The Micro and Macro Implications of Multidimensional Skill Uncertainty*

Piero De Dominicis†

July 16, 2026

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 implies that removing information frictions about skills would raise aggregate output by 5.2%. These gains are highly unequal across workers: the worst-matched within occupation would earn over 20% more under full information, while the best-matched would earn about 3% more. I then evaluate how much of these gains a feasible information policy can deliver. Calibrated to the treatment effect of a field experiment, a skill assessment delivered at labor-market entry raises aggregate output by 0.91%, but its value decays sharply when the same assessment is delivered later in the career. Finally, job transformation induced by Large Language Model adoption raises long-run output, and raises the output gain available from removing information frictions from 5.2% to 5.9%.

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 and CEF 2026 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 more 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 returns are highest when the assessment is delivered before labor-market entry, when workers have not yet accumulated task-specific experience and switching costs are low: the same assessment delivered two years later produces gains that are more than halved, and within a decade shrinks to a small fraction of its at-entry value. 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 friction. 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) is the first to study the experimentation value of occupations, showing that occupational choice and wage growth in the NLSY are consistent with workers trading off information acquisition against wage growth; 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, differently from 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

[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.

shock.

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.

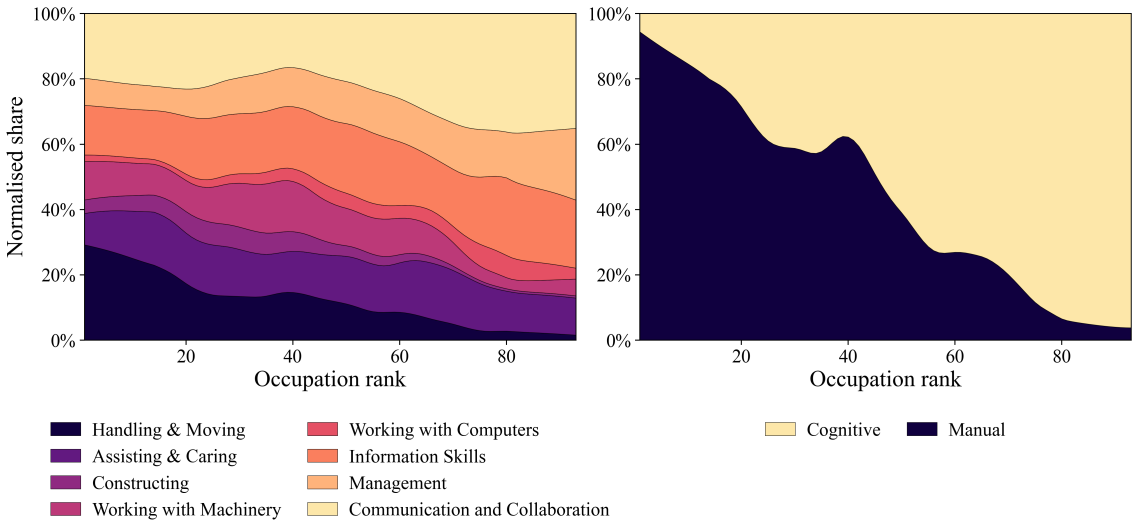
I use two related samples throughout the paper, each serving a different purpose. Detailed descriptions of the sample construction and their comparison are provided in Appendix E. The large sample is designed to maximize the number of observations and representativeness for descriptive facts and baseline 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

[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—recurring payments such as meal allowances, tenure-related premiums, and other regular components of remuneration—and is then divided by the worker's normal hours, defined as the contractual monthly hours fixed by the applicable collective agreement, firm rules, or individual contract. The measure is "straight-time" in the sense that both earnings and hours exclude overtime: overtime pay is omitted from the earnings concept and overtime hours from the hours base.

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 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:

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

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 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

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

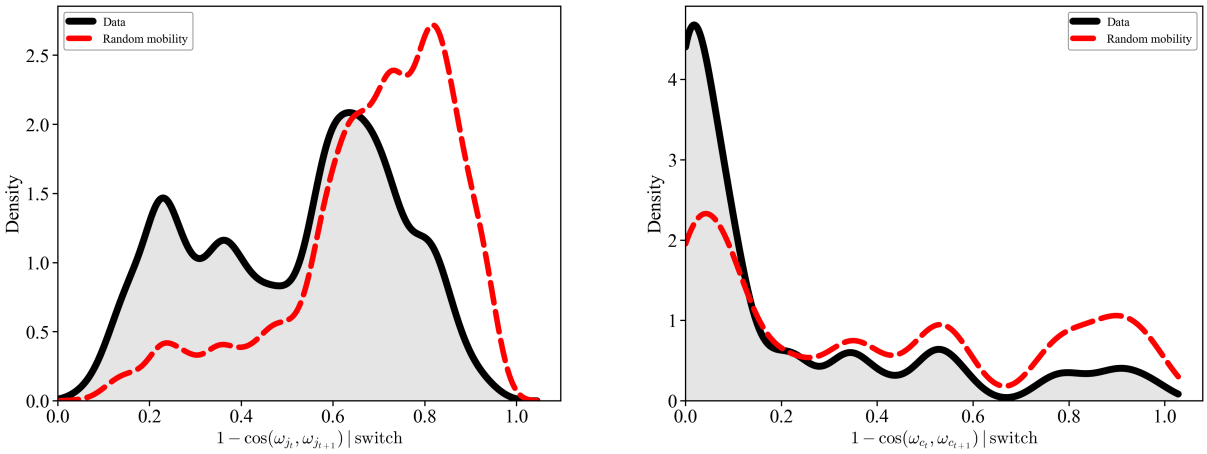
where $\boldsymbol{\omega}_j$ collects the skill weights of occupation j, and $\text{Sim}(j, j') \in [0, 1]$ equals 1 for identical and 0 for orthogonal skill mixes.[7] The mean skills distance among actual occupation switchers is 0.52, with a standard deviation of 0.22. The closest common move is between labourers 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 labourers 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); mean distance 0.52 vs. 0.68; Kolmogorov–Smirnov $p < 0.001$). 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:

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

[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 28 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)

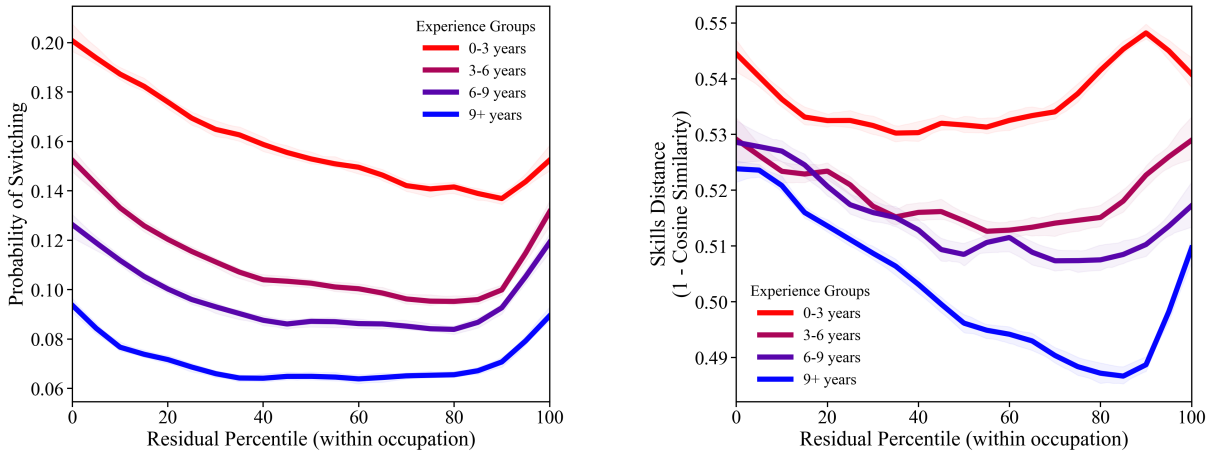
(b) Cognitive–Manual task distance (occupation clusters)

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.

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 $\widehat{\epsilon}_{it} = \widehat{\mu}_{ij} + \widehat{\nu}_{it}$. 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.

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

[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.



(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.

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.

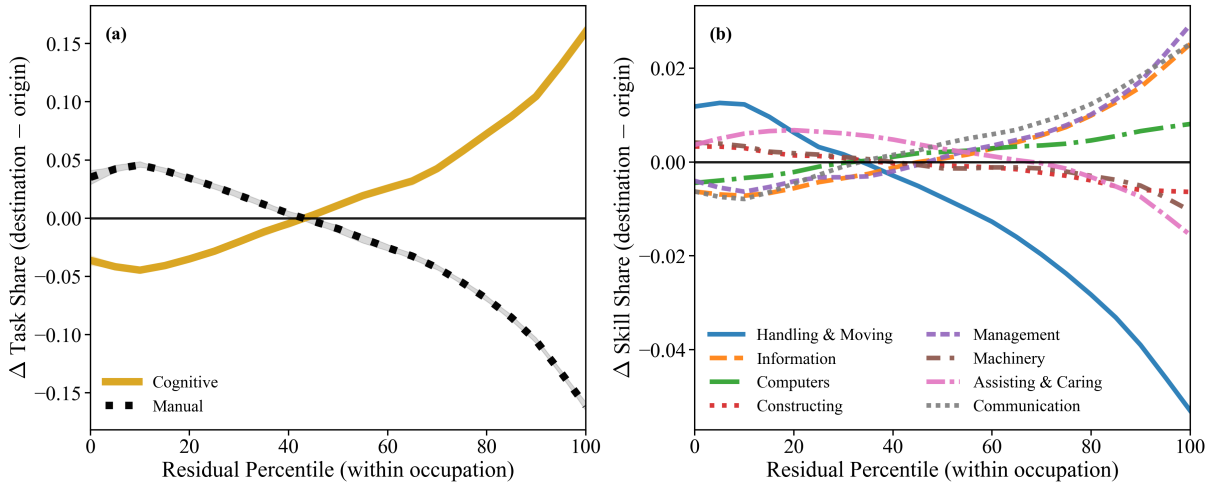


Figure 4: *Panel (a)*: Change in cognitive and manual task shares (vertical axis) by residualized within-occupation wage percentile (horizontal axis) for occupational movers. Change is $\Delta\text{Task-Share}_p^k = \text{LOWESS}_p\left(\omega^k_{j'(i)} - \omega^k_{j(i)}\right)$, the smoothed difference 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 29). 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 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 32).

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

[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. Under the piece-rate assumption (Section 4.2), wages equal realized output up to the occupation price, so this concavity at the log-output level maps directly into the well-documented concave life-cycle profile of wages.

the corresponding primitive task-level elasticity. The occupation-level human-capital bundle aggregates the task-level functions via the Cobb-Douglas weights:

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

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

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

[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.

subject to the occupation production function

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

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 $\boldsymbol{\lambda}_j$ govern how skills map into productivity 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} inherited from output.

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. This is most transparent at entry: all workers in a given education group enter with the same prior beliefs and no experience, so without the shocks every group would make a single, deterministic entry 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

parametrize it as a single utility constant:

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

A move pays κ_0 utils regardless of where the worker comes from, where she goes, and how much experience she has accumulated; staying is free. The cost captures, in reduced form, the disruption of changing occupation—the loss of occupation-specific match value, the search and retraining effort, the psychic cost of the move—that is not already priced by the wage equation. It is deliberately parsimonious: 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.[13] Under log utility and the wage equation (14), expected utility is linear in expected log productivity:

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

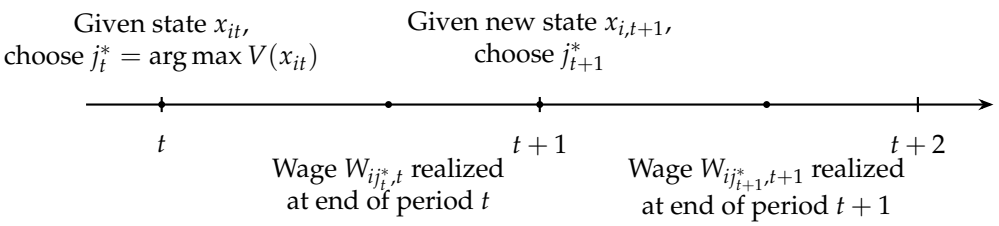


Figure 5: Timing of choices and wage realizations.

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

$$\begin{aligned} V_t\left(g, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\right) = \max_{j \in \mathcal{J}} \Big\{ & \mathbb{E}_{\boldsymbol{\xi}, \varepsilon_{jt}}\left[\log W_{ijt} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t\right] + \eta_{jg} - \mathcal{C}\left(j, d_{t-1}, e_t\right) + \epsilon_{jt} \\ & + \beta \, \mathbb{E}_{\boldsymbol{\xi}, s_{j_t t}, \epsilon_{t+1}}\left[V_{t+1}\left(g, \tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, j; \{p_{j\ell}\}_{\ell \geq t+1}\right)\right] \Big\}, \end{aligned}$$

$$\begin{aligned} \text{subject to:} \quad \boldsymbol{\tau}_{t+1} &= \boldsymbol{\tau}_t + \boldsymbol{\omega}_j, \\ s_{jt} &= \boldsymbol{\lambda}_j' \boldsymbol{\xi} + \varepsilon_{jt}, \\ \tilde{\boldsymbol{\xi}}_{t+1} &= \tilde{\boldsymbol{\xi}}_t + \boldsymbol{K}_{jt}\left(s_{jt} - \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_t\right), \\ \tilde{\boldsymbol{\Sigma}}_{t+1} &= \left(\boldsymbol{I}_{|\mathcal{K}|} - \boldsymbol{K}_{jt} \boldsymbol{\lambda}_j'\right) \tilde{\boldsymbol{\Sigma}}_t, \\ \text{given} \quad \left(\tilde{\boldsymbol{\xi}}_0, \tilde{\boldsymbol{\Sigma}}_0, \boldsymbol{\tau}_0, d_{-1}\right) &= \left(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \mathbf{0}, \varnothing\right) \text{ for an entrant of group } g. \end{aligned} \tag{26}$$

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

and the likelihood well-behaved for estimation.

[13] Richer specifications are nested in the model—for instance letting the cost vary with total experience, $\kappa_0 + \kappa_1 e_{it}$, or with the direction of the task-content change—and are straightforward to estimate. I do not pursue them here: the wage equation already generates an experience- and direction-dependent opportunity cost of moving through the loss of task-specific human capital, and a flexible utility cost on top of it is only weakly separated from that channel.

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:

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

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, observable component of the state (23): it does not evolve, but it shifts the payoff of every occupation through the amenities η_{jg}. Value functions and choice probabilities are therefore group-specific, and two workers with identical beliefs, task experience and origin occupation but different education need not choose alike. In the empirical implementation the choice probabilities are accordingly computed within education group; pooling across groups would confound the amenity differences η_{jg} with the choice probabilities themselves.

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 $x = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$ (excluding prices). Entrants are characterized by group masses π_g (with $\sum_{g \in \mathcal{G}} \pi_g = 1$) and initial distributions $\pi_g^0(\cdot)$ with $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}) = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ and $\boldsymbol{\tau} = \mathbf{0}$.

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). Workers face a constant exogenous per-period exit hazard δ of leaving the labor market (death or retirement). This exit risk is not internalized in the worker's problem—it does not enter the discount factor the worker optimizes against, so the operative discount remains β—and serves only to generate a well-defined stationary age-experience distribution: workers who exit are replaced one-for-one by entrants drawn from the same group-g prior $\boldsymbol{\xi} \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. I choose δ to match the average career length in the estimation sample, which yields $\delta = 0.08$ per period.

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 determined by the task technology (5), 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 (incorporating experience accumulation (10) and Bayesian updating (18)), 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: workers sort on beliefs about latent skills, and those beliefs are formed from the same wage and occupational histories observed by the econometrician. Conditional on the worker's history, occupational choice therefore selects on no additional unobserved permanent skill component, converting the selection problem into a control-function problem in which the worker's history is the relevant control.

