occupation is observed both as an origin and as a destination, and that both stays and moves are observed, the dynamic choice parameters $\big(\{\eta_{jg}\}_{j\in\mathcal{J},g\in\mathcal{G}}, \kappa_0\big)$ are identified. Theorem 2 in Appendix A.2 states this formally.

5.3 Estimation: Two-Stage Estimation Procedure

Model parameters are estimated using a two–stage procedure. In the first stage I estimate the wage equation parameters and the group–specific distribution of latent skills using an Expectation–Maximization (EM) algorithm. A key object in the wage equation is the occupation price p_j. I assume that wages are generated by a unique stationary equilibrium (the one observed in the data), so that occupation prices are constant over time. In the empirical implementation I estimate p_j using occupation fixed effects, so that the wage equation becomes

$$w_{ijt} = p_j + \boldsymbol{\lambda}_j'\boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_{it}) + \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}(0, \varsigma^2_{\varepsilon j}), \tag{34}$$

where p_j is treated as an estimable nuisance parameter (occupation price). Latent log skills satisfy $\boldsymbol{\xi}_i \mid g(i) \sim \mathcal{N}(\boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)})$. The first stage yields the wage–side parameter vector

$$\boldsymbol{\Gamma} \equiv \Big\{(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g\in\mathcal{G}}, (\boldsymbol{\lambda}_j)_{j\in\mathcal{O}}, (\delta_j)_{j\in\mathcal{O}}, (\varsigma^2_{\varepsilon j})_{j\in\mathcal{O}}, (p_j)_{j\in\mathcal{O}}\Big\}.$$

In the second stage I estimate the choice–side parameters using the Rust (1987) Nested Fixed Point (NFXP) algorithm,

$$\boldsymbol{\Delta} \equiv \Big\{ (\eta_{jg})_{j\in\mathcal{O},\, g\in\mathcal{G}}, \kappa_0 \Big\},$$

taking $\widehat{\boldsymbol{\Gamma}}$ as given and setting the taste-shock scale $\varrho = 0.271$ to match the average occupational mobility rate in the large sample (see Section 5 for discussion). The complete–data likelihood for worker i factors into a wage component and a choice component:

$$L_i(w_i, d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) = L_i^{w}(w_i; \boldsymbol{\Gamma}) \cdot L_i^{c}(d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}), \tag{35}$$

where $w_i = \{w_{it_0(i)}, \ldots, w_{iT(i)}\}$ and $d_i = \{d_{it_0(i)}, \ldots, d_{iT(i)}\}$ denote observed wage and choice histories[17].

The wage likelihood integrates out latent skills,

$$L_i^{w}(w_i; \boldsymbol{\Gamma}) = \int \left[\prod_{t=t_0(i)}^{T(i)} \prod_{j\in\mathcal{O}} f_{\varepsilon j}\Big(w_{ijt} - p_j - \boldsymbol{\lambda}_j' \boldsymbol{\xi} - h_j(\boldsymbol{\tau}_{it}) \Big)^{d_{ijt}} \right] \phi\Big(\boldsymbol{\xi}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)} \Big)\, d\boldsymbol{\xi}, \tag{36}$$

where $f_{\varepsilon j}$ is the normal density with variance $\varsigma_{\varepsilon j}^2$ and $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal density. The choice likelihood separates the first observed choice (which uses the group prior) from subsequent choices:

$$\begin{aligned} L_i^{c}(d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) = & \underbrace{\prod_{j\in\mathcal{O}} \Big[\mathbb{P}\Big(d_{it_0(i)} = j \mid \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}, \boldsymbol{\tau}_{it_0(i)}, d_{i,t_0(i)-1}; \boldsymbol{\Gamma}, \boldsymbol{\Delta} \Big) \Big]^{d_{ij,t_0(i)}}}_{\text{first observed choice}} \\ & \times \underbrace{\prod_{t=t_0(i)+1}^{T(i)} \prod_{j\in\mathcal{O}} \Big[\mathbb{P}\big(d_{it} = j \mid \tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}, \boldsymbol{\tau}_{it}, d_{it-1}; \boldsymbol{\Gamma}, \boldsymbol{\Delta} \big) \Big]^{d_{ijt}}}_{\text{subsequent choices}}, \end{aligned} \tag{37}$$

where CCPs are given by (27) and $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ are the Bayesian posterior moments implied by $\boldsymbol{\Gamma}$. Because $L_i = L_i^{w} \cdot L_i^{c}$, the log–likelihood is additively separable:

$$\ell_i(w_i, d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) = \log L_i^{w}(w_i; \boldsymbol{\Gamma}) + \log L_i^{c}(d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}),$$

[17] Starting from $L_i(w_i, d_i; \Gamma, \Delta) = \int p(w_i, d_i, \xi \mid g(i); \Gamma, \Delta)\, d\xi$, the chain rule gives

$$p(w_i, d_i, \xi \mid g(i)) = \phi(\xi; \mu_{g(i)}, \Sigma_{g(i)}) \prod_t p(w_{it} \mid d_{it}, \xi, \tau_{it}; \Gamma) \prod_t p(d_{it} \mid \mathcal{H}_{it}; \Gamma, \Delta).$$

By timing/independence, conditional on (d_{it}, ξ, τ_{it}) wages depend only on Γ, while choices depend on history only through $(\tilde{\xi}_{it}, \tilde{\Sigma}_{it}, \tau_{it}, d_{it-1})$, hence $\prod_t p(d_{it} \mid \mathcal{H}_{it}; \Gamma, \Delta)$ does not depend on ξ. Therefore,

$$\begin{aligned} L_i(w_i, d_i; \Gamma, \Delta) &= \int \phi(\xi; \mu_{g(i)}, \Sigma_{g(i)}) \prod_t p(w_{it} \mid d_{it}, \xi, \tau_{it}; \Gamma) \prod_t p(d_{it} \mid \mathcal{H}_{it}; \Gamma, \Delta)\, d\xi \\ &= \left[\prod_t p(d_{it} \mid \mathcal{H}_{it}; \Gamma, \Delta) \right] \int \phi(\xi; \mu_{g(i)}, \Sigma_{g(i)}) \prod_t p(w_{it} \mid d_{it}, \xi, \tau_{it}; \Gamma)\, d\xi \\ &\equiv L_i^{c}(d_i; \Gamma, \Delta) \cdot L_i^{w}(w_i; \Gamma). \end{aligned}$$

which permits sequential estimation: first obtain $\widehat{\boldsymbol{\Gamma}} = \arg\max_{\boldsymbol{\Gamma}} \sum_i \log L_i^w(w_i; \boldsymbol{\Gamma})$, then conditional on $\widehat{\boldsymbol{\Gamma}}$ compute $\widehat{\boldsymbol{\Delta}} = \arg\max_{\boldsymbol{\Delta}} \sum_i \log L_i^c(d_i; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta})$.

5.4 Estimation Stage 1: Expectation–Maximization Algorithm

Starting from an initial guess $\boldsymbol{\Gamma}^{(0)}$, the EM algorithm iterates:

- **E–step.** For current $\boldsymbol{\Gamma}^{(q)}$ and observed histories (w_i, d_i), compute the posterior distribution of $\boldsymbol{\xi}_i$,
$$f(\boldsymbol{\xi}_i \mid w_i, d_i; \boldsymbol{\Gamma}^{(q)}),$$
and form the expected complete–data log–likelihood.
- **M–step.** Update $\boldsymbol{\Gamma}^{(q+1)}$ by maximizing the expected complete–data log–likelihood. In practice this step updates $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, the occupation loadings $(\boldsymbol{\lambda}_j, h_{jk}(\cdot))$, the shock variances $(\varsigma_{\varepsilon j}^2)$, and the occupation prices (p_j). The M-step explicitly operationalizes the control-function identification argument of Section 5.1: by regressing log wages on the E-step posterior mean $\hat{\boldsymbol{\zeta}}_i^q$ and on the regressors entering $h_j(\boldsymbol{\tau}_{it})$, it absorbs selection on beliefs through the observed history.

Iterations continue until $\|\boldsymbol{\Gamma}^{(q+1)} - \boldsymbol{\Gamma}^{(q)}\|_2$ is small. Because EM guarantees only local convergence, I use multiple starting values and select the solution with the highest likelihood.

5.5 Estimation Stage 2: Nested Fixed Point Algorithm

In the second stage I estimate $\boldsymbol{\Delta}$ by maximizing the choice part of the log–likelihood given $\widehat{\boldsymbol{\Gamma}}$:

$$\widehat{\boldsymbol{\Delta}} = \arg\max_{\boldsymbol{\Delta}} \; \ell^c(\widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta}), \tag{38}$$

where

$$\ell^c(\widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta}) = \sum_{i=1}^{N} \Bigg\{ \sum_{j \in \mathcal{O}} d_{ij,t_0(i)} \log \mathbb{P}\Big(d_{it_0(i)} = j \mid \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}, \boldsymbol{\tau}_{it_0(i)}, d_{i,t_0(i)-1}; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta} \Big) + \sum_{t=t_0(i)+1}^{T(i)} \sum_{j \in \mathcal{O}} d_{ijt} \log \mathbb{P}\Big(d_{it} = j \mid \tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}, \boldsymbol{\tau}_{it}, d_{it-1}; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta} \Big) \Bigg\}.$$

The NFXP algorithm nests the solution of the worker dynamic problem inside the maximization over $\boldsymbol{\Delta}$. Fix a candidate $\boldsymbol{\Delta}$ and define the state $\boldsymbol{x} \equiv (g, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$, on which the value function is solved group by group. The inner step solves for the integrated ex–ante value function $W(\cdot; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta})$ as the fixed point of the integrated Bellman operator:

$$W(\boldsymbol{x}) = \varrho \log \sum_{j \in \mathcal{O}} \exp\left(\frac{\tilde{w}_j(\boldsymbol{x}) + \eta_{jg} - \mathcal{C}(j, d_{-1}) + \beta\, \mathbb{E}\Big[W(\boldsymbol{x}') \mid \boldsymbol{x}, d = j; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta} \Big]}{\varrho} \right), \tag{39}$$

where the expectation is taken over next-period state $\boldsymbol{x}'$ implied by the law of motion under

choice j (including belief updating). Given the fixed point $W(\cdot; \widehat{\mathbf{\Gamma}}, \mathbf{\Delta})$, the CCPs used in the likelihood are computed from (27).

The outer step updates $\mathbf{\Delta}$ to increase (38) using a numerical optimizer. The procedure alternates between (i) solving the fixed point in (39) for W at the current $\mathbf{\Delta}$ and (ii) updating $\mathbf{\Delta}$ given the resulting CCPs, until convergence.

5.6 Externally Set and Calibrated Parameters

Several parameters are set externally or calibrated to match aggregate statistics, as summarized in Table 1 in the Appendix. The discount factor is set to $\beta = 0.9$; education group shares π_g are set to the estimation sample composition; and task importance weights $\boldsymbol{\omega}_j$ are taken from Mihaylov and Tijdens (2019). The per-period exit hazard $\delta = 0.08$ is chosen to match the average career length in the estimation sample; as explained in Section 4.6, it is exogenous, not internalized by workers, and serves only to close the stationary distribution.

The CES elasticity of substitution is set to $\sigma = 1.57$. The aggregator weights $\{\alpha_j\}$ are calibrated so that, evaluated at the Stage-1 prices, the CES inverse demand (29) holds exactly while the final-good price index (31) equals one. Inverting the CES demand (29) leaves the CES weights proportional to $P_j Y_j^{1/\sigma}$ up to a single scalar; we fix that scalar by normalizing the aggregate price level $P(\hat{P}) = 1$ and we obtain the following closed form pricing equation

$$\alpha_j = \frac{P_j\, Y_j^{1/\sigma}}{\left(\sum_{k \in \mathcal{J}} P_k\, Y_k\right)^{1/\sigma}}, \qquad \text{equivalently} \qquad \alpha_j^{\sigma} = s_j\, P_j^{\sigma-1}, \tag{40}$$

where $P_j = \exp(\hat{p}_j)$ is occupation j's good price estimated in Stage 1, Y_j its output (the observed occupation wage bill deflated by P_j, because under piece-rate pricing each occupation's revenue equals its wage bill), and $s_j = P_j Y_j / \sum_k P_k Y_k$ its wage-bill share. Because $\sum_j \alpha_j^{\sigma} P_j^{1-\sigma} = \sum_j s_j = 1$, this delivers $P(\hat{P}) = 1$ exactly.

6 Estimation Results

6.1 Stage 1 Estimation Results: Wage Equation Parameters

The EM algorithm jointly estimates occupation-specific skill and experience elasticities ($\boldsymbol{\lambda}_j$, $\boldsymbol{\beta}_j$), idiosyncratic wage variances, occupation prices, and education-specific skill distributions $(\boldsymbol{\mu}_g, \mathbf{\Sigma}_g)$. Full estimation results are reported in Table 2 in the Appendix.

Three key findings emerge from estimation of the wage equation parameters. First, occupations differ substantially in the productivity of cognitive and manual skills, both in their skill loadings λ_{jk} and in their returns to experience[18]. Second, the estimated group skill distributions $(\boldsymbol{\mu}_g, \mathbf{\Sigma}_g)$ (Figure 6) determine the initial comparative advantage of each education group, which is what shapes entrants' initial beliefs. Mean skills rise with education in both dimensions and the cognitive comparative advantage widens: low-educated workers sit below the

[18] Occ_0 is the cognitive anchor and Occ_1 the manual anchor.

origin, medium-educated workers near it, and high-educated workers above it and tilted toward cognitive skill. The estimated cross-skill correlations are weak in every education group, so workers are not strongly specialized ex ante and enter the labor market with genuine uncertainty about where their comparative advantage lies. These estimated distributions imply that higher educated workers have (and perceive) higher skills in cognitive tasks, making them more likely to start their career in more cognitive occupations, consistent with Deming (2025). Third, occupations have heterogeneous returns to both manual and cognitive task experience, and returns to cognitive experience are generally steeper than returns to manual experience (Figure 7).

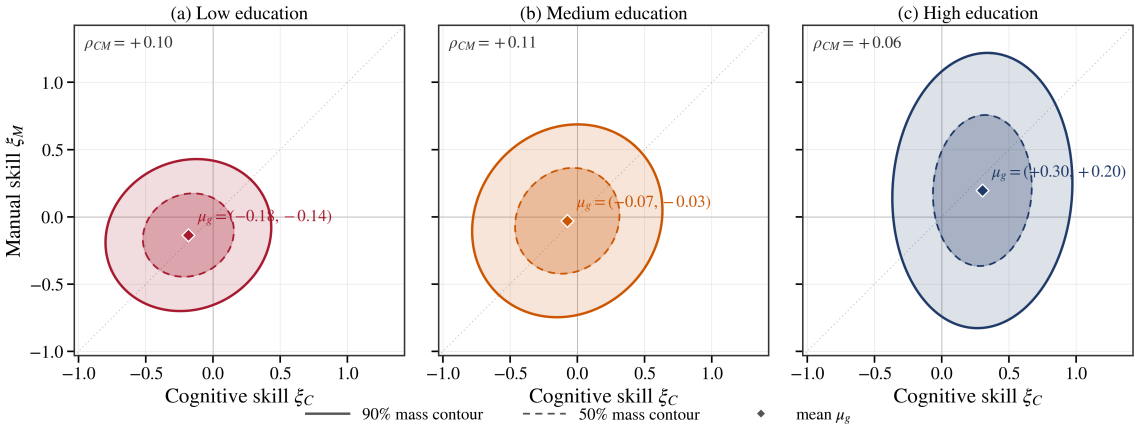


Figure 6: Estimated skill distributions by education group. Each panel (low, medium, high education) plots the joint prior distribution of cognitive skill ξ_C (horizontal axis) and manual skill ξ_M (vertical axis) implied by the estimated group means and covariances $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. The shaded regions show the 50% (dashed) and 90% (solid) probability-mass contours, the diamond marks the group mean $\boldsymbol{\mu}_g$, and the dotted line marks $\xi_C = \xi_M$; the within-group cognitive-manual correlation ρ_{CM} is reported in each panel.

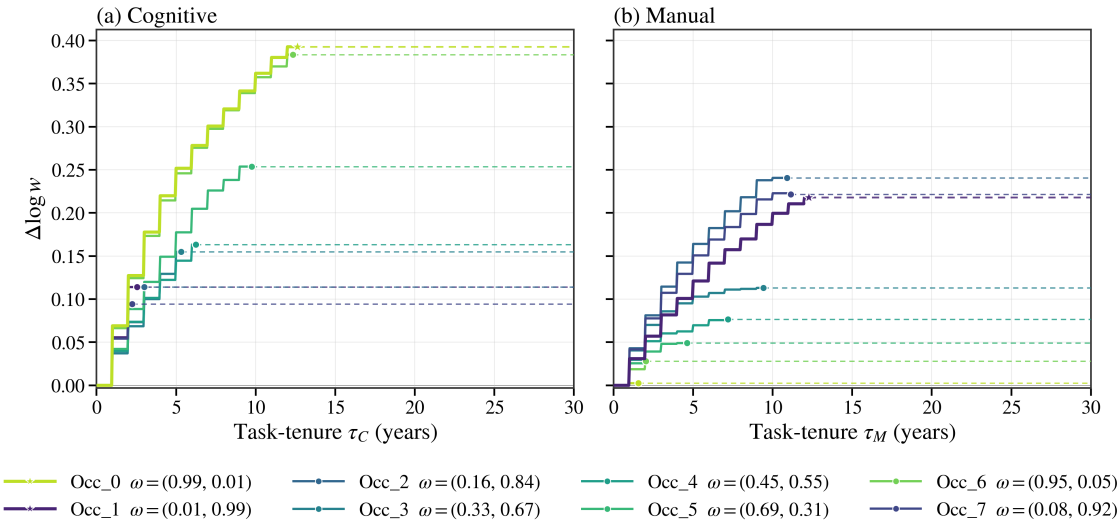


Figure 7: Estimated returns to task-specific experience by occupation. Each line plots the cumulative log-wage return to task experience, $\sum_b \beta_{jkb} \mathbf{1}[\tau \geq b]$, as task tenure τ accumulates from 0 to 30 years. Returns are estimated as a step function of tenure with annual bins, with the last bin being defined by the 95th percentile of observed task-specific tenure within that occupation. *Panel (a)*: returns to cognitive experience τ_C. *Panel (b)*: returns to manual experience τ_M. The marker on each curve is the occupation's 95th percentile of observed task tenure. Occupations are labeled by their task weights $\boldsymbol{\omega}_j = (\omega_{jC}, \omega_{jM})$.

6.2 Stage 2 Estimation Results: Dynamic Choice Parameters

The NFXP estimates are reported in Table 3 in the Appendix. The taste-shock scale is set to $\varrho = 0.271$ to match the average occupational mobility rate in the large sample, rather than estimated jointly by maximum likelihood: as discussed in Section 5, in dynamic learning models this parameter is weakly identified from continuation values without exogenous payoff shifters, so I pin it with this external moment while still allowing within-group heterogeneity in initial choices. The utility switching cost is moderate—a single constant of roughly one log point ($\kappa_0 = 1.103$)—suggesting that the lifecycle decline in occupational mobility is driven primarily by task-specific human capital accumulation and learning rather than by prohibitive switching costs. Occupation- and education-specific amenities η_{jg} display considerable heterogeneity. In the model, amenities act as compensating differentials that rationalize persistent occupational sorting patterns that cannot be explained by wage differentials alone. Potentially, these parameters capture other frictions not considered in the model. For example, if access to some occupations requires a specific educational degree, η_{jg} would capture the fact that certain groups of workers do not move to these occupations despite having attractive wages.

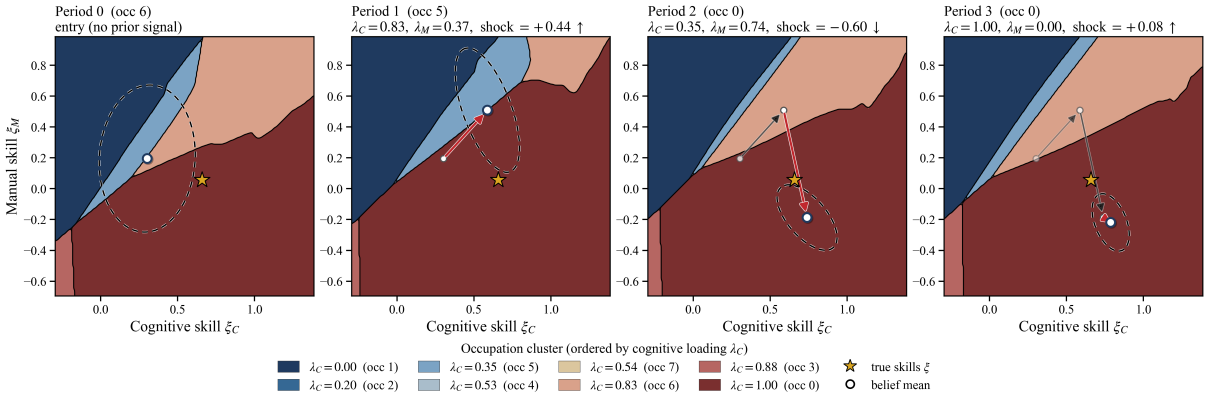
6.3 Occupational Mobility and the Role of Learning

6.3.1 Learning Generates Directionality in the Task Space

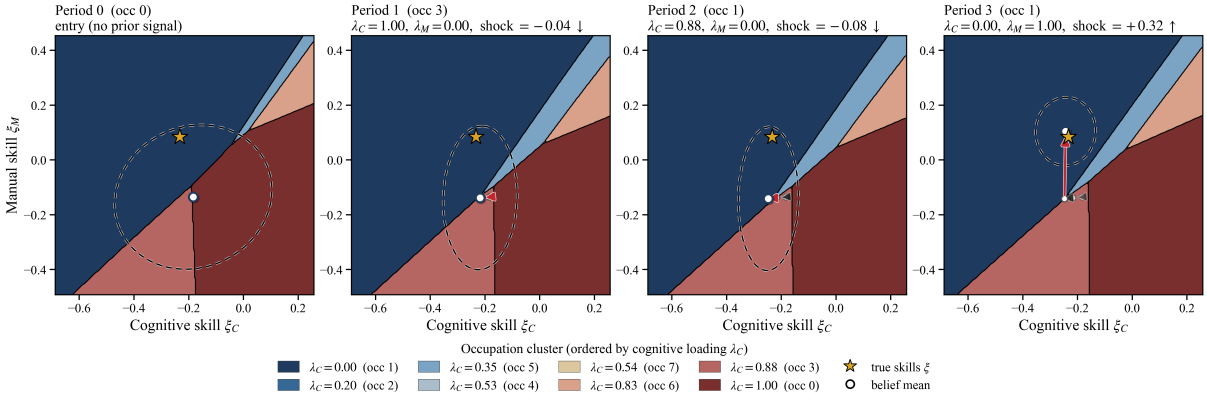
Before assessing the model's fit, I simulate a simplified version of the model in which only learning operates, with no amenities, switching costs, human capital accumulation, or taste shocks. This simplified model is useful for isolating the role of learning in shaping occupational mobility, and I then add these components back in to assess their relative importance. In what follows, I illustrate the mechanism through which learning shapes occupational mobility. Figure 8 plots, in the space of latent skills (ξ_C, ξ_M), each worker's belief about her own comparative advantage as it evolves during her first few periods in the labor market. Movements of the white dot trace changes in beliefs, the star marks the worker's true skills, and the arrows show the moves generated by those belief changes. Each shaded region gives the best occupation for that combination of beliefs.

Panel (a) of Figure 8 shows the sequence of choices of a worker who starts in the mixed occupation and whose true skills lie in the cognitive-intensive part of the map. After period 0, this worker receives a positive signal in the mixed occupation, so she updates her beliefs on both task dimensions and moves diagonally toward the upper-right quadrant of the skill space. She then receives a negative signal while still in a mixed occupation that loads more heavily on the manual task. In fact, she revises her mean manual belief downward and her mean cognitive belief slightly upward, because the correlation between the two skills turns negative after a few signals. This sequence of signals pushes the worker toward the cognitive-intensive occupations.

Panel (b) shows a similar pattern for a worker who starts in the cognitive-intensive occupation and ends in the manual-intensive occupation, where her comparative advantage lies. Unlike the previous case, this worker sorts into highly specialized occupations; in fact, her moves are almost always fully horizontal or vertical, meaning that she updates along only one skill



(a) Worker starting in a mixed occupation and ending in a cognitive-intensive one



(b) Worker starting in a cognitive occupation and ending in a manual-intensive one

Figure 8: Belief dynamics in the space of latent skills for two simulated workers over their first four periods; each row is one worker. Horizontal axis: cognitive skill ξ_C; vertical axis: manual skill ξ_M. The shaded regions give the occupation chosen as a function of the belief mean with amenities set to zero, ordered by cognitive loading λ_C (legend). The star is the worker's true skills $\boldsymbol{\xi}$, the circle her posterior belief mean, and the dashed ellipse her posterior covariance. The red arrow is the belief-mean update from the period's wage signal and the gray arrows trace past updates; panel headers report the period, the occupation held, its loadings (λ_C, λ_M), and the realized signal.

dimension. For example, between periods 2 and 3 she receives a positive signal in the manual-intensive occupation that pushes her toward the manual-intensive occupation, where she settles for the rest of her career.

6.3.2 Model vs. Data

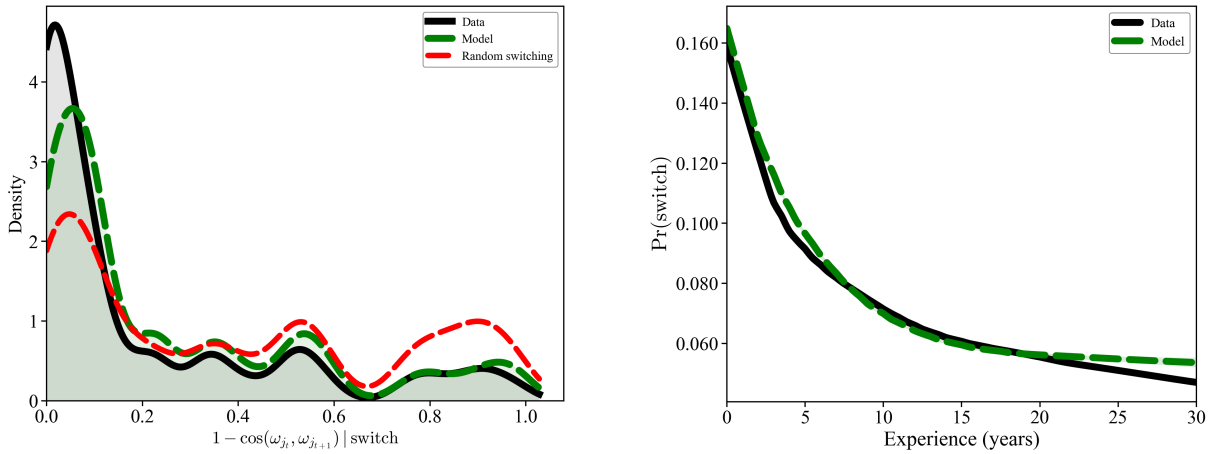
I now assess how well the estimated model reproduces the empirical patterns of occupational mobility documented in Section 3.[19]

[19]Throughout this section a worker's position in the wage distribution is measured by a worker–occupation fixed effect $\widehat{\alpha}_{ij}$, estimated from

$$\log w_{ijt} = \alpha_{ij} + \phi_t + f_j(\text{tenure}_{ijt}) + \nu_{ijt},$$

where ϕ_t are year effects and f_j is an occupation-specific cubic in occupational tenure. It is the empirical counterpart of the model's match quality μ_{ij}: it nets out the price of the occupation and the transitory wage shock, leaving the component of the wage specific to the worker–occupation pair, and workers are ranked on it within (occupation, year) cells.

In Figure 9, Panel (a) shows the distribution of the task-content distance of occupational switches. The model reproduces the empirical distribution closely, and the comparison with a benchmark in which switchers are reassigned at random across occupations makes clear that this is not mechanical: random switching would place far too much mass on large task moves and too little on the short-distance moves that we can observe in the data. Workers move to occupations that are close to the one they leave. Panel (b) shows the hazard of switching by experience: mobility is high at entry, around 16%, and falls to about 5% after two decades, and the model matches both the level and the decay. It is worth stressing that this hazard rate is not strictly targeted by some specific parameters of the model: its decline over experience in the labor market arises naturally from the interaction of learning and task-specific experience. Early in the career workers have larger uncertainty about their comparative advantage in tasks, so the option value of moving is high and workers have accumulated little task-specific human capital to lose by moving. As wage signals accumulate, beliefs converge on the truth and the informational return to trying a different occupation falls; at the same time, task-specific experience builds up, so the opportunity cost of moving increases as moving now would destroy human capital.



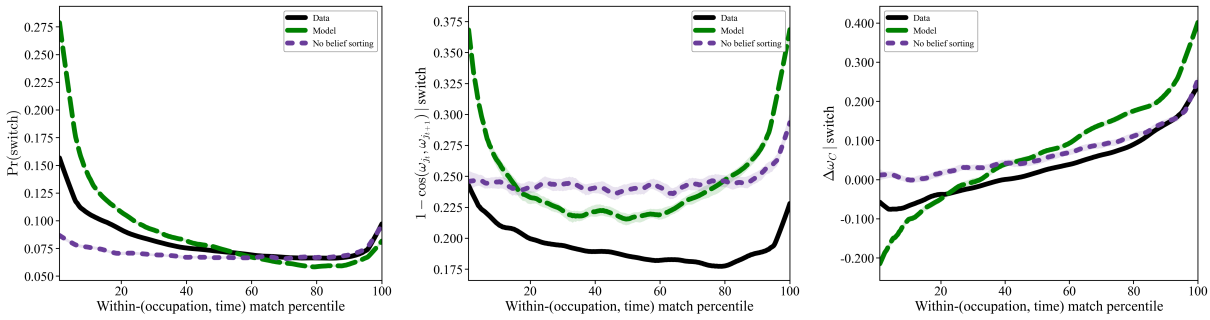
(a) Task-content distance of switches

(b) Hazard of switching by experience

Figure 9: Occupational mobility, model versus data. *Panel (a)*: distribution of the task-content distance between origin and destination occupation, $1 - \cos(\boldsymbol{\omega}_{j_t}, \boldsymbol{\omega}_{j_{t+1}})$, conditional on switching, in the data (black), in the model (green), and under a benchmark in which switchers are reassigned to occupations at random (red). *Panel (b)*: probability of switching occupation by years of potential experience, in the data (black) and in the model (green).

Figure 10 shows how occupational mobility varies with the match-quality percentile within occupation. Each panel plots a moment against the worker's within-(occupation, time) match percentile, comparing data, the estimated model, and a counterfactual in which workers do not sort on their beliefs. Panel (a) shows the U-shape of occupational mobility (Groes et al. 2015): the probability of switching occupation in the next period is higher for workers with low and high match quality within their occupation. Panel (b) is the task content of the switch: workers at both ends of the match-quality distribution make larger moves in task space than workers in the middle. Panel (c) is the direction of the move: poorly matched workers move toward more manual-intensive occupations ($\Delta\omega_C < 0$), well-matched workers toward more cognitive-intensive ones ($\Delta\omega_C > 0$). The model reproduces all three, though it overpredicts the level of

mobility and the size of task moves at the very bottom of the distribution.



(a) Probability of switching (b) Task distance of the switch (c) Direction of the switch

Figure 10: Occupational mobility by within-(occupation, time) match percentile (horizontal axis), in the data (black, solid), in the estimated model (green, dashed), and in a counterfactual in which workers do not sort on their beliefs (purple, dotted). *Panel (a)*: probability of switching occupation. *Panel (b)*: task-content distance of the switch, $1 - \cos(\boldsymbol{\omega}_{j_t}, \boldsymbol{\omega}_{j_{t+1}})$. *Panel (c)*: change in the cognitive task content of the occupation held, $\Delta\omega_C$; positive values are moves toward more cognitive-intensive occupations.