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]Under standard regularity conditions on the induced Markov process, 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 laid out above. 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 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—all stated formally in Appendix A.1—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 proof is deferred to Appendix A.1; it factorizes the observed joint density of wages and choices into the conditional choice probabilities and the potential-wage path density, and then identifies the wage-equation parameters from the moments of the potential-wage path recovered by inverse-probability weighting. More intuitively, conditional on history, selection is on observables, so the identified CCPs serve as inverse-probability weights that recover the potential wage path under any supported occupational sequence. Under the linear-Gaussian wage equation this recovered path is multivariate normal: its means identify occupation prices plus group skill means, and its within-worker covariances identify the latent skill covariance and the occupation-specific loadings, with repeated observations in the same occupation separating the permanent component from transitory noise and switchers identifying how different occupations load on the common skill vector. When $K = 2$, one anchor occupation fixes the cognitive coordinate and one fixes the manual coordinate; covariances of all other occupations with these two anchors recover their cognitive and manual loading vectors.

The 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 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 are all group-specific and every comparison below is made within a group. With β fixed externally and ϱ calibrated, the only unknowns left anywhere in the model are the amenities η_{jg} and the single switching cost κ_0 of (25). The identifying content of the data is the level and composition of mobility across destinations. Under the Type-I extreme value assumption the observed log-odds recover differences in value indices,

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

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, since $W(x) = v_j(x) - \varrho \log P_j(x)$ 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: it is small where j is chosen often and large where j is chosen rarely, so choice frequencies measure how far each occupation's value lies below $W(x)$.

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, workers have already solved their dynamic problem, so their observed choices, summarized by the conditional choice probabilities, already encode their value functions; the continuation value can therefore be expressed through the CCPs identified from the data rather than through the unknown structural parameters. Collecting terms, the value index is affine 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 competing with η_{jg} and κ_0; they are data, up to a known linear loading on those two parameters.

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 separate empirical identification of ϱ is fragile without payoff variation that moves current utility but not the continuation value. In the baseline I therefore pin ϱ to the average occupational mobility rate in the large sample rather than estimating it jointly by maximum likelihood, which yields $\varrho = 0.271$. 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

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

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 $\sigma_{\varepsilon j}^2$ 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:

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

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 $\boldsymbol{\xi}_i$,

$$f(\boldsymbol{\xi}_i \mid w_i, d_i; \boldsymbol{\Gamma}^{(q)}),$$

[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,

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

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 $(\varsigma^2_{\varepsilon j})$, 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 $\boldsymbol{x} \equiv (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$. 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 $W(\cdot; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta})$, the CCPs used in the likelihood are computed from (27).

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

5.6 Externally Set and Calibrated Parameters

Several parameters are set externally or calibrated to match aggregate statistics rather than estimated jointly with the dynamic model, 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 ω_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 (29) leaves the weights proportional to $P_j Y_j^{1/\sigma}$ up to a single scalar, and fixing that scalar by $P(\hat{P}) = 1$ yields the closed form

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

where $P_j = \exp(\hat{p}_j)$ is occupation j's Stage-1 price, Y_j its output (the observed occupation wage bill deflated by P_j, since 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, so I do not impose the simplex constraint $\sum_j \alpha_j = 1$; the scale of $\{\alpha_j\}$ is itself the numeraire, and the observed cross-section of prices and quantities is an exact stationary equilibrium of the aggregate block.

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 $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. Full estimation results are reported in Table 2 in the Appendix.

Three patterns are central to the counterfactual analysis. First, occupations differ substantially in how heavily they weight the cognitive and manual task dimensions, both in their skill loadings λ_{jk} and in their returns to experience; Occ_0 is the cognitive anchor and Occ_1 the manual anchor. Second, the estimated group skill distributions $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ 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 origin, medium-educated workers near it, and high-educated workers above it and tilted toward cognitive skill, consistent with Deming (2025). At the same time, 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 within the (ξ_C, ξ_M) plane their comparative advantage lies (Figure 6). These distributional differences rationalize the key empirical patterns: low-wage workers switch to exploit comparative advantage, making large task changes, while high-wage workers specialize further in cognitive tasks. Third, returns to cognitive task experience are generally steeper than returns to manual task experience (Figure 7)—an alignment with the skill loadings whereby occupations that load on a given task dimension reward both the initial endowment and accumulated experience in that dimension, generating substantial heterogene-

ity in wage profiles across worker types.

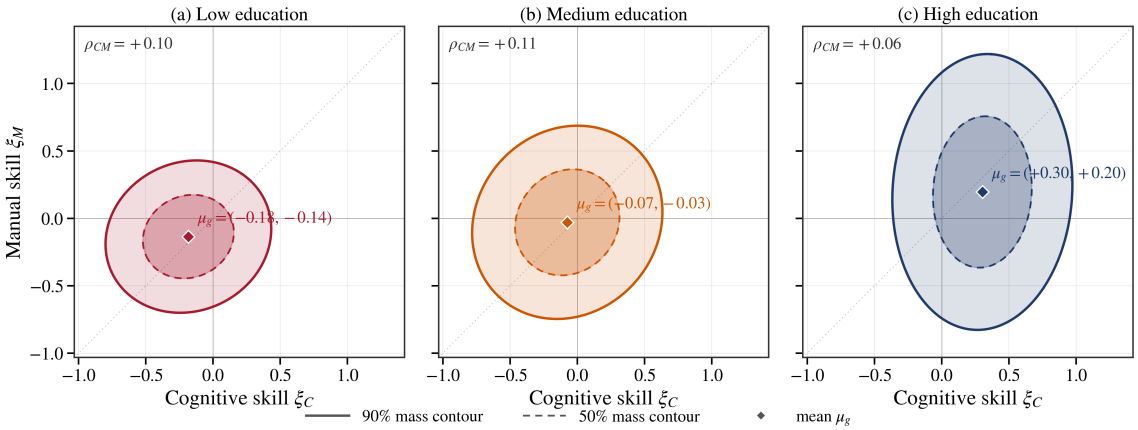


Figure 6: Estimated skill distributions by education group. Each panel (low, medium, high education) plots the joint prior distribution of cognitive skill ξ_C (horizontal axis) and manual skill ξ_M (vertical axis) implied by the estimated group means and covariances $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. 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.

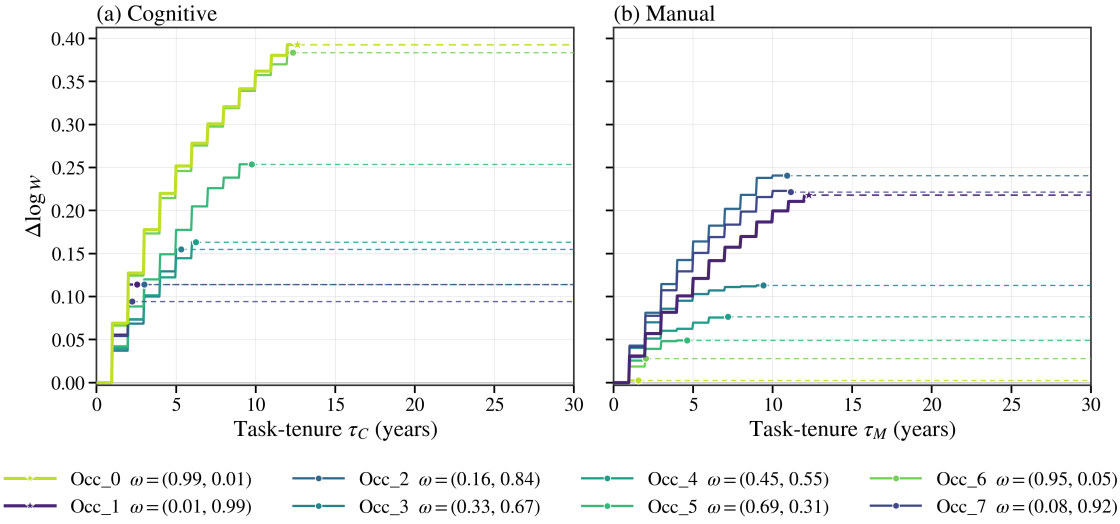


Figure 7: Estimated returns to task-specific experience by occupation. Each line plots the cumulative log-wage return to task experience, $\sum_b \beta_{jkb} \mathbf{1}[\tau \geq b]$, as task tenure τ accumulates from 0 to 30 years. Returns are estimated as a step function of tenure with annual knots; each curve is the resulting staircase, drawn solid up to the occupation's 95th percentile of observed tenure and held flat (dashed) thereafter. *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 not separately 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, rationalizing sorting patterns that cannot be explained by wage differentials alone.

6.3 The Experimentation Value of Occupations

Beyond wages and amenities, occupations differ in how much they let workers learn about their own comparative advantage, and the estimated model lets me measure this directly. For a worker in state $\boldsymbol{x} = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1}, g)$, I define the experimentation value of occupation j as the part of its continuation value that is generated by the wage signal the occupation produces, that is, the change in the worker's continuation value from observing the current signal:

$$EV_j^{\text{info}}(\boldsymbol{x}) = \beta\Big\{\mathbb{E}_s\big[V\big(\tilde{\boldsymbol{\xi}}_j'(s), \tilde{\boldsymbol{\Sigma}}_j', \boldsymbol{\tau} + \boldsymbol{\omega}_j, j\big)\big] - V\big(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau} + \boldsymbol{\omega}_j, j\big)\Big\}, \tag{41}$$

where V is the value function of Section 4 (equation (26)) and $(\tilde{\boldsymbol{\xi}}_j', \tilde{\boldsymbol{\Sigma}}_j')$ are the posterior belief moments after the Kalman update induced by the signal. Both terms hold the current payoff, the experience update $\boldsymbol{\tau} + \boldsymbol{\omega}_j$, and next period's origin occupation fixed, so current wages, switching costs, and amenities cancel: EV_j^{info} isolates the value of the information the occupation delivers. It nets two forces. A non-negative news component rewards the informativeness of the signal—because the value function is convex in the belief mean, resolving uncertainty is valuable, a Blackwell argument—while a non-positive depletion component subtracts the option value consumed as the signal draws down the worker's remaining stock of learnable uncertainty ($\tilde{\boldsymbol{\Sigma}}_j' \preceq \tilde{\boldsymbol{\Sigma}}$).[18] Because posterior beliefs evolve only along a single direction, I evaluate both components on the solved value function by Gauss–Hermite quadrature, at no additional simulation cost.

Figure 8 traces the experimentation value over the life cycle from model simulations, evaluating EV^{info} at the realized beliefs of the occupation each worker holds along a simulated career and expressing it as a share of the occupation's expected one-period (log) wage. The value of experimentation is large at labor-market entry—worth about 9% of a period's expected wage for high-educated workers, 4% for medium-educated, and 1.5% for low-educated—and collapses thereafter, falling by roughly four-fifths within the first year and converging toward zero, and across education groups, by mid-career. Two features stand out. First, learning is overwhelm-

[18] The two forces are the familiar Oi–Hartman–Abel logic (Oi 1961; Hartman 1972; Abel 1983) applied to careers. The worker cannot predict her own wage signal, so it is a pure gamble on tomorrow's beliefs; and because she can act on whatever she learns—moving toward cognitive work on good cognitive news and toward manual work on bad—she gains from that gamble, just as a competitive firm with convex profits prefers a volatile price to a stable one. That is the news component, and it is always positive. But uncertainty is also an asset here: it is what makes tomorrow's signals worth having, and reading today's signal spends part of it. That is the depletion component, and it is always negative. Appendix A.4 states both formally and shows that the news sign holds unconditionally, while signing depletion requires that the worker value the uncertainty she has left—automatic with a single skill, and true throughout the states used here.

ingly an early-career phenomenon: once workers have accumulated about a decade of task experience their beliefs are close to the truth, and the marginal wage signal is almost worthless. Second, the education gradient at entry is a depletion story. All groups enjoy a large news component (7–11% of the expected wage), but for low- and medium-educated workers it is nearly offset by an equally large depletion term, whereas high-educated workers—who begin with the most diffuse priors relative to their skill returns—retain most of the news as net value. The value of occupational experimentation is thus concentrated exactly where the model locates the cost of imperfect information: among young workers with diffuse beliefs. This is also why the skill-information intervention analyzed in the next section is most effective when delivered at entry. The same measure also ranks occupations by how much workers learn from them: as I show in Appendix B.4, the most informative occupations at entry are the manual-intensive ones—above all the manual anchor—and they are most informative for high-educated workers, who are ex ante the most uncertain about their manual skill. It is worth keeping the magnitudes in perspective, however. Occupational choice in the model is driven by much more than the informational content of wages—expected earnings, occupation amenities η_{jg}, idiosyncratic taste shocks, and switching costs—so relative to the total lifetime value of an occupation the experimentation value is modest: expressed as a fraction of the conditional (lifetime) value function it never exceeds about 2%, even at entry for the high-educated (Appendix B.4). The experimentation motive tilts early-career occupational choices at the margin rather than dominating them.

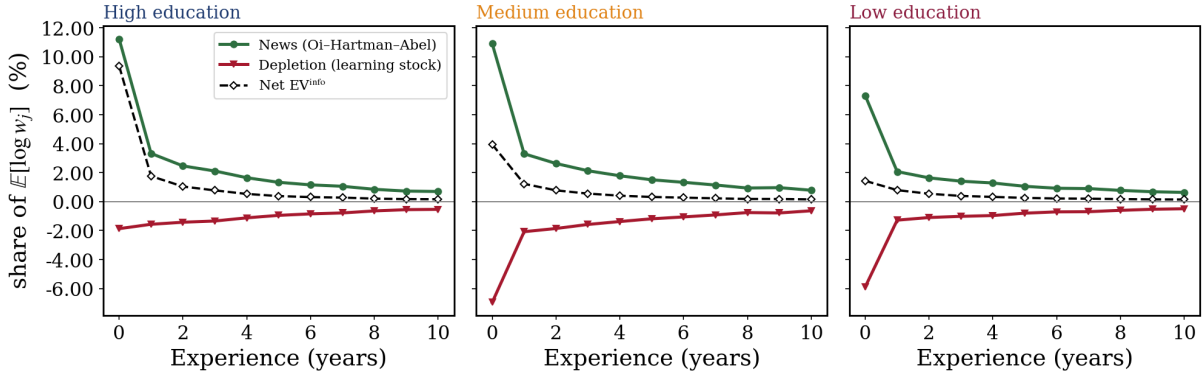
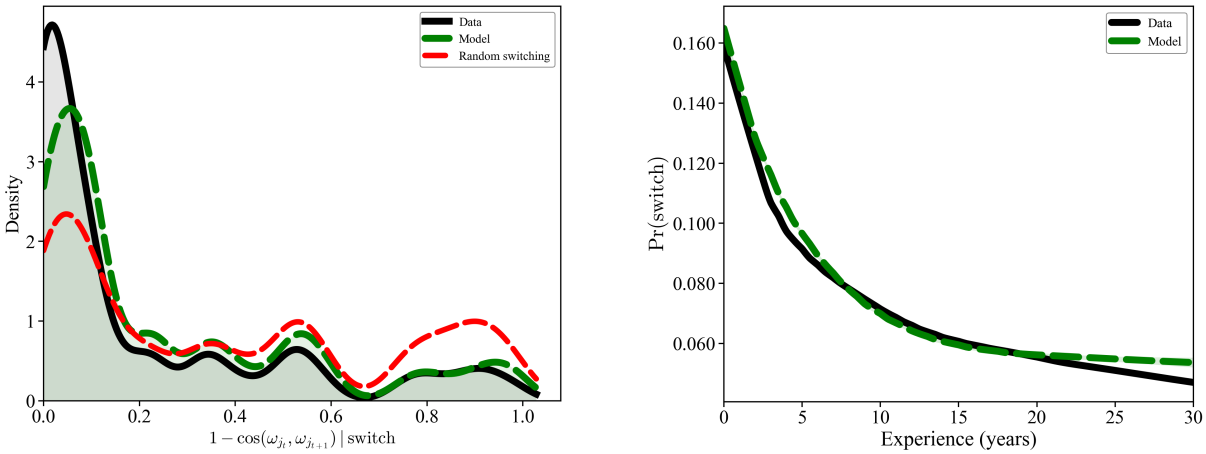


Figure 8: Experimentation value of occupations over the life cycle, by education group. For each education group, the curves plot the experimentation value EV^{info} of the occupation held—evaluated at workers' realized beliefs along a simulated career and averaged within bins of accumulated experience, expressed as a percentage of the occupation's expected one-period log wage—decomposed into the non-negative news component (Oi–Hartman–Abel, green) and the non-positive depletion component (the drawdown of the stock of learnable uncertainty, red), with the net value shown as the dashed line. Taste-shock scale $\varrho = 0.271$.

6.4 Occupational Mobility and the Role of Learning

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

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 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 dominate 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 profile is not targeted, and that no force in the model mechanically produces it: the declining hazard arises naturally from the interaction of learning and task-specific experience. Early in the career beliefs are diffuse, 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).

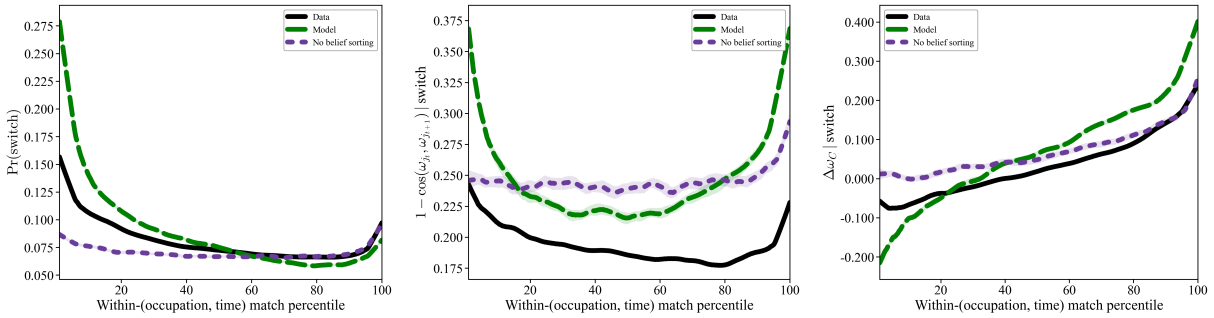
Figure 10 shows how occupational mobility varies with the match-quality percentile within

[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.

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 belief sorting, the mobility and the task content of the occupational moves of low-match-quality workers flatten: the switching of poorly matched workers disappears entirely, and with it the left-skewness of the U-shape. The model overshoots the height of that left tail, but without learning there is no left tail at all—the counterfactual sits below the data at every percentile up to the median, and the entire asymmetry of the U-shape is a learning phenomenon. Learning is even more important for 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 change the task content of their job by more. This is exactly what learning about comparative advantage delivers: a worker who receives a bad wage signal revises her beliefs about where her advantage lies and, to act on that revision, moves toward a better match by reaching for an occupation with a substantially different task composition rather than a neighboring one. Without belief sorting this link is severed and the U-shape in task distance 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 advantage. Beliefs are rational in the sense that the prior coincides with the true within-group population distribution of $\boldsymbol{\xi}_i$, so entrants are correct on average; but each worker's own skill vector is only one draw from that distribution, and any given individual's prior can lie far from her realization.

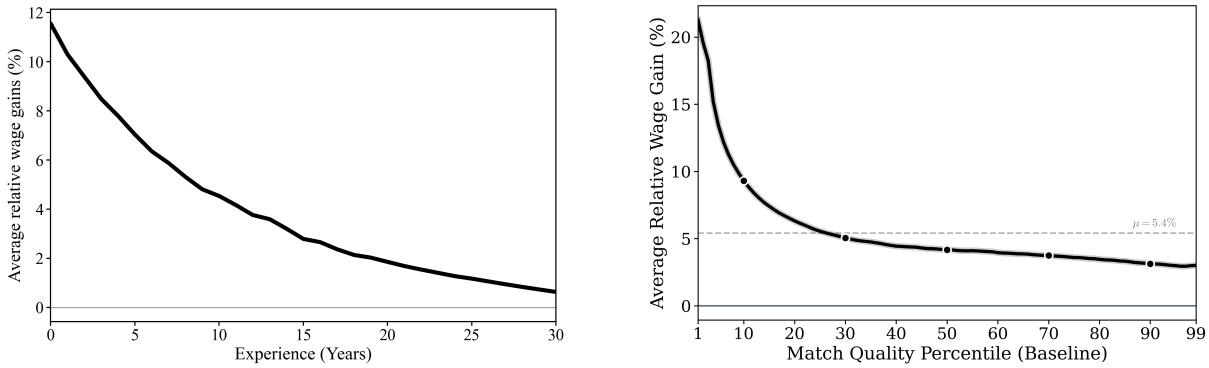
I construct a counterfactual general equilibrium in which workers have perfect beliefs from the start of their careers: each worker's initial beliefs equal her true skills, and she faces no uncertainty.[20] In both the baseline and counterfactual economies, workers face identical switching costs, idiosyncratic 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.

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 sharply as workers learn and re-sort, falling to about 2% by twenty years of experience and fading toward zero (about 0.3%) by the end of the working life, while remaining positive throughout. Mismatch is therefore largely an early-career phenomenon: the cost of imperfect information is borne mostly in the first decade, when beliefs are diffuse and workers are still discovering their comparative advantage, and it is progressively resolved—rather than reversed—as learning and re-sorting bring imperfect-information workers closer to their perfect-information counterparts. Learning, imperfectly transferable experience, and switching costs still interact to create path dependence, so that some workers never fully converge to their optimal career, but on average the residual gain closes gradually and does not turn negative.

On average, this reallocation operates almost entirely through better sorting rather than through

[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.



(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.

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 31 in the Appendix.

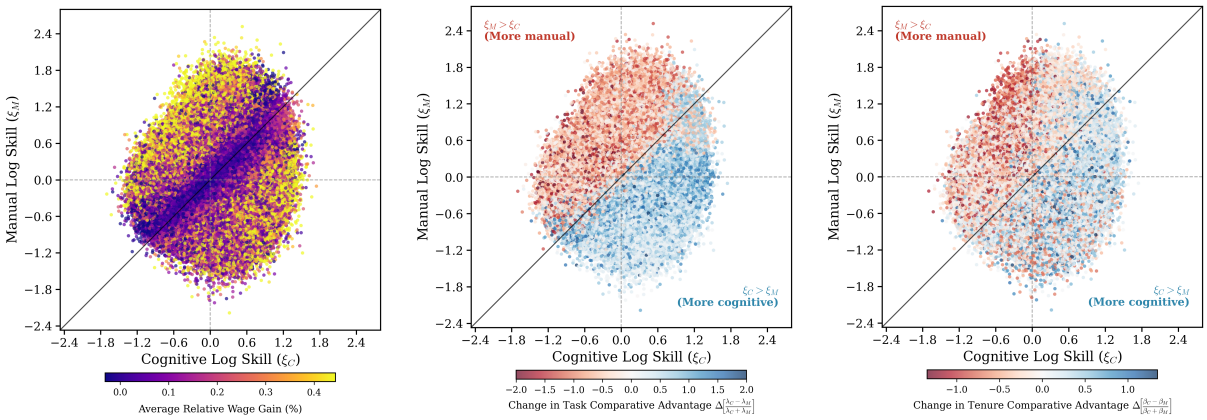
Additionally, gains are largest for the workers who are worst matched in the baseline. Panel (b) plots the average wage gain against the worker's position in the within-occupation match-quality distribution. The gain falls steeply in match quality: workers in the bottom percentile of their occupation's 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, gain most 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. Spe-

[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.



(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.

cialists have high returns to correct sorting: a worker with $\xi_C \gg \xi_M$ loses substantially if she mistakenly enters a manual-intensive occupation, and when beliefs are imprecise she may fail to recognize her comparative advantage and sort into the wrong occupation for several years. Generalists, in contrast, are approximately equally productive across occupations, so the cost of sorting "incorrectly" is small. This finding has implications for the design of career guidance policies: interventions that provide better information about skills, such as aptitude testing, structured internships, or early career mentoring, may yield the largest returns for workers with pronounced but unrecognized comparative advantage.

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 diagonal pattern 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. Taken together, the three panels confirm the model's core mechanism: imperfect beliefs lead workers to sort on perceptions rather than on true comparative advantage, and removing this friction lets them specialize in their strongest dimension—which is worth most to those whose advantage is most pronounced.

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 computerised assessment covering numeracy, communication, and concept formation skills, followed by individualised feedback about their measured standing. Treated workers update their beliefs about their own skills and subsequently shift their job search and occupational aspirations toward careers that better match their assessed comparative advantage. A practical attraction of this class of intervention is that it is cheap to deploy[23]. The headline estimate is that the share of workers whose stated career aspiration is aligned with their measured comparative advantage rises by 13.7 percentage points under treatment.

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):

$$s^0_{i,k} \;=\; \xi_{i,k} + \zeta_{i,k}, \qquad \zeta_{i,k} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_\zeta), \qquad k = 1, \ldots, K. \tag{42}$$

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

$$\tilde{\boldsymbol{\xi}}_{i0} = \boldsymbol{\mu}_{g(i)} + \boldsymbol{\Sigma}_{g(i)} \left(\boldsymbol{\Sigma}_{g(i)} + \sigma^2_\zeta \boldsymbol{I}_K \right)^{-1} \left(\boldsymbol{s}^0_i - \boldsymbol{\mu}_{g(i)} \right), \tag{43}$$

$$\tilde{\boldsymbol{\Sigma}}_{i0} = \boldsymbol{\Sigma}_{g(i)} - \boldsymbol{\Sigma}_{g(i)} \left(\boldsymbol{\Sigma}_{g(i)} + \sigma^2_\zeta \boldsymbol{I}_K \right)^{-1} \boldsymbol{\Sigma}_{g(i)}. \tag{44}$$

The worker is myopic about the arrival of this signal: she uses the same value function $V(\cdot)$ as in the baseline throughout her life, and after the assessment all subsequent wage signals, Bayesian updating, occupational choices, and task-specific experience accumulation proceed exactly as in the baseline economy. The intervention therefore operates through a single channel: it hands each entrant a tighter prior before her first occupational choice.

The signal variance σ^2_ζ, common across skill dimensions, is the only free parameter of the experiment. I calibrate it externally so that, in the model, the intervention raises the share of workers whose beliefs about their comparative advantage match their true comparative advantage by 13.7 percentage points, matching the treatment effect of Kiss et al. (2023) on the quintile-based aligned-comparative-advantage indicator.[24] The target is the model counterpart: the differ-

[23] 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.

[24] I follow their Section 2.6 definition. A large reference pool of latent types $\boldsymbol{\xi}$ is drawn from the unconditional population mixture—the model analogue of their Harambee jobseeker pool—and used

ence between post- and pre-assessment alignment rates.[25] 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.

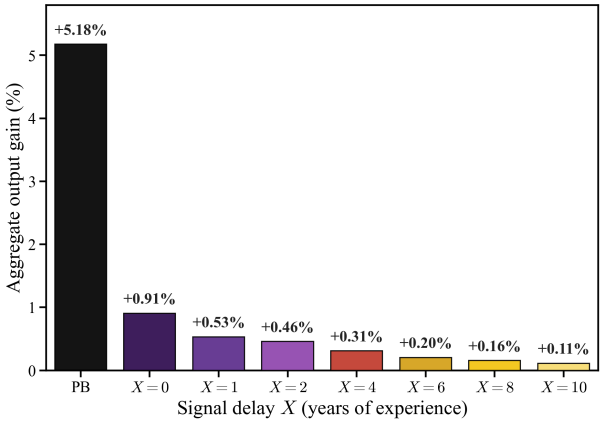


Figure 13: Aggregate output gain (percent) relative to the baseline economy when the calibrated skill assessment is delivered $X \in \{0, 1, 2, 4, 6, 8, 10\}$ years after labor-market entry, holding the calibrated signal precision σ_ζ^2 fixed at its at-entry value. Each bar is a separately solved stationary 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.

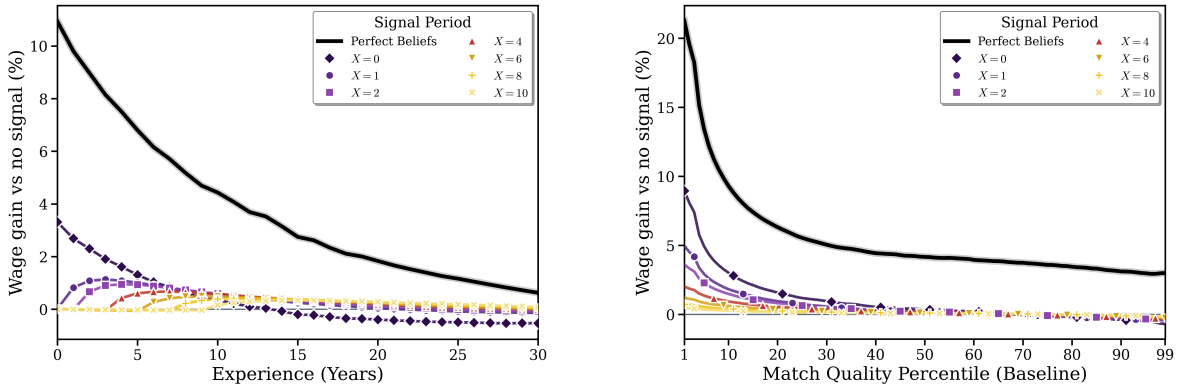
Relative to the baseline economy, providing entrants with the calibrated skill signal raises aggregate output by 0.91%. Set against the 5.18% that full information would deliver (the leftmost bar of Figure 13), a single noisy assessment at entry recovers $0.91/5.18 \approx 18\%$ of the output gain available from resolving skill uncertainty entirely. The wage gains are strongly front-loaded: average wages are about 3.3% above the baseline at entry, then decline as better-matched entrants accumulate experience and baseline workers catch up through their own wage-signal learning, crossing zero after roughly thirteen years and settling mildly negative (-0.5%) thereafter (Figure 14, Panel (a)). The perfect-information benchmark in the same panel gives the scale: a single noisy assessment at entry buys about 3.3% against the 11.6% that full knowledge of one's skills would deliver, that is, a little under a third of the value of resolving skill uncertainty entirely. The gain comes almost entirely from better matching on skills at the first occupational choice; the late-career dip reflects the cross-task experience mechanism emphasized earlier, as the policy induces early specialization and so forgoes some of the exploratory experience that baseline workers accumulate under their diffuse prior.

The assessment does not help everyone equally, and Panel (b) shows who it helps. Plotting the wage gain against the worker's position in the baseline match-quality distribution, the entire effect is concentrated at the bottom: workers in the lowest percentiles—those the baseline economy has sorted into occupations that fit them poorly—gain about 9%; the gain fades to roughly zero by the fortieth percentile, and above the median the assessment is worth noth-

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).

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

ing at all. The perfect-information benchmark in the same panel makes the comparison stark: a single noisy assessment recovers a little under half of the perfect-information gain for the worst-matched workers (9% against 21%), and essentially none of it for the well-matched. This is what one should expect from an information policy—it repairs mistakes, and workers who were not going to make one have nothing to gain from it.