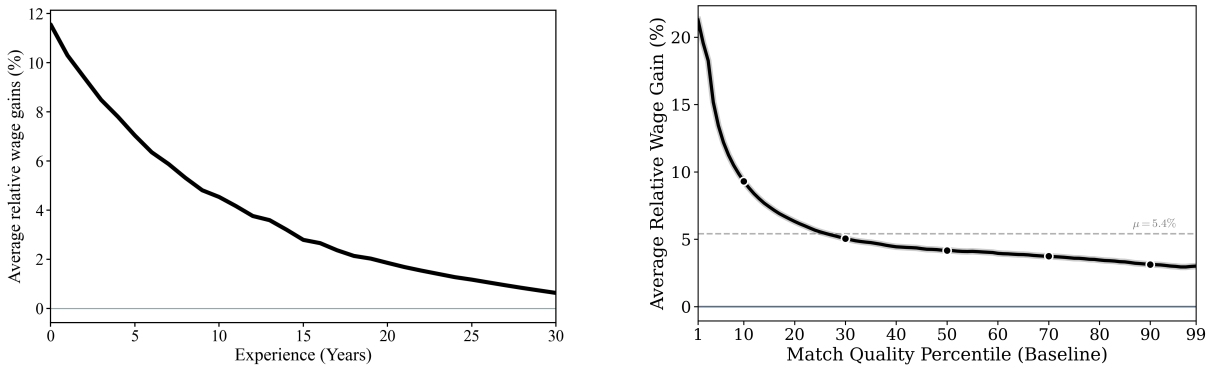
The counterfactual line in each panel shows that learning mainly drives mobility at the bottom. Switching off learning in the model, occupational mobility and the task content of the occupational moves of low-match-quality workers flatten: the left-skewness of the U-shape of occupational mobility disappears. Learning is even more important for the intensity of the change in task content conditional on switching occupation (Panel (b)). Workers with low match quality within their occupation do not only switch more often; when they switch, they also move to more distant occupations in the task space. Learning about comparative advantage in tasks is a useful mechanism to generate this change in task content conditional on changing occupation as a function of the within-occupation match quality. While some workers may be low ability overall, workers at the bottom within occupation are more likely to be mismatched, that is, they are not in the right match given their set of skills. In the model, low match quality implies negative signals that make workers update their beliefs about their skills with a specific directionality, as shown in Figure 8. Without belief sorting the intensity of this directionality at the bottom vanishes: switchers move roughly as far in task space regardless of how well matched they were. Panel (c) makes the same point with the direction of the move: without learning, poorly matched workers are no longer the ones who reallocate toward manual work, and the directionality that motivates the whole exercise is gone. Note that the counterfactual retains switching costs, amenities, taste shocks, and task-specific human capital accumulation; none of them, alone or together, generates any of the three gradients. Match quality predicts mobility, and the size and direction of the task change it induces, only because workers read their wages as signals about their comparative advantage and act on what they learn.

7 The Consequences of Skill Uncertainty

I now use the estimated model to quantify the wage cost of imperfect information about skills. In the model, mismatch arises because the initial prior $\mathcal{N}(\boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)})$ is diffuse relative to true heterogeneity in $\boldsymbol{\xi}_i$: entrants sort on group means rather than on individual comparative ad-

vantage. Beliefs are rational in the sense that the prior beliefs that workers have when starting their career coincide with the true within-group population distribution of $\boldsymbol{\xi}_i$. While beliefs are correct on average, a worker's own skill vector is only one draw from that distribution, and any given individual's realization can lie far from her prior.

I construct a counterfactual general equilibrium in which workers have perfect beliefs from the start of their careers: each worker's initial beliefs equal her true skills, and she faces no uncertainty.[20] In both the baseline and counterfactual economies, workers face identical switching costs, idiosyncratic taste shocks, and wage noise. The only difference is the distribution of beliefs across simulated workers at labor market entry. This exercise quantifies the wage cost of imperfect information about skills, that is, the wage loss associated with sorting based on imperfect rather than perfect knowledge of one's comparative advantage. Comparing the two stationary equilibria, giving workers perfect knowledge of their comparative advantage at entry raises aggregate output by 5.2%.[21] Figure 11 then traces the individual wage gains from removing these frictions, over the life cycle and across the match-quality distribution within occupation.



(a) Wage gains over the life cycle

(b) Wage gains by match-quality percentile

Figure 11: Wage gains from removing information frictions. *Panel (a)*: average relative wage gain over the life cycle (horizontal axis: years of potential experience; vertical axis: percent wage gain). *Panel (b)*: average relative wage gain by the worker's within-(time, occupation) wage-match percentile in the baseline economy (horizontal axis); the dashed horizontal line marks the unconditional mean gain, $\mu = 5.4\%$. Shaded bands are bootstrap confidence intervals.

This output gain reflects the misallocation of workers to occupations that do not match their true comparative advantage under imperfect information. The wage gains are concentrated at the start of the career and flatten over the lifecycle. At labor market entry, the average worker would earn about 11.6% more if she knew her true skills. This gap narrows as workers learn and change occupation accordingly, falling to about 2% after twenty years of experience and fading toward zero (about 0.3%) after 30 years. Mismatch driven by imperfect information is largely an early-career phenomenon: the cost of imperfect information is borne mostly in the

[20] In practice, $\tilde{\boldsymbol{\xi}}_i = \boldsymbol{\xi}_i$ and $\tilde{\sigma}^2_{ik} \approx 0$ for all workers i and skill dimensions k.

[21] Aggregate output in the stationary equilibrium is $Y = \left(\sum_{j \in \mathcal{O}} \alpha_j Y_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where $Y_j = \int \exp(\varepsilon)\, Y_j(x)\, \mathbf{1}_{\{d=j\}}\, \pi^*(x)\, dx\, dH_j(\varepsilon)$ is occupation j's aggregate output, $Y_j(x)$ is individual output from the task technology (5), and π^* is the stationary distribution of worker states. The mismatch cost is computed as $(Y^{PI} - Y^{II})/Y^{II}$, where Y^{PI} and Y^{II} are aggregate output in the perfect and imperfect information stationary equilibria, respectively, both solved at the same model parameters.

first decade, when there is more skill uncertainty and workers are still discovering their comparative advantage, and it is progressively resolved as learning and re-sorting bring imperfect-information workers closer to their perfect-information counterparts. Learning, imperfectly transferable experience, and the cost of switching interact to create a strong path dependence, so that some workers do not fully converge to their optimal career in the imperfect information scenario, but on average the residual gain closes gradually.

On average, this reallocation operates almost entirely through better sorting rather than through experience accumulation in the right dimension. Decomposing the average lifecycle wage gains across the two scenarios, I show that they are mainly driven by better matching on comparative advantage, decaying in parallel with it from about 11.6% at entry toward zero late in the career.[22] The full decomposition is reported in Figure 29 in the Appendix.

Additionally, gains are largest for the workers who are worst matched in the baseline. Panel (b) plots the average wage gain against the worker's position in the within-occupation match-quality distribution. The gain falls steeply in match quality: workers in the bottom percentile of their occupation's match-quality distribution would earn over 20% more under perfect information, against roughly 3% at the top, with an unconditional mean of 5.4%. Perfect information therefore does not raise all wages proportionately: it repairs the bottom of the match-quality distribution, where imperfect beliefs have led workers into occupations that fit them poorly.

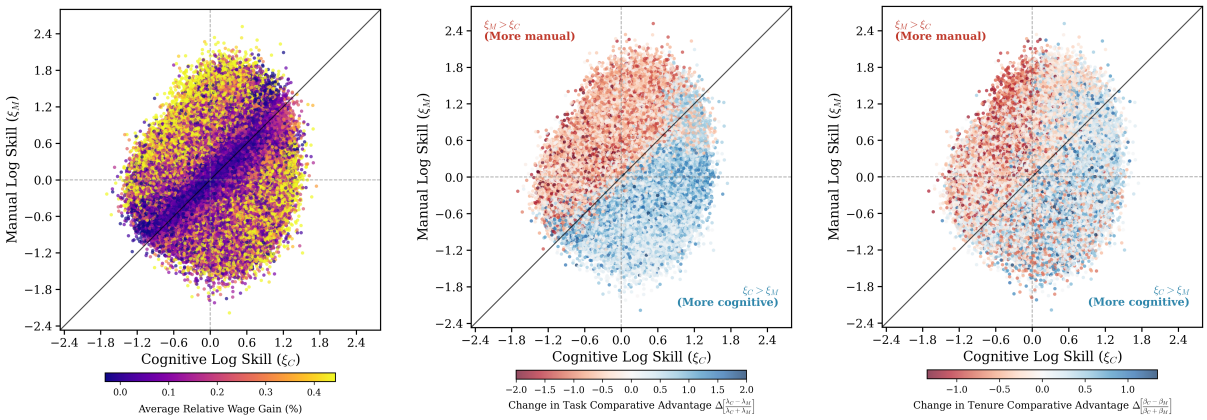
7.1 Who Benefits Most, and Why: Sorting on Comparative Advantage

Importantly, not all workers reallocate when provided with perfect information. The model includes other important determinants of occupational choice—idiosyncratic taste shocks, compensating differentials, and switching costs—that have been shown to matter for labor mobility. Under perfect information, a sizeable share of workers would choose a different initial occupation, and many more would switch occupations at some point during their career. Among those who do reallocate, the wage gains are substantial, and Figure 12 shows both who they are and how they move.

Panel (a) plots the lifecycle-average wage gain of each reallocated worker in the space of true skills. Specialists, workers with a strong comparative advantage in one dimension, have the largest gains from perfect information, while generalists, workers near the 45° line with similar cognitive and manual skills, gain relatively little. This has an intuitive economic interpretation[23]. Specialists have high returns to correct sorting: a worker with substantially larger cognitive skill loses if she mistakenly enters a manual-intensive occupation, or even a more balanced occupation. Generalists, in contrast, are less likely to be mismatched on comparative advantage, so the cost of sorting "incorrectly" is smaller than for specialists. This finding has important implications for the design of career guidance policies: interventions that provide better information about skills, such as aptitude testing, rotational programs, or early career mentoring, may deliver the largest wage gains for workers with a strong comparative advan-

[22]Notice that the residual may change because while I keep luck fixed, the variance of the idiosyncratic shocks changes, as it is also driven by the occupational choice.

[23]This result is likely driven by the rationality assumption, together with the normalization of the population distribution of skills.



(a) Average wage gain (b) Change in task-loading (c) Change in tenure-return

Figure 12: Wage gains and reallocation in the space of true skills when information frictions are removed. In each panel a point is a worker who chooses a different occupation under perfect than under imperfect beliefs, positioned by her true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis); the solid line is the 45° line $\xi_C = \xi_M$. *Panel (a)*: color is the worker's average relative wage gain over the life cycle, in percent. *Panel (b)*: color is the change, from the imperfect- to the perfect-information economy, in the task-loading comparative advantage of the occupation held, $\Delta\big[(\lambda_C - \lambda_M)/(\lambda_C + \lambda_M)\big]$. *Panel (c)*: the same object for the tenure-return comparative advantage, $\Delta\big[(\beta_C - \beta_M)/(\beta_C + \beta_M)\big]$. In Panels (b) and (c) blue denotes a move toward a more cognitive occupation and red a move toward a more manual one.

tage.

Panels (b) and (c) show the mechanism behind these gains. They plot the same workers, now colored by the direction in which they reallocate: blue workers move toward occupations with more cognitive comparative advantage, red workers toward more manual ones. A sharp separation along the 45 degree line emerges. Workers with $\xi_C > \xi_M$ (cognitive specialists, below the 45° line) systematically move to more cognitive occupations when given perfect information, while workers with $\xi_M > \xi_C$ move to more manual ones. Panel (b) shows this for the task loadings $\boldsymbol{\lambda}_j$; Panel (c) shows an analogous but weaker and noisier pattern for tenure returns, where workers with cognitive comparative advantage move toward occupations with higher returns to cognitive experience. This suggests that while the contribution of experience to the aggregate wage gain is small, task-specific experience still plays a role in generating the wage costs that keep workers from switching to the right occupation as they discover their true comparative advantage.

7.2 The Effects of Information Interventions

I now use the model to assess how much of the output and wage losses can be recovered by implementing realistic information interventions, and how their effectiveness changes when they are provided at different stages of the career. I take as empirical anchor the field experiment of Kiss et al. (2023). The paper studies two related information interventions; I calibrate the model intervention to the effects estimated in the first experiment, in which young jobseekers receive a short computerized assessment covering numeracy, communication, and concept formation skills. The information treatment consists of individualized feedback about where the treated worker stands in the skill distribution of treated individuals. This information treatment is

therefore informative about workers' comparative advantage in skills. Treated workers update their beliefs about their own skills and re-direct their search toward occupations that better match their assessed comparative advantage. A desirable feature of this class of intervention is that it is relatively cheap to deploy[24]. The estimated treatment effects amount to an increase of 13.7 percentage points in the share of workers whose stated career aspiration is aligned with their measured comparative advantage.

I embed this intervention in the model as follows. At entry ($t = 0$), before making her first occupational choice, each worker i receives K independent noisy Gaussian signals, one for each latent skill dimension (in the estimated model, $K = 2$, corresponding to the cognitive and manual dimensions):

$$s^0_{i,k} \;=\; \xi_{i,k} + \zeta_{i,k}, \qquad \zeta_{i,k} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\zeta), \qquad k = 1, \ldots, K. \tag{41}$$

Stacking the signals into $\boldsymbol{s}^0_i = (s^0_{i,1}, \ldots, s^0_{i,K})'$, we have $\boldsymbol{s}^0_i = \boldsymbol{\xi}_i + \boldsymbol{\zeta}_i$ with $\boldsymbol{\zeta}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2_\zeta \boldsymbol{I}_K)$. Because the prior $\boldsymbol{\xi}_i \mid g \sim \mathcal{N}(\boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)})$ and the signal vector are jointly Gaussian, the post-assessment belief is the standard closed-form Gaussian update

$$\tilde{\boldsymbol{\xi}}_{i0} = \boldsymbol{\mu}_{g(i)} + \boldsymbol{\Sigma}_{g(i)} \left(\boldsymbol{\Sigma}_{g(i)} + \sigma^2_\zeta \boldsymbol{I}_K \right)^{-1} \left(\boldsymbol{s}^0_i - \boldsymbol{\mu}_{g(i)} \right), \tag{42}$$

$$\tilde{\boldsymbol{\Sigma}}_{i0} = \boldsymbol{\Sigma}_{g(i)} - \boldsymbol{\Sigma}_{g(i)} \left(\boldsymbol{\Sigma}_{g(i)} + \sigma^2_\zeta \boldsymbol{I}_K \right)^{-1} \boldsymbol{\Sigma}_{g(i)}. \tag{43}$$

The worker is myopic about the arrival of this signal and uses the same value function $V(\cdot)$ as in the baseline economy. The signal variance σ^2_ζ is common across all skill dimensions and is the only free parameter of the experiment. I calibrate the noise of the signal to match the average treatment effect of Kiss et al. (2023). Provided before entering the labor market, the information treatment in the model also raises the share of workers whose beliefs about their comparative advantage match their true comparative advantage by 13.7 percentage points.[25] The target is the model counterpart: the difference between post- and pre-assessment alignment rates.[26] The calibration yields $\sigma_\zeta \approx 0.452$. I then feed this value into the estimated model and re-simulate the economy, holding all other parameters at their estimated values.

Relative to the baseline economy, providing entrants with the calibrated skill signal raises aggregate output by 0.91%, that is around 17.5% of the total misallocation generated by information frictions about skills. The wage gains arising from the information treatment are strongly front-loaded in the career: average wages are about 3.3% above the baseline at entry, then de-

[24]The field-experiment variable cost reported by Kiss et al. (2023) is $45.74 USD PPP per worker. Converting to 2020 EUR via the OECD USD–EUR PPP rate ($1 = € 0.92) yields € 42.08, and deflating to 1995 real EUR via the Portuguese CPI (2020–to–1995 deflator of 0.65) gives € 27.35 per worker, which is the per-worker policy cost expressed in the model's wage units (real hourly EUR 1995) and is therefore directly comparable to the hourly wage gains reported in the paper.

[25]I follow their Section 2.6 definition. A large reference pool of latent types $\boldsymbol{\xi}$ is drawn from the mixture distribution of education groups and used to compute the marginal quintile cut-points of each skill dimension. A worker is then assigned to one of five quintile bins on each skill according to her true type, her prior mean, and her post-assessment posterior mean. Her belief is aligned if the skill in which her believed quintile is higher coincides with the skill in which her true quintile is higher; workers whose true type falls in the same bin on both skills (no unique comparative advantage) are dropped, as in Kiss et al. (2023).

[26]The function mapping signal noise to the average-treatment-effect is monotone, as the posterior collapses to the truth as $\sigma^2_\zeta \to 0$ and to the prior as $\sigma^2_\zeta \to \infty$, so this moment identifies σ^2_ζ.

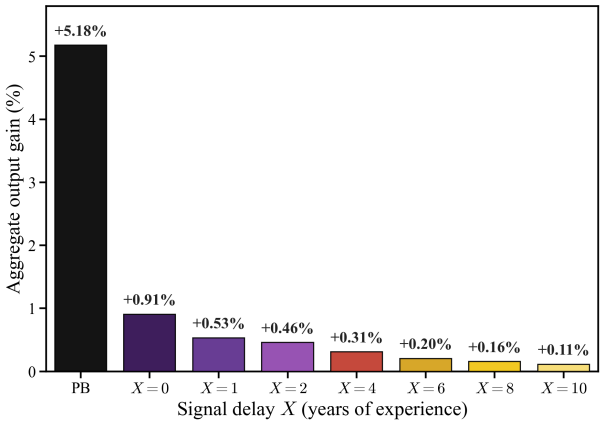
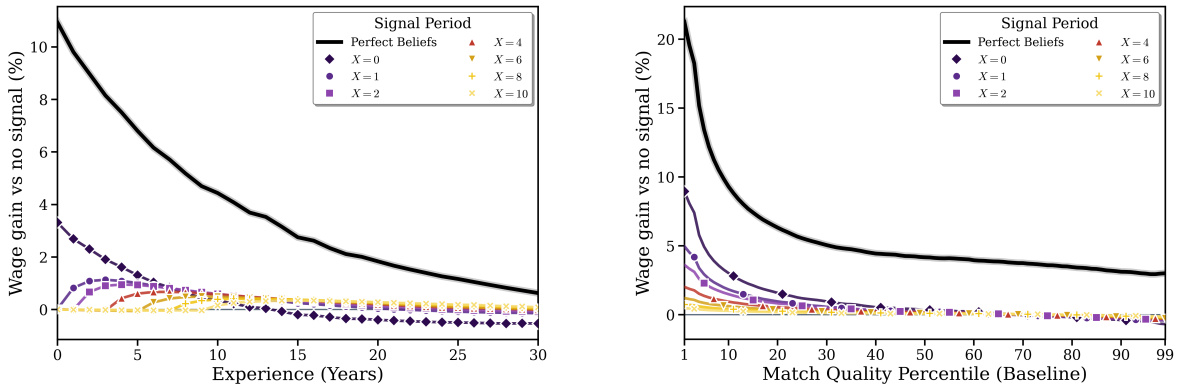


Figure 13: Aggregate output gain (percent) relative to the baseline economy when the calibrated skill assessment is delivered $X \in \{0, 1, 2, 4, 6, 8, 10\}$ years after labor-market entry, holding the calibrated signal precision σ_ζ^2 fixed at its at-entry value. Each bar is a separately solved stationary equilibrium. The leftmost bar (PB) is the output gain from giving workers perfect information about their skills at entry, the upper bound the assessment could in principle recover.

cline as better-matched entrants accumulate experience and baseline workers catch up after updating their beliefs and sorting accordingly, crossing zero after roughly thirteen years and settling mildly negative (-0.5%) thereafter (Figure 14, Panel (a)). These gains come almost entirely from better matching on skills at the first occupational choice; the late-career reduction in the gap between the two economies reflects the cross-task experience mechanism emphasized earlier, as the policy induces early specialization and so reduces some of the exploratory experience that baseline workers accumulate under imperfect information. In other words, imperfect information may generate mismatch, but conditional on reaching their ideal career path at some point, cross-returns to different task experience may actually be beneficial.

The treatment does not help everyone equally, and Panel (b) in Figure 14 shows how the gains are distributed. The wage gains are concentrated at the bottom of the match quality distribution within occupation: workers in the lowest percentiles in the baseline economy, that are more likely to be mismatched, gain about 9%; the gain fades to roughly zero by the fortieth percentile. A single noisy assessment provided early in the career recovers a little under half of the perfect-information gain for the worst-matched workers (9% against 21%), and essentially no gains for the well-matched.

A natural question is how the value of the policy depends on when it is delivered. I therefore repeat the experiment delivering the signal in (41) at $t \in \{0, 1, 2, 4, 6, 8, 10\}$ periods after entry, holding its precision σ_ζ^2 fixed and keeping workers myopic about when it will arrive. The later the assessment comes, the lower are the realized output and wage gains: the output gain is largest at entry, drops by roughly half within the first two years, and is nearly gone by the end of the first decade. The information treatment, when provided at labor market entry, increases wages for the first years compared to the baseline economy. A delayed assessment generates only a small hump once it arrives, mainly because task-specific experience, which is imperfectly transferable, contributes to creating a locked-in effect. Even if there is an occupation in which workers' skills might be more productive, switching would make the worker lose some task-specific experience. This implicit cost of switching is increasing over labor market experience and directly reduces the effectiveness of the signal if delayed.



(a) Wage gains over the life cycle

(b) Wage gains by match-quality percentile

Figure 14: Wage gains from the skill assessment, delivered $X \in \{0, 1, 2, 4, 6, 8, 10\}$ years after labor-market entry, holding the calibrated signal precision σ_ζ^2 fixed. *Panel (a)*: life-cycle profile of average wage gains (percent) relative to the no-signal baseline, by delay, with the perfect-information counterfactual (solid black) as a benchmark; horizontal axis is years of experience. *Panel (b)*: average wage gain relative to the no-signal baseline by the worker's within-(time, occupation) wage-match percentile in the baseline economy, for each delay, with the perfect-information counterfactual (solid black) as a benchmark.

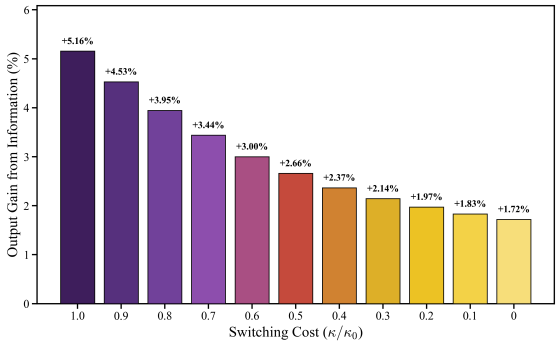
In Appendix C.6 I study an experiment that directs each labor market entrant, for one period only, into the occupation that is most informative about her comparative advantage, after which she chooses freely. The forced exploration reduced wages early in the career but improved learning and sorting from the second period onward. The large initial reduction is driven not only by forcing workers into occupations for which they are poorly matched, but also by the downward-sloping demand curve for occupation-specific goods. Forcing everyone to move into their most informative occupation pushes prices down in the informative occupations while raising prices elsewhere. Overall, aggregate output rises by 0.65%.

7.3 Switching Costs and Information Frictions

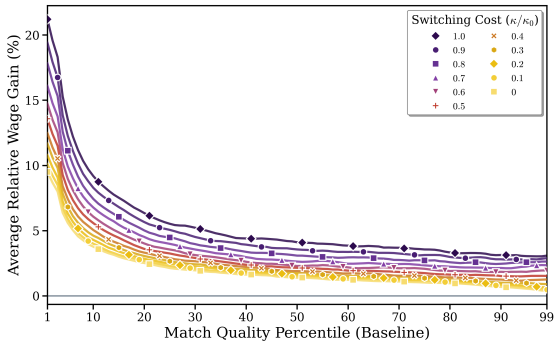
The cost of imperfect information depends on how easily workers can switch once they learn where their best fit is. If occupational mobility were free, a worker who discovers she is mismatched could simply switch at any point in her career, even if learning is slow. Switching costs and information frictions are thus complementary sources of misallocation, and this subsection examines how they interact. I reduce the switching cost κ_0 from its estimated value to zero in steps of ten percent of the baseline, re-solving the stationary equilibrium at each level, and I run two comparisons.

The first holds the switching cost fixed and compares the imperfect-information economy to its perfect-information counterpart, that is, it measures the cost of information frictions at each level of mobility friction. This cost falls monotonically as mobility becomes cheaper (Figure 15): the output gap between the perfect- and imperfect-information economies declines from about 5.2% at the estimated switching cost to under 2% when moving is free (Panel (a)), and the gain from perfect information shrinks at every point of the match-quality distribution as the switching cost falls (Panel (b)). This result is intuitive: lower moving cost allow workers to

switch once their learn their types, so misallocation generated by information frictions is lower.



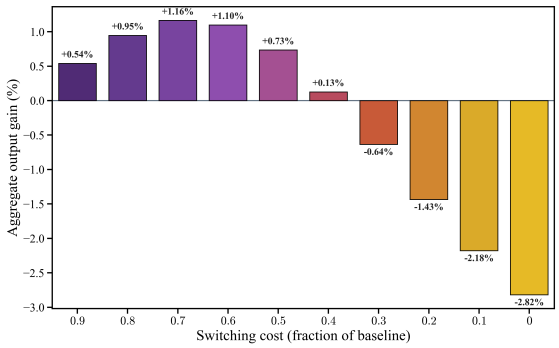
(a) Output gain from perfect information



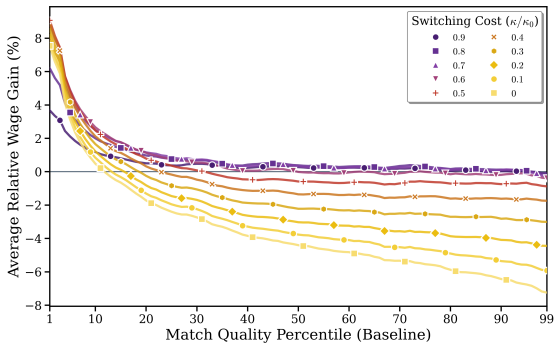
(b) Wage gain from perfect information by match quality

Figure 15: The cost of information frictions at different levels of the switching cost. At each value of κ/κ_0 both the imperfect- and the perfect-information stationary equilibria are re-solved, and the figure reports the gap between them. *Panel (a)*: aggregate output gain from moving to perfect information, $(Y^{PI} - Y^{II})/Y^{II}$, in percent, by κ/κ_0. *Panel (b)*: average relative wage gain from perfect information by the worker's within-occupation match-quality percentile, by κ/κ_0. The corresponding lifecycle wage profiles are in Appendix C.5.

The second comparison is across the imperfect-information economies themselves, and asks how aggregate output responds to the switching cost that workers actually face. Here output is not monotonically decreasing in the switching cost (Figure 16): it first rises as mobility becomes cheaper, with a peak when switching cost are 70% of the estimated value, and then falls, turning negative and reaching a loss of about 2.8% when switching is completely free[27].



(a) Aggregate output



(b) Wage gains by match quality

Figure 16: Aggregate output and wage gains across imperfect-information economies with different switching costs. Each curve or bar is a separately solved stationary equilibrium, indexed by the switching cost as a fraction of its estimated value ($\kappa/\kappa_0 = 1$ is the baseline, 0 is costless mobility). *Panel (a)*: percent change in CES aggregate output relative to the baseline economy. *Panel (b)*: average relative wage gain over the baseline by the worker's within-occupation match-quality percentile, by κ/κ_0. The corresponding lifecycle wage profiles are in Appendix C.5.

Two forces produce this "inverted U" pattern. A small reduction in the switching cost raises

[27] Notice that in these counterfactuals, the implicit cost driven by human capital accumulation is still present, so there are still implicit cost of moving.

output, because workers can freely reallocate towards occupations in which they have a comparative advantage as they learn. They often benefit from the returns to their accumulated experience along the way: an occupation may expose a worker to tasks she would not have performed elsewhere, building experience that remains valuable in the occupation she eventually moves to. A large reduction, however, makes workers excessively responsive to idiosyncratic taste shocks. This excessive churning pushes them into occupations where their skills and human capital are not particularly productive. Panel (b) of Figure 16 shows who bears the resulting losses: they fall for the well-matched workers, at the top of the within-occupation match-quality distribution. The poorly matched workers gain from cheaper mobility almost at every switching-cost level; but a worker who was already well matched has nothing to gain from moving. The aggregate output loss from excessive churning is therefore driven by the top of the match-quality distribution. These two comparisons bring different insight on the nature and causes of misallocation: lower cost of moving always reduces the misallocation generated by information frictions, but some frictions are useful to make workers accumulating task-specific human capital.

We close this subsetion by providing two practical examples on what type of frictions the switching cost captures. First, if occupations are concentrated in different local labor markets then workers have to change location in order to switch occupation, so part of κ_0 may reflect the monetary and psychic cost of geographic mobility (Kennan and Walker 2011). Reducing switching costs in this case can be thought as providing moving vouchers to mismatched workers. Second, reallocating across occupations often requires passing through an unemployment spell, as in the island and sectoral-reallocation tradition in which workers move by first becoming unemployed (Lucas and Prescott 1974; Alvarez and Shimer 2011; Pilossoph 2014; Chodorow-Reich and Wieland 2020). In this case, κ_0 represents the cost of being unemployed, and we can think of switching costs reduction as policies that affect unemployment benefits (duration, overall amounts etc...).

8 Artificial Intelligence and Skill Uncertainty

The estimated model is well suited to study the interaction of technological change and skill uncertainty. I focus on the job transformation (Freund and Mann 2025) induced by Large Language Model (LLM) adoption, and I model it as a task-biased shock that lowers the importance of cognitive tasks across occupations. This affects the time workers spend on cognitive tasks in proportion to each occupation's exposure to large language models. I simulate counterfactual scenarios with and without imperfect information to show how output and the cost of mismatch driven by information frictions change not only in the long run, but also on impact and along the transition path.

8.1 AI Shock Design

I measure each occupation's exposure to AI following the framework of Eloundou et al. (2024), which treats a task as exposed if access to a large language model would cut the time needed to complete it by at least half. I apply the rubric to the 296 Level-3 skills of the ESCO classification,

assigning each skill to one of four exposure categories with weights $\varphi_{E0} = 0$, $\varphi_{E1} = 1$, and $\varphi_{E2} = \varphi_{E3} = 0.5$. Manual skills are unexposed ($\varphi = 0$), since language models cannot perform physical tasks. Averaging skill exposure by ESCO importance within each occupation and aggregating to the $K = 8$ clusters yields a cognitive exposure $\chi_{jC} \in [0, 1]$ for each cluster, with $\chi_{jM} = 0$; exposure is concentrated in cognitive-intensive clusters.

AI lowers the time spent on the cognitive task in proportion to exposure. Through the model's optimal time allocation, the post-shock task weights are

$$\omega_{jC}^{\text{AI}} = \omega_{jC}\,(1 - \chi_{jC}), \qquad \omega_{jM}^{\text{AI}} = \omega_{jM}, \tag{44}$$

renormalized so that $\omega_{jC}^{\text{AI}} + \omega_{jM}^{\text{AI}} = 1$, so the manual weight rises mechanically. Because the effective wage parameters scale with the task weights ($\lambda_{jk} = \omega_{jk} a_{jk}$, $\beta_{jk} = \omega_{jk} b_{jk}$, $q_{jk} = \omega_{jk} c_{jk}$), the shock scales the skill loadings, the linear experience returns, and the quadratic experience terms by the common ratio $\omega_{jk}^{\text{AI}}/\omega_{jk}$. Lower cognitive weight therefore reduces the return to cognitive skills and experience. Additionally, because task-specific experience accumulates in proportion to the task weights (equation (10), now with $\boldsymbol{\omega}_j^{\text{AI}}$), a decrease in cognitive task importance implies less time spent on cognitive tasks and therefore slower accumulation of human capital along that dimension. Manual-intensive clusters are left largely untouched.

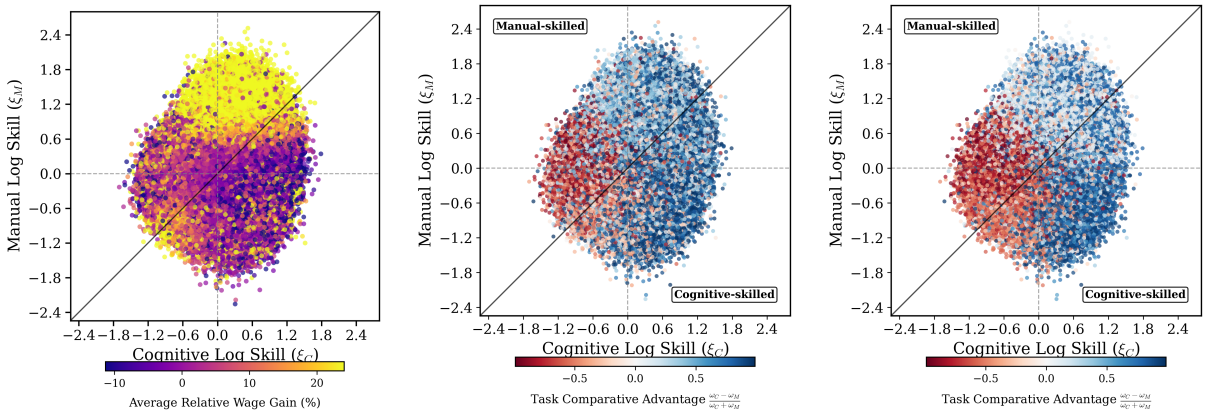
8.2 The Direct Effects of AI

I compare the pre- and post-shock stationary equilibria under imperfect information. AI raises aggregate productivity, but its wage effects are heterogeneous and polarizing. Figure 17 maps these effects onto the space of true skills and reveals the mechanism.

Panel (a) plots the average lifecycle wage gain from AI for each worker at her true skills. Workers with strong manual comparative advantage—high-ability manual workers in the upper-left region where $\xi_M > \xi_C$—gain the most, because AI raises the relative return to the manual skills in which they have a strong comparative advantage. Cognitive specialists in the lower-right region lose as cognitive tasks, in which they have a strong comparative advantage, become less important. Generalists near the 45° line are roughly unaffected.

Panels (b) and (c) trace these wage effects to occupational reallocation, showing the net task content of the occupations a worker holds over the life cycle before and after the shock. Workers, especially those with a manual comparative advantage, sort into more manual-intensive occupations after AI (the red region expands). Cognitive specialists instead remain in cognitive-intensive occupations but earn lower returns there. On the other hand, high ability generalists benefit by moving towards more balanced occupations.