(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.

The at-entry variant assumes that the assessment is delivered exactly at $t = 0$, just before the first occupational choice. 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 (42) 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 (Figure 14). The returns fall sharply the later the assessment comes: 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 wage profiles in Panel (a) show why an at-entry assessment is special: it is the only one that lifts wages in a worker's first years in the labor market, when mismatch is most acute, whereas a delayed assessment produces only a small hump once it arrives and never recovers the value lost by coming late. Two forces drive the decay. Task-specific experience, accumulated under a diffuse prior and only imperfectly transferable, leaves a delayed worker partly locked into an occupation that does not fit her; and her own wage-signal learning has by then already brought her beliefs closer to the truth, so an external assessment tells her less than it would have at entry. The upshot is a clear timing lesson: skill-profiling tools are most valuable delivered just before the first occupational choice, not after it.

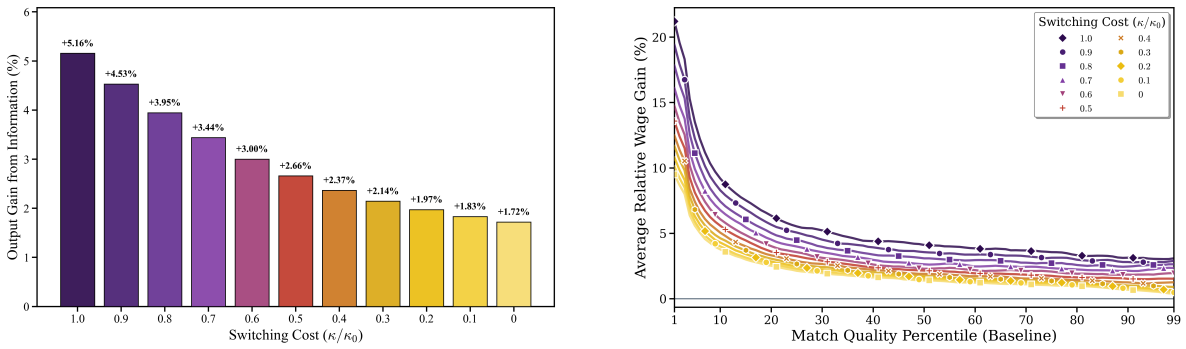
In Appendix C.6 I study an experiment that directs each labor market entrant, for one period only, into the occupation that is most informative about her comparative advantage, after which she chooses freely. The forced exploration is costly on impact—entrants are bunched away from their preferred occupations and their wages fall—but the sharper beliefs it pro-

duces improve sorting for the rest of the career, and on net aggregate output rises by about 0.65%. The value of learning is thus large enough that even compelling workers to experiment early can be output-enhancing.

7.3 Switching Costs and Information Frictions

The cost of imperfect information depends on how easily workers can act on what they learn. If mobility were free, a worker who discovers she is mismatched could simply leave, so the mistake would be transitory; if it is costly, the same mistake is locked in for years. 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)). The result is intuitive: cheaper mobility is a partial substitute for information, because it lets workers repair a bad initial choice once their beliefs sharpen, so being uninformed costs less when the consequences of being wrong are less persistent.



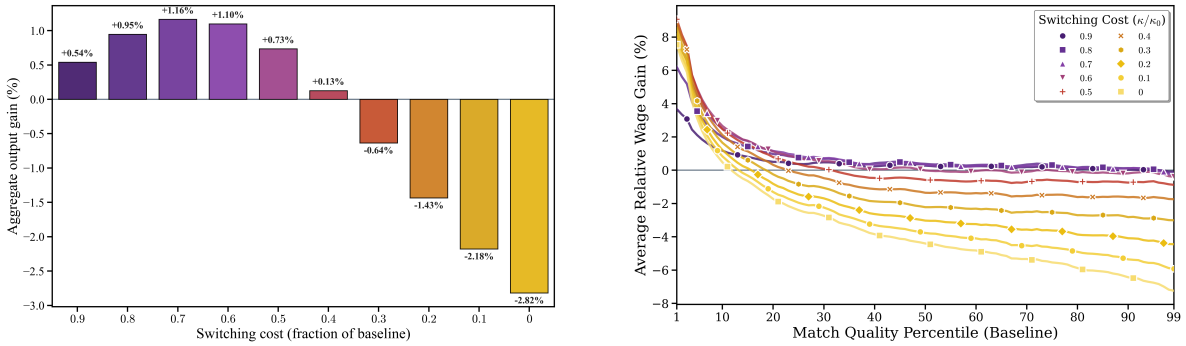
(a) Output gain from perfect information

(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 rises as mobility becomes cheaper, peaks at a moderate reduction, and then falls, turning negative and reaching a loss of

about 2.8% when switching is free.



(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 the inverted U. A small reduction in the switching cost raises output, because workers are freer to reallocate toward the occupations in which they have a comparative advantage as they discover it. 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, so they begin to churn across occupations for reasons unrelated to their skills. This excessive churning pushes them into occupations where their task-specific human capital is not particularly productive, destroying experience faster than better sorting can create value. Panel (b) of Figure 16 shows who bears the resulting losses: they fall on the well-matched workers, at the top of the within-occupation match-quality distribution. The poorly matched gain from cheaper mobility at every switching-cost level, because they have a bad match to escape; but a worker who was already well matched has nothing to gain from moving and everything to lose, and under costless mobility, taste shocks pull her out of a good occupation and into worse ones, so her wage falls by up to 6–8%. The aggregate output loss from excessive churning is therefore driven by the top of the match-quality distribution, not the bottom. The two comparisons answer different questions and do not conflict: cheaper mobility always reduces the cost of not knowing one's skills, but making it too cheap still lowers output, because the churning it induces destroys task-specific experience for informed and uninformed workers alike.

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 Lan-

guage 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

$$\omega_{jC}^{\text{AI}} = \omega_{jC}\,(1 - \chi_{jC}), \qquad \omega_{jM}^{\text{AI}} = \omega_{jM}, \tag{45}$$

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.

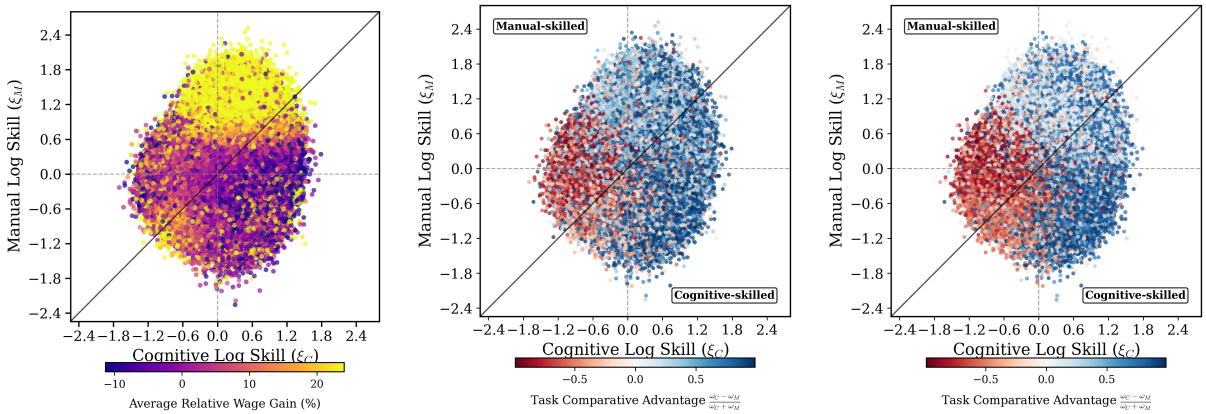
8.2 The Direct Effects of AI

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 are advantaged. Cognitive specialists in the lower-right region lose, as their primary skill dimension is directly devalued. 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. Comparing the two, workers—especially those with manual comparative advantage—sort into

more manual-intensive occupations after AI (the red region expands), where their now-more-valuable manual skills are better rewarded. Cognitive specialists largely remain in cognitive-intensive occupations but earn lower returns there. The workers who gain most are precisely those who hold, or can move toward, manual comparative advantage.



(a) Average wage gain from AI

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

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

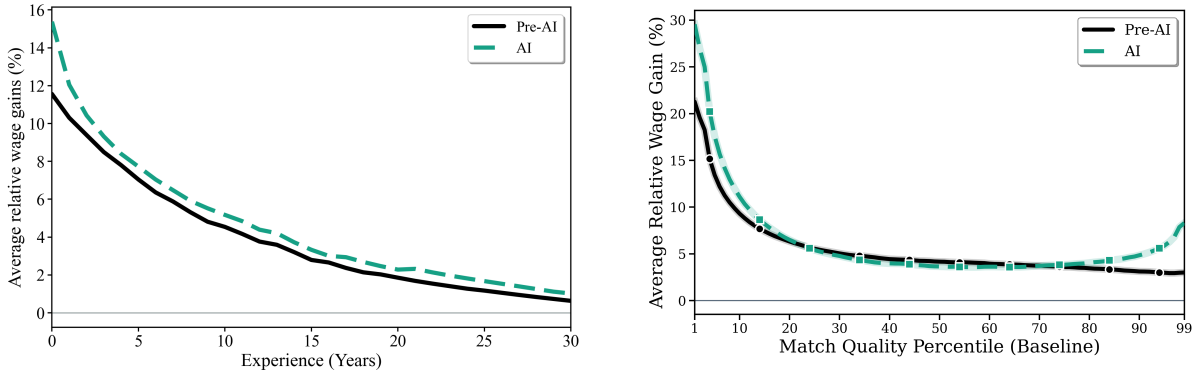
Figure 17: The direct effects of AI in the space of true skills. In each panel a point is a simulated worker placed at her true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis); the solid line is the 45° line $\xi_C = \xi_M$, which separates cognitive specialists (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.

These skill-space effects 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 36 in the Appendix).

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 value of resolving skill uncertainty, and it does so throughout the career. Panel (a) shows that the wage gain from perfect information is higher under AI 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



(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.

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 the two curves also separate: well-matched workers, who under the pre-AI technology had little to gain from better information, now gain visibly more (roughly 8% against 3% in the top percentiles). In the middle of the distribution the two curves are essentially indistinguishable.

8.4 The Gains from Removing Information Frictions Along the Transition Path

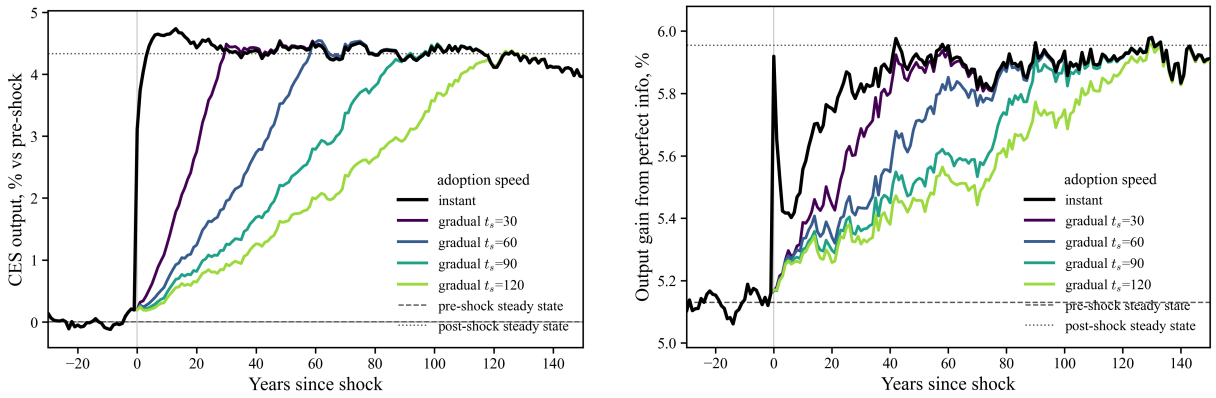
The counterfactuals above compare stationary equilibria and therefore speak to the long-run incidence of the AI shock, abstracting from the adjustment dynamics that follow it. I now study the output consequences of AI along the transition path, tracing how the economy moves from the pre-shock to the post-shock equilibrium and how the gains from removing information frictions evolve along the way.

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 unanticipated until it arrives, but from the moment it hits, workers have perfect foresight over the entire subsequent path of prices and technology. I consider two conventions for how the new technology materializes. Under instantaneous adoption it jumps to its post-shock level on impact; under gradual adoption it diffuses over time, so that the task content of occupations—and the associated returns to skills and experience—adjusts smoothly toward the post-shock economy. When the transition is gradual, workers still have perfect foresight over the entire diffusion

path: they know at $t = 0$ how the technology will unfold and optimize against the whole anticipated sequence of prices and task weights. In both cases the equilibrium price path is solved by backward–forward shooting, holding the CES weights and the per-period numeraire fixed exactly as in the stationary problem; the full procedure is described in Appendix C.3.

This experiment isolates the transitional consequences of AI that the stationary comparison cannot capture: the response of aggregate output on impact and its subsequent convergence to the new long-run level, and the path of the output gain from removing information frictions as the economy adjusts.

Figure 19 reports the transition. 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 is output-improving at every adoption speed, but the timing of the gains depends sharply 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 the short-run path is very different—output barely moves on impact and then climbs monotonically through the diffusion window—but 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. The speed of adoption thus governs when the gains arrive, not how large they ultimately are.

Second, AI raises the gain from removing information frictions (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. The increase opens immediately and is permanent. Why the shock raises the gain is a question I do not resolve here; I document it as a robust feature of the transition

rather than tracing it to a specific channel. Slower diffusion delays the increase, so that the gain tracks the technology as it arrives, but does not avoid it. Removing information frictions is thus worth more in the AI economy than before the shock, reinforcing the value of the information interventions studied in Section 7.2.

In short, the AI shock raises long-run output by about 4.3% at every adoption speed, with faster adoption front-loading the gains, while the output gain available from removing information frictions rises permanently from about 5.2% to roughly 5.9%.

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 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 an information policy that reduces this uncertainty and ask how much of the lost output it can recover. Calibrated to the 13.7 percentage-point treatment effect of Kiss et al. (2023), a skill assessment delivered at entry raises aggregate output by 0.91%, with the same incidence as full information: the worst-matched workers gain the most, the best-matched almost nothing. Its value decays sharply when the assessment is delivered later, once task-specific experience is sunk, so such programs are most valuable before the first occupational choice. Reducing the cost of switching occupations helps too, but only up to a point: modest reductions let workers act on what they learn and raise output, whereas eliminating switching costs induces excessive churning that destroys task-specific experience and lowers it.

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 be-

tween 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

Abel, A. B. (1983), 'Optimal investment under uncertainty', *American Economic Review* **73**(1), 228–233.

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.

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.

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.

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

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.

Hartman, R. (1972), 'The effects of price and cost uncertainty on investment', *Journal of Economic Theory* **5**(2), 258–266.

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.

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.

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.

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.

Oi, W. Y. (1961), 'The desirability of price instability under perfect competition', *Econometrica* **29**(1), 58–64.

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.

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.

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.

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, once the worker's history is controlled for, occupational choice does not reveal any additional private information about permanent skill.

The wage equation is

$$w_{it}(j) = p_{jt} + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}, \tag{46}$$

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 I also observe, so conditional on $\mathcal{H}_{it}$ their choice reveals nothing further about permanent skill. Assumption 3 guarantees that the wage distributions the argument needs are observed. Assumption 4 provides the repeated measurements that separate permanent skill from transitory noise: experience variation within an occupation–year separates the returns to task experience from the occupation price, and mobility across occupations delivers the cross-occupation wage covariances. Assumption 5 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 the location, and the base experience level the split between the price of an occupation and the return to experience in it.

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

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

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:

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

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

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

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. So

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

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:

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

Conditional on their history, the workers who choose j are not a skill-selected subset of those who could have chosen it. All the selection sits in the choice probabilities (Bunting et al. 2024).

Dividing by the choice probabilities therefore isolates the potential wage path: along any supported sequence $d^T = (d_1, \ldots, d_T)$,

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

Everything on the right is in the data, so the distribution of potential log wages along every supported path is identified. Workers who happen to follow a given path are upweighted by how unlikely that path was for them, which undoes the selection created by their own choices; after reweighting I can treat the path as assigned. 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,

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

is jointly Gaussian, because it is a linear map of the Gaussian skill vector $\boldsymbol{\xi}_i$ and the Gaussian wage shocks. Its mean and covariance are therefore all I need, and they have a simple structure: the same worker's wages in different periods share her permanent skill $\boldsymbol{\xi}_i$ but not her wage shocks, which are independent over time. What is left after reweighting is thus a standard repeated-measurement problem—repeated wages measure the same latent skill vector with different loadings 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:

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

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 outright if prices are constant). 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. The contrast between the last two is what will separate them. Without the weights each of these would be computed on a selected sample—the workers who *chose* that occupation, or that pair of occupations—and would not be a moment of the group.

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,

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

This is where the anchors earn their name: because each loads on a single skill, the mean wage in anchor a_k is just the price of that occupation plus the group's mean skill in dimension k. 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 the anchor wages: two wages of the same worker in the anchors, at two different dates, share nothing but her permanent skill, 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

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

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 wage equation is identified, the only unknowns left in the model are the amenities $\{\eta_{jg}\}$ and the switching cost κ_0. This subsection shows that both are identified from the observed choice frequencies. Throughout, $\mathcal{J}$ is the set of occupations, $0 \in \mathcal{J}$ the base occupation, $g \in \mathcal{G}$ the education group, and for a state x I write d for the origin occupation d_{t-1} contained in x and $\tilde{w}_j(x) \equiv \mathbb{E}[\log W_{ijt} \mid x]$ for the expected log wage in occupation j. The choice-specific value index and the choice probabilities are

$$v_j(x) = \tilde{w}_j(x) + \eta_{jg} - \mathcal{C}(j,d) + \beta\, \mathbb{E}\big[W(x') \mid x,\, d_t = j\big], \qquad P_j(x) = \frac{\exp(v_j(x)/\varrho)}{\sum_{m\in\mathcal{J}} \exp(v_m(x)/\varrho)},$$

where $W(x) = \varrho \log \sum_m \exp(v_m(x)/\varrho)$ is the ex-ante value of state x, before the taste shocks are drawn.

I make the following assumptions.

Assumption 6 (Stationarity and known primitives) *The data are generated by the stationary equilibrium of Section 4.6, so prices are constant and W, v_j and P_j do not depend on t. The wage equation and the wage-signal distribution are identified (Theorem 1), so the expected wage $\tilde{w}_j(x)$ and the transition law $\mathcal{P}(x' \mid x, d_t = j)$ are known. The discount factor $\beta \in (0,1)$ and the taste-shock scale $\varrho > 0$ are fixed; in the baseline ϱ is calibrated to the average mobility rate. The state space is compact and $\tilde{w}_j$ is bounded on it, so W is the unique bounded fixed point of the Bellman operator (Lemma 2).*

Assumption 7 (Positive choice probabilities) *The choice probabilities $P_j(x) = \Pr(d_t = j \mid x_{it} = x)$ are identified from the joint distribution of choices and states, and are bounded away from zero, $P_j(x) \geq \underline{p} > 0$, on the states used below.*

Assumption 8 (Origins, destinations, stays and moves) *For every group g: (i) every occupation $j \neq 0$ is observed as a destination from some origin $d \notin \{j, 0\}$, so that j and the base occupation are reached from a common origin; and (ii) some origin d is observed with both a stay and a move. No variation in task experience or in the direction of the move is needed.*

The argument has two parts. The first shows that continuation values are not free parameters: the same logit structure that generates the choice probabilities can be inverted to express the ex-ante value in terms of them, so the value index becomes a known object plus a linear function of (η_{jg}, κ_0). The second part then reads the two parameters off the observed log-odds, and it does so sequentially: comparing two destinations reached from a common origin identifies the amenities, because both moves pay the same switching cost and it cancels; comparing a move with a stay from the same origin then identifies the switching cost, since the amenities are by then known.

Lemma 1 (CCP inversion along the stay path) *Let Assumptions 6–7 hold. Then, for every supported x and every j,*

$$W(x) = \nu_j(x) - \varrho \log P_j(x),$$

and the value index is affine in $(\{\eta_{jg}\}, \kappa_0)$ with known coefficients,

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

$$\Psi_j(x) \equiv \mathbb{E}\left[\sum_{s \geq 1} \beta^{s-1} \Big(\tilde{w}_j(x_{t+s}) - \varrho \log P_j(x_{t+s}) \Big) \;\Big|\; x_t = x,\ d_{t+s-1} = j \ \ \forall s \geq 1 \right],$$

where $\Psi_j(x)$ is the discounted stream generated along the path on which the worker takes j and never leaves it. It is finite, and it is known: it involves only the identified wage equation, the known transition law, and the observed choice probabilities.

Proof. The first display is the inversion of Hotz and Miller (1993): taking logs of the logit choice probability gives $\varrho \log P_j(x) = \nu_j(x) - \varrho \log \sum_m \exp(\nu_m(x)/\varrho) = \nu_j(x) - W(x)$. The correction $-\varrho \log P_j(x)$ is the gap between the ex-ante value and the value of committing to j: it is small where j is chosen often and large where j is chosen rarely, so choice frequencies measure how far each occupation's value lies below $W(x)$.

For the second display, fix x and j and consider a state x_{t+s}, $s \geq 1$, reached after j has been chosen in every period since t. The origin recorded in that state is j, so choosing j again is a *stay* and costs nothing: $\mathcal{C}(j, j) = 0$. Applying the inversion at x_{t+s},

$$W(x_{t+s}) = \tilde{w}_j(x_{t+s}) + \eta_{jg} - \varrho \log P_j(x_{t+s}) + \beta\, \mathbb{E}[W(x_{t+s+1}) \mid x_{t+s}, d = j],$$

and iterating forward under the known transition law,

$$\mathbb{E}[W(x_{t+1}) \mid x, d_t = j] = \mathbb{E}\left[\sum_{s=1}^{S} \beta^{s-1} \Big(\tilde{w}_j(x_{t+s}) + \eta_{jg} - \varrho \log P_j(x_{t+s}) \Big) \right] + \beta^S\, \mathbb{E}[W(x_{t+S+1})].$$

The group is fixed, so η_{jg} is constant along the path and $\sum_{s \geq 1} \beta^{s-1} \eta_{jg} = \eta_{jg}/(1-\beta)$. The remainder $\beta^S \mathbb{E}[W(x_{t+S+1})]$ vanishes because W is bounded and $\beta < 1$, and the series converges because $\tilde{w}_j$ is bounded and $|\varrho \log P_j| \leq \varrho |\log \underline{p}|$. Hence $\mathbb{E}[W(x_{t+1}) \mid x, d_t = j] = \Psi_j(x) + \eta_{jg}/(1-\beta)$; substituting into $\nu_j(x)$ and using $1 + \beta/(1-\beta) = 1/(1-\beta)$ gives the result. □

The intuition is that pricing an occupation by the counterfactual "take j and never leave" path is convenient precisely because staying is free: the switching cost disappears from the continuation value altogether, and the amenity is collected every period, so it enters as a permanent flow $\eta_{jg}/(1-\beta)$. The $-\varrho \log P_j$ terms convert the value of *committing* to j into the value of behaving optimally, using only observed choice frequencies. Continuation values are therefore not free parameters that could absorb amenities or switching costs; they are known functions of the data plus a known loading on the two parameters. This forward expansion is the representation of the value function as a sequence of choice probabilities and flow payoffs of Arcidiacono and Miller (2011), obtained by applying the Hotz–Miller inversion recursively along a chosen action path. The finite-dependence version of that representation, in which the sum terminates after finitely many periods, is not available here: beliefs and task tenure never re-merge once

two different occupations are chosen, so no two state paths coincide. I therefore use the infinite version along the stay path, on which the switching cost is never paid.

Theorem 2 (Identification of the dynamic choice parameters) *Let Assumptions 6–8 hold, let the switching cost be* $C(j,d) = \kappa_0 \mathbf{1}_{\{j \neq d\}}$ *as in* (25)*, and normalize* $\eta_{0g} = 0$ *for every* g*. Then* $\big(\{\eta_{jg}\}_{j\in\mathcal{J},g\in\mathcal{G}}, \kappa_0\big)$ *is identified.*

Proof. By the lemma, the observed log-odds between any two occupations are

$$\varrho \log \frac{P_j(x)}{P_m(x)} = \underbrace{\big[\tilde{w}_j(x) - \tilde{w}_m(x)\big] + \beta\big[\Psi_j(x) - \Psi_m(x)\big]}_{\text{known}} + \frac{\eta_{jg} - \eta_{mg}}{1-\beta} - \kappa_0\Big(\mathbf{1}_{\{j\neq d\}} - \mathbf{1}_{\{m\neq d\}}\Big), \quad (47)$$

so the only unknowns in the log-odds are the amenities and the switching cost.

Identification of the amenities η_{jg}. Fix a group g and an occupation $j \neq 0$, and take a state whose origin is neither j nor the base, $d \notin \{j, 0\}$, as guaranteed by Assumption 8(i). Going to j and going to 0 are then both *moves*, so both pay κ_0 and it cancels exactly in (47)—in the flow payoff and, by the lemma, in the continuation value as well. Setting $m = 0$ and using $\eta_{0g} = 0$,

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

Everything on the right is known, so the amenities are identified: how often workers leaving a common origin go to j rather than to the base occupation, once expected wages and continuation values are netted out, is exactly how much more they like j.

Identification of the switching cost κ_0. Now take an origin d at which both a stay and a move to some $j \neq d$ are observed, as guaranteed by Assumption 8(ii). Setting $m = d$ in (47), the indicator difference is one, so

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

Every term on the right is now known—expected wages from Theorem 1, the Ψ's from the lemma, the choice probabilities from the data, and the amenities from the previous paragraph—so κ_0 is identified. It is the part of the reluctance to move that the wages, the amenities and the continuation values cannot explain. □

Remark. The parameter vector has $|\mathcal{G}|(|\mathcal{J}|-1)+1 = 3\times 7+1 = 22$ free elements, and the proof uses exactly 22 log-odds equations: one per (j,g) with $j \neq 0$, and a single move-versus-stay comparison. A minimal set of comparisons therefore suffices; all remaining origin–destination pairs, states and stay/move contrasts generate further copies of (47) in the same unknowns and overidentify the parameters, and the NFXP likelihood uses all of them. Two things are worth noting. First, no variation in task experience or in the direction of the move is needed anywhere: a cost that varied with experience, destination or direction would not cancel when comparing two destinations, and amenities and switching costs would then have

to be separated jointly. This is the identification price of the richer switching-cost specifications nested in (25). Second, the amenities enter the value index as a permanent flow, scaled by $1/(1-\beta)$, whereas κ_0 is a one-off cost with a loading of one. With $\beta = 0.9$ the amenities equal $(1-\beta) = 0.1$ times the log-odds residual: they are tightly determined by the data, but proportionally sensitive to the externally fixed discount factor.

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. □

A.4 The News–Depletion Decomposition of EV^{info}

Fix a state $\boldsymbol{x} = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1}, g)$ and an occupation j. The wage signal $s = \boldsymbol{\lambda}_j'\boldsymbol{\xi} + \varepsilon_j$ moves the worker's beliefs by the Kalman update

$$\tilde{\boldsymbol{\xi}}_j'(s) = \tilde{\boldsymbol{\xi}} + \boldsymbol{K}_j(s - \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\xi}}), \qquad \tilde{\boldsymbol{\Sigma}}_j' = \tilde{\boldsymbol{\Sigma}} - \frac{\tilde{\boldsymbol{\Sigma}}\boldsymbol{\lambda}_j\boldsymbol{\lambda}_j'\tilde{\boldsymbol{\Sigma}}}{\boldsymbol{\lambda}_j'\tilde{\boldsymbol{\Sigma}}\boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2}, \qquad \boldsymbol{K}_j = \frac{\tilde{\boldsymbol{\Sigma}}\boldsymbol{\lambda}_j}{\boldsymbol{\lambda}_j'\tilde{\boldsymbol{\Sigma}}\boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2}. \tag{48}$$

It does two things: it moves her best guess about her skills, and it lowers her uncertainty about them. Adding and subtracting $V(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}_j', \boldsymbol{\tau} + \boldsymbol{\omega}_j, j)$ in (41) separates the two:

$$\begin{aligned} EV_j^{\text{info}}(\boldsymbol{x}) = {} & \underbrace{\beta\Big\{\mathbb{E}_s\big[V\big(\tilde{\boldsymbol{\xi}}_j'(s),\, \tilde{\boldsymbol{\Sigma}}_j',\, \boldsymbol{\tau} + \boldsymbol{\omega}_j,\, j\big)\big] - V\big(\tilde{\boldsymbol{\xi}},\, \tilde{\boldsymbol{\Sigma}}_j',\, \boldsymbol{\tau} + \boldsymbol{\omega}_j,\, j\big)\Big\}}_{N_j(\boldsymbol{x})\ \text{(news)}} \\ & + \underbrace{\beta\Big\{V\big(\tilde{\boldsymbol{\xi}},\, \tilde{\boldsymbol{\Sigma}}_j',\, \boldsymbol{\tau} + \boldsymbol{\omega}_j,\, j\big) - V\big(\tilde{\boldsymbol{\xi}},\, \tilde{\boldsymbol{\Sigma}},\, \boldsymbol{\tau} + \boldsymbol{\omega}_j,\, j\big)\Big\}}_{D_j(\boldsymbol{x})\ \text{(depletion)}}. \end{aligned} \tag{49}$$

News moves the belief mean at the post-signal uncertainty; depletion lowers the uncertainty at an unchanged mean.

The news term is nonnegative, and this can be shown formally.

Lemma 3 (Convexity in beliefs and the news effect) *V is convex in $\tilde{\boldsymbol{\xi}}$, and $N_j(\boldsymbol{x}) \geq 0$ for every state and every occupation.*

Proof *Each occupation's expected flow payoff, $p_j + \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\xi}} + h_j(\boldsymbol{\tau}) - \kappa_0\mathbf{1}_{\{j \neq d_{-1}\}} + \eta_{jg}$, is linear in $\tilde{\boldsymbol{\xi}}$, and the posterior covariance does not enter it. Write the signal as $s = \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\xi}} + u_j$: the surprise u_j is mean-zero noise whose distribution does not depend on $\tilde{\boldsymbol{\xi}}$, so by (48) tomorrow's belief $\tilde{\boldsymbol{\xi}} + \boldsymbol{K}_j u_j$ is today's*

belief shifted by noise that is independent of it, while tomorrow's uncertainty is deterministic. Convexity in $\tilde{\boldsymbol{\xi}}$ therefore passes through the expectation over u_j, and $V = \varrho \log \sum_m \exp(v_m/\varrho)$ is convex and increasing in each v_m, so V is convex in $\tilde{\boldsymbol{\xi}}$; the Bellman operator preserves the property, so the fixed point of Lemma 2 has it. For the sign, posterior means are martingales—the worker cannot forecast her own signal, $\mathbb{E}_s[\tilde{\boldsymbol{\xi}}'_j(s)] = \tilde{\boldsymbol{\xi}}$—so Jensen's inequality applied to the convex map $\tilde{\boldsymbol{\xi}} \mapsto V(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}'_j, \boldsymbol{\tau} + \boldsymbol{\omega}_j, j)$ gives $N_j(\boldsymbol{x}) \geq 0$. □

The intuition is that the signal is a pure spread of tomorrow's belief around today's, and the worker's payoff is convex in beliefs because she can act on what she learns—moving toward cognitive work on good cognitive news and toward manual work on bad. A convex payoff gains from a spread, for the same reason a competitive firm with convex profits prefers a volatile price to a stable one. This is the Oi–Hartman–Abel effect (Oi 1961; Hartman 1972; Abel 1983), and it requires no restriction on $\tilde{\boldsymbol{\Sigma}}$.