These effects in the skill space aggregate into the familiar polarization pattern of the routinization literature (Autor et al. 2003): across the baseline wage distribution the wage gains are U-shaped—workers in the middle lose while those at the bottom and top gain—and employment shifts out of cognitive-intensive clusters and into manual-intensive ones, whose occupation prices rise in general equilibrium (Figure 34 in the Appendix).



(a) Average wage gain from AI

(b) Net task content of occupations held, pre-AI

(c) Net task content of occupations held, post-AI

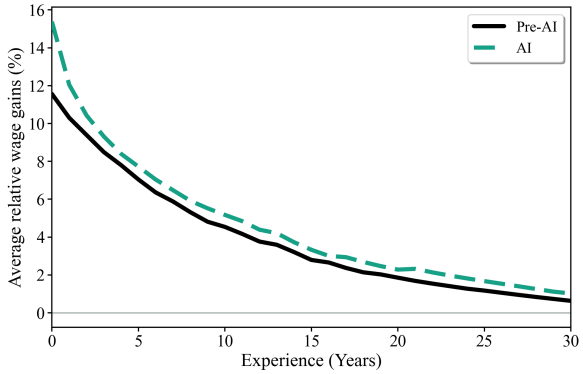
Figure 17: The direct effects of AI in the space of true skills. In each panel a point is a simulated worker placed at her true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis); the solid line is the 45° line $\xi_C = \xi_M$, which separates cognitive specialists (below) from manual specialists (above). *Panel (a)*: color is the worker's average relative wage gain from AI over the life cycle, in percent (yellow = gains, dark = losses). *Panels (b) and (c)*: color is the average net task comparative advantage $(\omega_C - \omega_M)/(\omega_C + \omega_M)$ of the occupations the worker holds over her career—blue = cognitive-intensive, red = manual-intensive—in the pre-AI (b) and post-AI (c) stationary equilibrium.

8.3 AI and the Implications of Skill Uncertainty

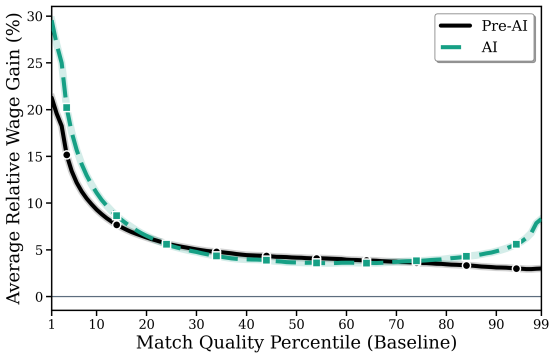
The analysis so far compares stationary equilibria and therefore captures the long-run incidence of AI in a world with information frictions. I now ask how AI changes the implications of skill uncertainty itself, that is, how it reshapes the wage cost of imperfect information studied in Section 7. Figure 18 recomputes the wage gain from giving workers perfect information about their skills, once in the baseline economy and once in the post-AI economy, along the same two dimensions as Figure 11: over the life cycle and across the match-quality distribution. In the aggregate, the output loss generated by imperfect information about skills increases from 5.2% in the baseline to 5.9% under AI.

AI raises the cost of skill uncertainty over the whole career path. Panel (a) shows that the wage gain from perfect information is higher in the economy affected by the AI shock at every level of experience: at labor-market entry it rises from about 11.5% to about 15.5%, and the wedge between the two curves persists as experience accumulates. By raising the dispersion of returns across occupations, AI makes correct sorting more valuable, so the cost of holding the wrong occupation under imperfect beliefs is higher.

Panel (b) shows who accounts for this increase. The extra cost of skill uncertainty under AI is concentrated at the two ends of the match-quality distribution. For the worst-matched workers within occupation, the gain from perfect information rises from about 21% to about 30%. At the top of the distribution AI also increase the gains of removing information frictions: wage gains for well-matched workers within occupation now would gain 8% compared 3% in the pre-AI scenario.



(a) Wage gains over the life cycle: pre-AI vs AI



(b) Wage gains by match-quality percentile: pre-AI vs AI

Figure 18: The wage cost of skill uncertainty before and after AI. Both panels report the average relative wage gain from moving to perfect information, computed separately in the pre-AI (black, solid) and post-AI (teal, dashed) stationary equilibria. *Panel (a)*: by years of potential experience; the vertical gap between the two curves is the additional value of information created by AI. *Panel (b)*: by the worker's within-(time, occupation) wage-match percentile in the corresponding baseline economy. Shaded bands are bootstrap confidence intervals.

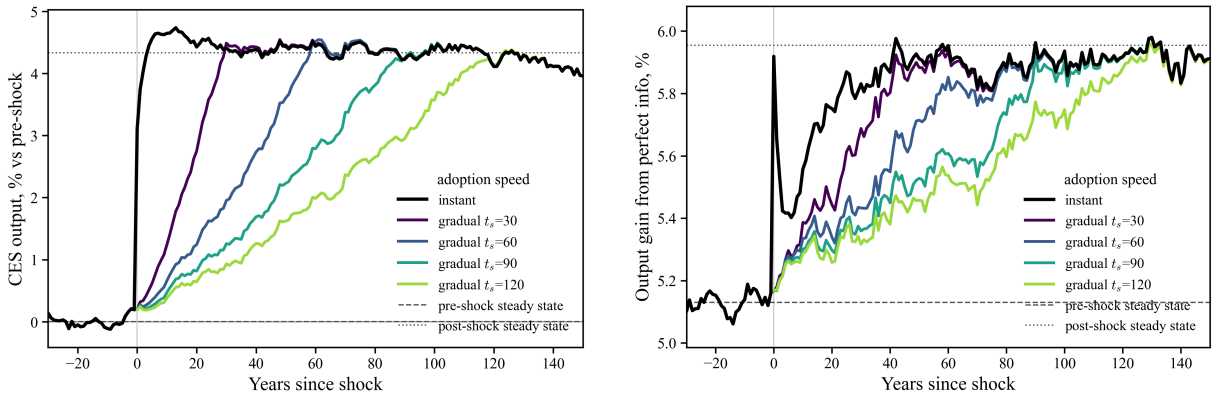
8.4 Output and Misallocation along the Transition Path

The counterfactuals above compare stationary equilibria and therefore speak to the long-run incidence of the AI shock, abstracting from the adjustment dynamics that follow it. I now study the effects of the AI shock along the transition path, in particular I compute the transition path between the pre-shock to the post-shock equilibrium and to understand the consequences of information frictions not only in the long-run, but also on impact and along the transition path.

I model the AI shock—the LLM-exposure shock introduced above, which lowers the cognitive-task content of occupations in proportion to their exposure—as an MIT shock: it is an unanticipated shock and workers have perfect foresight over the entire subsequent path of prices and technology. I consider MIT shocks that take different amounts of time for the technology to adjust to the post-shock scenario. Under instantaneous AI adoption, the economy immediately jumps to its post-shock level at time 0; under gradual AI adoption the economy takes some time to adjust, so that the task content of occupations and the associated returns to skills and experience adjust smoothly toward the post-shock economy. When the transition is gradual, workers still have perfect foresight over the entire transition path: they know at $t = 0$ how the technology will be adopted over time unfold and optimize against the whole anticipated sequence of prices and task weights. In both cases the path for equilibrium prices is solved via a backward–forward shooting algorithm, holding the CES weights and the per-period numeraire fixed exactly as in the stationary problem; the algorithm used to compute this transition path is described in Appendix C.3.

This experiment captures the response of aggregate output and misallocation driven by information frictions about skills in the new long-run stationary equilibrium and along the transition path. Figure 19 reports output and misallocation along the transition paths for different adoption speed. Panel (a) plots aggregate output in the baseline imperfect-information economy, for an instantaneous shock and for gradual diffusion over $t_s \in \{30, 60, 90, 120\}$ periods. Panel (b) plots the corresponding output gain from moving to the perfect-information economy, that is,

the gain from removing information frictions along the transition.



(a) Aggregate output

(b) Output gain from perfect information

Figure 19: Aggregate output along the AI transition. *Panel (a)*: CES aggregate output in the baseline imperfect-information economy, percent deviation from the pre-shock steady state. *Panel (b)*: the output gain from moving to the perfect-information economy, $(Y^{PI} - Y^{II})/Y^{II}$, that is the gain from removing information frictions. Horizontal axis: years since the shock. Lines correspond to different adoption speeds—instantaneous (black) or gradual diffusion over $t_s \in \{30, 60, 90, 120\}$ periods; the dashed and dotted lines mark the pre- and post-shock steady states.

Two results stand out. First, the AI shock increases output at every adoption speed, but the timing of the gains depends on how fast the technology diffuses (Panel (a)). Under instantaneous adoption, aggregate output jumps to about +4.3% above its pre-shock level on impact, overshoots slightly to roughly +4.7% about a decade out as workers reallocate, and then settles at its new long-run level of about +4.3%. Under gradual diffusion, output barely moves on impact and then climbs monotonically as task importance within occupations changes. Every adoption speed converges to the same long-run gain, which is pinned down by the post-shock ergodic distribution and is therefore independent of the speed of adoption.

Second, AI increases misallocation generated by information frictions about skills (Figure 19 Panel (b)). The output gain from moving to perfect information rises from about 5.2% before the shock to roughly 5.9% in the new long run. Slower diffusion delays this increase in misallocation, at the cost of slowing down output growth. In short, the AI shock raises long-run output by about 4.3% at every adoption speed, with faster adoption front-loading the aggregate output gains, while the output gain available from removing information frictions rises permanently from about 5.2% to roughly 5.9%.

From an aggregate perspective, this counterfactual exercise suggests that policymakers may want to incentivize faster AI adoption, as faster adoption front-loads aggregate output gains. However, we may want to do so while also delivering the information interventions studied in Section 7.2, because while AI adoption raises aggregate output, it also raises misallocation generated by information frictions.

9 Conclusion

I develop and estimate a general equilibrium dynamic Roy model in which workers have unobserved multidimensional skills, accumulate task-specific experience, and learn about their comparative advantage through noisy wage signals. Estimated on Portuguese administrative data with eight occupations and two skill dimensions (cognitive and manual) the model reproduces the U-shaped and directional occupational mobility documented by Groes et al. (2015). It interprets the moves of workers with low match quality within their occupation, those most likely to be mismatched, as driven by learning: upward mobility of well-matched workers reflects experience accumulation and non-wage factors, while the downward, task-reallocating moves at the bottom arise only when workers update their beliefs from wages.

Removing information frictions about skills raises aggregate output by 5.2%. The individual wage gain from perfect information is largest at entry, about 11.6%, and falls to roughly zero after thirty years. Almost all of the early wage gain comes from better sorting on skills rather than from additional experience accumulation, and it is heterogeneous across workers within the same occupation: those at the bottom of the within-occupation match-quality distribution, who are most likely to be mismatched, would earn over 20% more under perfect information, against about 3% for those at the top. Specialists, that is, workers with a strong comparative advantage, are the ones who would have the highest wage gains had they held correct beliefs.

I then study policies that can reduce this uncertainty and ask how much of the lost output it can recover. Calibrated to the average treatment effect of Kiss et al. (2023), an information treatment about comparative advantage in skills delivered at entry raises aggregate output by 0.91%, with the same incidence as the full information scenario: the worst-matched workers within occupation gain the most, the best-matched almost nothing. The effectiveness of this information treatment decays sharply when the assessment is delivered later in the career, due to the presence of imperfectly transferable human capital. Reducing the cost of switching occupations also increases output and wages, but not if the decrease is too large: modest reductions of the switching cost allow workers to reallocate more after learning their comparative advantage, but eliminating switching costs induces excessive churning that destroys task-specific experience and lowers aggregate output and wages.

Finally, I use the model to study how information frictions shape occupational reallocation to technology shocks that change the task content of occupations. I simulate the job transformation induced by Large Language Model (LLM) adoption: an AI shock that lowers time spent on cognitive tasks in proportion to each occupation's exposure, modeled as an unanticipated change with perfect foresight over a transition path that can be instantaneous or gradual. The shock raises long-run aggregate output (+4.3%) at every adoption speed, with faster adoption front-loading the gains. Crucially, it makes skill uncertainty more costly: the output gap between the imperfect- and full-information economies widens along the transition, from about 5.2% before the shock to roughly 5.9% in the new long run.

References

Acemoglu, D. and Autor, D. (2011), Skills, tasks and technologies: Implications for employment and earnings, *in* 'Handbook of Labor Economics', Vol. 4, Elsevier, pp. 1043–1171.

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

Althoff, L. and Reichardt, H. (2026), Task-specific technical change and comparative advantage, CESifo Working Paper 12403, CESifo.
URL: *https://ssrn.com/abstract=6084170*

Altonji, J. G. and Pierret, C. R. (2001), 'Employer learning and statistical discrimination', *The Quarterly Journal of Economics* **116**(1), 313–350.

Alvarez, F. and Shimer, R. (2011), 'Search and rest unemployment', *Econometrica* **79**(1), 75–122.

Antonovics, K. and Golan, L. (2012), 'Experimentation and job choice', *Journal of Labor Economics* **30**(2), 333–366.

Arcidiacono, P., Aucejo, E., Maurel, A. and Ransom, T. (2025), 'College attrition and the dynamics of information revelation', *Journal of Political Economy* **133**(1), 53–110.

Arcidiacono, P. and Miller, R. A. (2011), 'Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity', *Econometrica* **79**(6), 1823–1867.

Autor, D. H. and Dorn, D. (2013), 'The growth of low-skill service jobs and the polarization of the us labor market', *American economic review* **103**(5), 1553–1597.

Autor, D. H. and Handel, M. J. (2013), 'Putting tasks to the test: Human capital, job tasks, and wages', *Journal of Labor Economics* **31**(S1), S59–S96.

Autor, D. H., Katz, L. F. and Kearney, M. S. (2006), 'The polarization of the us labor market', *American economic review* **96**(2), 189–194.

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.

Baley, I., Figueiredo, A. and Ulbricht, R. (2022), 'Mismatch cycles', *Journal of Political Economy* **130**(11), 2943–2984.

Bunting, J., Diegert, P. and Maurel, A. (2024), Heterogeneity, uncertainty and learning: Semiparametric identification and estimation, Technical report, National Bureau of Economic Research.

Camargo, B., Lange, F. and Pastorino, E. (2022), On the role of learning, human capital, and performance incentives for wages, NBER Working Paper 30653, National Bureau of Economic Research.

Card, D., Cardoso, A. R. and Kline, P. (2016), 'Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women', *The Quarterly journal of economics* **131**(2), 633–686.

Chodorow-Reich, G. and Wieland, J. (2020), 'Secular labor reallocation and business cycles', *Journal of Political Economy* **128**(6), 2245–2287.

de Melo, R. L. and Papageorgiou, T. (2025), Occupational choice, human capital and learning: A multi-armed bandit approach, Technical Report 1076, Boston College Department of Economics. Revised August 1, 2025.

de Paula, Á., Gualdani, C., Pastorino, E. and Salgado, S. (2025), Uncertainty, learning about productivity, and human capital acquisition: A reassessment of sorting, Technical report, Working Paper.

Deming, D. J. (2025), Why do wages grow faster for educated workers?, NBER Working Paper 31373, National Bureau of Economic Research.

Dix-Carneiro, R. (2014), 'Trade liberalization and labor market dynamics', *Econometrica* **82**(3), 825–885.

Dvorkin, M. A. and Monge-Naranjo, A. (2019), Occupation mobility, human capital and the aggregate consequences of task-biased innovations, Working Paper 2019-013, Federal Reserve Bank of St. Louis.

Eloundou, T., Manning, S., Mishkin, P. and Rock, D. (2024), 'Gpts are gpts: Labor market impact potential of llms', *Science* **384**(6702), 1306–1308.

Farber, H. S. and Gibbons, R. (1996*a*), 'Learning and wage dynamics', *The Quarterly Journal of Economics* **111**(4), 1007–1047.

Farber, H. S. and Gibbons, R. (1996*b*), 'Learning and wage dynamics', *The Quarterly Journal of Economics* **111**(4), 1007–1047.

Freund, L. B. and Mann, L. F. (2025), Job transformation, specialization, and the labor market effects of ai, CESifo Working Paper 12072, CESifo. Available at SSRN: https://ssrn.com/bstract=5466890.
URL: *https://ssrn.com/abstract=5466890*

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 Katz, L. F. (1991), 'Layoffs and lemons', *Journal of Labor Economics* **9**(4), 351–380.

Gibbons, R., Katz, L. F., Lemieux, T. and Parent, D. (2005*a*), 'Comparative advantage, learning, and sectoral wage determination', *Journal of Labor Economics* **23**(4), 681–724.

Gibbons, R., Katz, L. F., Lemieux, T. and Parent, D. (2005*b*), 'Comparative advantage, learning, and sectoral wage determination', *Journal of labor economics* **23**(4), 681–724.

Gibbons, R. and Waldman, M. (1999*a*), Careers in organizations: Theory and evidence, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Part B, Elsevier, pp. 2373–2437.

Gibbons, R. and Waldman, M. (1999*b*), 'A theory of wage and promotion dynamics inside firms', *The Quarterly Journal of Economics* **114**(4), 1321–1358.

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

Gibbons, R. and Waldman, M. (2006), 'Enriching a theory of wage and promotion dynamics inside firms', *Journal of Labor Economics* **24**(1), 59–108.

Goller, D., Brox, E. and Wolter, S. C. (2026), Imperfect self-knowledge about skills and skill mismatch, Technical report, University of Zurich, Department of Business Administration (IBW).

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), 210–44.

Harris, M. and Holmström, B. (1982), 'A theory of wage dynamics', *The Review of Economic Studies* **49**(3), 315–333.

Heckman, J. J. and Honore, B. E. (1990), 'The empirical content of the roy model', *Econometrica: Journal of the Econometric Society* pp. 1121–1149.

Heckman, J. J., Stixrud, J. and Urzua, S. (2006), 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor economics* **24**(3), 411–482.

Hincapié, A. (2020), 'Entrepreneurship over the life cycle: Where are the young entrepreneurs?', *International Economic Review* **61**(2), 617–681.

Hotz, V. J. and Miller, R. A. (1993), 'Conditional choice probabilities and the estimation of dynamic models', *Review of Economic Studies* **60**(3), 497–529.

Hsieh, C.-T., Hurst, E., Jones, C. I. and Klenow, P. J. (2019), 'The allocation of talent and u.s. economic growth', *Econometrica* **87**(5), 1439–1474.

James, J. (2011), Ability matching and occupational choice, Working Paper 11-25, Federal Reserve Bank of Cleveland.

James, J. (2012), Learning and occupational sorting, Working Paper 12-25, Federal Reserve Bank of Cleveland.

Jovanovic, B. (1979), 'Job matching and the theory of turnover', *Journal of Political Economy* **87**(5), 972–990.

Kambourov, G. and Manovskii, I. (2009), 'Occupational mobility and wage inequality', *The Review of Economic Studies* **76**(2), 731–759.

Kennan, J. and Walker, J. R. (2011), 'The effect of expected income on individual migration decisions', *Econometrica* **79**(1), 211–251.

Kiss, A., Garlick, R., Orkin, K. and Hensel, L. (2023), 'Jobseekers' beliefs about comparative advantage and (mis) directed search'.

Lange, F. (2007*a*), 'The speed of employer learning', *Journal of Labor Economics* **25**(1), 1–35.

Lange, F. (2007*b*), 'The speed of employer learning', *Journal of Labor Economics* **25**(1), 1–35.

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

Lise, J. and Postel-Vinay, F. (2020), 'Multidimensional skills, sorting, and human capital accumulation', *American Economic Review* **110**(8), 2328–76.

Lucas, R. E. and Prescott, E. C. (1974), 'Equilibrium search and unemployment', *Journal of Economic Theory* **7**(2), 188–209.

Mihaylov, E. and Tijdens, K. G. (2019), Measuring the routine and non-routine task content of 427 four-digit isco-08 occupations, Technical report, Tinbergen Institute discussion paper.

Miller, R. A. (1984), 'Job matching and occupational choice', *Journal of Political economy* **92**(6), 1086–1120.

Neal, D. A. (1999), 'The complexity of job mobility among young men', *Journal of Labor Economics* **17**(2), 237–261.

Papageorgiou, T. (2014), 'Learning your comparative advantages', *Review of Economic Studies* **81**(3), 1263–1295.

Pastorino, E. (2024), 'Careers in firms: The role of learning about ability and human capital acquisition', *Journal of Political Economy* **132**(6), 1994–2073.

Pavan, R. (2011), 'Career choice and wage growth', *Journal of Labor Economics* **29**(3), 549–587.

Pilossoph, L. (2014), A multisector equilibrium search model of labor reallocation. Working paper.

Poletaev, M. and Robinson, C. (2008), 'Human capital specificity: Evidence from the DOT and displaced worker surveys, 1984–2000', *Journal of Labor Economics* **26**(3), 387–420.

Roy, A. D. (1951), 'Some thoughts on the distribution of earnings', *Oxford economic papers* **3**(2), 135–146.

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

Sorkin, I. (2018), 'Ranking firms using revealed preference', *The quarterly journal of economics* **133**(3), 1331–1393.

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.
URL: *https://doi.org/10.2307/2118478*

Traiberman, S. (2019), 'Occupations and import competition: Evidence from denmark', *American Economic Review* **109**(12), 4260–4301.

A Appendix: Proofs

A.1 Identification of the Wage Equation Parameters under Learning

This subsection proves identification of the wage equation in the presence of multidimensional skills that are unobserved both by the econometrician and by the workers. The argument follows the pure-learning identification logic of Bunting et al. (2024), adapted to the multidimensional skill environment of this paper. The key idea is that workers sort into occupations based on beliefs about their latent skills, but those beliefs are formed from the same wage and occupational histories observed by the econometrician. Hence, controlling for the worker's history control for belief-based selection as workers do not sort on any additional private information with respect to what the econometrician can observe.

The wage equation is

$$w_{it}(j) = p_{jt} + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}, \tag{45}$$

where p_{jt} is the occupation-time price, $h_j(\boldsymbol{\tau}_{it})$ is the human-capital component, $\boldsymbol{\lambda}_j$ is the occupation loading on the latent skill vector, $\boldsymbol{\xi}_i$ is fixed over the worker's life, and ε_{ijt} is the transitory wage shock.

I make the following assumptions.

Assumption 1 (Linear-Gaussian skills and wage shocks) *Latent skills are Gaussian within group, $\boldsymbol{\xi}_i \mid g(i) = g \sim N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ with $\boldsymbol{\Sigma}_g \succ 0$, and are fixed over the worker's life. Wage shocks are $\varepsilon_{ijt} \sim N(0, \varsigma^2_{\varepsilon j})$ with $\varsigma^2_{\varepsilon j} > 0$, independent across workers and periods, independent of $\boldsymbol{\xi}_i$, and realized after the occupational choice is made.*

Assumption 2 (Pure-learning selection) *Let*

$$\mathcal{H}_{it} = \left(g(i),\, d_i^{t-1},\, w_i^{t-1},\, \boldsymbol{\tau}_i^t,\, \{p_{j\ell}\}_{j\in\mathcal{J},\ell\geq t}\right)$$

be the history observed by the econometrician when worker i chooses d_{it}. Workers observe no private signal about $\boldsymbol{\xi}_i$ beyond this history:

$$d_{it} \;\perp\!\!\!\perp\; \left(\boldsymbol{\xi}_i,\, \{\varepsilon_{ijs} : s \geq t,\, j \in \mathcal{J}\}\right) \;\mid\; \mathcal{H}_{it}.$$

Assumption 3 (Overlap) *Every occupation used in the identifying moments is chosen with positive probability in the histories used:* $\Pr(d_{it} = j \mid \mathcal{H}_{it}) > 0$.

Assumption 4 (Repeated observations and mobility) *Within each occupation j, task experience $\boldsymbol{\tau}_{it}$ varies across the workers observed in the same occupation and year. Across occupations, workers are mobile enough that, for each occupation j and each anchor $a_k \in \mathcal{J}_A = \{a_1, \ldots, a_K\}$, wages in j and in a_k are observed for the same worker at two distinct dates.*

Assumption 5 (Normalizations) *The K anchor occupations load on a single skill each, $\boldsymbol{\lambda}_{a_k} = \boldsymbol{e}_k$ for $k = 1, \ldots, K$; the location of the skill distribution is fixed by $\sum_{g\in\mathcal{G}} \pi_g \boldsymbol{\mu}_g = \mathbf{0}$, with π_g the population share of group g; and the returns to task experience are measured from a base level, $h_j(\boldsymbol{\tau}^0) = 0$ for every j.*

Assumption 2 is the central restriction. Workers sort into occupations on the basis of their beliefs, but those beliefs are formed from histories that the econometrician also observes, so conditional on $\mathcal{H}_{it}$ their choice has no additioanl information about unobserved skills. Assumption 3 guarantees that inverse-probability weights used to control for selection are well identified. Assumption 4 ensure that workers are observed for enough periods and that there are enough movers across occupations to identify the various components of the wage equation: experience variation within an occupation–year separates the returns to task experience from the occupation price; mobility across occupations pin down the wage covariances that will in turn the multidimensional skill permanent match component and the stationary price vector separately idiosyncratic noise. Assumption 5 is needed to separately identify each components of the wage equation. In particular it fixes the rotation, location and scale of the latent skill space, which the wage equation leaves free: the anchors pin down the rotation, the zero-mean condition on the population of workers the location, and the base experience level the split between the price of an occupation and the return to experience.

Theorem 1 (Identification of the wage equation parameters under learning) *Suppose Assumptions 1– 5 hold. Then the wage-equation parameters*

$$\left\{ (p_{jt}, \boldsymbol{\lambda}_j, \delta_j, \varsigma^2_{\varepsilon j})_{j\in\mathcal{J},t},\ (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g\in\mathcal{G}} \right\}$$

are identified from the joint distribution of observed wages, occupational choices, groups, and predetermined state variables.

Proof of Theorem 1.

I first show how the observed history can be used as an inverse-probability weight to control for selection into occupations created by beliefs. Fix a worker in group g, and consider a realized occupational history

$$d^t = (d_1, \ldots, d_t).$$

After subtracting the occupation price and human-capital component, each past wage is a noisy signal of the worker's latent skill:

$$S_{is} \equiv w_{is}(d_s) - p_{d_s s} - h_{d_s}(\boldsymbol{\tau}_{is}) = \boldsymbol{\lambda}'_{d_s}\boldsymbol{\xi}_i + \varepsilon_{id_s s}.$$

Stack the signals observed before period t into

$$\boldsymbol{S}_{i,t-1} = (S_{i1}, \ldots, S_{i,t-1})'.$$

Let $\boldsymbol{\Lambda}_{i,t-1}$ be the matrix whose sth row is $\boldsymbol{\lambda}'_{d_s}$, and let

$$\boldsymbol{\Omega}_{i,t-1} = \operatorname{diag}\left(\varsigma^2_{\varepsilon d_1}, \ldots, \varsigma^2_{\varepsilon d_{t-1}}\right).$$

Because the prior over $\boldsymbol{\xi}_i$ is Gaussian and the wage signals are linear and Gaussian, the posterior is Gaussian:

$$\boldsymbol{\xi}_i \mid \mathcal{H}_{it}, g(i) = g \sim N\left(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}\right),$$

where

$$\tilde{\boldsymbol{\Sigma}}_{it} = \left(\boldsymbol{\Sigma}_g^{-1} + \boldsymbol{\Lambda}'_{i,t-1}\boldsymbol{\Omega}^{-1}_{i,t-1}\boldsymbol{\Lambda}_{i,t-1}\right)^{-1},$$

and

$$\tilde{\boldsymbol{\xi}}_{it} = \tilde{\boldsymbol{\Sigma}}_{it}\left(\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\mu}_g + \boldsymbol{\Lambda}_{i,t-1}'\boldsymbol{\Omega}_{i,t-1}^{-1}\boldsymbol{S}_{i,t-1}\right).$$

These are also the belief objects that enter the worker's state x_{it}. Conditional on the history, the potential log wage in occupation j is normally distributed according to

$$w_{it}(j) \mid \mathcal{H}_{it}, g(i) = g \sim N\left(p_{jt} + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\xi}}_{it},\ \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\Sigma}}_{it}\boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2\right).$$

The history is therefore a sufficient summary of what is known about latent skill. Note that the worker's beliefs and mine coincide only if her prior is the true skill distribution. Identification of the wage equation does not require this: it is enough that the worker holds no private information beyond what I observe.

The observed wage is of course selected: workers choose occupations on the basis of beliefs about their skills. But those beliefs are generated by the observed history, so once I condition on $\mathcal{H}_{it}$, the choice reveals nothing further about $\boldsymbol{\xi}_i$ or about future wage shocks (Assumption 2). Let $h_t(j \mid \mathcal{H}_{it}) \equiv \Pr(d_{it} = j \mid \mathcal{H}_{it})$ be the conditional choice probability, which is identified from the joint distribution of histories and choices. To see how it works as a weight, write the joint density sequentially, using the timing of the model: the worker enters period t with history $\mathcal{H}_{it}$, chooses her occupation, and only then observes her wage. Hence, we have

$$f_{w^T,D^T|g}(w^T,d^T) = \prod_{t=1}^{T} \underbrace{h_t(d_t \mid \mathcal{H}_{it})}_{\text{observed choices}}\ \underbrace{f_{w_t|d_t,\mathcal{H}_{it},g}(w_t)}_{\text{observed wages}} = \prod_{t=1}^{T} \underbrace{h_t(d_t \mid \mathcal{H}_{it})}_{\text{selection}}\ \underbrace{f_{w_t(d_t)|\mathcal{H}_{it},g}(w_t)}_{\text{potential wages}},$$

The second equality is where the potential wage appears. The wage I observe in j is the potential wage in j, $w_{it} = w_{it}(d_{it})$; and since $w_{it}(j) = p_{jt} + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j'\boldsymbol{\xi}_i + \varepsilon_{ijt}$ depends only on $(\boldsymbol{\xi}_i, \varepsilon_{ijt})$, which Assumption 2 makes independent of the choice given the history, the conditioning on $d_t = j$ can be dropped:

$$f_{w_t|d_t=j,\mathcal{H}_{it},g}(w) = f_{w_t(j)|d_t=j,\mathcal{H}_{it},g}(w) = f_{w_t(j)|\mathcal{H}_{it},g}(w).$$

Conditional on their history, the workers who choose j are not from a pool of workers that have selected set of skills. All the selection sits in the choice probabilities that, infact, we can use as weights to control for this belief-based selection mechanism. Dividing by the choice probabilities therefore allow us to obtain the potential wage path along any supported sequence $d^T = (d_1, \ldots, d_T)$,

$$\underbrace{\prod_{t=1}^{T} f_{w_t(d_t)|\mathcal{H}_{it},g}(w_t)}_{\text{potential wage path}} = \frac{\overbrace{f_{w^T,D^T|g}(w^T,d^T)}^{\text{observed}}}{\underbrace{\prod_{t=1}^{T} h_t(d_t \mid \mathcal{H}_{it})}_{\text{identified from choices}}}.$$

Everything on the right is in the data, so the distribution of potential log wages along every supported path is identified. This weighting scheme, ensure that workers are weighted by how likely/unlikely that path was for them i.e. workers who believe that they have high cognitive skills are more likely to self-select into occupations with a high cognitive task content. Every moment I use below is a moment of this reweighted distribution.

Under the wage equation, the potential wage path associated with a sequence d^T,

$$w_i(d^T) = \big(w_{i1}(d_1), \ldots, w_{iT}(d_T)\big)',$$

is jointly Gaussian, because it is a linear map of the Gaussian skill vector $\boldsymbol{\xi}_i$ and the Gaussian wage shocks. To characterize the path of potential wages, the mean and covariance are therefore all I need, and repeated measurement ensure I can identify them: for a worker observed multiple times, the observed wages share her permanent skill $\boldsymbol{\xi}_i$ while shocks are independent over time. What is left after reweighting is thus the same latent skill vector, with different loadings if the worker switches occupation, and independent noise. I write the three moments below, after netting out the returns to task experience.

Identification of the returns to task experience δ_j. Because $h_j(\cdot)$ is a known function up to unknown parameters, let $h_j(\boldsymbol{\tau}) = \boldsymbol{\delta}_j' \boldsymbol{D}(\boldsymbol{\tau})$ be a linear-in-parameters representation, where

$$\boldsymbol{D}(\boldsymbol{\tau})$$

is a vector of indicators for the task-specific tenure bins, and $\boldsymbol{\delta}_j$ stacks the corresponding coefficients. Using the reweighted potential wage distribution, compare two wages for the same worker in the same occupation j at two dates t and s:

$$w_{it}(j) - w_{is}(j) = (p_{jt} - p_{js}) + \big[h_j(\boldsymbol{\tau}_{it}) - h_j(\boldsymbol{\tau}_{is})\big] + \varepsilon_{ijt} - \varepsilon_{ijs}.$$

Comparing a worker with herself, in the same occupation, differences out her permanent skill $\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i$; the price is absorbed by occupation-time effects (and cancels out if prices are constant i.e. if the model is generated by a stationary equilibrium). What is left is the return to the task experience she accumulated in between. The wage shocks are mean zero and independent of predetermined experience, so $\boldsymbol{\delta}_j$ is identified, provided experience varies across the workers observed in the same occupation and year—otherwise experience is indistinguishable from calendar time. With $h_j(\cdot)$ in hand, define the human-capital-adjusted log wage

$$Z_{it}(j) \equiv w_{it}(j) - h_j(\boldsymbol{\tau}_{it}).$$

Then

$$Z_{it}(j) = p_{jt} + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}.$$

Everything that follows uses the means and covariances of $Z_{it}(j)$, which is the wage stripped of experience and therefore a direct measurement of skill.