The depletion term is nonpositive. Reading the signal always lowers the worker's uncertainty—strictly so unless the signal is uninformative about the skills she is paid for, that is, unless $\tilde{\boldsymbol{\Sigma}}\boldsymbol{\lambda}_j = \mathbf{0}$—and uncertainty is an asset in this model: it is what makes future wage signals worth having. Spending it therefore costs her the option value it carried, so $D_j(\boldsymbol{x}) \leq 0$. With a single skill this is immediate: a worker who is more uncertain learns more from today's wage *and* still has more left to learn tomorrow, so more uncertainty is unambiguously better and consuming it is unambiguously a cost. With two skills there is a caveat: what is valuable is not uncertainty in total but uncertainty in the direction that matters for her choice, and being more uncertain overall can shift *which* skill she ends up learning about. I therefore do not claim the sign in general. I compute both components directly on the solved value function, and in every simulated state underlying Figure 8 the decomposition comes out with $N_j \geq 0 \geq D_j$.

B Appendix: Estimation Results

B.1 Externally Set and Calibrated Parameters

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 $\boldsymbol{\omega}_j$ 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.

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.

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.

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.

B.3 Stage 2 Estimation Results: Dynamic Choice Parameters

B.4 The Informativeness of Occupations at Entry

The experimentation value defined in Section 6.3 also ranks occupations by how much a worker learns from working in them. Figure 20 evaluates EV_j^{info} at the entry state of each education group—a single deterministic point per group, at the rational prior $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ with zero experience—and reports it, together with its news/depletion decomposition, occupation by occupation, as a share of each occupation's expected one-period log wage.

The most informative occupations are the manual-intensive ones. Across all three education groups the single most informative occupation at entry is the manual anchor Occ_1 ($\boldsymbol{\omega}_1 = (0.01, 0.99)$), and for high-educated workers the next most informative are the other manual-loading clusters, Occ_2 and Occ_7. The pattern is far more pronounced for the high-educated—for whom one year in the manual anchor is worth about 27% of a period's expected wage—than for medium- or low-educated workers. The reason lies in the structure of prior uncertainty: high-educated entrants carry by far the largest prior variance in *manual* skill (Table 2), so an occupation that loads on the manual dimension resolves the most uncertainty for them. For medium- and low-educated workers the manual anchor Occ_1 and the cognitive anchor Occ_0 are the two most informative occupations, and informativeness is uniformly lower. The decomposition also clarifies why the net value is small, or even negative, in several manual occupations for medium- and low-educated workers: the news component is large everywhere, but for these groups it is nearly offset by the depletion term—the option value that the first signal consumes.

Figure 21 re-expresses the same experimentation values as a fraction of each occupation's con-

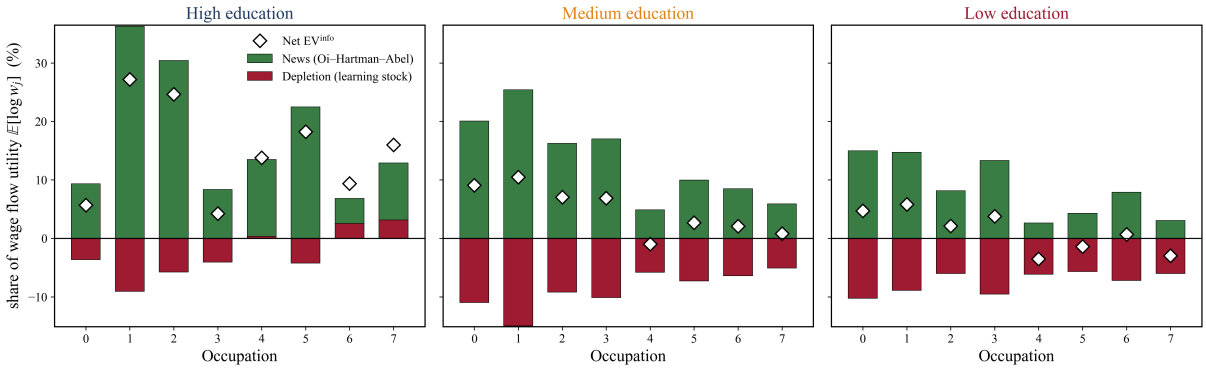


Figure 20: Experimentation value of occupations at entry, by education group. Each panel plots, for one education group, the experimentation value EV_j^{info} of each occupation evaluated at the group's entry state, as a percentage of the occupation's expected one-period log wage, decomposed into the non-negative news (Jensen, green) and non-positive depletion (red) components; white diamonds mark the net value. Occupations are indexed 0–7 as in Table 2 (Occ_0 the cognitive anchor, Occ_1 the manual anchor). Taste-shock scale $\varrho = 0.271$.

ditional value function—the worker's expected lifetime value from choosing that occupation. On this scale the magnitudes are small: even for high-educated entrants the most informative occupation is worth only about 2% of the occupation's lifetime value, and for medium- and low-educated workers the figure is well below 1%. This confirms that, while the experimentation motive matters at the margin for early-career choices, it is a small component of the overall value of an occupation relative to expected wages, amenities, taste shocks, and switching costs.

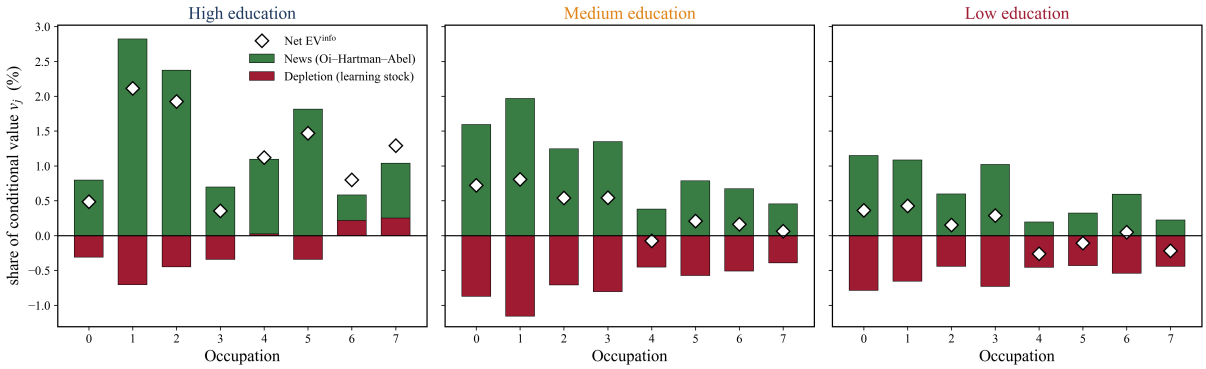


Figure 21: Experimentation value of occupations at entry as a fraction of lifetime value, by education group. Same object as Figure 20, but with EV_j^{info} expressed as a percentage of the occupation's conditional (alternative-specific, lifetime) value v_j rather than of the expected one-period wage; news (Jensen, green) and depletion (red) components with the net value as white diamonds. Taste-shock scale $\varrho = 0.271$.

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 $\boldsymbol{\xi}_i \mid \mathcal{I}_{it} \sim \mathcal{N}(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$.

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

$$y_{jk} = \ell_{jk}\, \theta_{ik}^{a_{jk}}\, H_{jk}(\tau_{ikt}), \qquad H_{jk}(\tau) \;\equiv\; \exp\!\big(b_{jk}\,\tau + c_{jk}\,\tau^2\big), \tag{50}$$

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 $\sum_{k=1}^{K} \omega_{jk} = 1$:

$$Y_{ijt} = \prod_{k=1}^{K} \left(\frac{y_{jk}}{\omega_{jk}}\right)^{\omega_{jk}}. \tag{51}$$

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

$$\log Y_{ijt} = \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} + \sum_{k=1}^{K} \omega_{jk}\, a_{jk}\, \xi_{ik} + \sum_{k=1}^{K} \omega_{jk} \big[b_{jk}\,\tau_{ikt} + c_{jk}\,\tau_{ikt}^2\big] - \sum_{k=1}^{K} \omega_{jk} \log \omega_{jk}. \tag{52}$$

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

$$\mathbb{E}\big[\log Y_{ijt} \,\big|\, \tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}\big] = \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} + \sum_{k=1}^{K} \omega_{jk}\, a_{jk}\, \tilde{\xi}_{ikt} + \sum_{k=1}^{K} \omega_{jk} \big[b_{jk}\,\tau_{ikt} + c_{jk}\,\tau_{ikt}^2\big] - \sum_{k=1}^{K} \omega_{jk} \log \omega_{jk}. \tag{53}$$

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

$$\max_{\{\ell_{jk}\}} \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} \quad \text{s.t.} \quad \sum_{k=1}^{K} \ell_{jk} = 1. \tag{54}$$

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

$$\frac{\omega_{jk}}{\ell_{jk}} = \mu \quad \Rightarrow \quad \ell_{jk} = \frac{\omega_{jk}}{\mu}. \tag{55}$$

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

$$\ell_{jk}^{*} = \omega_{jk}. \tag{56}$$

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 Cobb-Douglas structure makes the belief-dependent terms in expected log output additively separable from $\{\ell_{jk}\}$, so they drop out of the ℓ-optimization and the posterior $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ plays no role in task assignment.

Extensions with CES aggregation or dynamic task assignment break this separability and make the allocation belief-dependent; I discuss these directions in the footnote accompanying the main-text statement of $\ell^*_{jk} = \omega_{jk}$.

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—their overall scale is the price-level numeraire—and any candidate price vector is renormalized so that the dual 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 equilibrium by a damped iteration on prices: each round solves the worker's problem at the current prices, simulates the cross-section those prices induce, and moves prices toward the levels that clear the market. Throughout, workers discount at the constant β and do not internalize the exogenous exit probability δ.

Fix a price vector $\tilde{\boldsymbol{p}}$. I solve the worker's dynamic program once—obtaining the value-function (Smolyak) coefficients and the implied policies—and simulate the economy it induces with common random numbers held fixed across rounds. Workers exit at rate δ and are replaced by entrants drawn from the group priors; after discarding an initial burn-in, this death-and-rebirth turnover erases initial conditions and the simulated panel approximates the ergodic cross-section $\pi^\star(\tilde{\boldsymbol{p}})$.

Occupation outputs are then formed by choice-probability aggregation. Rather than assigning each worker to a single occupation, I let her contribute to every occupation in proportion to her conditional choice probability—written $\mathbb{P}_{ij}(\tilde{\boldsymbol{p}})$ to distinguish it from the occupation price P_j—which is the penalized softmax over the price-dependent choice values,

$$\begin{aligned} \mathbb{P}_{ij}(\tilde{\boldsymbol{p}}) &= \frac{\exp\big(\nu_{ij}(\tilde{\boldsymbol{p}})/\varrho\big)}{\sum_{k\in\mathcal{J}}\exp\big(\nu_{ik}(\tilde{\boldsymbol{p}})/\varrho\big)}, \\ \nu_{ij}(\tilde{\boldsymbol{p}}) &= p_j + \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\xi}}_i + h_j(\boldsymbol{\tau}_i) + \eta_{j,g(i)} - \mathcal{C}\left(j, d_{i,-1}, e_i\right) + \beta\,\mathbb{E}\left[V(\boldsymbol{x}') \mid \boldsymbol{x}_i, j\right], \end{aligned} \tag{57}$$

where the choice value is evaluated at the worker's beliefs, the switching cost enters as the penalty, and the price p_j acts directly through the expected wage. 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 (Jensen) correction $\mathbb{E}[e^\varepsilon] = e^{\varsigma_{\varepsilon j}^2/2}$, is $\overline{W}_{ij}(\tilde{\boldsymbol{p}}) = \exp\big(p_j + \boldsymbol{\lambda}_j'\boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_i) + \tfrac{1}{2}\varsigma_{\varepsilon j}^2\big)$, so occupation output is the probability-weighted sum of these wages deflated by the price,

$$Y_j(\tilde{\boldsymbol{p}}) = \frac{1}{P_j}\sum_i \mathbb{P}_{ij}(\tilde{\boldsymbol{p}})\,\overline{W}_{ij}(\tilde{\boldsymbol{p}}) = \sum_i \mathbb{P}_{ij}(\tilde{\boldsymbol{p}})\,\exp\Big(\boldsymbol{\lambda}_j'\boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_i) + \tfrac{1}{2}\varsigma_{\varepsilon j}^2\Big), \tag{58}$$

with $P_j = e^{p_j}$, so the price cancels and output is measured in efficiency units. Aggregating with probabilities rather than with realized discrete choices smooths how output responds to prices and sharply reduces simulation noise.

Given these outputs, CES inverse demand returns the market-clearing prices $\hat{p}_j = \alpha_j\,(Y/Y_j)^{1/\sigma}$, with aggregate output $Y = \left(\sum_j \alpha_j Y_j^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}$, which I renormalize to the numeraire, $\hat{\boldsymbol{p}} \leftarrow \hat{\boldsymbol{p}}/P(\hat{\boldsymbol{p}})$. A damped step then updates the guess, $\tilde{\boldsymbol{p}} \leftarrow (1-\zeta)\,\tilde{\boldsymbol{p}} + \zeta\log\hat{\boldsymbol{p}}$ with $\zeta \in (0,1)$, and

the iteration repeats until prices stop moving. Because the random numbers are held fixed across rounds, the price-to-output map is smooth and the damped iteration converges reliably; if prices oscillate near the end I stabilize by averaging the last few iterates. The procedure is summarized in the algorithm box below.

Algorithm: Stationary Equilibrium (price tâtonnement)

Inputs: the estimated wage and skill parameters $\widehat{\mathbf{\Gamma}} = \big\{(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g\in\mathcal{G}},\ (p_j, \boldsymbol{\lambda}_j, \boldsymbol{\delta}_j, \varsigma^2_{\varepsilon j})_{j\in\mathcal{J}}\big\}$ and dynamic-choice parameters $\widehat{\mathbf{\Delta}} = \big\{(\eta_{j,g}), \kappa_0\big\}$; 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 $\mathbf{\Phi}^\star$, 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 $\mathbf{\Phi}^{(n)}$ 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 (57) and the output $Y_j = \sum_i \mathbb{P}_{ij} \exp\big(\boldsymbol{\lambda}_j'\boldsymbol{\xi}_i + h_j(\boldsymbol{\tau}_i) + \frac{1}{2}\varsigma^2_{\varepsilon j}\big)$ as in (58).
- **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.** $\tilde{\boldsymbol{p}}^{(n+1)} = (1-\zeta)\,\tilde{\boldsymbol{p}}^{(n)} + \zeta \log \hat{\boldsymbol{p}}$.

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

C.3 Transition Path Algorithm

A transition path is the perfect-foresight response of the economy to an unanticipated technology shock at $t = 0$, linking the pre-shock stationary equilibrium to the post-shock one over a horizon T. The unknown is now the entire path of occupation log prices $\{\tilde{\boldsymbol{p}}_t\}_{t=0}^{T-1}$ that clears the market in every period, under the same fixed weights $\{\alpha_j\}$ and the same per-period numeraire $P(\tilde{\boldsymbol{p}}_t) = 1$ as in the stationary problem, with workers again discounting at the constant β and not internalizing exit. The logic mirrors the stationary algorithm—solve, simulate, aggregate, update—now applied to a price path by backward–forward shooting, with damping $\zeta \in (0, 1)$.