The moments of the adjusted wage. Three moments of the reweighted distribution of adjusted wages are identified, and they are all I use below. Write

$$\pi_i(d^T) \equiv \prod_{t=1}^{T} h_t(d_t \mid \mathcal{H}_{it})$$

for the probability that worker i follows the path d^T, and recall that each moment of the potential wage path is a π-weighted moment of the observed data.

The first is the mean adjusted wage in occupation j, group g, period t,

$$m_{jgt} \equiv \mathbb{E}\left[\frac{\mathbf{1}\{d_{it} = j\}\, Z_{it}}{\pi_i(d^T)} \;\middle|\; g(i) = g\right] = \mathbb{E}[Z_{it}(j) \mid g(i) = g] = p_{jt} + \boldsymbol{\lambda}_j' \boldsymbol{\mu}_g.$$

Only the workers observed in j at t enter, each counted by the inverse of the probability that she got there.

The second is the covariance between wages in occupations j and ℓ at two different dates, computed on the workers observed in j at t *and* in ℓ at s, reweighted the same way,

$$C_{j\ell,g} \equiv \mathbb{E}\left[\frac{\mathbf{1}\{d_{it} = j,\, d_{is} = \ell\}\, (Z_{it} - m_{jgt})(Z_{is} - m_{\ell gs})}{\pi_i(d^T)} \;\middle|\; g(i) = g\right] = \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_\ell, \qquad t \neq s,$$

which contains only permanent skill, since the transitory shocks are independent across periods. The third is the same-period variance in occupation j,

$$V_{jg} \equiv \mathbb{E}\left[\frac{\mathbf{1}\{d_{it} = j\}\, (Z_{it} - m_{jgt})^2}{\pi_i(d^T)} \;\middle|\; g(i) = g\right] = \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2,$$

which contains skill and noise together. Comparing the weighted covariance $C_{j\ell,g}$ and the weighted variance V_{jg} allow to identify the variance of skills of each group g and the occupation-specific noise separately. Without the weights each of these would be computed on a selected sample—the workers who chose that occupation, or that pair of occupations.

Identification of the skill distribution: the means $\boldsymbol{\mu}_g$ and the anchor prices. Let the K anchor occupations be

$$\mathcal{J}_A = \{a_1, \ldots, a_K\},$$

with normalized loadings

$$\boldsymbol{\lambda}_{a_k} = \boldsymbol{e}_k.$$

Stack the mean adjusted wages of the anchor occupations:

$$\boldsymbol{m}_{Agt} = \left(m_{a_1 gt}, \ldots, m_{a_K gt}\right)'.$$

Also define the vector of anchor prices

$$\boldsymbol{p}_{At} = \left(p_{a_1 t}, \ldots, p_{a_K t}\right)'.$$

Because the anchor loading matrix is the identity, the mean adjusted wages identify the sum the mean skill of group g in dimension k with the anchor price in the normalized occupation

$$\boldsymbol{m}_{Agt} = \boldsymbol{p}_{At} + \boldsymbol{\mu}_g.$$

The price is common across groups, the skill mean is not. Using the normalization

$$\sum_{g \in \mathcal{G}} \pi_g \boldsymbol{\mu}_g = \mathbf{0},$$

we obtain

$$\boldsymbol{p}_{At} = \sum_{g \in \mathcal{G}} \pi_g \boldsymbol{m}_{Agt}.$$

Then

$$\boldsymbol{\mu}_g = \boldsymbol{m}_{Agt} - \boldsymbol{p}_{At}.$$

The anchor prices and the group skill means are therefore identified: differences in mean anchor wages across groups reveal how skilled each group is, and the normalization says where zero is. Since skills do not change over the life cycle, the same $\boldsymbol{\mu}_g$ must come out at every date—an overidentifying restriction.

Identification of the skill distribution: the covariance matrices $\boldsymbol{\Sigma}_g$. Let $\boldsymbol{C}_{AA,g}$ be the $K \times K$ covariance matrix whose (k, ℓ) element is

$$C_{a_k a_\ell, g} = \operatorname{Cov}\left(Z_{it}(a_k), Z_{is}(a_\ell) \mid g(i) = g\right), \qquad t \neq s.$$

Using the anchor normalization,

$$C_{a_k a_\ell, g} = \boldsymbol{e}_k' \boldsymbol{\Sigma}_g \boldsymbol{e}_\ell.$$

Therefore,

$$\boldsymbol{C}_{AA,g} = \boldsymbol{\Sigma}_g.$$

So the skill covariance matrix is read directly off wages in the anchor occupations: two wages of the same worker in the anchors in two different periods depend only on the worker's permanent skills, and their covariance is exactly $\boldsymbol{\Sigma}_g$.

Identification of the skill loadings $\boldsymbol{\lambda}_j$. For any non-anchor occupation j, define the vector of covariances between occupation j and the anchors:

$$\boldsymbol{C}_{jA,g} = \left(C_{ja_1,g}, \ldots, C_{ja_K,g}\right)'.$$

From the covariance formula,

$$\boldsymbol{C}_{jA,g}' = \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g.$$

Since $\boldsymbol{\Sigma}_g$ has already been identified,

$$\boldsymbol{\lambda}_j' = \boldsymbol{C}_{jA,g}' \boldsymbol{\Sigma}_g^{-1}.$$

The loadings of every occupation are therefore identified: how strongly a worker's wage in occupation j co-moves with her wage in each anchor tells me how much j pays for each skill. The anchors themselves are fixed by normalization. With more than one group, the same $\boldsymbol{\lambda}_j$ must be recovered from each of them—a further overidentifying restriction.

Identification of the occupation prices p_{jt}. Once $\boldsymbol{\lambda}_j$ and $\boldsymbol{\mu}_g$ are known, the mean adjusted wage equation

$$m_{jgt} = p_{jt} + \boldsymbol{\lambda}_j' \boldsymbol{\mu}_g$$

identifies the occupation-time price:

$$p_{jt} = m_{jgt} - \boldsymbol{\lambda}_j' \boldsymbol{\mu}_g.$$

Prices are thus identified for every occupation and period: once I know what an occupation pays for skill, and how much skill each group has, whatever is left in the mean wage is the price. The same p_{jt} must emerge from every group, since all groups face the same prices.

Identification of the transitory wage variances $\varsigma^2_{\varepsilon j}$. Finally, the adjusted wage variance satisfies

$$V_{jg} = \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j + \varsigma^2_{\varepsilon j}.$$

The first term is now known, since $\boldsymbol{\lambda}_j$ and $\boldsymbol{\Sigma}_g$ have been identified. Subtracting it,

$$\varsigma^2_{\varepsilon j} = V_{jg} - \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j.$$

And what remains of the wage variance, once the part explained by skill is taken out, is the transitory noise.

Combining the arguments above identifies

$$\left\{ (p_{jt}, \boldsymbol{\lambda}_j, \boldsymbol{\delta}_j, \varsigma^2_{\varepsilon j})_{j \in \mathcal{J}, t},\ (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g \in \mathcal{G}} \right\}.$$

This proves Theorem 1. □

Specialization to two skills. In the empirical model there are two task-specific skills, cognitive and manual. Let a_C denote the cognitive anchor occupation and a_M the manual anchor occupation. The normalization is

$$\boldsymbol{\lambda}_{a_C} = (1, 0)', \qquad \boldsymbol{\lambda}_{a_M} = (0, 1)'.$$

The covariance matrix of latent skills for group g is identified from distinct-date covariances of adjusted wages in the two anchor occupations:

$$\boldsymbol{\Sigma}_g = \begin{pmatrix} C_{a_C a_C, g} & C_{a_C a_M, g} \\ C_{a_M a_C, g} & C_{a_M a_M, g} \end{pmatrix}.$$

For any other occupation j, its cognitive and manual loadings are recovered from its covariances with the two anchors:

$$\boldsymbol{\lambda}_j' = \left(C_{j a_C, g}, C_{j a_M, g} \right) \boldsymbol{\Sigma}_g^{-1}.$$

Thus, in the two-skill version of the model, the cognitive and manual task returns of every occupation are identified from how adjusted wages in that occupation covary with adjusted wages in the cognitive and manual anchor occupations.

A.2 Identification of the Dynamic Choice Parameters

Once the parameters in the wage equation are identified, the remaining parameters to be identified are the occupation specific amenities $\{\eta_{jg}\}$ and the switching cost κ_0. Let x denote the state in equation (23), and let g and $d = d_{t-1}$ denote the group and origin occupation contained in x. Write

$$\tilde{w}_j(x) \equiv \mathbb{E}[\log W_{ijt} \mid x].$$

Now we define the following "assignment" operator Q_j for a function f. This operator computes the value of the function f given the expected state transition conditional on assignment to occupation j at time t, that is

$$(Q_j f)(x) \equiv \int f(x')\, \mathcal{P}(dx' \mid x, d_t = j).$$

Then, we denote with Q_j^s the same operator when the assignment to j is repeated for s consecutive periods. Therefore $Q_j^s f(x)$ propagates the state forward for s periods under repeated choice of occupation j. Notice that this choice is defined exogeneously, that is we are forcing $d_{im} = j$ for any m from period t to period s. The conditional value function can be written using this operator as

$$v_j(x) = \tilde{w}_j(x) + \eta_{jg} - \kappa_0 \mathbf{1}_{\{j \neq d\}} + \beta (Q_j W)(x).$$

The conditional choice probabilities are

$$P_j(x) = \frac{\exp(v_j(x)/\varrho)}{\sum_{m \in \mathcal{J}} \exp(v_m(x)/\varrho)}.$$

With a Type-I extreme-value assumption and the re-centering of the shocks, the integrated value function is

$$W(x) = \varrho \log \sum_{m \in \mathcal{J}} \exp\left(\frac{v_m(x)}{\varrho}\right).$$

Define the choice-specific value correction

$$\psi_j(x) \equiv W(x) - v_j(x).$$

This term measures the value lost if the worker is required to commit to j in the current period, before observing the taste shocks, rather than being free to choose optimally, and future choices remain optimal. This is the probability correction term of Section 5.2. Below, we use the Hotz and Miller (1993) inversion argument to show that under the multinomial logit assumption we can recover these terms directly from the choice probabilities, that is $\psi_j(x) = -\varrho \log P_j(x)$. The correction $\psi_j = -\varrho \log P_j$ can be understood as a penalty for using j as the reference action. Equivalently, it is the gap between the ex-ante value of behaving optimally and the value of committing to j: it measures how much value is lost by committing to j rather than allowing the worker to choose optimally. From the CCP representation this term has an intuitive revealed-preference argument: it is small when j is chosen often, and so it is optimal, and large when j is rarely chosen, and is not optimal.

Assumption 6 (Stationarity and known primitives) *The data are generated by the stationary equilibrium of Section 4.6. The wage equation and wage-signal distribution are identified, so $\tilde{w}_j$ and Q_j are known for every occupation j. The discount factor $\beta \in (0,1)$ and taste-shock scale $\varrho > 0$ are fixed. The conditions of Lemma 2 hold, so the Bellman equation has a unique integrated value function W.*

Assumption 7 (Support along the controlled path and limit-conditions) *For every supported state x and every occupation j, the wage equation, CCPs, and transition law are identified at the states reached under $Q_j^s(x,\cdot)$, for every $s \geq 1$, and $P_j > 0$ on those states. Moreover,*

$$\sum_{s=1}^{\infty} \beta^{s-1} Q_j^s \left[|\tilde{w}_j| + \varrho |\log P_j| \right](x) < \infty, \qquad \lim_{S\to\infty} \beta^S Q_j^{S+1} |W|(x) = 0.$$

I write $x_{dg} = (\cdot, d, g)$ for a state in which the worker last period choice was occupation d and belongs to group g, with $\cdot$ collecting the remaining state variables (mean and variance-covariance of the beliefs distribution, and task-specific tenure).

Assumption 8 (Comparison support) *For every group g and occupation $j \neq 0$, there is a supported state x_{dg} with $d \notin \{j, 0\}$. In addition, there is a supported state x_{dg} whose occupation d admits a destination $j \neq d$. Thus observing CCPs from a similar state permit a comparison between two moves from a common origin for each (j,g), and at least one comparison between moving to any occupation and staying in the current occupation.*

The identification argument has two steps. First, the Hotz–Miller inversion and the known transition law recover continuation values from future wages and observed choice probabilities. Using the identified wage parameters and the observed wages, CCPs, and the known transition law recover the continuation value rather than leaving it as an additional free parameter. Second, comparing two moves from a common origin identifies amenities, after which comparing a move with a stay identifies the switching cost.

I now show how using a standard inversion argument à la Hotz and Miller (1993) allow us to recover continuation value.

Lemma 1 (CCP inversion along the j-assignment path) *Under Assumptions 6–7, for every supported state x and occupation j,*

$$\psi_j(x) \equiv W(x) - \nu_j(x) = -\varrho \log P_j(x).$$

Moreover,

$$\nu_j(x) = \underbrace{\tilde{w}_j(x) + \beta \Psi_j(x)}_{\text{known}} + \frac{\eta_{jg}}{1-\beta} - \kappa_0 \mathbf{1}_{\{j \neq d\}},$$

where

$$\Psi_j(x) \equiv \sum_{s=1}^{\infty} \beta^{s-1} \left[Q_j^s \left(\tilde{w}_j + \psi_j \right) \right](x).$$

Because $\psi_j = -\varrho \log P_j$, the term $\Psi_j(x)$ depends only on identified wages and observed choice probabilities, and is therefore known.

Proof. Taking logs of the logit CCP gives

$$\varrho \log P_j(x) = \nu_j(x) - \varrho \log \sum_{m \in \mathcal{J}} \exp\left(\frac{\nu_m(x)}{\varrho}\right) = \nu_j(x) - W(x),$$

so that $\psi_j(x) = W(x) - \nu_j(x) = -\varrho \log P_j(x)$. This identity holds for every occupation j, whether or not j is the worker's optimal or realized choice: a low choice probability means that choosing j has a value far below $W(x)$, hence a large correction. Now fix a state x and an occupation j. Repeated assignment to j generates a path along which no switching cost is ever paid, since after the initial choice the worker starts and stays in the same occupation. At every state y reached along this controlled path, substitute $W(y) = \nu_j(y) + \psi_j(y)$ into the conditional value index to obtain a recursion for the ex-ante value function through the assignment operator,

$$W(y) = \tilde{w}_j(y) + \eta_{jg} + \psi_j(y) + \beta(Q_j W)(y).$$

Applying the assignment operator Q_j at the initial state and iterating this identity S times yields

$$\begin{aligned}(Q_j W)(x) = \sum_{s=1}^{S} \beta^{s-1} \left[Q_j^s\left(\tilde{w}_j + \psi_j\right)\right](x) \\ + \frac{1-\beta^S}{1-\beta}\eta_{jg} + \beta^S (Q_j^{S+1} W)(x).\end{aligned}$$

When $S \to \infty$, Assumption 7 gives

$$(Q_j W)(x) = \Psi_j(x) + \frac{\eta_{jg}}{1-\beta}.$$

Substituting this expression into the choice-specific value index and using $1 + \beta/(1-\beta) = 1/(1-\beta)$ gives

$$\nu_j(x) = \tilde{w}_j(x) + \beta \Psi_j(x) + \frac{\eta_{jg}}{1-\beta} - \kappa_0 \mathbf{1}_{\{j \neq d\}}.$$

□

The repeated-j path applies the correction term at every future state. The path is convenient because, after the initial choice, every subsequent assignment to j is a stay, so the switching cost is paid only on the initial move, whereas the amenity is received every period and enters with loading $1/(1-\beta)$.

This is a special case of the forward CCP representation in Arcidiacono and Miller (2011): the general representation may average over several future reference actions, whereas here all the reference weight is placed on j. Consequently Q_j^s weights future states through their transition probabilities, while the choice probabilities enter only through the correction $\psi_j = -\varrho \log P_j$.

Theorem 2 (Identification of the dynamic choice parameters) *Let Assumptions 6–8 hold and normalize $\eta_{0g} = 0$ for every g. Then*

$$\left(\{\eta_{jg}\}_{j \neq 0,\, g \in \mathcal{G}}, \kappa_0\right)$$

is identified.

Proof. By Lemma 1, the index $\tilde{w}_j(x) + \beta\Psi_j(x)$ is known, and the lemma implies

$$\varrho \log \frac{P_j(x)}{P_m(x)} - \left[\left(\tilde{w}_j(x) + \beta\Psi_j(x)\right) - \left(\tilde{w}_m(x) + \beta\Psi_m(x)\right)\right] = \frac{\eta_{jg} - \eta_{mg}}{1-\beta} - \kappa_0 \left(\mathbf{1}_{\{j\neq d\}} - \mathbf{1}_{\{m\neq d\}}\right). \tag{46}$$

Amenities. Fix g and $j \neq 0$, and evaluate equation (46) at a supported state x_{dg} with $d \notin \{j, 0\}$, as guaranteed by Assumption 8. Because both j and 0 are then moves, their switching costs cancel. Comparing occupation j with the baseline normalized occupation where $\eta_{0g} = 0$ yields

$$\eta_{jg} = (1-\beta)\left\{\varrho \log \frac{P_j(x_{dg})}{P_0(x_{dg})} - \left[\left(\tilde{w}_j(x_{dg}) + \beta\Psi_j(x_{dg})\right) - \left(\tilde{w}_0(x_{dg}) + \beta\Psi_0(x_{dg})\right)\right]\right\}.$$

Thus common-origin comparisons between two moves identify every amenity relative to the base occupation.

Switching cost. At a supported state x_{dg} whose occupation d admits a destination $j \neq d$, compare moving to j with staying in d. Equation (46) gives

$$\kappa_0 = \frac{\eta_{jg} - \eta_{dg}}{1-\beta} - \varrho \log \frac{P_j(x_{dg})}{P_d(x_{dg})} + \left[\left(\tilde{w}_j(x_{dg}) + \beta\Psi_j(x_{dg})\right) - \left(\tilde{w}_d(x_{dg}) + \beta\Psi_d(x_{dg})\right)\right].$$

All terms on the right-hand side are now known, so κ_0 is identified. Intuitively, wages, amenities, or continuation values all imply some switching behavior across occupations and if workers do not move enough then this implies that there must be some switching costs preventing workers from moving. □

Remark. A just-identifying subset contains

$$|\mathcal{G}|(|\mathcal{J}| - 1) + 1$$

log-odds equations: one move-versus-move comparison for each (j, g) with $j \neq 0$, and one move-versus-stay comparison for the common switching cost. In the application this is $3 \times 7 + 1 = 22$. Other supported states and destination comparisons provide overidentifying restrictions and are used by the NFXP likelihood.

The simple comparison argument relies on the switching cost being common across destinations and paid only when the worker moves. No additional exclusion-style variation in task experience or move direction is needed beyond the support required to identify the CCPs and controlled paths. If switching costs varied with destination, direction, or experience, they would no longer cancel in the first step and would have to be identified jointly with amenities. Finally, amenities are flow payoffs and therefore enter with loading $1/(1-\beta)$, whereas κ_0 is a one-time cost with loading one.

A.3 Bellman Fixed Point

This subsection shows that the worker's dynamic problem has a unique solution. The argument is useful because the identification proof for the dynamic choice parameters uses the fact that, given amenities and switching costs, continuation values are pinned down by the model.

To keep notation compact, write x_t for the full state

$$x_t = \left(g, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}, \{p_{j\ell}\}_{j\in\mathcal{J},\ell\geq t}\right).$$

This is the same state as in equation (23). The price path is part of the state, so the future evolution of prices is known from the worker's point of view.

Define the deterministic part of the value of choosing occupation j as

$$\begin{aligned} v_{jt}(x_t; V) = \mathbb{E}\left[\log W_{ijt} \mid x_t\right] + \eta_{jg} - \mathcal{C}(j, d_{t-1}) \\ + \beta \mathbb{E}[V(x_{t+1}) \mid x_t, d_t = j]. \end{aligned}$$

The expectation is taken over the wage signal that updates beliefs. The current taste shocks are integrated out.

The Bellman operator is

$$(\mathcal{T}V)(x_t) = \varrho \log\left[\sum_{j\in\mathcal{J}} \exp\left(\frac{v_{jt}(x_t; V)}{\varrho}\right)\right].$$

The additive Euler-constant term from the Type-I extreme value distribution is omitted because it is common to all choices and does not affect the CCPs. Adding it back would only add a constant to the value function and would not change the argument.

Lemma 2 (Bellman contraction) *Suppose $\beta \in (0,1)$, $\varrho > 0$, expected log wages and switching costs are bounded on the state space used in estimation, and $\mathcal{P}_t(x_{t+1} \mid x_t, d_t = j)$ is a valid transition probability for each occupation j. Then the Bellman operator $\mathcal{T}$ has a unique fixed point. Hence the integrated value function V is uniquely determined.*

Proof.

Take two possible value functions, V and $\widetilde{V}$. We want to show that applying the Bellman operator brings them closer together.

For a fixed state x_t and occupation j, the only difference between $v_{jt}(x_t; V)$ and $v_{jt}(x_t; \widetilde{V})$ is the continuation value:

$$v_{jt}(x_t; V) - v_{jt}(x_t; \widetilde{V}) = \beta \mathbb{E}\left[V(x_{t+1}) - \widetilde{V}(x_{t+1}) \mid x_t, d_t = j\right].$$

The absolute value of this difference is bounded by

$$\left|v_{jt}(x_t; V) - v_{jt}(x_t; \widetilde{V})\right| \leq \beta \|V - \widetilde{V}\|_\infty.$$

This holds for every occupation j.

The log-sum expression in the Bellman operator cannot magnify differences by more than the

largest difference across occupations. Therefore,

$$\left|(\mathcal{T}V)(x_t) - (\mathcal{T}\widetilde{V})(x_t)\right| \leq \beta\|V - \widetilde{V}\|_\infty.$$

Taking the largest difference across all states gives

$$\|\mathcal{T}V - \mathcal{T}\widetilde{V}\|_\infty \leq \beta\|V - \widetilde{V}\|_\infty.$$

Since $\beta < 1$, repeated application of the Bellman operator brings any two starting value functions closer together. Therefore the Bellman equation has one and only one fixed point. This fixed point is the integrated value function used in the occupational choice problem. □

B Appendix: Estimation Results

B.1 Externally Set and Calibrated Parameters

Table 1: Externally Set Parameters

Parameter	Description	Empirical Moment	Value
Preferences & Demographics			
β	Discount factor		0.900
δ	Per-period exit hazard	Avg. career length (est. sample)	0.080
Education Distribution			
Low education (L)	Population share	Empirical shares	36.4%
Medium education (M)	Population share	Empirical shares	33.3%
High education (H)	Population share	Empirical shares	30.3%
Human Capital Accumulation			
$\boldsymbol{\omega}_j$	Task importance weights	Mihaylov and Tijdens (2019)	
CES Final-Good Aggregator			
σ	Elasticity of substitution	Althoff and Reichardt (2026)	1.57
α_j	CES aggregator weights	Stage-1 prices & outputs	Calibrated[a]
P	Aggregate price level	Final-good numeraire	$\equiv 1$

Notes: Most parameters are set (rather than calibrated) to match aggregate statistics or for computational tractability. The per-period exit hazard δ is exogenous and not internalized by workers; it only generates the stationary age-experience distribution, and is chosen to match the average career length in the estimation sample. [a]The CES aggregator weights α_j are calibrated in closed form as $\alpha_j = p_j Y_j^{1/\sigma} / (\sum_k p_k Y_k)^{1/\sigma}$ (equivalently $\alpha_j^\sigma = s_j p_j^{\sigma-1}$, with wage-bill share s_j), where p_j is the Stage-1 occupation price and Y_j the occupation output (wage bill deflated by p_j). This makes the dual CES price index $P = (\sum_j \alpha_j^\sigma p_j^{1-\sigma})^{1/(1-\sigma)} = 1$ at the Stage-1 prices. The simplex constraint $\sum_j \alpha_j = 1$ is not imposed; the scale of $\{\alpha_j\}$ is the final-good numeraire, so the estimated price vector is itself the stationary equilibrium.

The discount factor $\beta = 0.9$ is a standard value in the dynamic discrete choice literature. The per-period exit hazard $\delta = 0.08$ (death or retirement) is chosen to match the average career length in the estimation sample; it is exogenous, not internalized by workers, and only serves to generate the stationary age-experience distribution (Section 4.6). Education-gender shares are set to match the estimation sample: low-educated workers comprise 36.4% (24.4% male, 12.0% female), medium-educated 33.3% (18.5% male, 14.7% female), and high-educated 30.3%

(13.0% male, 17.4% female). The task importance weights ω_j are from Mihaylov and Tijdens (2019); ω_{jk} governs the time a worker in occupation j spends on task k, which in turn determines how task-specific experience accumulates. The CES elasticity of substitution is $\sigma = 1.57$, and the aggregator weights are calibrated in closed form to $\alpha_j = p_j Y_j^{1/\sigma} / (\sum_k p_k Y_k)^{1/\sigma}$ (equivalently $\alpha_j^\sigma = s_j p_j^{\sigma-1}$ with wage-bill share s_j), where Y_j is occupation output (the wage bill deflated by p_j). This makes the dual price index $P = (\sum_j \alpha_j^\sigma p_j^{1-\sigma})^{1/(1-\sigma)}$ equal to one at the Stage-1 prices; the simplex constraint $\sum_j \alpha_j = 1$ is not imposed, and the scale of $\{\alpha_j\}$ is the final-good numeraire.

B.2 Stage 1 Estimation Results: Wage Technology and Skill Distributions

Table 2 reports the full estimation results. Panel A shows the skill loadings λ_{jk}, the idiosyncratic wage variances ς_j^2, and the log occupation prices p_j. For identification, Occ_0 is the cognitive anchor ($\lambda_0 = (1, 0)$) and Occ_1 the manual anchor ($\lambda_1 = (0, 1)$). Occupations differ sharply in their loadings: some load mainly on cognitive skills (Occ_0, Occ_3, Occ_6), others mainly on manual skills (Occ_1, Occ_2, Occ_5), and Occ_4 and Occ_7 load comparably on both. Returns to task-specific experience are estimated as a flexible step function of tenure with annual knots rather than a parametric polynomial; the estimated profiles are reported in Figure 7 and are aligned with the skill loadings, in that occupations that value a given skill dimension also reward experience in it. Panel B reports the task importance weights ω_{jk} from Mihaylov and Tijdens (2019) under the $K = 8$ cluster partition, which govern time allocation within occupations and the accumulation of task-specific experience. Panel C shows the estimated skill distributions by education group (genders pooled), normalized so that the sample-share-weighted mean of skills across groups is zero, $\sum_g \pi_g \boldsymbol{\mu}_g = \mathbf{0}$. Mean skills rise with education and the cognitive comparative advantage widens.

B.3 Stage 2 Estimation Results: Dynamic Choice Parameters

C Appendix: Model

This appendix provides additional derivations and describes the computational algorithms used to solve the model.

C.1 Firm's Time Allocation Problem

This section derives the optimal allocation of worker time across tasks within an occupation. Following Althoff and Reichardt (2026), the firm controls the worker's time allocation statically, choosing $\{\ell_{jk}\}_{k=1}^K$ to maximize expected log output conditional on the symmetric posterior $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ shared with the worker, where the expectation is taken under the learning block's posterior $\boldsymbol{\xi}_i \mid \mathcal{I}_{it} \sim \mathcal{N}(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$.

Table 2: Stage 1: Expectation-Maximization Estimates of Wage Equation Parameters

Panel A: Loadings, Shock Variances, and Occupational Prices

Occupation	$\lambda_{j,C}$	$\lambda_{j,M}$	ς_j^2	p_j
Occ_0	1.000	0.000	0.0235	1.538
Occ_1	0.000	1.000	0.0200	1.443
Occ_2	0.202	0.826	0.0197	1.442
Occ_3	0.875	0.000	0.0271	1.520
Occ_4	0.527	0.567	0.0210	1.481
Occ_5	0.350	0.737	0.0211	1.496
Occ_6	0.827	0.375	0.0264	1.522
Occ_7	0.541	0.491	0.0218	1.455

Panel B: Task Importance Weights by Occupation

Occupation	$\omega_{j,C}$ (Cognitive)	$\omega_{j,M}$ (Manual)
Occ_0	0.988	0.012
Occ_1	0.009	0.991
Occ_2	0.162	0.838
Occ_3	0.327	0.673
Occ_4	0.448	0.552
Occ_5	0.692	0.308
Occ_6	0.946	0.054
Occ_7	0.082	0.918

Panel C: Skills Distribution by Education Group

	Low (L)	Medium (M)	High (H)
$\boldsymbol{\mu_g} = \begin{pmatrix} \mu_C \\ \mu_M \end{pmatrix}$	$\begin{pmatrix} -0.183 \\ -0.136 \end{pmatrix}$	$\begin{pmatrix} -0.075 \\ -0.030 \end{pmatrix}$	$\begin{pmatrix} 0.302 \\ 0.196 \end{pmatrix}$
$\boldsymbol{\Sigma}_g = \begin{pmatrix} \sigma_C^2 & \rho\sigma_C\sigma_M \\ \rho\sigma_M\sigma_C & \sigma_M^2 \end{pmatrix}$	$\begin{pmatrix} 0.082 & 0.007 \\ 0.007 & 0.069 \end{pmatrix}$	$\begin{pmatrix} 0.109 & 0.012 \\ 0.012 & 0.112 \end{pmatrix}$	$\begin{pmatrix} 0.097 & 0.008 \\ 0.008 & 0.227 \end{pmatrix}$
Implied correlation ρ_{CM}	0.099	0.109	0.055

Panel A reports the skill loadings λ_{jk}, the idiosyncratic wage–shock variance ς_j^2, and the log occupation prices p_j (the occupation intercept in the wage equation). By normalization *Occ_0* is the cognitive anchor ($\lambda_{0,C} = 1$, $\lambda_{0,M} = 0$) and *Occ_1* the manual anchor ($\lambda_{1,C} = 0$, $\lambda_{1,M} = 1$). Returns to task-specific experience are estimated as a flexible step function of tenure with annual knots; the estimated profiles are shown in Figure 7. Panel B reports task importance weights ω_{jk} from the Mihaylov and Tijdens (2019) classification under the $K = 8$ cluster partition; these govern task aggregation and time allocation within occupations. Panel C reports EM estimates of prior means $\boldsymbol{\mu}_g$ and covariance matrices $\boldsymbol{\Sigma}_g$ of cognitive (ξ_C) and manual (ξ_M) latent skills by education group $g \in \{L, M, H\}$ (genders pooled), with implied correlation ρ_{CM} below. Skill means are identified up to location and anchored by the sample-share-weighted normalization $\sum_g \pi_g \boldsymbol{\mu}_g = \mathbf{0}$, with π_g the estimation-sample group shares.

Table 3: Stage 2: NFXP Estimates of Dynamic Choice Parameters

Switching cost and taste-shock scale	
κ_0 (switching cost)	1.103
ϱ (matched)	0.271

Occupation–education amenities η_{jg}

Occupation	Low (L)	Medium (M)	High (H)
Occ_0 (base)	0.000	0.000	0.000
Occ_1	0.169	0.088	−0.219
Occ_2	0.188	0.119	−0.034
Occ_3	0.176	0.123	0.116
Occ_4	0.242	0.211	0.198
Occ_5	0.191	0.143	0.133
Occ_6	0.087	0.047	0.017
Occ_7	0.204	0.134	0.061

Notes: The utility switching cost is a single constant κ_0, paid whenever the worker changes occupation and independent of the origin, the destination, and accumulated experience. Amenities η_{jg} are occupation- and education-specific, normalized so that the baseline occupation Occ_0 has $\eta = 0$ within each education group (21 free amenities). The taste-shock scale ϱ is not separately identified from continuation values without exogenous payoff shifters and is fixed to match the average occupational mobility rate in the large sample.

Task-level output of worker i in occupation j at time t is

$$y_{jk} = \ell_{jk}\, \theta_{ik}^{a_{jk}}\, H_{jk}(\tau_{ikt}), \qquad H_{jk}(\tau) \;\equiv\; \exp\!\big(b_{jk}\,\tau + c_{jk}\,\tau^2\big), \tag{47}$$

where $\ell_{jk} \in [0,1]$ is the fraction of unit time allocated to task k, subject to $\sum_{k=1}^{K} \ell_{jk} = 1$; a_{jk} is the primitive task-level elasticity of output with respect to latent skill; and $H_{jk}(\cdot)$ is the multiplicative task-level human-capital function, with primitive coefficients (b_{jk}, c_{jk}) inside the exponent. Tasks aggregate according to the normalized Cobb-Douglas technology with weights $\omega_{jk} > 0$ satisfying $\sum_{k=1}^{K} \omega_{jk} = 1$:

$$Y_{ijt} = \prod_{k=1}^{K} \left(\frac{y_{jk}}{\omega_{jk}} \right)^{\omega_{jk}}. \tag{48}$$

Taking logs and using $\xi_{ik} \equiv \log \theta_{ik}$,

$$\log Y_{ijt} = \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} + \sum_{k=1}^{K} \omega_{jk}\, a_{jk}\, \xi_{ik} + \sum_{k=1}^{K} \omega_{jk} \big[b_{jk}\, \tau_{ikt} + c_{jk}\, \tau_{ikt}^2 \big] - \sum_{k=1}^{K} \omega_{jk} \log \omega_{jk}. \tag{49}$$

Log output is linear in $\boldsymbol{\xi}_i$, so taking the expectation under the symmetric posterior replaces ξ_{ik} with $\tilde{\xi}_{ikt}$:

$$\mathbb{E}\big[\log Y_{ijt} \,\big|\, \tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}\big] = \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} + \sum_{k=1}^{K} \omega_{jk}\, a_{jk}\, \tilde{\xi}_{ikt} + \sum_{k=1}^{K} \omega_{jk} \big[b_{jk}\, \tau_{ikt} + c_{jk}\, \tau_{ikt}^2 \big] - \sum_{k=1}^{K} \omega_{jk} \log \omega_{jk}. \tag{50}$$

Only the first term depends on $\{\ell_{jk}\}$. The firm's problem therefore reduces to

$$\max_{\{\ell_{jk}\}} \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} \quad \text{s.t.} \quad \sum_{k=1}^{K} \ell_{jk} = 1. \tag{51}$$

Forming the Lagrangian $\mathcal{L} = \sum_k \omega_{jk} \log \ell_{jk} - \mu\left(\sum_k \ell_{jk} - 1\right)$, the first-order conditions are

$$\frac{\omega_{jk}}{\ell_{jk}} = \mu \quad \Rightarrow \quad \ell_{jk} = \frac{\omega_{jk}}{\mu}. \tag{52}$$

Summing over k and using $\sum_k \ell_{jk} = \sum_k \omega_{jk} = 1$ yields $\mu = 1$, so the optimal time allocation is

$$\ell^*_{jk} = \omega_{jk}. \tag{53}$$

Key property: the optimal time allocation is *independent of worker skills, experience, and beliefs about skills*, and depends only on the task weights ω_{jk}. The assumption of Cobb-Douglas aggregation of tasks, with a unitary elasticity of substitution, makes the beliefs in expected log output additively separable from $\{\ell_{jk}\}$, so beliefs do not play any role in the task assignment. Generalizing the production function to a CES aggregator in tasks, or a dynamic task assignment chosen by the firm break this separability and make the allocation belief-dependent but are beyond the scope of this paper.

C.2 Stationary Equilibrium Algorithm

A stationary equilibrium is a constant vector of occupation log prices $\tilde{\boldsymbol{p}} = (\log P_j)_{j \in \mathcal{J}}$ at which every occupation's goods market clears. The CES aggregator weights $\{\alpha_j\}$ are held fixed at their pre-shock calibration and any candidate price vector is renormalized so that the aggregate price index equals one, $P(\tilde{\boldsymbol{p}}) = \left(\sum_j \alpha_j^\sigma P_j^{1-\sigma}\right)^{1/(1-\sigma)} = 1$ (Section 4.6). I compute the stationary equilibrium of the model by iterating on the time-independent price vector: at each iteration, I solves the worker's dynamic problem given the current prices, simulates a synthetic panel, and compute the update for the price vector using the labor demand equation. When simulating, workers are replaced at rate δ[28] and are replaced by entrants drawn from the group priors. After discarding an initial burn-in, this death-and-rebirth process makes initial conditions irrelevant and the simulated panel approximates the ergodic cross-section $\pi^\star(\tilde{\boldsymbol{p}^*})$ resulting from the equilibrium price vector $\boldsymbol{p}^*$.

In the simulation, occupation specific output are aggregate using these shares as weights, which are basically CCPs when there is a large number of workers. To reduce the noise induce by numerical simulations, I do not assign each worker to a unique occupation as this would generate discrete jumps for small moves in the price vector. Rather, I let her contribute to every occupation in proportion to her conditional choice probability ($\mathbb{P}_{ij}(\tilde{\boldsymbol{p}})$) that are given by

$$\begin{aligned} \mathbb{P}_{ij}(\tilde{\boldsymbol{p}}) &= \frac{\exp\left(\nu_{ij}(\tilde{\boldsymbol{p}})/\varrho\right)}{\sum_{k \in \mathcal{J}} \exp\left(\nu_{ik}(\tilde{\boldsymbol{p}})/\varrho\right)}, \\ \nu_{ij}(\tilde{\boldsymbol{p}}) &= p_j + \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_i + h_j(\boldsymbol{\tau}_i) + \eta_{j,g(i)} - \mathcal{C}\left(j, d_{i,-1}, e_i\right) + \beta\, \mathbb{E}\left[V(\boldsymbol{x}') \mid \boldsymbol{x}_i, j\right]. \end{aligned} \tag{54}$$

Output, by contrast, is produced at the worker's true skills $\boldsymbol{\xi}_i$: her expected wage in occupation j, in levels and with the log-normal correction for noise $\mathbb{E}[e^\varepsilon] = e^{\varsigma^2_{\varepsilon j}/2}$, is $\overline{W}_{ij}(\tilde{\boldsymbol{p}}) = \exp\left(p_j + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_i) + \frac{1}{2}\varsigma^2_{\varepsilon j}\right)$. Intuitively, occupation output is the probability-weighted sum of these

[28] This is not internalized by workers when choosing dynamically, therefore the dynamic choice problem remains effectively infinite horizong.

wages deflated by the price in that occupation,

$$Y_j(\tilde{\boldsymbol{p}}) = \frac{1}{P_j} \sum_i \mathbb{P}_{ij}(\tilde{\boldsymbol{p}})\, \overline{W}_{ij}(\tilde{\boldsymbol{p}}) = \sum_i \mathbb{P}_{ij}(\tilde{\boldsymbol{p}})\, \exp\Big(\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_i) + \tfrac{1}{2}\varsigma_{\varepsilon j}^2\Big), \tag{55}$$

with $P_j = e^{p_j}$, so the price cancels and output is properly measured in efficiency units. Using probabilities for aggregation, rather then realized discrete choices, smooth how output responds to prices and sharply reduces simulation noise.

After having computed these occupation-specific output and the overall aggregate output, CES inverse demand returns the updating price vector for each occupation. A damped step then updates the guess and the iteration repeats until the price vector stop moving. The procedure is summarized in the algorithm box below.

Algorithm: Stationary Equilibrium (price tâtonnement)

Inputs: the estimated wage and skill parameters $\widehat{\boldsymbol{\Gamma}} = \big\{(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g\in\mathcal{G}},\ (p_j, \boldsymbol{\lambda}_j, \boldsymbol{\delta}_j, \varsigma_{\varepsilon j}^2)_{j\in\mathcal{J}}\big\}$ and dynamic-choice parameters $\widehat{\boldsymbol{\Delta}} = \big\{(\eta_{j,g}), \kappa_0\big\}$; CES weights $\{\alpha_j\}$ and elasticity σ; discount β, exit probability δ, taste scale ϱ; damping ζ, tolerance *tol*.

Output: equilibrium log prices $\tilde{\boldsymbol{p}}^\star$, value coefficients $\boldsymbol{\Phi}^\star$, outputs $\{Y_j\}$, ergodic cross-section $\pi^\star$.

Initialize $\tilde{\boldsymbol{p}}^{(0)}$ at the Stage-1 prices, renormalized to $P(\tilde{\boldsymbol{p}}^{(0)}) = 1$, where $P(\boldsymbol{p}) = \big(\sum_j \alpha_j^\sigma p_j^{1-\sigma}\big)^{1/(1-\sigma)}$.

Repeat for $n = 0, 1, \ldots$ until $\max_{j\in\mathcal{J}} |\tilde{p}_j^{(n+1)} - \tilde{p}_j^{(n)}| < tol$:

- **Solve.** Compute $\boldsymbol{\Phi}^{(n)}$ as the fixed point of the integrated Bellman operator at prices $\tilde{\boldsymbol{p}}^{(n)}$ (discount β; δ not internalized).
- **Simulate.** Draw a worker panel with common random numbers; workers exit at rate δ and are replaced by entrants from the group priors; discard the burn-in to obtain the ergodic cross-section.
- **Aggregate.** For each worker i and occupation j, form the choice probability $\mathbb{P}_{ij}(\tilde{\boldsymbol{p}}^{(n)})$ of (54) and the output $Y_j = \sum_i \mathbb{P}_{ij}\, \exp\big(\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_i) + \tfrac{1}{2}\varsigma_{\varepsilon j}^2\big)$ as in (55).
- **Prices.** Set $\hat{p}_j = \alpha_j\, (Y/Y_j)^{1/\sigma}$, with $Y = \big(\sum_j \alpha_j Y_j^{(\sigma-1)/\sigma}\big)^{\sigma/(\sigma-1)}$, and renormalize $\hat{\boldsymbol{p}} \leftarrow \hat{\boldsymbol{p}} / P(\hat{\boldsymbol{p}})$.
- **Update.** $\tilde{\boldsymbol{p}}^{(n+1)} = (1-\zeta)\, \tilde{\boldsymbol{p}}^{(n)} + \zeta \log \hat{\boldsymbol{p}}$, with $\zeta \in (0, 1)$

Return $\tilde{\boldsymbol{p}}^\star$, $\boldsymbol{\Phi}^\star$, $\{Y_j\}$, $\pi^\star$.

C.3 Transition Path Algorithm

Now I describe how to compute the perfect-foresight transition path to an unanticipated technology shock at $t = 0$. To solve for the transition path, we need to compute the sequence of price vector determined by the sequence of competitive equilibria (non-stationary) of the economy along the transition to the new stationary price vector. Each of these competitive equilibria is computed under the same fixed weights $\{\alpha_j\}$ and the same per-period numeraire $P(\tilde{\boldsymbol{p}}_t) = 1$. The computation logic mirrors that of the stationary algorithm (solve, simulate, aggregate, update), but now applied to the whole price path by backward–forward shooting algorithm.

The technology shock considered in the experiment changes the task weights ω_j, with AI lowering the cognitive-task weight in proportion to each occupation's exposure; the wage parameters inherit the change through the effective mapping, $\lambda'_{jk} = \lambda_{jk}\,\omega'_{jk}/\omega_{jk}$, and the log human-capital bundle h_j inherits the new task weights $\boldsymbol{\omega}'_j$, while prices, noise variances, amenities, and switching costs are unchanged. Let Λ collect the parameters of this technology. We compute both an "instantaneous" technology adoption, where the production function technology mapping skills and experience to output jumps from Λ_{pre} to Λ_{post} at $t = 0$; and a more "gradual" technology adoption, where the change in the technology parameters are more gradual, that is $\Lambda_t = (1 - s_t)\Lambda_{\text{pre}} + s_t\Lambda_{\text{post}}$ with $s_t = \min(t/t_{\text{shock}}, 1)$. In the latter, workers have perfect foresight on the whole sequence of technology change.

I first compute the two fixed points that determine pre- and post-shock stationary competitive equilibria, that are those determining the equilibrium price vector and the dynamic program of the workers. For each of the two endpoints I obtain the tuple $(\tilde{\boldsymbol{p}}_{\text{pre}}, \boldsymbol{\Phi}_{\text{pre}})$ and $(\tilde{\boldsymbol{p}}_{\text{post}}, \boldsymbol{\Phi}_{\text{post}})$, where $\boldsymbol{\Phi}$ collects the Smolyak value-function coefficients that solve the Bellman equations at these two stationary points. The terminal value of the transition path for prices is pinned at $\tilde{\boldsymbol{p}}_{\text{post}}$. Given a guess for the price path, I solve for the optimal coefficients approximating the value functions by backward induction from the terminal condition $\boldsymbol{\Phi}_{T-1} = \boldsymbol{\Phi}_{\text{post}}$,

$$\boldsymbol{\Phi}_t = \Pi\Big[\max_{j\in\mathcal{J}}\Big\{r_j(\boldsymbol{x};\tilde{\boldsymbol{p}}_t,\Lambda_t) + \beta\,\mathbb{E}\big[V_{t+1}(\boldsymbol{x}') \mid \boldsymbol{x}, j\big]\Big\}\Big], \qquad t = T-2,\ldots,0, \tag{56}$$

where Π is the Chebyshev/Smolyak inversion[29], the flow value uses the period-t prices and technology, and the expectation integrates the Gaussian belief update by Gauss–Hermite quadrature. I then forward simulate a single panel where workers solve the time-varying dynamic program (so the value function is now not stationary), prices, and technology, and occupation outputs Y_{jt} are formed by exactly the price-dependent choice-probability aggregation described before. This is now applied period by period as we are solving for a sequence of competitive equilibria. Finally, I use the per-period CES inverse demand to get the updating price sequence, which I renormalize to the numeraire in each period. The path is iterated until the largest price change across iterations and occupations falls below a certain tolerance level. The procedure is summarized in the algorithm box below.

[29]The model is solved via collocation, hence I obtain the value function on the collocation nodes, and each node has a Chebyshev polynomial coefficient that can be directly obtained by inverting the function.

Algorithm: Transition Path (backward–forward shooting)

Inputs: pre- and post-shock technologies $\Lambda_{\text{pre}}, \Lambda_{\text{post}}$; CES weights $\{\alpha_j\}$ and elasticity σ; discount β, exit probability δ; horizon T, relaxation τ, damping ζ, tolerance *tol*.

Output: equilibrium price path $\{\tilde{\boldsymbol{p}}_t^\star\}_{t=0}^{T-1}$, value coefficients $\{\boldsymbol{\Phi}_t\}$, outputs $\{Y_{jt}\}$.

Endpoints. Solve the pre- and post-shock stationary equilibria (stationary algorithm above) for $(\tilde{\boldsymbol{p}}_{\text{pre}}, \boldsymbol{\Phi}_{\text{pre}})$ and $(\tilde{\boldsymbol{p}}_{\text{post}}, \boldsymbol{\Phi}_{\text{post}})$.

Initialize the path $\tilde{\boldsymbol{p}}_t^{(0)} = \tilde{\boldsymbol{p}}_{\text{post}} + (\tilde{\boldsymbol{p}}_{\text{pre}} - \tilde{\boldsymbol{p}}_{\text{post}})\, e^{-t/\tau}$, pin $\tilde{\boldsymbol{p}}_{T-1} = \tilde{\boldsymbol{p}}_{\text{post}}$, and renormalize each row to $P = 1$.

Repeat for $n = 0, 1, \ldots$ until $\max_{t<T-1,\, j\in\mathcal{J}} |\tilde{p}_{jt}^{(n+1)} - \tilde{p}_{jt}^{(n)}| < tol$:

- **Backward.** Set $\boldsymbol{\Phi}_{T-1} = \boldsymbol{\Phi}_{\text{post}}$; for $t = T-2, \ldots, 0$, $\boldsymbol{\Phi}_t = \Pi\big[\max_{j\in\mathcal{J}}\{r_j(\boldsymbol{x}; \tilde{\boldsymbol{p}}_t, \Lambda_t) + \beta\, \mathbb{E}[V_{t+1}(\boldsymbol{x}') \mid \boldsymbol{x}, j]\}\big]$.
- **Forward.** Simulate one panel with common random numbers: burn in under Λ_{pre}, then from $t = 0$ apply $\{\boldsymbol{\Phi}_t, \tilde{\boldsymbol{p}}_t, \Lambda_t\}$ and form Y_{jt} by the aggregation (54)–(55), period by period.
- **Prices.** Set $\hat{p}_{jt} = \alpha_j\, (Y_t / Y_{jt})^{1/\sigma}$, renormalize $\hat{\boldsymbol{p}}_t \leftarrow \hat{\boldsymbol{p}}_t / P(\hat{\boldsymbol{p}}_t)$, and pin the terminal date to $\tilde{\boldsymbol{p}}_{\text{post}}$.
- **Update.** On the non-terminal block, $\tilde{\boldsymbol{p}}_t^{(n+1)} = (1-\zeta)\, \tilde{\boldsymbol{p}}_t^{(n)} + \zeta \log \hat{\boldsymbol{p}}_t$ (optional Anderson acceleration).

Return $\{\tilde{\boldsymbol{p}}_t^\star\}$, $\{\boldsymbol{\Phi}_t\}$, $\{Y_{jt}\}$.

C.4 Taste Shocks and the Entry Occupational Choice

This appendix illustrates the role of the idiosyncratic taste shocks ϵ_{ijt} in determining occupational choice at entry in the labor market. At the beginning of the career, every worker in education group g shares the same prior beliefs $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ and has no experience, so the systematic part of the choice value is identical within the group. In the deterministic limit $\varrho \to 0$ the entry choice is therefore degenerate: every worker of a specific group would enter the same occupation and workers paths would diverge only later, through the wage signals they happen to draw and the experience they accumulate. Panel (a) of Figure 20 maps this deterministic assignment as a function of entry beliefs $(\tilde{\xi}_C, \tilde{\xi}_M)$; the white marker locates each group's prior mean $\boldsymbol{\mu}_g$.

The taste shocks break this degenerate choice at labor market entry. With taste shocks the entry choice noisy around the occupation that would have been chosen in the degenerate case. Workers from the same group, that have identical beliefs at entry, spread across neighboring occupations. Panel (b) of Figure 20 plots the expected cognitive task intensity $\mathbb{E}[\omega_C]$ of the entry occupation for a simulate panel of workers. The dispersion in the occupational choice still depends on the initial prior: Panel (c) reports the probability of the most likely entry occupation, $\max_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\xi}})$, which falls along the indifference lines where a worker is nearly indifferent between occupations, and rises toward one where comparative advantage is strong. For example with the estimated priors, low-educated workers would always enter in occupation 1, that has a very high manual content. However, with taste shocks, they may end up in any occupation but are more likely to end up in either occupation 3 with $\omega_M = 0.67$ or oc-

cupation 1 with $\omega_M = 0.99$. Over the life cycle taste shocks also allow the model to rationalize occupational choices and moves that its systematic component alone would not predict.

C.5 Switching Costs: Lifecycle Wage Profiles

This appendix reports the lifecycle wage profiles behind the switching-cost counterfactuals of Section 7.3, which the main text summarizes by match-quality percentile. Figure 21 shows, for each switching-cost level κ/κ_0, the average wage gain over the life cycle: Panel (a) is the gain from moving to perfect information at each switching cost (the lifecycle counterpart of Figure 15), and Panel (b) is the gain relative to the estimated baseline economy (the lifecycle counterpart of Figure 16). In Panel (b), the wage gain of reducing switching costs gradually become more negative over the lifecycle: due to the presence of idiosyncratic taste shocks, reducing switching costs churning workers keep restarting their task-specific experience, and these losses compound over the career; when the reduction is extreme, wage losses materialize almost immediately.

C.6 Forced Exploration at the Beginning of the Career

The switching-cost counterfactual shows that frictions to mobility are a key driver of persistent mismatch, but even with low switching costs the baseline information structure limits how quickly workers can sort optimally. A complementary policy question is whether a targeted intervention at the very start of the career can accelerate learning and improve lifetime sorting and wages. I study a one-period "forced exploration" policy: in their first period, workers are assigned to the occupation that maximizes the information gain about their latent skills. I measure the learning value of occupation j for entrants in group g as the mutual information between skills and the first wage signal,

$$IG_{jg}^{\text{entry}} \equiv I(\xi_i; s_{ij1} \mid g) = \frac{1}{2}\log\left(1 + \frac{\boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j}{\varsigma_{\varepsilon j}^2}\right), \tag{57}$$

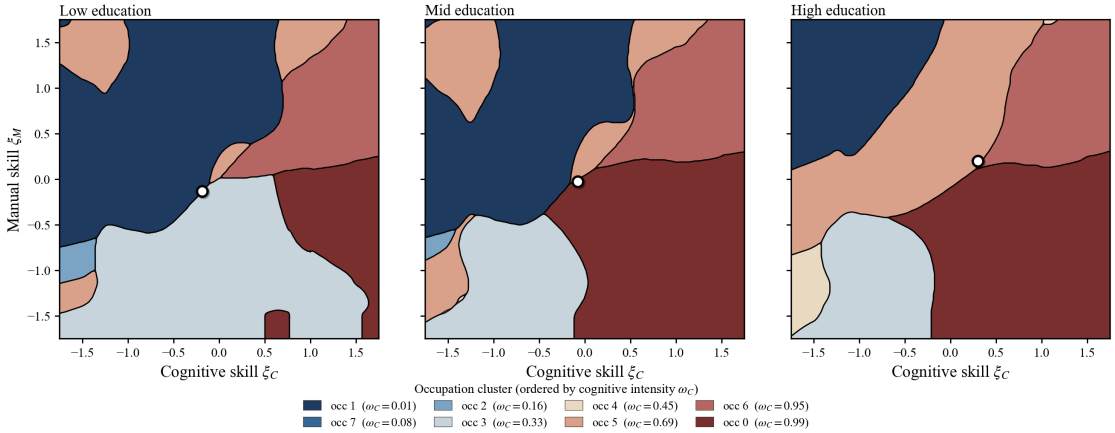
where $\boldsymbol{\lambda}_j$ are the skill loadings of occupation j, $\boldsymbol{\Sigma}_g$ is the prior skill covariance of group g, and $\varsigma_{\varepsilon j}^2$ is the occupation-specific wage noise variance. This has a natural signal-to-noise interpretation: an occupation is more informative when it loads strongly on dimensions about which entrants are ex ante uncertain and when wages are a precise signal of skills[30]. Each worker is assigned

[30]Mutual information is the reduction in uncertainty in the signal due to knowledge of latent skills,

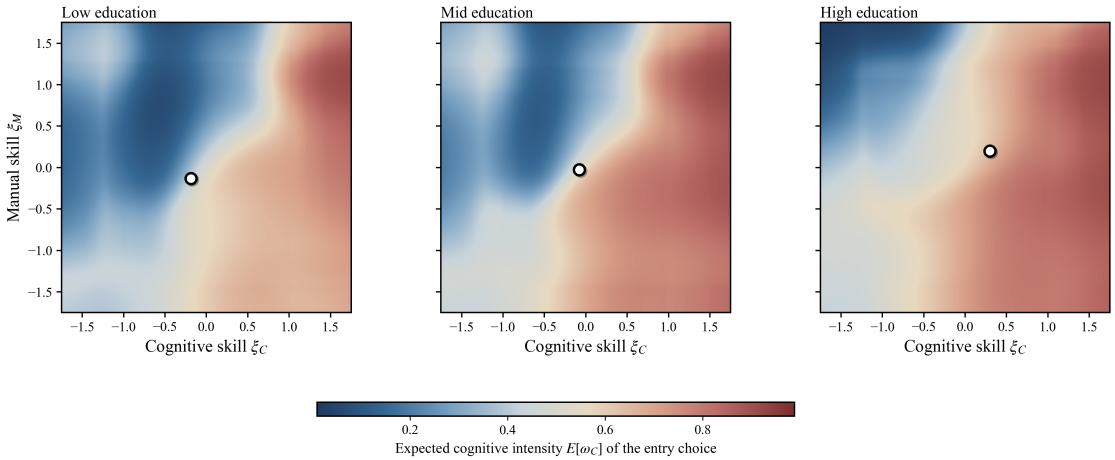
$$I(\xi_i; s_{ij1} \mid g) = h(s_{ij1} \mid g) - h(s_{ij1} \mid \xi_i, g),$$

where $h(Y) = \mathbb{E}[-\log f_Y(Y)]$ denotes differential entropy. Under the model's linear-Gaussian signal structure, $s_{ij1} = \lambda_j'\xi_i + \varepsilon_{ij1}, \quad \xi_i \mid g \sim N(\mu_g, \Sigma_g), \quad \varepsilon_{ij1} \sim N(0, \varsigma_{\varepsilon j}^2)$, the signal is scalar Gaussian with unconditional variance $\lambda_j'\Sigma_g\lambda_j + \varsigma_{\varepsilon j}^2$ and conditional variance $\varsigma_{\varepsilon j}^2$. Since the entropy of a scalar Gaussian $N(m, v)$ is $h(Y) = \frac{1}{2}\log(2\pi e\, v)$, it follows that

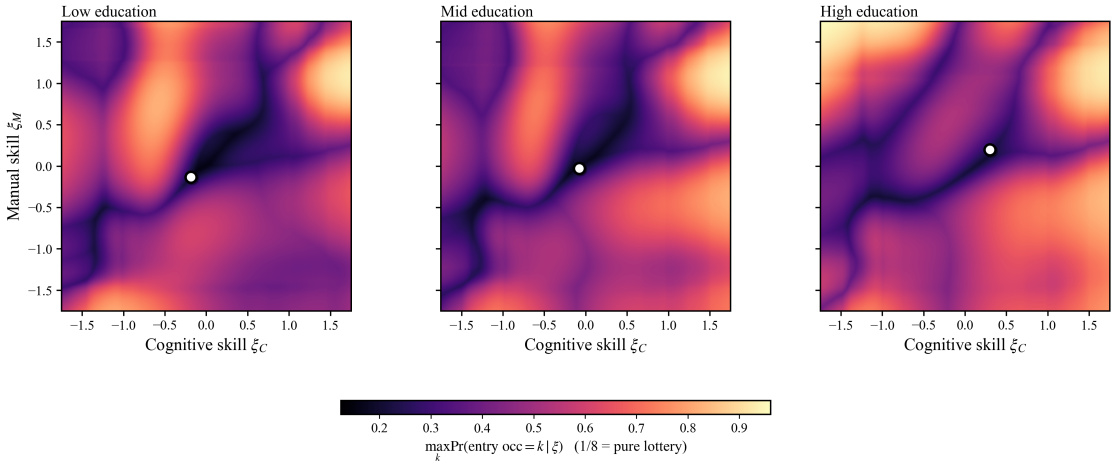
$$I(\xi_i; s_{ij1} \mid g) = \frac{1}{2}\log\Big(2\pi e\left[\lambda_j'\Sigma_g\lambda_j + \varsigma_{\varepsilon j}^2\right]\Big) - \frac{1}{2}\log\Big(2\pi e\,\varsigma_{\varepsilon j}^2\Big) = \frac{1}{2}\log\left(1 + \frac{\lambda_j'\Sigma_g\lambda_j}{\varsigma_{\varepsilon j}^2}\right).$$



(a) Deterministic entry assignment ($\varrho \to 0$)

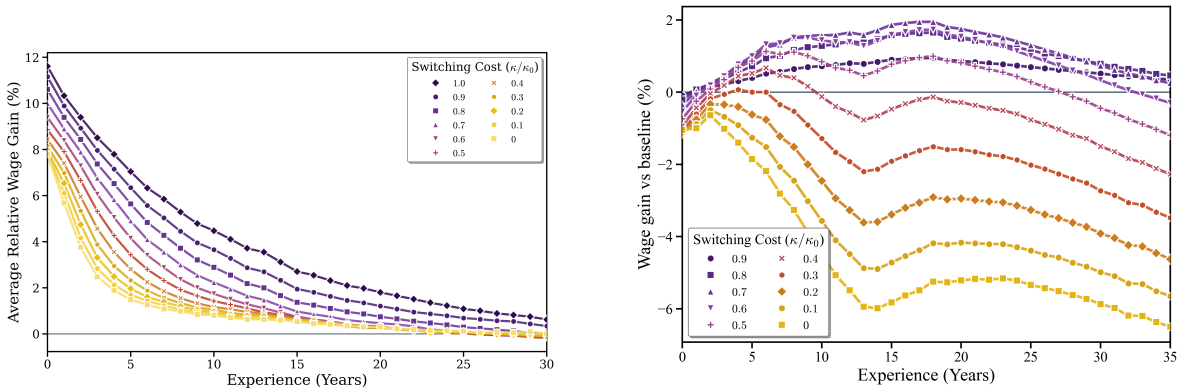


(b) Expected cognitive task intensity of the entry occupation, $\mathbb{E}[\omega_C]$



(c) Probability of the most likely entry occupation, $\max_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\xi}})$

Figure 20: The entry occupational choice over the belief space. In every panel the three sub-panels correspond to the three education groups, and each sub-panel maps the entry choice as a function of entry beliefs about cognitive skill ($\tilde{\xi}_C$, horizontal axis) and manual skill ($\tilde{\xi}_M$, vertical axis), holding experience at zero. The white marker locates the group's prior mean $\boldsymbol{\mu}_g$, the point at which all members of the group actually enter. *Panel (a)* shows the occupation cluster that maximizes the systematic choice value, i.e. the choice in the deterministic limit $\varrho \to 0$; clusters are ordered by cognitive task intensity ω_C (legend). *Panel (b)* shows the task intensity ω_C of the entry occupation in expectation over the taste shocks, $\mathbb{E}[\omega_C] = \sum_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\xi}})\, \omega_{kC}$. *Panel (c)* shows the probability mass placed on the single most likely entry occupation; the value $1/8$ corresponds to a uniform lottery over the eight clusters. Panels (b) and (c) are evaluated at the estimated taste-shock scale $\varrho = 0.271$.



(a) Gain from perfect information, by κ/κ_0

(b) Gain over the baseline economy, by κ/κ_0

Figure 21: Lifecycle wage profiles for the switching-cost counterfactuals. Horizontal axis: years of potential experience; vertical axis: average relative wage gain (percent). *Panel (a)*: wage gain from moving to perfect information at each switching cost κ/κ_0. *Panel (b)*: wage gain relative to the estimated baseline economy at each κ/κ_0.

to $\arg\max_j IG_{jg}^{\text{entry}}$ in their first period. From period 2 onward they choose freely under their updated beliefs. The assignment is welfare-neutral in the sense that workers receive the going market wage in the assigned occupation; the only constraint is on occupational choice in the first period.

The forced occupations are those that best reveal a worker's comparative advantage, and they turn out to be the two *anchor* occupations: the manual anchor (Occ_1) and, for a smaller share of workers, the cognitive anchor (Occ_0). This is intuitive: the anchors load on a single skill dimension so their wage is basically a noisy, but still more informative signal about one skill, compared to a signal about mixed skills. Figure 23 shows the occupational shares at entry, after one year, and after ten years, under forced exploration and in the baseline.

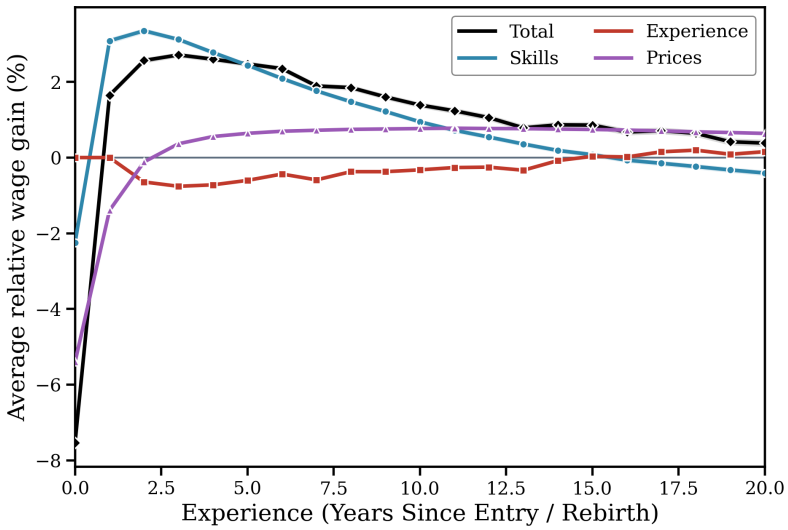


Figure 22: Lifecycle wage gain from forced exploration relative to the baseline economy, decomposed into channels: Total (black), Skills (blue), Experience (red), and Prices (purple). Horizontal axis: years of potential experience since entry; vertical axis: average relative wage gain (percent).

At entry, workers are concentrated in the two anchor occupations, differently from the base-

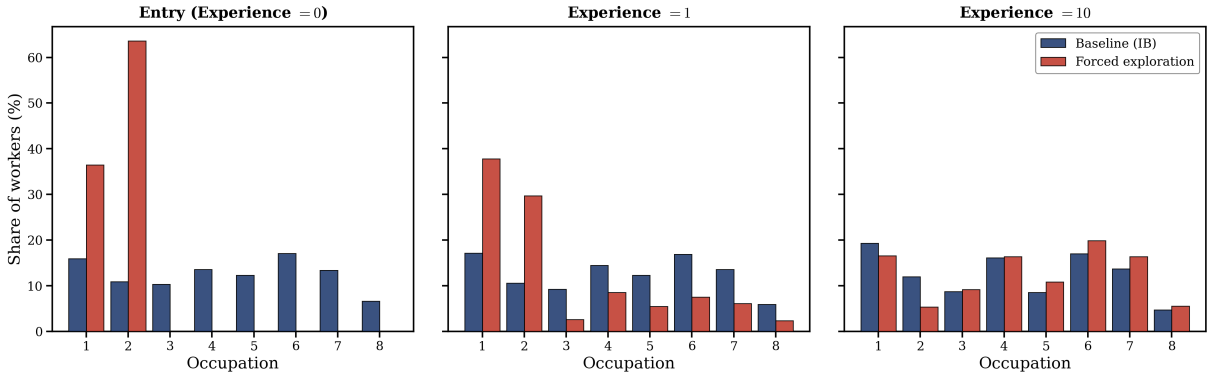


Figure 23: Share of workers in each occupation cluster (vertical axis) under forced exploration (red) and in the baseline imperfect-information economy (blue), at entry, after one year, and after ten years of experience. Occupations are numbered 1–8 for Occ_0–Occ_7.

line, where entrants are spread across all eight occupations (Figure 23). This forced bunching triggers a strong general-equilibrium price response: excess labor supply in the informative occupations drives their prices down while prices in the other occupations rise. This drop in prices produces a sharp wage loss on impact, about 7.6% at entry (Figure 22). From the second period workers choose freely, and they reallocate quickly to occupations that better match their comparative advantage in tasks. Infact, the Skills channel jumps to about +3.4% within two years and stays above the baseline for most of the career, suggesting that in this counterfactual workers are indeed better matched on skills. The total wage gain turns positive after the first year, peaking at roughly +2.7% around the third. As shown in Figure 23, the occupational distribution converges back toward the baseline distribution but workers are now better matched within each occupation. The net result is that the short-run wage loss more than pays for itself: aggregate output in the exploration economy is 0.65% higher than in the baseline. These results highlight a key asymmetry in the value of early information. This counterfactual experiment suggests that a policy that temporarily redirects workers toward their most informative occupation upon entering in the labor market can thus be both productivity-enhancing later in the career, at the cost of short-run wage losses.