The shock perturbs the task weights $\boldsymbol{\omega}_j$, 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 this technology. Under the "instantaneous" convention it jumps from Λ_{pre} to Λ_{post} at $t = 0$; under a "gradual" convention it diffuses, $\Lambda_t = (1 - s_t)\Lambda_{\text{pre}} + s_t\Lambda_{\text{post}}$ with $s_t = \min(t/t_{\text{shock}}, 1)$, and workers foresee the whole path. In either case the burn-in that builds

the initial cross-section is run under the pre-shock technology Λ_{pre}, so the economy departs from the correct pre-shock stationary distribution.

I first compute the two endpoints—the pre- and post-shock stationary equilibria $(\tilde{\boldsymbol{p}}_{\text{pre}}, \boldsymbol{\Phi}_{\text{pre}})$ and $(\tilde{\boldsymbol{p}}_{\text{post}}, \boldsymbol{\Phi}_{\text{post}})$, where $\boldsymbol{\Phi}$ collects the Smolyak value-function coefficients—and initialize the path by relaxing geometrically from the pre- toward the post-shock prices, with the terminal date pinned at $\tilde{\boldsymbol{p}}_{\text{post}}$. Given a guess for the path, a backward pass solves the value functions by induction from the terminal condition $\boldsymbol{\Phi}_{T-1} = \boldsymbol{\Phi}_{\text{post}}$,

$$\boldsymbol{\Phi}_t = \Pi\Big[\max_{j\in\mathcal{J}}\Big\{r_j(\boldsymbol{x};\tilde{\boldsymbol{p}}_t,\Lambda_t) + \beta\,\mathbb{E}\big[V_{t+1}(\boldsymbol{x}') \mid \boldsymbol{x}, j\big]\Big\}\Big], \qquad t = T-2,\dots,0, \tag{59}$$

where Π is the Chebyshev/Smolyak projection, the flow value uses the period-t prices and technology, and the expectation integrates the Gaussian belief update by Gauss–Hermite quadrature. A forward pass then simulates a single panel with common random numbers: after the pre-shock burn-in, workers face the time-varying policies, prices, and technology, and occupation outputs Y_{jt} are formed by exactly the price-dependent choice-probability aggregation of (57)–(58), applied period by period along the path. Finally, per-period CES inverse demand gives implied prices, which I renormalize to the numeraire and use to update the path with the same damped step as in the stationary case—keeping the terminal date pinned at $\tilde{\boldsymbol{p}}_{\text{post}}$, with an optional Anderson acceleration available. The path is iterated until the largest price change across periods and occupations falls below tolerance. The procedure is summarized in the algorithm box below.

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 $n = 0, 1, \dots$ 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 $t = T-2,\dots,0$, $\boldsymbol{\Phi}_t = \Pi\big[\max_{j\in\mathcal{J}}\{r_j(\boldsymbol{x};\tilde{\boldsymbol{p}}_t,\Lambda_t) + \beta\,\mathbb{E}[V_{t+1}(\boldsymbol{x}') \mid \boldsymbol{x}, j]\}\big]$.
- **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 (57)–(58), 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} through the entry occupational choice. At labor-market entry 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 member of a group enters the same occupation—the one that maximizes the systematic value at the group's prior—and workers diverge only later, through the wage signals they happen to draw and the experience they accumulate. Panel (a) of Figure 22 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 $\boldsymbol{\mu}_g$.

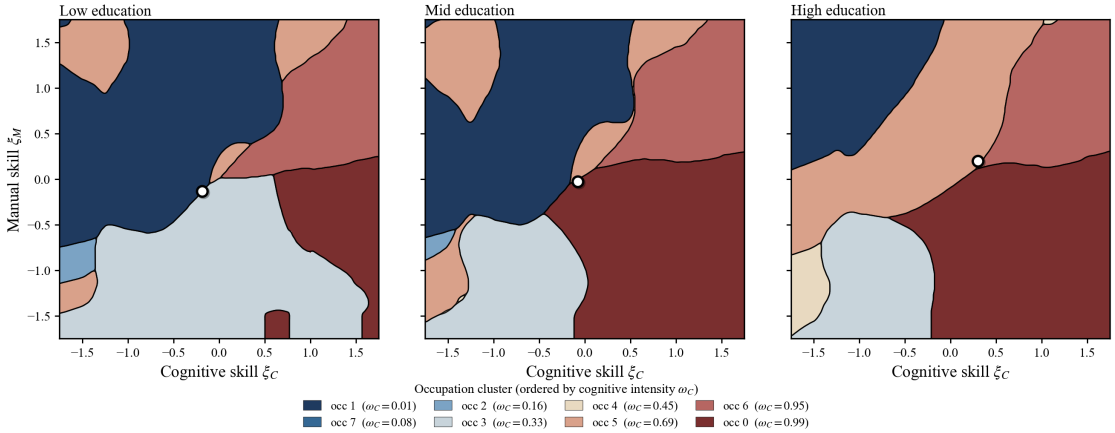
The taste shocks break this degeneracy by adding idiosyncratic noise to the choice in every period. With scale $\varrho > 0$ the entry choice becomes a smooth lottery over occupations, so workers with identical beliefs spread across neighboring clusters and the expected occupational content varies continuously over the belief space. Panel (b) of Figure 22 plots the expected cognitive task intensity $\mathbb{E}[\omega_C]$ of the entry occupation, the smooth counterpart of the sharp partition in Panel (a). The amount of noise is not uniform across workers: Panel (c) reports the probability of the most likely entry occupation, $\max_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\xi}})$, which falls toward $1/8$—a near-uniform lottery over the eight clusters—along the indifference ridges where a worker is nearly indifferent between adjacent occupations, and rises toward one where comparative advantage is strong. The taste shocks thus matter most for workers with balanced skills and least for specialists, and over the life cycle they let the model rationalize occupational choices and moves that its systematic component alone would not predict.

C.5 Switching Costs: Lifecycle Wage Profiles

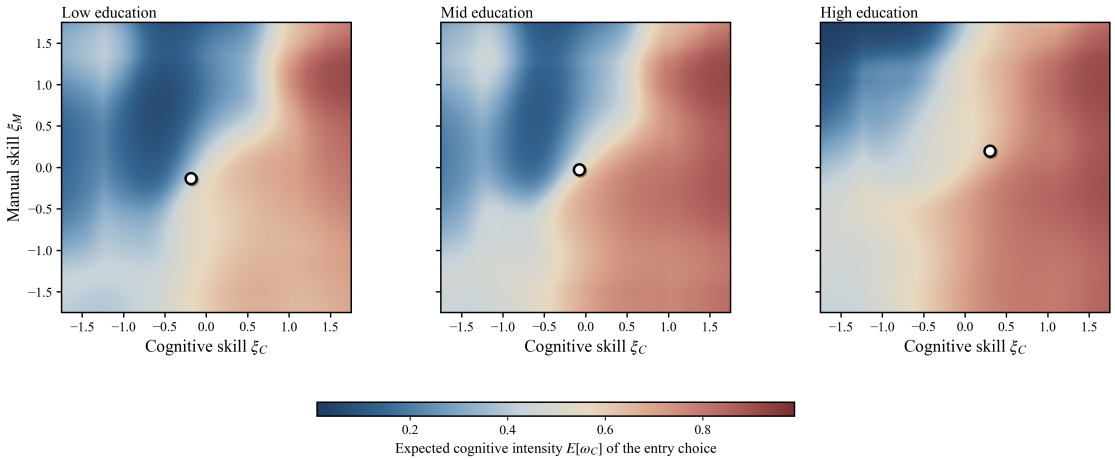
This appendix reports the lifecycle wage profiles behind the switching-cost counterfactuals of Section 7.3, which the main text summarizes by match-quality percentile. Figure 23 shows, for each switching-cost level κ/κ_0, the average wage gain over the life cycle: Panel (a) is the gain from moving to 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 losses under aggressive reductions widen with experience, the signature of a human-capital loss: churning workers keep restarting their task-specific experience, so the damage compounds over the career rather than being repaired by it.

C.6 Forced Exploration at the Beginning of the Career

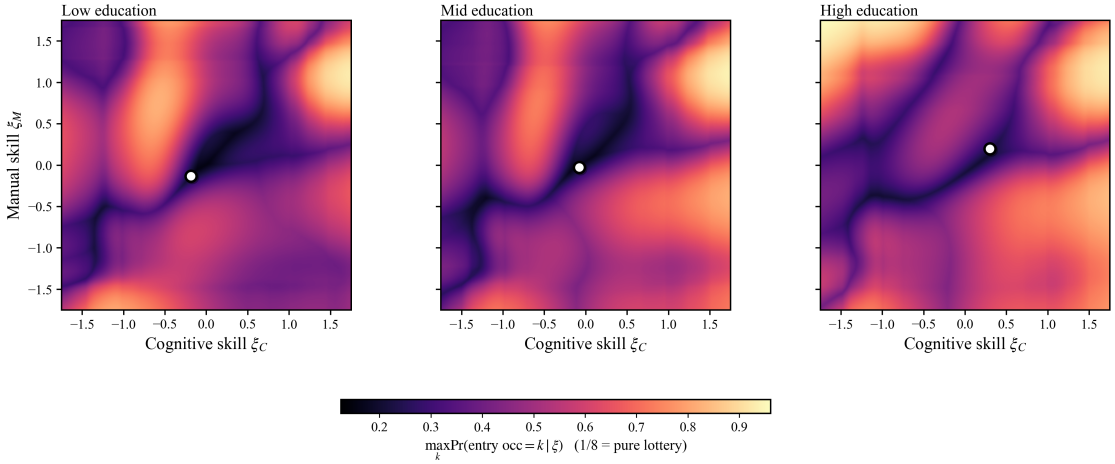
The switching-cost counterfactual shows that frictions to mobility are a key driver of persistent mismatch, but even with low switching costs the baseline information structure limits how quickly workers can sort optimally. A complementary policy question is whether a targeted intervention at the very start of the career—before any experience or wage signal has been observed—can accelerate learning and improve lifetime sorting. I study a one-period forced exploration policy: in their first period, workers are assigned to the occupation that maximizes



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

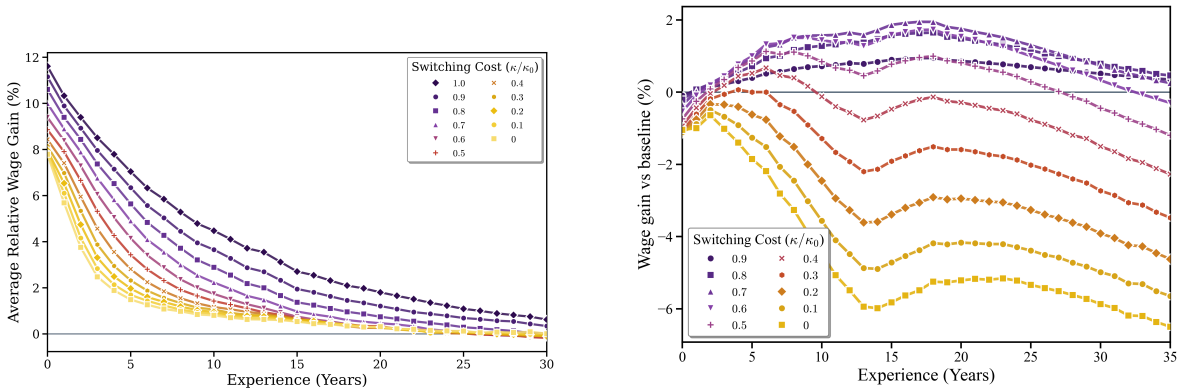


(b) Expected cognitive task intensity of the entry occupation, $\mathbb{E}[\omega_C]$



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

Figure 22: 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 23: 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.

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,

$$IG_{jg}^{\text{entry}} \equiv I(\xi_i; s_{ij1} \mid g) = \frac{1}{2}\log\left(1 + \frac{\boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j}{\varsigma_{\varepsilon j}^2}\right), \tag{60}$$

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[26]. Each worker is assigned 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—$\boldsymbol{\lambda}_0 = (1, 0)$ and $\boldsymbol{\lambda}_1 = (0, 1)$—so their wage is an undiluted signal about one skill,

[26] 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, $s_{ij1} = \lambda_j' \xi_i + \varepsilon_{ij1}$, $\xi_i \mid g \sim N(\mu_g, \Sigma_g)$, $\varepsilon_{ij1} \sim N(0, \varsigma_{\varepsilon j}^2)$, 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

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

which maximizes the signal-to-noise ratio of the first wage observation relative to the worker's prior uncertainty. Figure 25 shows the occupational shares at entry, after one year, and after ten years, under forced exploration and in the baseline.

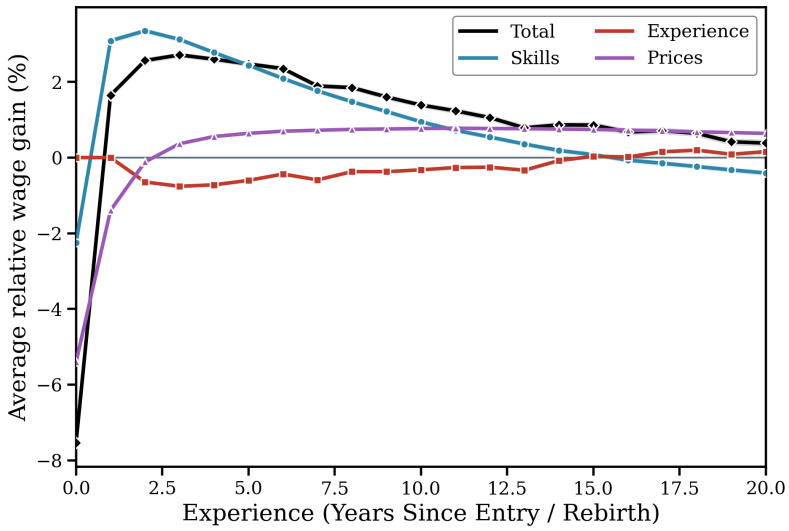


Figure 24: Lifecycle wage gain from forced exploration relative to the baseline economy, decomposed into channels: Total (black), Skills (blue), Experience (red), and Prices (purple). Horizontal axis: years of potential experience since entry; vertical axis: average relative wage gain (percent).

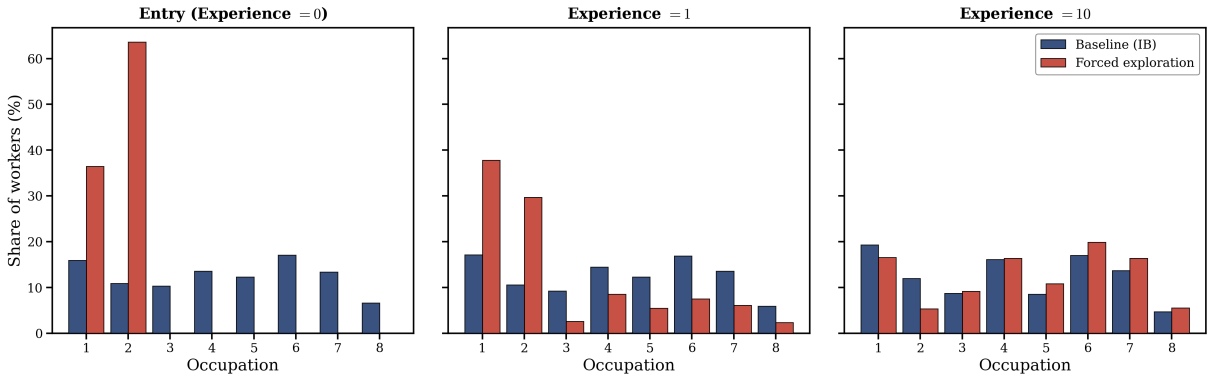


Figure 25: Share of workers in each occupation cluster (vertical axis) under forced exploration (red) and in the baseline imperfect-information economy (blue), at entry, after one year, and after ten years of experience. Occupations are numbered 1–8 for Occ_0–Occ_7.