D Appendix: Estimation and Numerical Methods

D.1 EM Algorithm: Computational Details

The EM algorithm described in this section follow the derivations from James (2012) and Hincapié (2020), and are adjusted for the structure of the model presented in the main text of this paper.

For worker i employed in occupation j at date t,

$$w_{ijt} \;=\; p_j \;+\; \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i \;+\; h_j(\boldsymbol{\tau}_{it}) \;+\; \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}\big(0, \varsigma_j^2\big), \tag{58}$$

where p_j is the occupation log price, $\boldsymbol{\xi}_i := \log \boldsymbol{\theta}_i \in \mathbb{R}^{|\mathcal{K}|}$ collects log latent skills, $\boldsymbol{\lambda}_j \in \mathbb{R}_+^{|\mathcal{K}|}$ are skill loadings, and $\boldsymbol{\tau}_{it} \in \mathbb{R}_+^{|\mathcal{K}|}$ are raw cumulative task tenures. The term $h_j(\boldsymbol{\tau}_{it})$ is parameterized

as a stepwise function, mapping the continuous tenure τ_{it} into tenure bins with corresponding dummy coefficients to capture the non-linear wage progression flexibly. Workers belong to observed groups $g(i) \in \{1, \dots, G\}$ defined by education. The skill prior is conditionally Gaussian within group,

$$\xi_i \mid g(i) = g \sim \mathcal{N}(\mu_g, \Sigma_g). \tag{59}$$

Because $g(i)$ is observed in the data, the model is a conditional Gaussian factor model. The group weights π_g are fixed at empirical sample shares and are not updated by EM.

The parameter vector iterated by the algorithm is

$$\Gamma^q = \Big\{ (\mu_g^q, \Sigma_g^q)_{g=1}^G,\ (p_j^q, \lambda_j^q, \delta_j^q, \varsigma_j^{2,q})_{j \in \mathcal{J}} \Big\},$$

with q indexing the EM iteration.

D.1.1 E-step

Conditional on Γ^q and on $g(i)$, the posterior of ξ_i is multivariate normal. Define the residual after netting out occupation prices and the tenure profile,

$$r_{it}^q \ := \ w_{ijt} \ - \ p_{j(i,t)}^q \ - \ h_{j(i,t)}^q(\tau_{it}), \tag{60}$$

Completing the square in ξ_i in the product of the wage likelihood implied by (58) and the group prior (59) delivers

$$V_i^q = \left((\Sigma_{g(i)}^q)^{-1} + \sum_{t=t_0(i)}^{T(i)} \frac{1}{\varsigma_{j(i,t)}^{2,q}} \lambda_{j(i,t)}^q (\lambda_{j(i,t)}^q)' \right)^{-1}, \tag{61}$$

$$\zeta_i^q = V_i^q \left((\Sigma_{g(i)}^q)^{-1} \mu_{g(i)}^q + \sum_{t=t_0(i)}^{T(i)} \frac{\lambda_{j(i,t)}^q \, r_{it}^q}{\varsigma_{j(i,t)}^{2,q}} \right), \tag{62}$$

$$\xi_i \mid \text{data}, \Gamma^q, g(i) \sim \mathcal{N}(\zeta_i^q,\ V_i^q). \tag{63}$$

I store the worker-specific pair (ζ_i^q, V_i^q) at the end of every E-step iteration. Equations (61)–(62) are themselves a $|\mathcal{K}|$-dimensional linear system per worker weighted by precision of the signals, solved in closed form by inverting the precision matrix in (61).

D.1.2 M-step

The M-step maximizes the expected complete-data log-likelihood over Γ while taking as given the posteriors (ζ_i^q, V_i^q) from the E-step. The complete-data log-likelihood sums, across workers, the log prior density of ξ_i in (59) and the log density of each wage observation in (58). Taking its expectation under the E-step posterior $\xi_i \mid w_i, d_i, \Gamma^q \sim \mathcal{N}(\zeta_i^q, V_i^q)$ gives

$$\begin{aligned} Q(\Gamma \mid \Gamma^q) = & -\tfrac{1}{2} \sum_i \Big[\log|2\pi\Sigma_{g(i)}| + \operatorname{tr}(\Sigma_{g(i)}^{-1} V_i^q) + (\zeta_i^q - \mu_{g(i)})' \Sigma_{g(i)}^{-1} (\zeta_i^q - \mu_{g(i)}) \Big] \\ & - \tfrac{1}{2} \sum_i \sum_t \Big[\log(2\pi\varsigma_j^2) + \frac{(w_{ijt} - p_j - \lambda_j' \zeta_i^q - h_j(\tau_{it}))^2 + \lambda_j' V_i^q \lambda_j}{\varsigma_j^2} \Big], \end{aligned} \tag{64}$$

with $j = j(i,t)$ being the occupation chosen by individual i in period t. This Q-function is additively separable into a prior block, separable across groups g, and a wage block, separable across occupations j; the M-step maximizes each block in closed form. Below we take the first-order conditions of (64) block by block: first for the prior moments, then for each occupation's wage parameters.

Prior moments optimization. Maximizing the prior block of (64) over $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ gives the standard updates

$$\boldsymbol{\mu}_g^{q+1} = \frac{1}{N_g} \sum_{i:\, g(i)=g} \boldsymbol{\zeta}_i^q, \tag{65}$$

$$\boldsymbol{\Sigma}_g^{q+1} = \frac{1}{N_g} \sum_{i:\, g(i)=g} \Big[\boldsymbol{V}_i^q + (\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_g^{q+1})(\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_g^{q+1})' \Big]. \tag{66}$$

The group weights π_g are held fixed at their sample frequencies throughout estimation.

To pin down the location of the latent factor, we normalize to zero the group-weighted population distribution of skills $\sum_g \pi_g \boldsymbol{\mu}_g = \mathbf{0}$. We implement that assumption at each iteration by subtracting $\boldsymbol{a}^{q+1} := \sum_g \pi_g \boldsymbol{\mu}_g^{q+1}$ from every $\boldsymbol{\mu}_g^{q+1}$ and adding it back to the occupation fixed effect (price). This adjustment removes the location indeterminacy of $\boldsymbol{\xi}_i$.

Wage equation parameters optimization. For each occupation j, I stack the regressors for each of the N_j worker–year observations that choose j. In the resulting matrix, each row represents a different observation within that occupation with a constant, the skills for that observation resulting from the E-step, and the task-specific tenure in the form of stepwise bins. Mathematically, we have

$$\boldsymbol{W}_j \;=\; \big[\, \mathbf{1},\;\; \boldsymbol{\zeta}^q,\;\; \boldsymbol{D}(\boldsymbol{\tau}) \,\big] \;\in\; \mathbb{R}^{N_j \times (1+|\mathcal{K}|+|\mathcal{B}|)}, \tag{67}$$

where $\boldsymbol{\zeta}^q$ is the posterior mean for each worker repeated vertically each time the same worker is observed, and $\boldsymbol{D}(\boldsymbol{\tau})$ collects the tenure bin indicators for task-specific experience with total dimension $|\mathcal{B}|$. The corresponding coefficient vector is

$$\boldsymbol{\vartheta}_j \;=\; \big[p_j,\; \boldsymbol{\lambda}_j',\; \boldsymbol{\delta}_j'\big]'.$$

The wage block of (64) is quadratic in $\boldsymbol{\vartheta}_j$, so updating the wage parameters is just a per-occupation least-squares regression of the stacked log wages $\boldsymbol{y}_j$ on $\boldsymbol{W}_j$. Its first-order conditions are given by the following closed form equations

$$\boldsymbol{\vartheta}_j^{q+1} \;=\; \boldsymbol{S}_j^{-1}\boldsymbol{b}_j, \qquad \boldsymbol{S}_j \;=\; \boldsymbol{W}_j'\boldsymbol{W}_j \;+\; \boldsymbol{\Delta}_j^q \qquad \boldsymbol{b}_j \;=\; \boldsymbol{W}_j'\boldsymbol{y}_j. \tag{68}$$

The only departure from textbook OLS is the correction $\boldsymbol{\Delta}_j^q = \sum_{t \in j} \boldsymbol{V}_{i(t)}^q$, which adds the summed posterior variances $\sum_{t \in j} \boldsymbol{V}_i^q$ to the $K \times K$ loading block of $\boldsymbol{S}_j$ and is zero elsewhere. This term is the curvature of the $\boldsymbol{\lambda}_j' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j$ correction in (64); it corrects the loadings $\boldsymbol{\lambda}_j$ for the residual posterior uncertainty about $\boldsymbol{\xi}_i$. Intuitively, residual wages are noisy measurement of multidimensional skills because there are the iid shocks and and the loadings multiply workers skills. If we had observed every worker for many periods moving across different occupations, then this uncertainty would be close to zero for everyone. This system of equations has dimension $1 + |\mathcal{K}| + |\mathcal{B}|$ and is solved separately for each occupation, delivering $(p_j^{q+1}, \boldsymbol{\lambda}_j^{q+1}, \boldsymbol{\delta}_j^{q+1})$ jointly.

The inclusion of the E-step posterior $\boldsymbol{\zeta}^q$ as a regressor in $\boldsymbol{W}_j$ representes the practical implementation of the control-function correction for selection on beliefs that the identification argument of Section 5.1 describe: regressing log wages on $(\mathbf{1}, \boldsymbol{\zeta}^q, \boldsymbol{D}(\boldsymbol{\tau}))$ per occupation absorbs endogenous occupational choice through the observed history, and so control for belief-based selection into occupation, while the augmentation $\boldsymbol{\Delta}_j^q$ corrects the loading block for residual posterior uncertainty.

Anchored normalization. The factor model is identified up to a $K \times K$ invertible transformation of $(\boldsymbol{\xi}, \boldsymbol{\lambda})$. To eliminate this rotation and scale indeterminacy, we pin the full loading vector at two anchor occupations: a cognitive anchor j_C and a manual anchor j_M. In the default implementation of the EM algorithm we set $\boldsymbol{C} = \boldsymbol{I}_K$, i.e. $\boldsymbol{\lambda}_{j_C} = (1, 0)$ and $\boldsymbol{\lambda}_{j_M} = (0, 1)$. This choice imposes four scalar restrictions (two scale, two rotation), which is exactly what is required to identify the orientation of the latent skill space. At the two anchor occupations we substitute $\boldsymbol{\lambda}_j$ out of (68) and solve the reduced system for $(p_j, \boldsymbol{\delta}_j)$ only.

Additionally, and to make sure that the parameters of the model have a meaningful economic content, we impose the skill loadings to be higher or equal than zero i.e. $\boldsymbol{\lambda}_j \geq \mathbf{0}$. This assumption ensure a meaningful ordinal ranking of skills: if $\lambda_{jk} < 0$ for any occupation j in task k then a very low skill worker in dimension k might have an unreasonable high productivity in occupation j.

Residual variance. Given the updated $(p_j^{q+1}, \boldsymbol{\lambda}_j^{q+1}, \boldsymbol{\delta}_j^{q+1})$, the variance update is

$$\varsigma_j^{2,q+1} \;=\; \frac{1}{N_j} \sum_{(i,t):\, j(i,t)=j} \Big[\big(r_{it}^{q+1} - (\boldsymbol{\lambda}_j^{q+1})' \boldsymbol{\zeta}_i^q\big)^2 + (\boldsymbol{\lambda}_j^{q+1})' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j^{q+1} \Big], \tag{69}$$

where r_{it}^{q+1} is the residual from (60) evaluated at the new wage parameters. The second term inside the bracket is the variance-of-the-projection correction $\boldsymbol{\lambda}_j' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j$ from (64) and ensures that posterior uncertainty about $\boldsymbol{\xi}_i$ is correctly transmitted into ς_j^2.

The algorithm iterates (61)–(63) and (65)–(69) until convergence. Two stopping rules are available. The default monitors the mean squared change in the flattened parameter vector across consecutive iterations and stops once this falls below 10^{-9}; an alternative monitors the absolute change in the Q-function with tolerance 10^{-4}.

D.1.3 Likelihood Evaluation

For model comparison and start selection, we evaluate the observed-data wage log-likelihood by integrating out $\boldsymbol{\xi}_i$ under the group prior. For a candidate parameter vector $\boldsymbol{\Gamma}$,

$$\tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \;=\; \int_{\mathbb{R}^{|\mathcal{K}|}} \left\{ \prod_{t=1}^{T_i} \phi\Big(w_{ij(i,t)t};\; p_{j(i,t)} + \boldsymbol{\lambda}_{j(i,t)}' \boldsymbol{\xi}_i + h_{j(i,t)}(\boldsymbol{\tau}_{it}),\; \varsigma_{j(i,t)}^2 \Big) \right\} \phi\big(\boldsymbol{\xi}_i; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\big)\, d\boldsymbol{\xi}_i, \tag{70}$$

where $\phi(\cdot\,; \mu, \varsigma^2)$ is the univariate normal density and $\phi(\cdot\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the multivariate normal density.

We approximate (70) by Monte Carlo. For each worker i we fix a sequence of standard-normal

draws $\boldsymbol{Z}_i^{(q)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and map them through the current prior,

$$\boldsymbol{\xi}_i^{(q)} = \boldsymbol{\mu}_{g(i)} + \boldsymbol{L}_{g(i)}\, \boldsymbol{Z}_i^{(q)}, \qquad \boldsymbol{L}_g \boldsymbol{L}_g' = \boldsymbol{\Sigma}_g,$$

and form the worker-level time-summed log contribution

$$S_i^{(q)}(\boldsymbol{\Gamma}) = \sum_{t=1}^{T_i} \log \phi\Big(w_{ij(i,t)t};\, p_{j(i,t)} + \boldsymbol{\lambda}_{j(i,t)}' \boldsymbol{\xi}_i^{(q)} + h_{j(i,t)}(\boldsymbol{\tau}_{it}),\, \varsigma_{j(i,t)}^2\Big).$$

The log-likelihood is then evaluated by the log-sum-exp identity,

$$\log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \;\approx\; \text{logsumexp}\big(\{S_i^{(q)}(\boldsymbol{\Gamma})\}_{q=1}^{Q}\big) - \log Q, \tag{71}$$

which is stable under reweighting at extreme parameter values[31]

D.2 NFXP Algorithm: Computational Details

D.2.1 Group–Occupation–Specific Integrated Value Functions and Fixed Point

In the dynamic model, the worker state at time t is

$$\boldsymbol{x}_t \;\equiv\; \big(g,\, \tilde{\boldsymbol{\xi}}_t,\, \tilde{\boldsymbol{\Sigma}}_t,\, \boldsymbol{\tau}_t,\, d_{t-1}\big) \in \mathcal{X},$$

the decision-relevant part of the state (23): the education group $g \in \mathcal{G}$, the posterior mean $\tilde{\boldsymbol{\xi}}_t \in \mathbb{R}^{|\mathcal{K}|}$ and covariance $\tilde{\boldsymbol{\Sigma}}_t \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{K}|}$ over latent log skills, task tenures $\boldsymbol{\tau}_t \in \mathbb{R}_+^{|\mathcal{K}|}$, and last period's occupation $d_{t-1} \in \mathcal{J}$ (occupation prices are constant in the stationary solve and are suppressed). To keep the occupation-specific value functions below transparent, we index the current occupational choice by $j \in \mathcal{J}$ and reserve m for the lagged occupation $d_{t-1} = m$. Conditional on choosing j at t, the realized log wage in (14) delivers the scalar signal, mirroring (17),

$$s_{jt} \;\equiv\; \log W_{jt} - p_{jt} - h_j(\boldsymbol{\tau}_t) = \boldsymbol{\lambda}_j' \boldsymbol{\xi} + \varepsilon_{jt}, \qquad \varepsilon_{jt} \sim \mathcal{N}(0, \varsigma_{\varepsilon j}^2), \tag{72}$$

and the Kalman recursion (the filter form of Bayes' rule (18)) reads

$$\tilde{\boldsymbol{\xi}}_{t+1} = \tilde{\boldsymbol{\xi}}_t + \boldsymbol{K}_{jt}\, z_{jt}, \qquad z_{jt} \equiv s_{jt} - \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_t \sim \mathcal{N}\big(0,\, s_{zz}^{(j)}\big), \tag{73}$$

$$\tilde{\boldsymbol{\Sigma}}_{t+1} = \big(\boldsymbol{I}_{|\mathcal{K}|} - \boldsymbol{K}_{jt} \boldsymbol{\lambda}_j'\big) \tilde{\boldsymbol{\Sigma}}_t, \qquad \boldsymbol{K}_{jt} = \frac{\tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_j}{s_{zz}^{(j)}}, \tag{74}$$

$$s_{zz}^{(j)} \equiv \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2, \qquad \boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\omega}_j. \tag{75}$$

The next-period covariance $\tilde{\boldsymbol{\Sigma}}_{t+1}$ and the next-period tenures $\boldsymbol{\tau}_{t+1}$ are both deterministic transitions given $(\boldsymbol{x}_t, j)$; the only random object is the scalar innovation z_{jt}, with predictive variance $s_{zz}^{(j)}$. Because ε_{jt} is the only stochastic shock between two consecutive choices, the next posterior mean $\tilde{\boldsymbol{\xi}}_{t+1}$ moves only along the direction defined by the Kalman gain $\boldsymbol{K}_{jt}$. Although the next period beliefs are multidimensional (as the number of skills), $\tilde{\boldsymbol{\xi}}_{t+1} \in \mathbb{R}^{|\mathcal{K}|}$, its conditional dis-

[31] Across iterations, we fix the commond random numbers ensures that differences in $\sum_i \log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma})$ across starts reflect an improvement in the fit of the wage model rather than simulation noise.

tribution is supported on a one-dimensional subspace of $\mathbb{R}^{|\mathcal{K}|}$. Therefore the expectation over the continuation value reduce to a one-dimensional integral over z_{jt}, regardless of the number of skills $|\mathcal{K}|$.

Univariate Gauss–Hermite quadrature. We use the change of variables $u = z/\sqrt{2\,s_{zz}^{(j)}}$ to map the Gaussian density into the Hermite weight e^{-u^2}:

$$\mathbb{E}\big[G(\tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, j) \,\big|\, \boldsymbol{x}_t, j\big] \;=\; \pi^{-1/2} \int_{\mathbb{R}} G\Big(\tilde{\boldsymbol{\xi}}_t + \sqrt{2\,s_{zz}^{(j)}}\,\boldsymbol{K}_{jt}\,u,\; \tilde{\boldsymbol{\Sigma}}_{t+1},\; \boldsymbol{\tau}_t + \boldsymbol{\omega}_j,\; j\Big)\, e^{-u^2}\, du. \tag{76}$$

Letting $\{(u_r, w_r)\}_{r=1}^{N}$ denote the N-node univariate Gauss–Hermite nodes and weights satisfying $\int_{\mathbb{R}} g(u) e^{-u^2} du \approx \sum_{r=1}^{N} w_r\, g(u_r)$, the continuation expectation is approximated by

$$\mathbb{E}\big[G(\boldsymbol{x}_{t+1}) \,\big|\, \boldsymbol{x}_t, j\big] \;\approx\; \pi^{-1/2} \sum_{r=1}^{N} w_r\, G\Big(\tilde{\boldsymbol{\xi}}_t + \sqrt{2\,s_{zz}^{(j)}}\,\boldsymbol{K}_{jt}\,u_r,\; \tilde{\boldsymbol{\Sigma}}_{t+1},\; \boldsymbol{\tau}_t + \boldsymbol{\omega}_j,\; j\Big). \tag{77}$$

The cost is therefore N function evaluations per $(\boldsymbol{x}_t, j)$ rather than $N^{|\mathcal{K}|}$, which would have been required by a tensor-product rule applied to the full posterior covariance.

Deterministic index and integrated value. In practice, rather than including the discrete states in the state space, we define group- and occupation-specific value functions, with only the continuous variables $\boldsymbol{x} \equiv (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau})$ remaining as arguments. Let

$$W_{mg}(\boldsymbol{x}) \;\equiv\; W(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1} = m,\, g), \qquad m \in \mathcal{J},\; g \in \mathcal{G},$$

denote the integrated value function conditional on being in occupation m at the start of the period and belonging to group g. For a current choice $j \in \mathcal{J}$, the choice-j specific deterministic index combines the expected log wage, the amenity, the switching cost, and the discounted continuation value:

$$\begin{aligned} \nu_{j|m,g}(\boldsymbol{x}) \;=\; & \underbrace{p_j + \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}} + h_j(\boldsymbol{\tau})}_{\text{expected log wage}} + \eta_{j,g} - \mathcal{C}(j,m) \\ & + \beta\, \mathbb{E}\big[\, W_{jg}(\boldsymbol{x}') \;\big|\; \boldsymbol{x}, j \,\big], \end{aligned} \tag{78}$$

where group g is fixed over the worker's life and is therefore carried unchanged into the continuation value W_{jg}, whose incumbent occupation is the current choice j.

With EV–I$(0, \varrho)$ taste shocks under the mean-zero normalization (so the Euler–Mascheroni constant is absorbed into the location parameter and no additive $\gamma\varrho$ remains), the occupation-specific integrated values satisfy a log-sum-exp fixed point. Substituting the univariate Gauss–Hermite approximation (77) for the continuation expectation in (78) and writing the deterministic flow utility as $u_{jg}(\boldsymbol{x}) \equiv p_j + \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}} + h_j(\boldsymbol{\tau}) + \eta_{j,g}$, the integrated Bellman equation reads, for

each $m \in \mathcal{J}$ and $g \in \mathcal{G}$,

$$W_{mg}(\boldsymbol{x}) \;= \varrho \, \log \sum_{j \in \mathcal{J}} \exp\Bigg(\frac{1}{\varrho} \Big[\, u_{jg}(\boldsymbol{x}) - \mathcal{C}(j,m) \\ + \beta \, \pi^{-1/2} \sum_{r=1}^{N} w_r \, W_{jg}\Big(\tilde{\boldsymbol{\xi}} + \sqrt{2\, s_{zz}^{(j)}}\, \boldsymbol{K}_j \, u_r, \; \tilde{\boldsymbol{\Sigma}}', \; \boldsymbol{\tau} + \boldsymbol{\omega}_j \Big) \, \Big] \Bigg), \qquad \boldsymbol{x} \in \mathcal{X}. \tag{79}$$

For each group g, collect the integrated values across incumbent occupations into the vector $\boldsymbol{W}_g(\boldsymbol{x}) = \big(W_{1g}(\boldsymbol{x}), \ldots, W_{|\mathcal{J}|g}(\boldsymbol{x})\big)'$. The right-hand side of (79) for a given incumbent m aggregates over all possible occupation choices $j \in \mathcal{J}$, and each choice carries its own continuation value W_{jg} evaluated with the transitions implied by choosing j, so every component of $\boldsymbol{W}_g$ enters the update of every other component. The $|\mathcal{J}|$ equations therefore constitute a single vector fixed point,

$$\boldsymbol{W}_g \;=\; \mathcal{T}(\boldsymbol{W}_g), \qquad \big[\mathcal{T}(\boldsymbol{W}_g)\big]_m(\boldsymbol{x}) = \varrho \, \log \sum_{j \in \mathcal{J}} \exp\Big(\tfrac{1}{\varrho} \big[\, u_{jg}(\boldsymbol{x}) - \mathcal{C}(j,m) + \beta \, \mathbb{E}\big(W_{jg}(\boldsymbol{x}') \mid \boldsymbol{x}, j\big) \, \big] \Big), \tag{80}$$

which must be solved jointly for all $|\mathcal{J}|$ components rather than one occupation at a time. By Lemma 2, $\mathcal{T}$ is a contraction of modulus at most β, so (80) has a unique fixed point $\boldsymbol{W}_g^{\star}$. The fixed point is solved separately for each $g \in \mathcal{G}$. In what follows we suppress the dependency on g to simplify the notation.

Smolyak (Sparse–Grid) Approximation of the Integrated Value Function The continuous state $\boldsymbol{x} = (\tilde{\boldsymbol{\xi}}, \boldsymbol{\tau}) \in \mathbb{R}^{2|\mathcal{K}|}$ can be high-dimensional. We approximate each value function $W_{mg}(\cdot)$ on a sparse grid using the Smolyak method (Malin et al. 2011) with Chebyshev polynomials. Because the construction is identical for every incumbent occupation m and group g, we describe it for a generic value function $W(\boldsymbol{x})$ and suppress the (m, g) index.

Fix lower and upper bounds for each component of $\boldsymbol{x}$. Let $d = 2|\mathcal{K}|$ be the continuous-state dimension and μ the Smolyak level. The sparse grid is

$$\mathcal{H}^{d,\mu} = \bigcup_{\mathbf{i}:\, |\mathbf{i}| = d+\mu} \Big(\mathcal{G}^{m(i_1)} \times \cdots \times \mathcal{G}^{m(i_d)} \Big), \qquad |\mathbf{i}| := i_1 + \cdots + i_d,$$

where $\mathcal{G}^{m(\cdot)}$ are nested univariate nodes (Chebyshev extrema) and $m(\cdot)$ is the node-count function. For a given multi-index $\mathbf{i}$, the tensor-product interpolant for W is

$$W^{\mathbf{i}}(\boldsymbol{x}; \boldsymbol{\phi}^{(\mathbf{i})}) = \sum_{l_1=1}^{m(i_1)} \cdots \sum_{l_d=1}^{m(i_d)} \phi^{(\mathbf{i})}_{l_1 \cdots l_d} \, T_{l_1}(x_1) \cdots T_{l_d}(x_d) = \boldsymbol{\phi}^{(\mathbf{i})\prime} \mathbf{T}(\boldsymbol{x}),$$

with $\mathbf{T}(\boldsymbol{x})$ the basis vector of multivariate Chebyshev polynomials. The Smolyak combination gives the global approximation

$$\widehat{W}^{\,d,\mu}(\boldsymbol{x}) = \sum_{d \le |\mathbf{i}| \le d+\mu} (-1)^{d+\mu-|\mathbf{i}|} \binom{d-1}{d+\mu-|\mathbf{i}|} \, W^{\mathbf{i}}(\boldsymbol{x}; \boldsymbol{\phi}^{(\mathbf{i})}). \tag{81}$$

Fixed point in coefficient space. Substituting the Smolyak interpolant $\widehat{W}_j(\boldsymbol{x}) = \boldsymbol{\phi}^{(j)\prime}\mathbf{T}(\boldsymbol{x})$ for the unknown continuation value W_j inside (79) recasts the integrated Bellman equation as a fixed point in the finite-dimensional Smolyak coefficients $\boldsymbol{\Phi} \equiv [\boldsymbol{\phi}^{(1)}, \ldots, \boldsymbol{\phi}^{(|\mathcal{J}|)}] \in \mathbb{R}^{M\times|\mathcal{J}|}$, whose columns index the incumbent occupation and where M is the number of Smolyak basis functions. For each incumbent $m \in \mathcal{J}$,

$$\boldsymbol{\phi}^{(m)\prime}\mathbf{T}(\boldsymbol{x}) \;= \varrho\, \log \sum_{j\in\mathcal{J}} \exp\left(\frac{1}{\varrho}\Bigg[\, u_j(\boldsymbol{x}) - \mathcal{C}(j,m) \right.$$
$$\left. + \beta\, \pi^{-1/2} \sum_{r=1}^{N} w_r\, \boldsymbol{\phi}^{(j)\prime}\mathbf{T}\Big(\tilde{\boldsymbol{\xi}} + \sqrt{2\, s_{zz}^{(j)}}\, \boldsymbol{K}_j\, u_r,\; \tilde{\boldsymbol{\Sigma}}',\; \boldsymbol{\tau} + \boldsymbol{\omega}_j\Big) \Bigg]\right), \qquad \forall\, \boldsymbol{x} \in \mathcal{X},\; m \in \mathcal{J}. \tag{82}$$

Both sides are now expressed in coefficient space, so the fixed point is over the $M \times |\mathcal{J}|$ matrix of Smolyak coefficients $\boldsymbol{\Phi}$ rather than over the function $W_m(\cdot)$.

To obtain the matrix-stacked map solved in practice, evaluate (82) at every node of the sparse grid. Let $\{\boldsymbol{x}^{(g)}\}_{g=1}^{G}$ denote the G grid points in $\mathcal{H}^{d,\mu}$, and let $\boldsymbol{y}^{(g)} \in [-1,1]^d$ be $\boldsymbol{x}^{(g)}$ rescaled to the Chebyshev domain. Define the basis matrix $\mathbf{T} \in \mathbb{R}^{M\times G}$ by $\mathbf{T}_{\ell,g} = T_\ell(\boldsymbol{y}^{(g)})$ (basis function $\ell = 1, \ldots, M$, grid point g) and the collocation pseudo-inverse

$$\mathbf{T}' \;\equiv\; (\mathbf{T}\mathbf{T}')^{-1}\mathbf{T} \;\in\; \mathbb{R}^{M\times G};$$

in the unisolvent case this is the same as $\mathbf{T}^{-1}$; I use the pseudo-inverse rather than direct inversion for numerical stability. Stacking the right-hand sides of (82) into the matrix $\widetilde{\mathbf{W}}(\boldsymbol{\Phi}) \in \mathbb{R}^{|\mathcal{J}|\times G}$ with entries $\widetilde{\mathbf{W}}_{mg}(\boldsymbol{\Phi}) = \widehat{W}_m(\boldsymbol{x}^{(g)}; \boldsymbol{\Phi})$, the fixed-point map in coefficient space is

$$\boldsymbol{\Phi} \;=\; \mathcal{F}(\boldsymbol{\Phi}) \;\equiv\; \mathbf{T}'\, \widetilde{\mathbf{W}}(\boldsymbol{\Phi})'. \tag{83}$$

We compute the fixed point $\boldsymbol{\Phi}^\star$ of (83) by successive substitution, compiled into a single JAX while-loop so that gradients of the outer likelihood flow through the converged coefficients.

D.2.2 Fixed-Point Algorithm for Integrated Value Function Smolyak Approximation

The inner loop of the NFXP procedure solves (83) by successive iteration on the Smolyak coefficient matrix $\boldsymbol{\Phi}$. Each iteration evaluates the right-hand side of (82) at every grid point, takes a Gauss–Hermite expectation over next-period beliefs, and updates the coefficients. The algorithm box below gives the steps.

Algorithm: Fixed-Point Iteration for the Integrated Value Function Smolyak Approximation

Inputs.
- Structural parameters $\boldsymbol{\Theta}$.
- Sparse-grid nodes $\{\boldsymbol{x}^{(g)}\}_{g=1}^{G} \subset \mathcal{H}^{d,\mu}$.
- Chebyshev basis $\mathbf{T} \in \mathbb{R}^{M\times G}$ and pseudo-inverse $\mathbf{T}' = (\mathbf{T}\mathbf{T}')^{-1}\mathbf{T}$.
- Univariate Gauss–Hermite nodes and weights $\{(u_r, w_r)\}_{r=1}^{N}$.
- Tolerance $\varepsilon > 0$; maximum iterations $K_{\max}$.

Output. Fixed-point Smolyak coefficient matrix $\boldsymbol{\Phi}^{\star} \in \mathbb{R}^{M\times|\mathcal{J}|}$ solving (83).

Initialization. Set $\boldsymbol{\Phi}^{(0)} \in \mathbb{R}^{M\times|\mathcal{J}|}$ (default: zeros).

Main loop. Repeat for $n = 0, 1, \ldots, K_{\max} - 1$ until convergence:
For each grid point g and prospective choice $j \in \mathcal{J}$:
- Update next-period belief covariance $\tilde{\boldsymbol{\Sigma}}'(\boldsymbol{x}^{(g)}, j)$ via Kalman update.
- Build N shifted next-state nodes $\boldsymbol{x}'_r$ via the Gauss–Hermite map [(77)].
- Evaluate Smolyak interpolant $\widehat{W}_j(\boldsymbol{x}'_r; \boldsymbol{\Phi}^{(n)}) = \boldsymbol{\phi}^{(n,j)\prime}\mathbf{T}(\boldsymbol{x}'_r)$.
- Form conditional values $\nu_{j|m}(\boldsymbol{x}^{(g)})$ for all incumbents $m \in \mathcal{J}$ [(82)].

Aggregate via log-sum-exp: $\widetilde{W}_m(\boldsymbol{x}^{(g)}) = \varrho \log \sum_j \exp\big(\nu_{j|m}(\boldsymbol{x}^{(g)})/\varrho\big)$.
Collocation update: $\boldsymbol{\Phi}^{(n+1)} = \mathbf{T}'\,\widetilde{\mathbf{W}}(\boldsymbol{\Phi}^{(n)})'$ [(83)].
Convergence: if $\left\|\boldsymbol{\Phi}^{(n+1)} - \boldsymbol{\Phi}^{(n)}\right\|_F < \varepsilon$, exit.

Return $\boldsymbol{\Phi}^{\star} = \boldsymbol{\Phi}^{(n+1)}$.

Numerical stability. For small ϱ the log-sum-exp aggregator is computed in shifted form with $v_{\max} = \max_j v_j$:

$$\varrho \log\Big(\sum_j e^{v_j/\varrho}\Big) \;=\; v_{\max} + \varrho \log\Big(\sum_j e^{(v_j - v_{\max})/\varrho}\Big),$$

which avoids overflow when one alternative dominates.

D.2.3 Policy Functions, CCPs, and Simulation

Given $\{\widehat{W}_m\}_{m\in\mathcal{J}}$, the CCPs at $(\boldsymbol{x}, m)$ follow from (84). To simulate a worker's path:

1. Given $(\tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, m)$, compute CCPs $\{\mathbb{P}(d_t = j \mid \boldsymbol{x}_t, m)\}_{j\in\mathcal{J}}$ and draw d_t (or use the argmax for policy).
2. Draw a wage residual $\varepsilon_{d_t,t} \sim \mathcal{N}(0, \varsigma_{d_t}^2)$ and set
$$\log W_{d_t t} = p_{d_t t} + \boldsymbol{\lambda}'_{d_t}\boldsymbol{\xi} + h_{d_t}(\boldsymbol{\tau}_t) + \varepsilon_{d_t,t}.$$
Equivalently, the signal in (72) is $s_{d_t t} = \log W_{d_t t} - p_{d_t t} - h_{d_t}(\boldsymbol{\tau}_t)$.
3. Update beliefs via (73)–(75) using $j = d_t$.
4. Update tenures: $\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\omega}_{d_t}$ (cf. (10)).
5. Update the discrete state: $m \leftarrow d_t$ (next period's lagged occupation).

D.2.4 Conditional Choice Probabilities and Likelihood

Once the fixed-point coefficients $\mathbf{\Phi}^\star$ are recovered, conditional choice probabilities and the choice likelihood follow in closed form. Given the integrated value functions $\{\widehat{W}_j(\boldsymbol{x};\mathbf{\Phi}^\star)\}_{j\in\mathcal{J}}$ and a worker arriving at state $\boldsymbol{x}$ with lagged occupation m, the multinomial-logit CCPs implied by the Type-I extreme value taste shocks are

$$\mathbb{P}\big(d_t = j \mid \boldsymbol{x}, m; \mathbf{\Theta}\big) \;=\; \frac{\exp\big(\nu_{j|m}(\boldsymbol{x})/\varrho\big)}{\sum_{\ell\in\mathcal{J}}\exp\big(\nu_{\ell|m}(\boldsymbol{x})/\varrho\big)}, \qquad j\in\mathcal{J}, \tag{84}$$

where $\nu_{j|m}(\boldsymbol{x}) \;=\; u_j(\boldsymbol{x}) - \mathcal{C}(j,m) + \beta\,\mathbb{E}\big[W_j(\boldsymbol{x}') \mid \boldsymbol{x}, j\big]$ is the choice-specific value of moving to occupation j from m.

Beliefs, state variable and choice likelihood evaluation Using parameter estimates for the wage equation, we can compute the exact sequence of beliefs for every worker observed in the estimation sample.

Evaluation of the choice likelihood on the estimation sample. Using parameters estimated in the first stage and imposing the rationality assuption, we can compute the exact sequence of beliefs for every worker in the estimation sample. Workers belonging to group g enter in the labor market with state $x_{i0}(putstateshere)$. After entering, residual wages obtained after residualizing for prices and tenures allow me to compute the same signal that the worker uses to update his beliefs. Therefore for each subsequent period we can compute the sequence of beliefs updating for the mean and the variance-covariance matrix (add sequence of beliefs mean and varcov here mu sigma t+1). Once we have computed the beliefs at each point in time for every worker we can actually use these state variables to compute the likelihood function.

Evaluation of the choice likelihood on the estimation sample. Using the parameters estimated in the first stage and imposing the rationality assumption, we can compute the exact sequence of beliefs for every worker in the estimation sample. Workers belonging to group g enter the labor market with state

$$\boldsymbol{x}_{i0} = \big(g(i),\, \tilde{\boldsymbol{\xi}}_{i0},\, \tilde{\mathbf{\Sigma}}_{i0},\, \boldsymbol{\tau}_{i0},\, d_{i,-1}\big), \qquad \tilde{\boldsymbol{\xi}}_{i0} = \hat{\boldsymbol{\mu}}_{g(i)}, \quad \tilde{\mathbf{\Sigma}}_{i0} = \hat{\mathbf{\Sigma}}_{g(i)}, \quad \boldsymbol{\tau}_{i0} = \mathbf{0}, \quad d_{i,-1} = -1,$$

that is, rational initial beliefs equal to the group prior, no accumulated task experience, and no incumbent occupation ($d_{i,-1} = -1$, so the first choice bears no switching cost). After entering, the residual wage obtained after netting out prices and tenures gives the same signal the worker uses to update her beliefs,

$$s_{it} \;\equiv\; w_{id_{it}t} - p_{d_{it}} - h_{d_{it}}(\boldsymbol{\tau}_{it}),$$

where d_{it} is her observed occupation at t. Therefore, for each subsequent period we can compute the belief sequence for the mean and the variance–covariance matrix by the Kalman re-

cursion (73)–(75) evaluated at the observed choice d_{it},

$$\tilde{\boldsymbol{\xi}}_{i,t+1} = \tilde{\boldsymbol{\xi}}_{it} + \boldsymbol{K}_{it}\big(s_{it} - \boldsymbol{\lambda}'_{d_{it}}\tilde{\boldsymbol{\xi}}_{it}\big), \quad \tilde{\boldsymbol{\Sigma}}_{i,t+1} = \big(\boldsymbol{I}_{|\mathcal{K}|} - \boldsymbol{K}_{it}\boldsymbol{\lambda}'_{d_{it}}\big)\tilde{\boldsymbol{\Sigma}}_{it},$$
$$\boldsymbol{K}_{it} = \frac{\tilde{\boldsymbol{\Sigma}}_{it}\boldsymbol{\lambda}_{d_{it}}}{\boldsymbol{\lambda}'_{d_{it}}\tilde{\boldsymbol{\Sigma}}_{it}\boldsymbol{\lambda}_{d_{it}} + \varsigma^2_{\varepsilon d_{it}}}.$$

Once we have computed the beliefs at each point in time for every worker, we can use these state variables to evaluate the likelihood function.

The fixed-point algorithm of Section D.2.2 returns the Smolyak coefficient matrix $\boldsymbol{\Phi}^\star$ that approximate the value function at the current parameter estimates. To evaluate the choice likelihood, we must compute CCPs for each observation in the estimation sample. Because we approximate the value function using Chebyshev polynomials, we can simply compute the value function at every observed state.

Stacking (84) across workers and periods yields

$$\ell\big(\boldsymbol{\Theta}_{\text{dyn}}; \hat{\boldsymbol{\Theta}}_{\text{wage}}\big) \;=\; \sum_{i=1}^{N}\sum_{t=1}^{T_i} \log \mathbb{P}\big(d_{it} \mid \boldsymbol{x}_{it}, j_{i,t-1}; \boldsymbol{\Phi}^\star(\boldsymbol{\Theta}_{\text{dyn}}), \hat{\boldsymbol{\Theta}}_{\text{wage}}\big). \tag{85}$$

The wage-block parameters $\hat{\boldsymbol{\Theta}}_{\text{wage}}$ enter (85) only through the choice-specific values inside $\boldsymbol{\Phi}^\star$ and are held fixed at their Stage 1 EM estimates. Entry rows contribute via $j_{i,-1} = -1$, in which case no switching cost is incurred.

D.2.5 NFXP Procedure

Stage 1 of the EM routine delivers the wage-block estimates $\hat{\boldsymbol{\Theta}}_{\text{wage}}$ and the means and covariances $\{\hat{\boldsymbol{\mu}}_g, \hat{\boldsymbol{\Sigma}}_g\}$ for each group g. Taking these parameters as given, and using them to build the path of beliefs for each worker as described above, the NFXP estimator searches for the dynamic-choice parameters $\boldsymbol{\Theta}_{\text{dyn}} = \big(\{\eta_{jg}\}_{j\neq 0,\, g}, \kappa_0\big)$ that maximize the choice log-likelihood (85), holding the discount factor β and the taste-shock scale ϱ fixed. For each candidate $\boldsymbol{\Theta}^{(k)}_{\text{dyn}}$ proposed by the outer optimizer, the procedure executes two steps in sequence.

Step 1 — Inner fixed point. Solve the integrated-value Bellman recursion

$$\boldsymbol{\Phi} \;=\; \mathcal{F}(\boldsymbol{\Phi}) \;\equiv\; \mathbf{T}'\,\widetilde{\mathbf{W}}(\boldsymbol{\Phi})',$$

i.e. (83), by successive substitution on the Smolyak coefficient matrix $\boldsymbol{\Phi}$, returning the converged coefficients $\boldsymbol{\Phi}^\star\big(\boldsymbol{\Theta}^{(k)}_{\text{dyn}}\big)$ via the algorithm of Section D.2.2.

Step 2 — Choice likelihood. Given $\boldsymbol{\Phi}^\star\big(\boldsymbol{\Theta}^{(k)}_{\text{dyn}}\big)$, evaluate the CCPs (84) at the observed worker states $\{\boldsymbol{x}_{it}\}$ using the approximated value function, and compute the choice log-likelihood $\ell\big(\boldsymbol{\Theta}^{(k)}_{\text{dyn}}; \hat{\boldsymbol{\Theta}}_{\text{wage}}\big)$ in (85).

These two steps must be computed for each likelihood evaluation. In the practical implementation, we use L-BFGS-B with JAX automatic differentiation, see Section D.2.6) uses ℓ and its gradient to propose the next candidate $\boldsymbol{\Theta}^{(k+1)}_{\text{dyn}}$, and Steps 1–2 are repeated. Iteration continues until $\boldsymbol{\Theta}^{(k)}_{\text{dyn}}$ converges.

D.2.6 Differentiable NFXP Implementation (JAX)

The estimation routine is implemented in JAX and runs on the GPU. For a given vector of dynamic choice parametersΘ_{dyn}, I solve the inner fixed point $\Phi^\star$ by iterating the Bellman map $\mathcal{F}$ until successive iterations change by less than 10^{-8}. I then take the gradient of the likelihood by implicitly differentiating the fixed-point condition $\Phi^\star = \mathcal{F}(\Phi^\star)$; this keeps the gradient exact while remaining fast and light on memory. The dynamic choice parameters are updated with L-BFGS-B using these gradients. A full evaluation, the inner solve plus the choice likelihood over $N \simeq 2.4$ million observations, takes about 31 minutes on a single NVIDIA H100 GPU node.

D.2.7 Two-Step Estimation Summary

The complete estimation procedure pairs the EM algorithm of Appendix D.1 (Stage 1, wage and skill block) with the NFXP procedure described above (Stage 2, dynamic-choice block). The algorithm box below summarizes the full pipeline.

Algorithm: Two-Step Estimation (EM + NFXP)

Inputs.

- Worker–occupation–time panel $\{(w_{it}, d_{it}, \boldsymbol{\tau}_{it}, d_{i,t-1}, g(i))\}$.
- Anchor occupations j_C, j_M; task weights $\{\omega_j\}_{j\in\mathcal{J}}$.
- Initialization rules and tolerances for both stages.

Outputs. $\hat{\Theta}_{\text{wage}},\ \{\hat{\mu}_g, \hat{\Sigma}_g\}_{g\in\mathcal{G}},\ \hat{\Theta}_{\text{dyn}}$.

Stage 1 — EM [Appendix D.1].

1. Initialize $(\Theta^{(0)}_{\text{wage}}, \mu_g^{(0)}, \Sigma_g^{(0)})$.
2. Repeat for $q = 0, 1, \ldots$ until $\text{MSD}\big(\Theta^{(q)}\big) < 10^{-9}$:
   - **E-step** [(61), (62)]: per-worker posteriors $(\hat{\zeta}_i^{(q)}, \hat{V}_i^{(q)})$.
   - **M-step** [(68), (69), (65)]: closed-form linear systems per occupation; group prior moments.
3. **Output:** $\hat{\Theta}_{\text{wage}},\ \{\hat{\mu}_g, \hat{\Sigma}_g\}_{g\in\mathcal{G}},\ \{(\hat{\zeta}_i, \hat{V}_i)\}_{i=1}^N$.

Stage 2 — NFXP on $\Theta_{\text{dyn}} = (\{\eta_{jg}\}, \kappa_0)$, with β and ϱ fixed [Sections D.2.2, D.2.4].

1. Build per-worker state path $\{x_{it}\}$: entry belief $(\hat{\mu}_{g(i)}, \hat{\Sigma}_{g(i)})$; observed-row belief from the forward Kalman recursion on wage residuals at the first-stage estimates; tenures $\boldsymbol{\tau}_{it}$.
2. Initialize $\Theta^{(0)}_{\text{dyn}}$ with normalization $\eta_{0,g} = 0$ for each $g \in \mathcal{G}$.
3. Repeat for $k = 0, 1, \ldots$ until L-BFGS-B convergence (ftol $< 10^{-6}$, gtol $< 10^{-5}$):
   - **Inner solve**: fixed point $\Phi^\star\big(\Theta^{(k)}_{\text{dyn}}\big)$ via the algorithm in Section D.2.2.
   - **CCPs** [(84)]: evaluate CCPs at panel state variables $\{x_{it}\}$ using the approximated value function.
   - **Likelihood** [(85)]: $\ell\big(\Theta^{(k)}_{\text{dyn}}\big)$.
   - **Outer step**: L-BFGS-B update with JAX autodiff gradients .
4. **Output:** $\hat{\Theta}_{\text{dyn}}$.

Return $\big(\hat{\Theta}_{\text{wage}},\ \{\hat{\mu}_g, \hat{\Sigma}_g\}_{g\in\mathcal{G}},\ \hat{\Theta}_{\text{dyn}}\big)$.

E Appendix: Data and Additional Figures

Samples and Selection Criteria. The empirical analysis and structural estimation rely on two related samples constructed from *Quadros de Pessoal*.

The *large sample* follows the standard selection criteria of Rute Cardoso and Portela (2009), Card et al. (2016), and Fonseca et al. (2018). It is used to document the overarching empirical patterns of occupational mobility at the 3-digit occupation level, providing reliable moments that serve as external validation targets for the structural model.

The *estimation sample* is designed specifically for structural estimation using the lifecycle model. Starting from the large sample, I apply the following additional selection criteria: (i) I keep only workers first observed after 1995, with age at entry below 30 and zero initial experience—ensuring I observe workers from the beginning of their careers; (ii) I retain each worker's first consecutive employment spell (allowing for the 2001 data gap), so that experience can be tracked precisely from entry; (iii) I require at least 5 distinct years of observation per worker; and (iv) I trim extreme hourly wages at the 1st and 99th percentiles.

Because the estimation sample focuses on young entrants, it has a different age and experience composition than the large sample. Table 4 compares the two samples side-by-side. Crucially, the empirical patterns of occupational mobility are qualitatively consistent across the two samples. The descriptive empirical analysis and the structural model comparison use related but not identical wage objects. In the empirical section, I use residualized wages in the large sample to isolate the component of within-occupation wage variation driven by unobserved fit or comparative advantage. By contrast, the structural model is estimated on a more restricted sample and a reduced occupational state space, so it is evaluated against moments based on raw within-cluster wage ranks in the estimation sample, which constitute the natural model-consistent empirical counterpart.

Table 4: Comparison of the Large Sample and the Estimation Sample

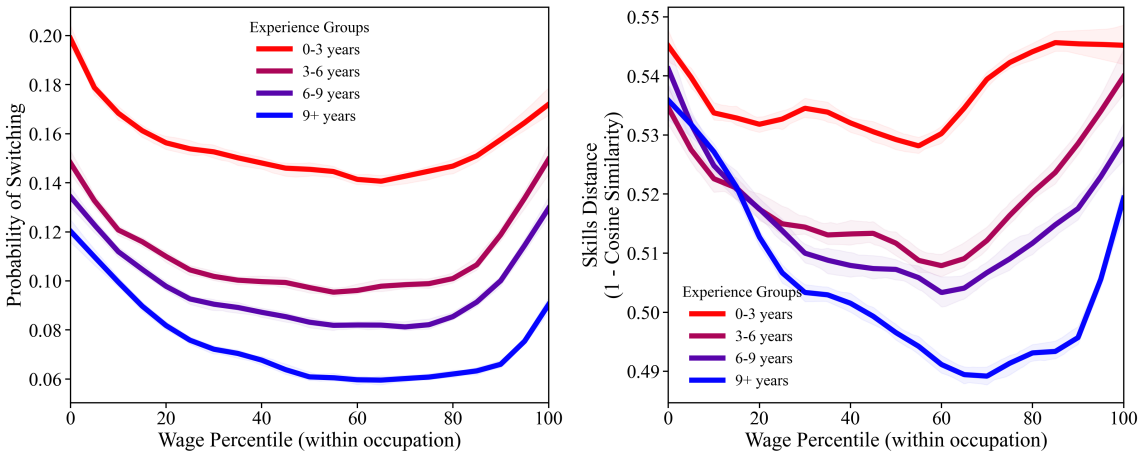
Large Sample		Estimation Sample	
Statistic	**Value**	**Statistic**	**Value**
Total observations	29,428,668	Total observations	2,316,035
Number of unique workers	4,425,614	Number of unique workers	283,765
Number of unique firms	618,009	Number of unique firms	131,626
Number of unique occupations	103	Number of unique occupations	101
Years covered	1995–2019	Years covered	1995–2019
Average real hourly wage (€)	6.32	Average real hourly wage (€)	5.92
Average occupational tenure (years)	4.22	Average occupational tenure (years)	6.11
Average total experience (years)	11.77	Average total experience (years)	7.29
Gender & Education (% share)		**Gender & Education (% share)**	
Men	56.90	Men	55.91
Women	43.10	Women	44.09
Low education (1–2)	61.61	Low education (1–2)	33.43
Medium education (3)	23.53	Medium education (3)	33.30
High education (4–5)	14.85	High education (4–5)	33.26

Notes: The large sample is designed to maximize representativeness and the number of observations for descriptive facts. The estimation sample is constructed from the large sample by applying the following additional restrictions: (i) entry age below 30 with zero initial experience; (ii) first consecutive employment spell only (the 2001 data gap is allowed); (iii) at least 5 distinct years observed per worker; (iv) hourly wages trimmed at the 1st and 99th percentiles. Low education = Ensino Básico; Medium = Ensino Secundário; High = Ensino Superior (Bachelor or more). Occupational tenure (end-of-spell) and total experience are computed as the maximum per worker for both samples.

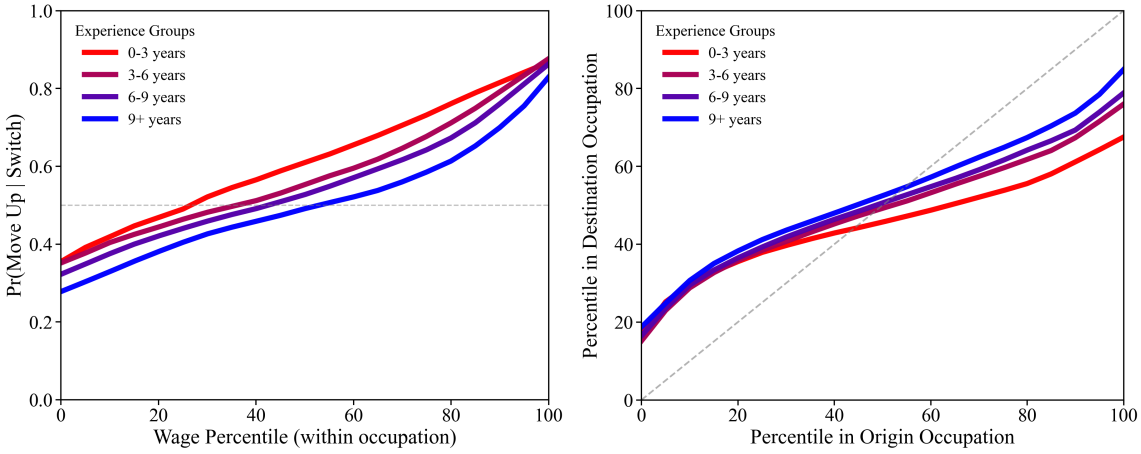
Task and Skill Classifications. The first classification, developed by Mihaylov and Tijdens (2019), classifies occupations into non-routine abstract, non-routine interactive, non-routine manual, routine cognitive, and routine manual tasks. The occupational code used in this paper is based on the ISCO88, embedded in the Portuguese code CNP1994. Starting in 2010, CPP2010 was introduced in *Quadros de Pessoal*. This code is based on ISCO08, which Mihaylov and Tijdens (2019) use. To ensure consistency across the classification change, I develop a mapping algorithm that combines the official crosswalk between CNP1994 and CPP2010 with empirical mappings based on workers who remained at the same firm during the 2009–2010 transition. This hybrid approach minimizes misclassification of occupational switches around the code change. I abstract from the routine dimension and focus on two broad categories: cognitive and manual skills. Cognitive skills sum non-routine abstract, routine cognitive, and non-routine interactive tasks; manual skills sum non-routine manual and routine manual tasks. I do this grouping because these are the most relevant dimensions to understand occupational mobility directionality, and in the quantitative section I estimate the model with two skill dimensions for tractability, ensuring a cleaner mapping between data and model.

Robustness: Raw Hourly Wages. The empirical analysis in Section 3 uses residualized wages as a proxy for unobserved worker ability or match quality. Figure 24 replicates the main empirical facts from that section using raw real hourly wages to construct the within-occupation percentiles. The patterns—the pseudo U-shape in switching, the directionality in

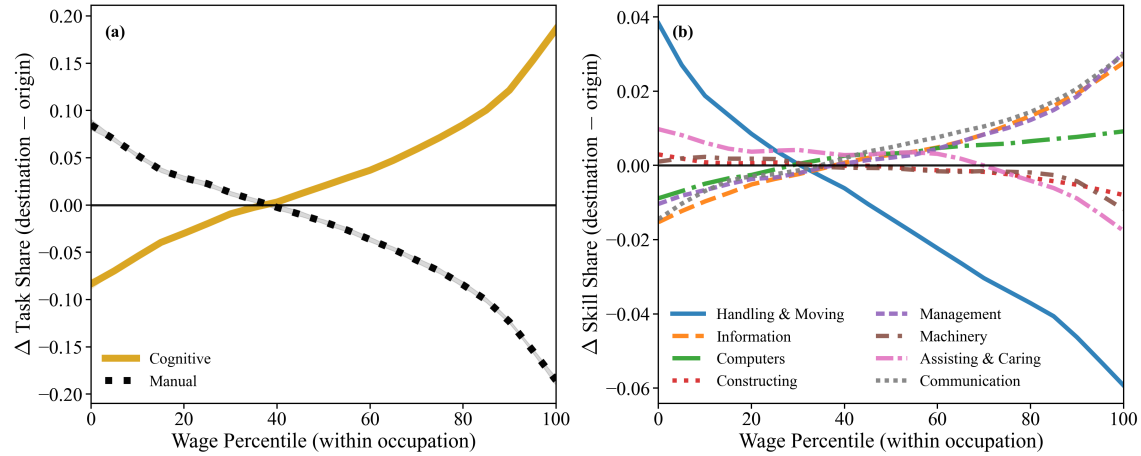
both wages and skills, and the task share changes—remain qualitatively robust when using raw wages instead of residualized wages.



(a) Probability of Switching & Skills Distance



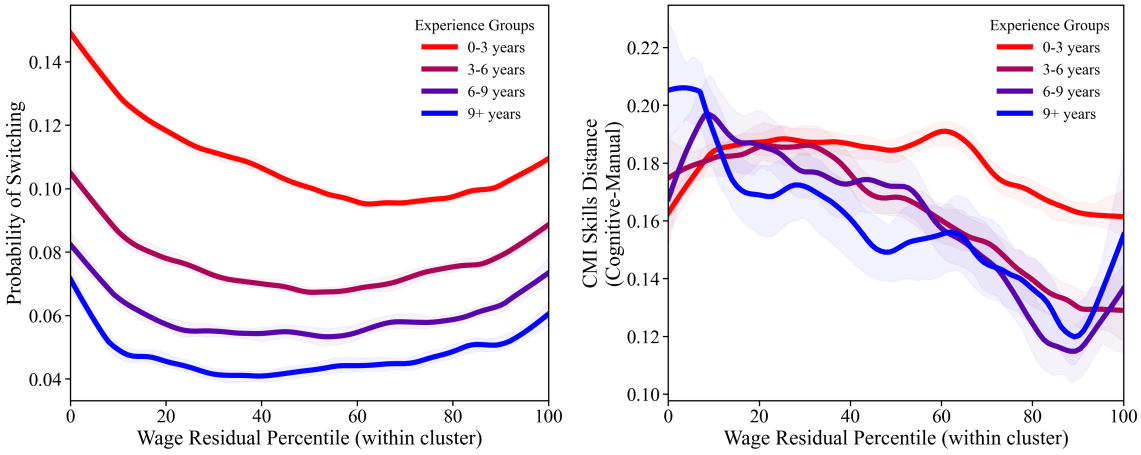
(b) Directionality by Experience



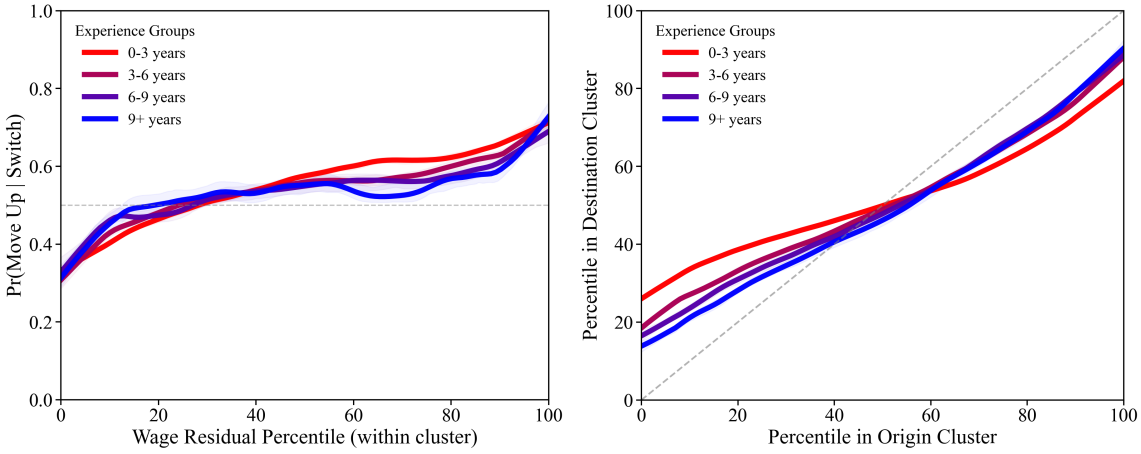
(c) Task Content Directionality (Cognitive/Manual and ESCO)

Figure 24: Robustness: Raw Wages. The figure replicates the main empirical facts using raw rather than residualized wages. Panel (a) shows the probability of switching occupations and skills distance by experience. Panel (b) shows directionality and rank-to-rank by experience group. Panel (c) shows the change in cognitive/manual and ESCO skill shares conditional on switching.

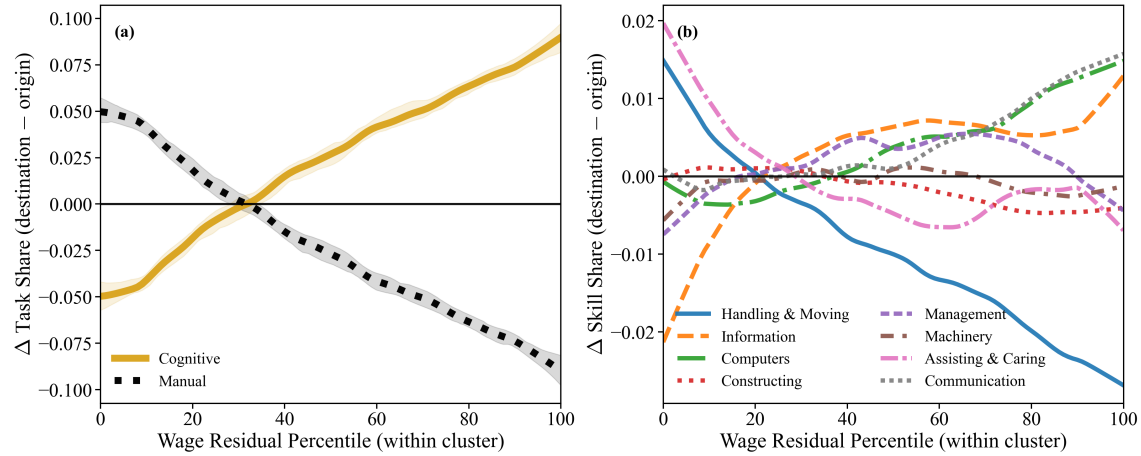
Robustness: Estimation Sample (Clusters). Furthermore, Figure 25 reproduces the main empirical analysis using the smaller estimation sample, where occupations are aggregated into ten task-based clusters. As discussed in Section 3, this aggregation mechanically lowers mobility levels, and the smaller sample size introduces more noise. Nevertheless, the figure shows that the estimation sample still broadly matches the original empirical patterns documented in the large sample, including the U-shape in switching and the dual directionality in wages and task content. Note that for this robustness check, task distance (Panel a, right plot) is computed using the Cognitive-Manual-Interpersonal (CMI) task space rather than the full ESCO skill space.



(a) Probability of Switching & Task Distance



(b) Directionality by Experience



(c) Task Content Directionality (Cognitive/Manual and ESCO)

Figure 25: Robustness: Estimation Sample (Clusters). The figure replicates the main empirical facts using the aggregated estimation sample and CMI task space. Panel (a) shows switching probabilities and CMI task distance. Panel (b) shows directionality and rank-to-rank by experience group. Panel (c) shows task content directionality.

Robustness: Skill Granularity and Distance Measure. Figure 26 examines the robustness of the skills distance patterns to the level of skill aggregation. The figure reports mean (LOWESS-smoothed) and median task distance by residualized within-occupation wage percentile and experience group, using the cosine distance measure of Gathmann and Schönberg (2010). The four rows correspond to the 8-dimension ESCO classification (Matrix 1.3), the 74-dimension ESCO classification (Matrix 2.3), the full 296-dimension ESCO classification (Matrix 3.3), and the two-dimensional Cognitive–Manual space. The declining pattern of task distance with match quality is robust across all granularity levels, although the range of variation narrows at higher dimensions due to the concentration of cosine distance in high-dimensional spaces.

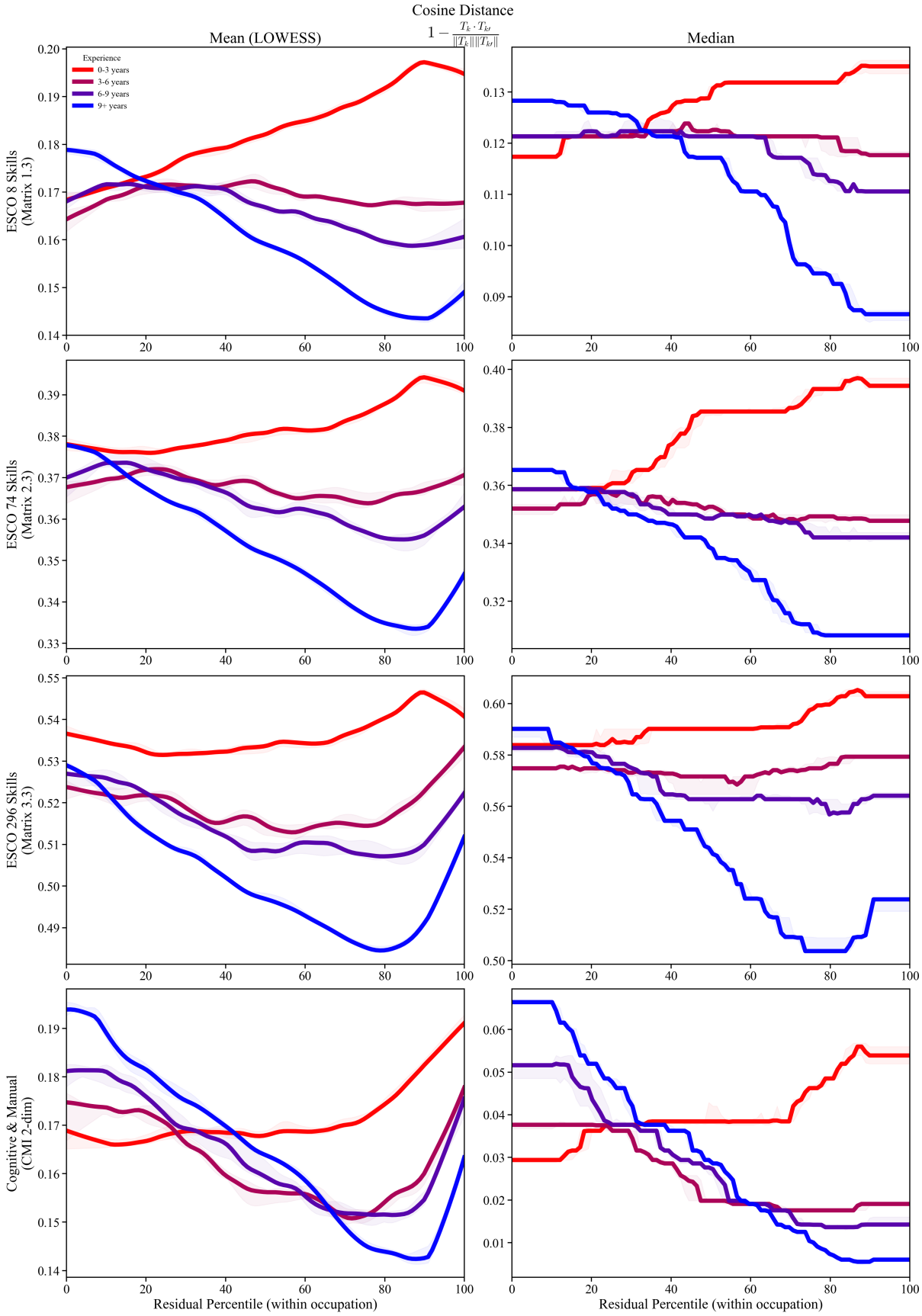
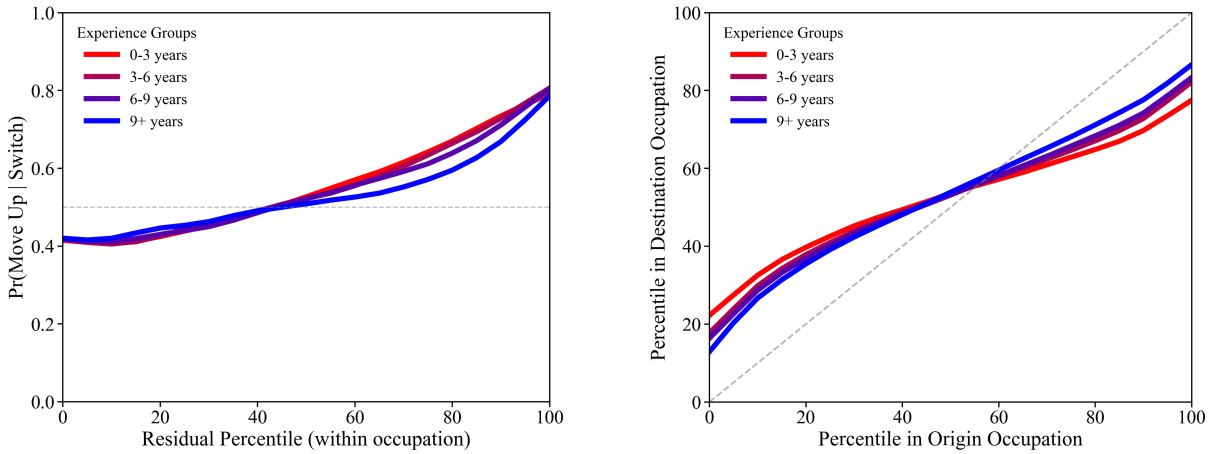


Figure 26: Robustness: Skill Granularity. Task distance (cosine distance, $1 - \text{Sim}(j, j')$, as in Gathmann and Schönberg 2010) by residualized within-occupation wage percentile and experience group. Left column: mean (LOWESS). Right column: median. Rows from top to bottom: ESCO 8 dimensions, ESCO 74 dimensions, ESCO 296 dimensions, Cognitive–Manual (2 dimensions).

Directionality and Rank-to-Rank Mobility by Experience. Figure 27 reports the directionality of occupational mobility documented in Section 3. Panel 27(a) shows that the probability of moving to a higher-wage occupation conditional on switching rises monotonically with the worker's residualized within-occupation wage percentile, with the slope flattening as experience accumulates. Panel 27(b) shows the corresponding rank-to-rank pattern: low-ranked workers in the origin occupation tend to land in higher relative positions at destination, while high-ranked workers fall toward the middle of the destination distribution. Both patterns weaken with experience, consistent with workers learning their comparative advantage and progressively sorting into better-matched occupations.



(a) Directionality by Experience

(b) Rank-to-Rank by Experience

Figure 27: *Panel 27(a)*: Probability of moving to a higher-wage occupation conditional on switching (vertical axis) as a function of residualized within-occupation wage percentile (horizontal axis), for different experience groups. *Panel 27(b)*: Percentile position in destination occupation (vertical axis) against percentile position in origin occupation (horizontal axis) for occupational switchers, by experience group.

Skills Distance: Summary Statistics. Table 5 reports summary statistics for the skills distance measure defined in equation (1). Panel A shows the distribution of pairwise distances across all 115 three-digit ISCO-88 occupations mapped to the ESCO skill classification (6,555 unique pairs). The mean pairwise distance is 0.71 (std. 0.19), indicating that the average pair of occupations has substantially different skill requirements. Panel B restricts to the 2.13 million actual occupation switches observed in the data. The mean distance for switchers is 0.52 (std. 0.22, median 0.58), well below the unconditional mean—confirming that workers tend to switch to occupations with similar skill profiles. The 10th percentile of switcher distances is 0.21, while the 90th percentile is 0.80.

The most distant occupation pair is glass and ceramics kiln operators and shoe cleaning and street services workers ($d = 0.99$). The most similar pair (excluding identical occupations) is stall and market salespersons and street food vendors ($d = 0.04$). Women switch to slightly more distant occupations on average (mean $d = 0.53$) than men (mean $d = 0.51$). Panel C reports the three most common occupational switches by education group. The most popular move for low-educated workers (education levels 1–2) is between mining and quarrying laborers and building frame workers ($d = 0.21$; 23,000 transitions), while the most common move for medium-educated workers (level 3) is between accounting clerks and bookkeepers ($d = 0.36$;

11,000 transitions). For high-educated workers (levels 4–5), the most frequent switches are between accounting clerks and bookkeepers ($d = 0.36$) and between physical science technicians and architects and engineers ($d = 0.43$). Panels D and E decompose the most common switches by gender and by gender $\times$ education. Most frequent move for men is between laborers and building frame workers, while for women the most frequent move is between accounting clerks and bookkeepers.

Table 5: Skills Distance Summary Statistics. Cosine distance $d(j, j') = 1 - \text{Sim}(j, j')$ computed using the 296-dimension ESCO skill classification. ISCO-88 three-digit occupation codes.

	Mean	Std	Median	p10	p25	p75	p90
Panel A: All occupation pairs (6,555 pairs)							
All pairs	0.71	0.19	0.76	0.44	0.63	0.85	0.91
Panel B: Actual switchers (2,127,097 transitions)							
All	0.52	0.22	0.58	0.21	0.33	0.69	0.80
Men	0.51	0.22	0.57	0.21	0.33	0.69	0.79
Women	0.53	0.22	0.58	0.22	0.34	0.69	0.81

Panel C: Most common occupation switch by group

Group	Occupation switch	d	N
Low educ. (1–2)	Labourers $\rightarrow$ Building frame workers	0.21	23,070
Medium educ. (3)	Accounting clerks $\rightarrow$ Bookkeepers	0.36	10,987
High educ. (4–5)	Accounting clerks $\rightarrow$ Bookkeepers	0.36	9,364
Men	Labourers $\rightarrow$ Building frame workers	0.21	24,311
Women	Accounting clerks $\rightarrow$ Bookkeepers	0.36	16,774
Men, Low	Labourers $\rightarrow$ Building frame workers	0.21	23,016
Men, Medium	Finance dealers $\rightarrow$ Shop salespersons	0.59	7,174
Men, High	Phys. science tech. $\rightarrow$ Engineers	0.43	5,224
Women, Low	Domestic helpers $\rightarrow$ Cooks	0.33	11,301
Women, Medium	Accounting clerks $\rightarrow$ Bookkeepers	0.36	7,570
Women, High	Accounting clerks $\rightarrow$ Bookkeepers	0.36	6,374

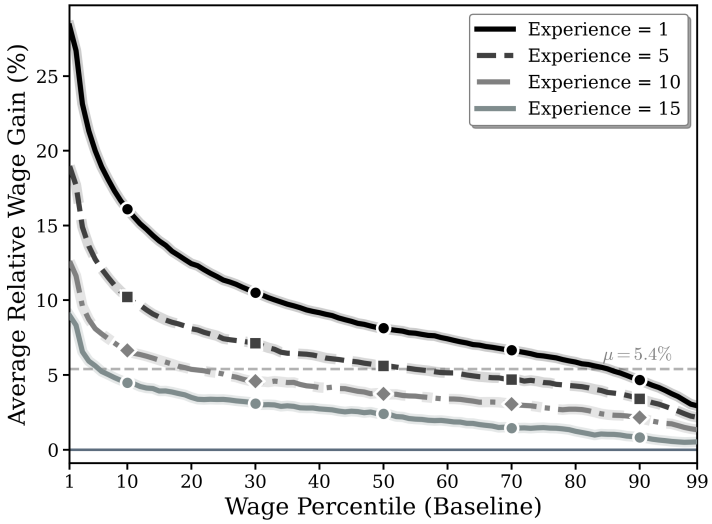


Figure 28: Average wage gain from removing information frictions by position in the overall wage distribution (horizontal axis: wage percentile; vertical axis: percent wage gain), with separate lines for different experience levels.

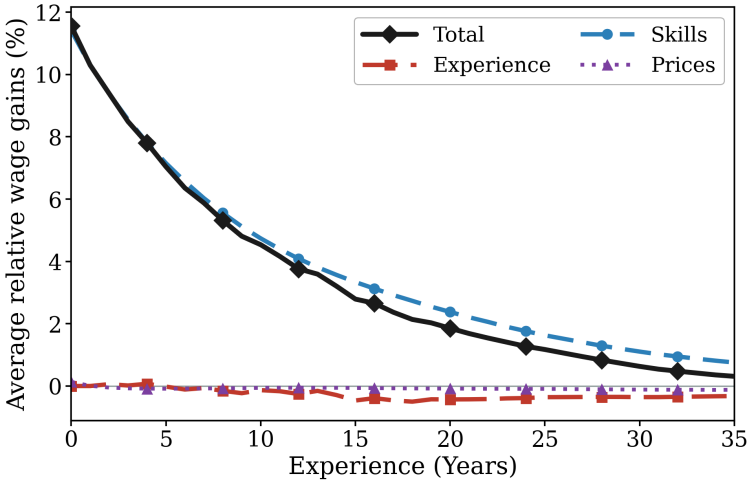
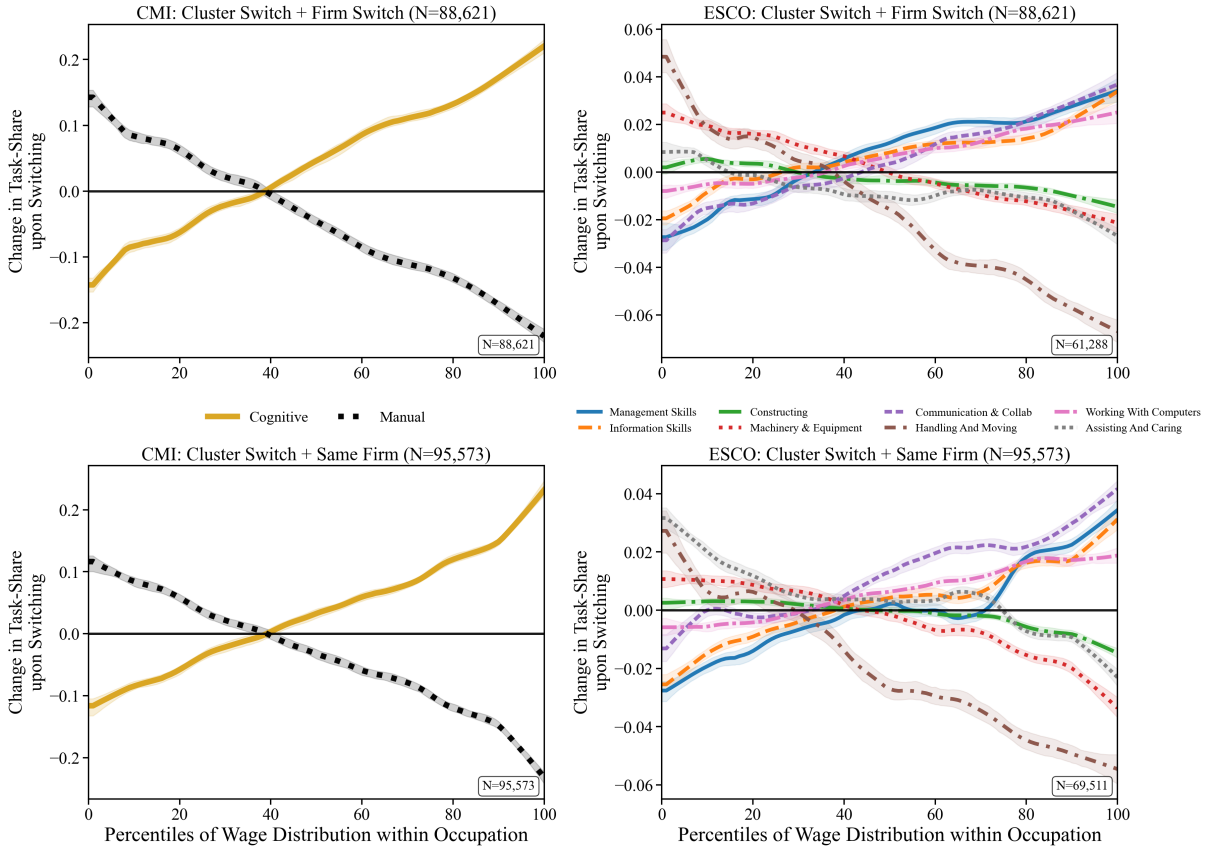


Figure 29: Decomposition of the lifecycle wage gain from removing information frictions (Figure 11, Panel (a)) into channels: Total (black), Skills (blue), Experience (red), and Prices (purple). Horizontal axis: years of potential experience; vertical axis: average relative wage gain (percent).

We construct our measure of earnings by summing a worker's base salary plus any regular earnings supplements and dividing the amount by his or her normal hours of work; this yields the same "straight time" hourly wage used in Card et al. (2016). We obtain real hourly wages by deflating our "straight time" using CPI Inflation with 1995 as base year. From the initial dataset we eliminate entries if the worker ID is reported multiple times per year; for each worker we drop the entire history if sex changes over time, age is inconsistent, or the worker is not aged between 20 and 55. We also drop observations if the hiring date for any job is missing or inconsistent across observations. If the individual is observed in two consecutive years at different firms, but the hiring date for the second job is the same as the hiring date for the first job, we drop the entire history for a worker. We also drop individuals who have missing education and we drop workers with education decreasing over time. The sample is further restricted



(a) Change in task content for occupation switchers using different task classifications

Figure 30: Change in task shares (vertical axis) by within-occupation wage percentile (horizontal axis) for occupation switchers, comparing those who stay at the same firm versus those who switch firms. Based on the cognitive/manual task classification and ESCO skill measures.

to full-time workers (30h per week or 130h per month) who earn at least the minimum wage. Following Fonseca et al. (2018), we drop workers in occupations and industries associated with the primary sector, given that the survey is optional for those. We also drop workers in public administration and defense[32] and drop a worker history if he/she has been in sectors with less than 100 workers reported or in occupations with less than 100 workers reported in a given year. As we are also interested in occupational mobility within firms, we drop workers in firms with less than 3 employees. Regarding industry codes, the CAE Rev. 3 (Classificação Portuguesa de Actividades Económicas, Revisão 3) replaced the CAE Rev 2.1 (Classificação Portuguesa de Actividades Económicas, Revisão 2.1) used since 1995[33]. As for occupations, the National Classification of Occupations 1994 (CNP/94) was used from 1994 until 2009 and was then replaced by the Portuguese Classification of Occupations 2010 (CPP/2010) that is embedded in the ISCO08 international classifications. I develop an algorithm to minimize the number

[32]These correspond to industries with NACE code A,B,C,L and Q before 2007, when the CAE rev2-rev2.1 was used in Portugal, and to the industries A,B,O and U after the CAE rev3 industry classification was introduced in Portugal, starting from 2007. Occupations dropped belong to the 2 digits group 11,61 and 92 for the CNP 1994 and 11, 61, 62, 63 and 92 for the CPP 2010.

[33]Between 2002 and 2003 the system of industry classification used for Quadros de Pessoal changed from CAE Rev 2 to CAE Rev 2.1. At 2 digits level, only two industries were introduced: 96 Activ. de produção de bens pelas famílias p/ uso próprio; 97 Activ. de produção de serviços pelas famílias p/ uso próprio.

of occupational switch misclassifications in the years around the occupational code change.

E.1 Occupation clustering

Let $i = 1, \ldots, n$ index three-digit ISCO occupations and let $c \in \{1, \ldots, K\}^n$ denote a partition. Each occupation has a standardized two-dimensional vector defining the task-content $x_i \in \mathbb{R}^2$ summarizing its cognitive and manual content, a mass weight ω_i equal to its share of person-periods in the panel, and pairwise mobility flow F_{ij} between different occupations measured in the large sample. The partition minimizes

$$L(c) \;=\; \underbrace{\sum_i \omega_i \, \|x_i - \mu_{c_i}\|^2}_{\text{task-content dispersion}} \;+\; \lambda \underbrace{\sum_{i,j} F_{ij} \, \mathbf{1}\{c_i = c_j\}}_{\text{absorbed flow}},$$

where μ_c is the ω-weighted centroid of cluster c. The first term ensures that occupations with similar task content together are clustered together; the second term penalizes observed worker flows are absorbed into a cluster, so that aggregated cluster-to-cluster transitions remain informative with respect to the original occupation-to-occupation mobility graph. In choosing the relative importance of the two measure for this loss function, I choose a $\lambda = 10$ considering the trade-off between within-cluster dispersion and cross-cluster mobility share. Importantly, there are two feasibility constraints that I impose while minimizing the above loss function: the cluster-level mobility graph must remain connected, and every non-anchor cluster must retain at least 5% of total mass. To make sure the interpretation and definition of the two anchor clusters has proper economic meaning I fix a subset of occupations which must be contained in each. These occupations are chosen based on their extreme task content, and are either fully manual or fully cognitive. The procedure is run from several starting points (ω-weighted k-means initializations) and the best partition is retained. Results with respect to the two metrics used in the loss function are plott in Figure 31. As displayed by the vertical dashed line, $K = 8$ is selected balancing parsimony in the number of cluster considered, within-cluster skill/task content dispersion and the cross-cluster mobility share. For larger K the measure improves but the number of observations included in each cluster cross teh 5% threshold, with potential implications for the identification of some of the cluster-specific component of the model. The clustering is implemented using the two different skill requirement and task content classification already described in the main body of the paper: an ESCO variant in which each granular ESCO skill is labeled as either cognitive or manual, and (C_i, M_i) is occupation i's share of cognitive vs. manual skills aggregated from those labels using its ESCO skill mix, and a task-based variant in which $(C_i^{\text{MT}}, M_i^{\text{MT}})$ is computed from the routine/non-routine task shares of Mihaylov and Tijdens (2019). Cluster labels are deterministic: the cognitive anchor is cluster 1, the manual anchor is cluster K, and the remaining clusters are ordered by ascending mean cognitive share.

In the context of a quantitative model of occupational mobility, the level of occupational clustering has first-order consequences for both estimation and the quantitative importance of the model's mechanisms, so it is important to discuss it explicitly. Grouping occupations helps identification by increasing the amount of within-occupation variation available to estimate task-specific parameters, especially returns to task tenure: if occupations are defined too gran-

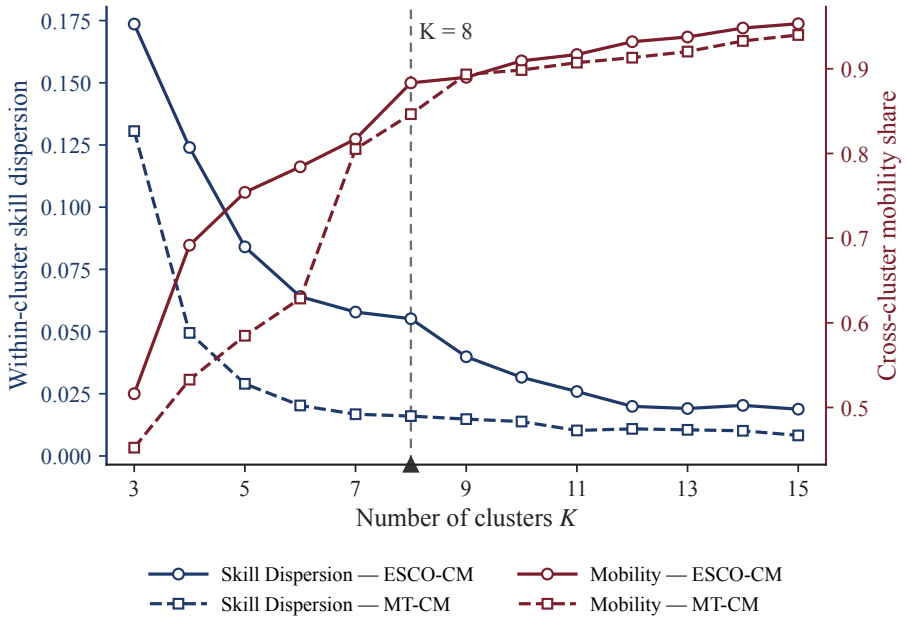


Figure 31: Headline diagnostic for the K-selection step of the anchored skill+mobility-penalty clustering. The horizontal axis is the number of clusters K on the 3–15 grid (with $\lambda_{\text{user}} = 10$ fixed at the empirical knee). Curves report the ω-weighted within-cluster skill dispersion and the cross-cluster mobility share at the L-minimizing partition, separately for the ESCO-based cognitive/manual shares (C_i, M_i) and the task-based cognitive/manual shares $(C_i^{\text{MT}}, M_i^{\text{MT}})$.

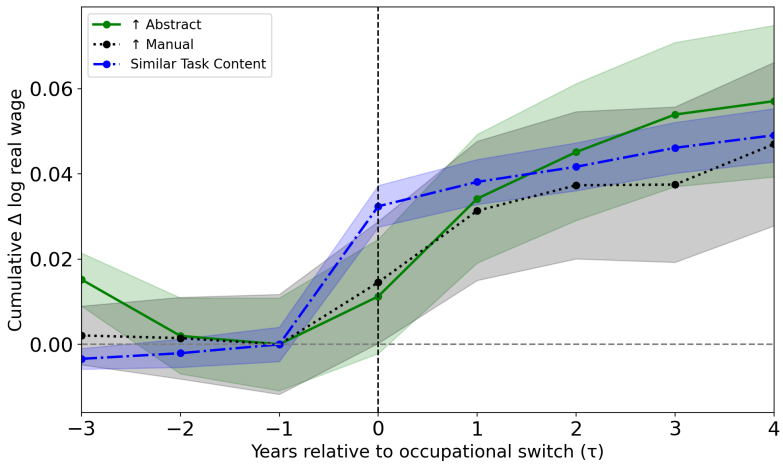
ularly, there may be too little variation in accumulated experience along each task dimension k within a given occupation to precisely identify the coefficients on task-specific human capital. Excessive granularity also creates sparse mobility patterns, weakens the connectivity of the occupational network, and produces small cells in which occupation-specific prices, skill loadings, experience returns, and residual variances are all estimated imprecisely; moreover, very fine occupations may be observationally distinct yet economically similar, adding parameters without adding much independent identifying variation. At the same time, in a discrete choice framework, clustering mechanically changes the worker's feasible choice set: when the model includes a very large number of occupations, even small shocks to wages, beliefs, or tastes can generate frequent switching across narrowly defined jobs, whereas broader clusters smooth these margins and force the model to interpret mobility as movement across economically meaningful alternatives.

E.2 Empirical evidence on wage changes upon occupational switches

E.2.1 Event–study with stayers as control group

This section estimates dynamic wage effects of occupational switches using a compact event–study in *first differences* with *stayers* as the reference group. Let t index calendar years and i workers. Define the log real hourly wage w_{it} and the one–year change

$$\Delta \log w_{it} \equiv \log w_{it} - \log w_{it-1}.$$



(a) Cumulative wage changes around occupational switches

Figure 32: Cumulative Δ log real wage (vertical axis) by event time relative to switch (horizontal axis: years from −3 to +4), normalized at $\tau = -1$. Lines: ↑Cognitive (green, solid), ↑Manual (black, dotted), Similar Task Content (blue, dash-dot). Shaded areas = 95% confidence bands clustered by worker.

For each worker–year t, mark whether the worker makes a *consecutive* occupation switch between t and $t+1$ and classify the switch by the dominant increase in the occupation's task indexes. We combine cognitive and interpersonal task content into a single "Cognitive" measure, yielding: (i) ↑Cognitive, (ii) ↑Manual; all other switches are coded as "Similar Task Content". Denote these mutually exclusive indicators by $D_{it}^{g} \in \{0,1\}$ for $g \in \{\text{Cog}, \text{Man}, \text{Other}\}$. Worker–years with no switch are *stayers* and form the omitted category.

Event time is anchored at the year t in which the (consecutive) switch occurs, so that outcomes at relative time τ are $\Delta \log w_{it+\tau}$. For each $\tau \in \{-3, -2, -1, 0, 1, 2, 3, 4\}$ we run a separate OLS:

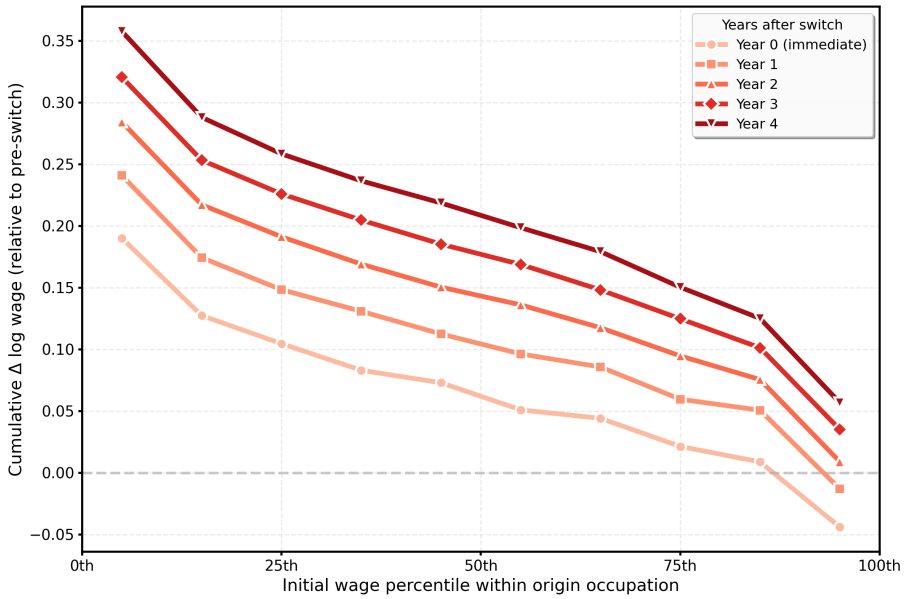
$$\underbrace{\left(\Delta \log w_{it+\tau}\right)^{\circ}}_{\text{demeaned by year } t} = \alpha_\tau + \sum_g \beta_{\tau g} D_{it}^{g} + \theta_\tau' X_{it}^{\circ} + \varepsilon_{it\tau}, \tag{86}$$

where X_{it} includes education fixed effects (baseline = level 1), experience and experience^2. A $^{\circ}$ superscript indicates demeaning by the calendar–year mean of the corresponding regressor (and of the dependent variable) computed in year t; this absorbs year fixed effects. Standard errors are clustered by worker.

The coefficients $\beta_{\tau g}$ measure, at each relative time τ, the mean difference in annual wage growth between type–g switchers and same–year stayers, conditional on controls. To present wage levels around the event we cumulate the estimated growth effects across relative time and normalize at $\tau = -1$, similar to Heise and Porzio (2022):

$$C_g(k) = \sum_{\tau=-3}^{k} \hat{\beta}_{\tau g} - \sum_{\tau=-3}^{-1} \hat{\beta}_{\tau g}, \qquad k \in \{-3, \ldots, 4\},$$

so that $C_g(-1) = 0$. Shaded 95% confidence bands in the figure are obtained by cumulating worker–clustered standard errors (normal approximation).



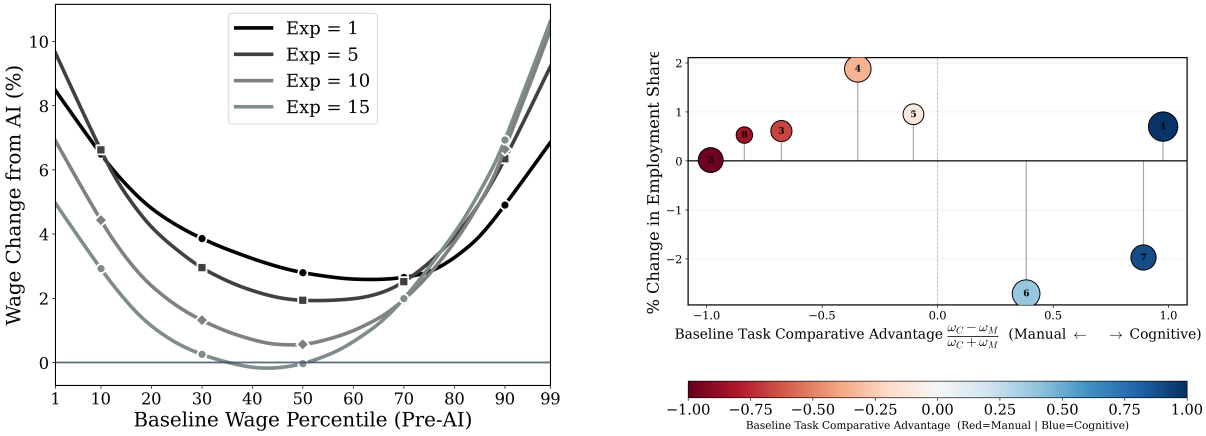
(a) Cumulative wage changes by initial wage percentile

Figure 33: Cumulative log wage change (vertical axis) by initial wage percentile within origin occupation (horizontal axis), for Years 0–4 after switching (separate lines). Workers at lower percentiles experience larger wage gains from switching.

E.2.2 Wage changes by initial position within occupation

Workers at different positions in their origin occupation's wage distribution experience markedly different wage trajectories upon switching. Figure 33 shows cumulative wage changes for switchers across the initial wage percentile distribution.

For each switcher, we calculate their wage percentile within their origin occupation (in the year before switching) and track cumulative log wage changes over the following four years. The plot reveals a strong negative relationship between the origin wage percentile within occupation and the wage change after switching. This pattern persists even after few years and suggests that occupational switching is particularly beneficial for workers who were poorly matched to their origin occupation.



(a) Wage gains from AI by baseline wage percentile

(b) Reallocation by task content of the occupation

Figure 34: *Panel (a)*: percent wage change from AI by baseline (pre-AI) wage percentile, for workers with different years of experience (separate lines). *Panel (b)*: percent change in each occupation cluster's employment share against its baseline task comparative advantage $(\omega_C - \omega_M)/(\omega_C + \omega_M)$ (manual to cognitive); marker color indicates the same comparative advantage and bubble size the baseline employment share. Employment flows toward manual-intensive clusters.

Appendix References

Althoff, L. and Reichardt, H. (2026), Task-specific technical change and comparative advantage, CESifo Working Paper 12403, CESifo.
URL: *https://ssrn.com/abstract=6084170*

Arcidiacono, P. and Miller, R. A. (2011), 'Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity', *Econometrica* **79**(6), 1823–1867.

Bunting, J., Diegert, P. and Maurel, A. (2024), Heterogeneity, uncertainty and learning: Semiparametric identification and estimation, Technical report, National Bureau of Economic Research.

Card, D., Cardoso, A. R. and Kline, P. (2016), 'Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women', *The Quarterly journal of economics* **131**(2), 633–686.

Fonseca, T., Lima, F. and Pereira, S. C. (2018), 'Job polarization, technological change and routinization: Evidence for portugal', *Labour Economics* **51**, 317–339.

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

Heise, S. and Porzio, T. (2022), Labor misallocation across firms and regions, Technical report, National Bureau of Economic Research.

Hincapié, A. (2020), 'Entrepreneurship over the life cycle: Where are the young entrepreneurs?', *International Economic Review* **61**(2), 617–681.

Hotz, V. J. and Miller, R. A. (1993), 'Conditional choice probabilities and the estimation of dynamic models', *Review of Economic Studies* **60**(3), 497–529.

James, J. (2012), Learning and occupational sorting, Working Paper 12-25, Federal Reserve Bank of Cleveland.

Malin, B. A., Krueger, D. and Kubler, F. (2011), 'Solving the multi-country real business cycle model using a smolyak-collocation method', *Journal of Economic Dynamics and Control* **35**(2), 229–239.

Mihaylov, E. and Tijdens, K. G. (2019), Measuring the routine and non-routine task content of 427 four-digit isco-08 occupations, Technical report, Tinbergen Institute discussion paper.

Rute Cardoso, A. and Portela, M. (2009), 'Micro foundations for wage flexibility: wage insurance at the firm level', *Scandinavian Journal of Economics* **111**(1), 29–50.