At entry, workers are concentrated in the two anchor occupations—roughly two thirds in the manual anchor and one third in the cognitive anchor—a sharp departure from the baseline, where entrants are spread across all eight clusters (Figure 25). This forced bunching triggers a strong general-equilibrium price response: excess labor supply in the assigned occupations drives their prices down while prices elsewhere rise, and it is this price movement, not bad sorting, that produces the sharp wage loss on impact—about 7.6% at entry, of which the Price channel alone accounts for more than 5 percentage points (Figure 24). From the second period workers choose freely, and they move fast: with much sharper beliefs about their comparative advantage, a large fraction immediately transitions into the occupation that fits their skills. The Skills channel jumps to about +3.4% within two years and stays above the baseline for most of the career, and the total wage gain turns positive after the first year, peaking at roughly

+2.7% around the third. The Price channel, initially deeply negative, normalizes within a few years and settles slightly positive as the economy adjusts to the new sorting pattern; the Experience channel is mildly negative, since the assigned first occupation is not where these workers would have started accumulating task-specific human capital. The occupational distribution converges back toward the baseline—by year ten the aggregate shares look similar—but workers are now far 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. Because the first wage signal is the most valuable—it arrives when prior uncertainty is highest—even a single period of optimally chosen exploration substantially improves the quality of all subsequent sorting decisions. A policy that temporarily redirects workers toward their most informative occupation at labor market entry can thus be both productivity- and welfare-enhancing, at the cost of short-run wage losses that the general equilibrium and individual wage gains quickly offset.

D Appendix: Estimation and Numerical Methods

D.1 EM Algorithm: Computational Details

The wage equation estimated by EM is the static cross-section of the structural model. For worker i employed in occupation j at date t,

$$w_{ijt} \;=\; p_j \;+\; \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i \;+\; h_j(\boldsymbol{\tau}_{it}) \;+\; \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}\big(0, \varsigma_j^2\big), \tag{61}$$

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 $\boldsymbol{\tau}_{it}$ into tenure bins with corresponding dummy coefficients to capture the non-linear wage progression flexibly. Workers belong to observed groups $g(i) \in \{1, \ldots, G\}$ defined by education. The skill prior is conditionally Gaussian within group,

$$\boldsymbol{\xi}_i \mid g(i) = g \;\sim\; \mathcal{N}\big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g\big). \tag{62}$$

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

$$\boldsymbol{\Gamma}^q = \Big\{ (\boldsymbol{\mu}_g^q, \boldsymbol{\Sigma}_g^q)_{g=1}^G,\; (p_j^q, \boldsymbol{\lambda}_j^q, \boldsymbol{\delta}_j^q, \varsigma_j^{2,q})_{j \in \mathcal{J}} \Big\},$$

with q indexing the EM iteration.

D.1.1 E-step

Conditional on $\mathbf{\Gamma}^q$ and on $g(i)$, the posterior of $\boldsymbol{\xi}_i$ is multivariate normal. Define the residual after netting out occupation prices and the tenure profile,

$$r_{it}^q \;:=\; w_{ijt} \;-\; p_{j(i,t)}^q \;-\; h_{j(i,t)}^q(\tau_{it}), \tag{63}$$

Completing the square in $\boldsymbol{\xi}_i$ in the product of the wage likelihood implied by (61) and the group prior (62) delivers

$$\boldsymbol{V}_i^q = \left((\boldsymbol{\Sigma}_{g(i)}^q)^{-1} + \sum_{t=t_0(i)}^{T(i)} \frac{1}{\varsigma_{j(i,t)}^{2,q}} \boldsymbol{\lambda}_{j(i,t)}^q (\boldsymbol{\lambda}_{j(i,t)}^q)' \right)^{-1}, \tag{64}$$

$$\boldsymbol{\zeta}_i^q = \boldsymbol{V}_i^q \left((\boldsymbol{\Sigma}_{g(i)}^q)^{-1} \boldsymbol{\mu}_{g(i)}^q + \sum_{t=t_0(i)}^{T(i)} \frac{\boldsymbol{\lambda}_{j(i,t)}^q \, r_{it}^q}{\varsigma_{j(i,t)}^{2,q}} \right), \tag{65}$$

$$\boldsymbol{\xi}_i \mid \text{data}, \mathbf{\Gamma}^q, g(i) \sim \; \mathcal{N}\big(\boldsymbol{\zeta}_i^q, \, \boldsymbol{V}_i^q\big). \tag{66}$$

The code stores the worker-specific pair $(\boldsymbol{\zeta}_i^q, \boldsymbol{V}_i^q)$ at the end of every E-step. Equations (64)–(65) are themselves a $|\mathcal{K}|$-dimensional precision-weighted linear system per worker, solved in closed form by inverting the precision matrix in (64).

The Q-function decomposes additively into a prior term and an occupation-by-occupation wage term. The wage term involves only the first and second posterior moments of $\boldsymbol{\xi}_i$: for any occupation-specific coefficient vector $\boldsymbol{\lambda}_j$,

$$E_q\big[(\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i)^2\big] \;=\; (\boldsymbol{\lambda}_j' \boldsymbol{\zeta}_i^q)^2 \;+\; \boldsymbol{\lambda}_j' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j, \tag{67}$$

which is what propagates the latent-skill uncertainty into the M-step.

D.1.2 M-step

The M-step splits into closed-form updates for the prior and a closed-form weighted least-squares update for each occupation's wage parameters.

Prior moments. Maximizing the expected complete-data log-likelihood over $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ gives the standard updates

$$\boldsymbol{\mu}_g^{q+1} = \frac{1}{N_g} \sum_{i:\, g(i)=g} \boldsymbol{\zeta}_i^q, \tag{68}$$

$$\boldsymbol{\Sigma}_g^{q+1} = \frac{1}{N_g} \sum_{i:\, g(i)=g} \Big[\boldsymbol{V}_i^q + (\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_g^{q+1})(\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_g^{q+1})' \Big]. \tag{69}$$

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

To pin down the location of the latent factor, we enforce the anchor $\sum_g \pi_g \boldsymbol{\mu}_g = \mathbf{0}$ 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 compensating the occupation intercept,

$$\boldsymbol{\mu}_g^{q+1} \;\leftarrow\; \boldsymbol{\mu}_g^{q+1} - \boldsymbol{a}^{q+1}, \qquad p_j^{q+1} \;\leftarrow\; p_j^{q+1} + (\boldsymbol{\lambda}_j^{q+1})' \boldsymbol{a}^{q+1}. \tag{70}$$

This translation leaves the wage equation invariant and removes the location indeterminacy of $\boldsymbol{\xi}_i$.

Wage equation, occupation by occupation. For each occupation j, stack the regressors for the N_j worker–year observations assigned to j into

$$\boldsymbol{W}_j \;=\; \big[\,\mathbf{1},\;\; \boldsymbol{\zeta}^{\,q},\;\; \boldsymbol{D}(\boldsymbol{\tau})\,\big] \;\in\; \mathbb{R}^{N_j \times (1+|\mathcal{K}|+|\mathcal{B}|)}, \tag{71}$$

where $\boldsymbol{\zeta}^{\,q}$ is the posterior mean repeated across the worker's observations, and $\boldsymbol{D}(\boldsymbol{\tau})$ collects the tenure bin indicators with total dimension $|\mathcal{B}|$. The corresponding coefficient vector is

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

The first-order conditions of the Q-function then take the form of augmented normal equations

$$\boldsymbol{S}_j\, \boldsymbol{\vartheta}_j^{\,q+1} \;=\; \boldsymbol{b}_j, \qquad \boldsymbol{S}_j \;=\; \boldsymbol{W}_j'\boldsymbol{W}_j \;+\; \boldsymbol{\Delta}_j^{\,q} \;+\; \rho\, \boldsymbol{I}, \qquad \boldsymbol{b}_j \;=\; \boldsymbol{W}_j'\boldsymbol{y}_j, \tag{72}$$

where $\boldsymbol{y}_j$ stacks the corresponding log wages and $\rho = 10^{-6}$ is a ridge for numerical stability. The matrix $\boldsymbol{\Delta}_j^{\,q}$ implements the posterior-variance correction of (67): it is zero everywhere except in the $K \times K$ block corresponding to $\boldsymbol{\lambda}_j$, which takes the value $\sum_t \boldsymbol{V}_{i(t)}^{\,q}$ summed over the observations assigned to j. Solving (72) delivers the full vector $(p_j^{\,q+1}, \boldsymbol{\lambda}_j^{\,q+1}, \boldsymbol{\delta}_j^{\,q+1})$ jointly.

Conformably with the parameter blocks of $\boldsymbol{\vartheta}_j = [p_j, \boldsymbol{\lambda}_j', \boldsymbol{\delta}_j']'$, and writing $\boldsymbol{\zeta}_{(j)}$ and $\boldsymbol{D}_{(j)}$ for the stacked posterior means and tenure indicators restricted to observations in j, the cross-product $\boldsymbol{S}_j$ has the block form

$$\boldsymbol{S}_j \;=\; \begin{pmatrix} N_j & \boldsymbol{\zeta}_{(j)}'\mathbf{1} & \boldsymbol{D}_{(j)}'\mathbf{1} \\ \boldsymbol{\zeta}_{(j)}'\mathbf{1} & \boldsymbol{\zeta}_{(j)}'\boldsymbol{\zeta}_{(j)} + \boldsymbol{\Delta}_j^{\,q} & \boldsymbol{\zeta}_{(j)}'\boldsymbol{D}_{(j)} \\ \boldsymbol{D}_{(j)}'\mathbf{1} & \boldsymbol{D}_{(j)}'\boldsymbol{\zeta}_{(j)} & \boldsymbol{D}_{(j)}'\boldsymbol{D}_{(j)} \end{pmatrix} \;+\; \rho\, \boldsymbol{I}, \qquad \boldsymbol{b}_j \;=\; \begin{pmatrix} \mathbf{1}'\boldsymbol{y}_j \\ \boldsymbol{\zeta}_{(j)}'\boldsymbol{y}_j \\ \boldsymbol{D}_{(j)}'\boldsymbol{y}_j \end{pmatrix}. \tag{73}$$

The posterior-variance correction $\boldsymbol{\Delta}_j^{\,q} = \sum_{t \in j} \boldsymbol{V}_{i(t)}^{\,q}$ enters only the $(\boldsymbol{\lambda}, \boldsymbol{\lambda})$ diagonal block. The linear system $\boldsymbol{S}_j \boldsymbol{\vartheta}_j^{\,q+1} = \boldsymbol{b}_j$ is solved per occupation in closed form.

The inclusion of the E-step posterior $\boldsymbol{\zeta}^{\,q}$ as a regressor in $\boldsymbol{W}_j$ is the control-function correction for selection on beliefs that the identification argument of Section 5.1 prescribes: regressing log wages on $(\mathbf{1}, \boldsymbol{\zeta}^{\,q}, \boldsymbol{D}(\boldsymbol{\tau}))$ per occupation absorbs endogenous occupational choice through the sufficient history, and 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 `standard` mode we set $\boldsymbol{C} = \boldsymbol{I}_K$, i.e. $\boldsymbol{\lambda}_{j_C} = (1, 0)$ and $\boldsymbol{\lambda}_{j_M} = (0, 1)$. In the alternative `task_based` mode we set $\boldsymbol{\lambda}_{j_k} = \boldsymbol{\omega}_{j_k}$ for $k \in \{C, M\}$, where $\boldsymbol{\omega}_{j_k}$ is the task-content vector of the anchor occupation. Either 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 (72) and solve the reduced system for $(p_j, \boldsymbol{\delta}_j)$ only.

Sign restrictions. The skill loadings $\boldsymbol{\lambda}_j \geq \mathbf{0}$ are imposed by default. Because the system (72)

is quadratic in ϑ_j with the inequality constraint active on at most 2^K subsets of components, the constrained M-step is solved by closed-form active-set enumeration: for each of the 2^K candidate active sets we fix the corresponding entries of λ_j at zero, solve the reduced normal equations for the remaining free parameters, and retain the candidate with the highest Q-value among those that satisfy $\lambda_j \geq \mathbf{0}$.

Residual variance. Given the updated $(p_j^{q+1}, \lambda_j^{q+1}, \delta_j^{q+1})$, the variance update is

$$\varsigma_j^{2,q+1} = \frac{1}{N_j} \sum_{(i,t):\, j(i,t)=j} \left[\left(r_{it}^{q+1} - (\lambda_j^{q+1})' \zeta_i^q \right)^2 + (\lambda_j^{q+1})' V_i^q \lambda_j^{q+1} \right], \tag{74}$$

where r_{it}^{q+1} is the residual from (63) evaluated at the new wage parameters. The second term inside the bracket is the variance-of-the-projection correction from (67) and ensures that posterior uncertainty about ξ_i is correctly transmitted into ς_j^2.

Cascade of closed-form linear systems. Taken together, the M-step is a cascade of small closed-form operations, with no inner numerical optimization:

1. For each non-anchor occupation $j \in \mathcal{J}$: solve the linear system (72)–(73) for ϑ_j^{q+1}.
2. For each anchor occupation $j \in \{j_C, j_M\}$: λ_j is pinned by the normalization, so substitute it out of (72) and solve the residual linear system for (p_j, δ_j).
3. For sign-constrained variants: enumerate the 2^K active-set patterns on $\lambda_j \geq \mathbf{0}$, solve the reduced linear system in closed form for each pattern, and retain the feasible candidate with the highest Q-value.
4. For each j: closed-form residual variance $\varsigma_j^{2,q+1}$ via (74).
5. For each group g: closed-form posterior moments $(\mu_g^{q+1}, \Sigma_g^{q+1})$ via (68)–(69), followed by the location-anchor translation (70).
6. Group weights π_g are held fixed at their sample frequencies.

Combined with the per-worker linear system in the E-step, every iteration of the algorithm reduces to a sequence of small dense linear solves and weighted averages.

D.1.3 Convergence and Multi-Start Strategy

The algorithm iterates (64)–(66) and (68)–(74) 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}. The maximum number of iterations is set to 2,000.

Each Python process runs a single start ($N_{\text{starts}} = 1$). Multi-start aggregation is implemented externally via SLURM array jobs that launch independent processes from different seeds. The selected estimator is the run that delivers the highest observed-data log-likelihood, computed as described next.

D.1.4 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 $\boldsymbol{\Gamma}$,

$$\tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \;=\; \int_{\mathbb{R}^{|\mathcal{K}|}} \left\{ \prod_{t=1}^{T_i} \phi\Big(w_{ij(i,t)t};\, p_{j(i,t)} + \boldsymbol{\lambda}'_{j(i,t)}\boldsymbol{\xi}_i + h_{j(i,t)}(\boldsymbol{\tau}_{it}),\, \varsigma^2_{j(i,t)}\Big) \right\} \phi\big(\boldsymbol{\xi}_i; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\big)\, d\boldsymbol{\xi}_i, \quad (75)$$

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

We approximate (75) 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

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

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

$$\log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \;\approx\; \operatorname{logsumexp}\big(\{S_i^{(q)}(\boldsymbol{\Gamma})\}_{q=1}^{Q}\big) - \log Q, \quad (76)$$

which is stable under reweighting at extreme parameter values. Fixing the underlying standard-normal draws across candidate parameters (common random numbers) ensures that differences in $\sum_i \log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma})$ across starts reflect fit rather than simulation noise.

D.2 NFXP Algorithm: Computational Details

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

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

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

with 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 $d_{t-1} \in \mathcal{J}$. Choosing $m \in \mathcal{J}$ incurs a contemporaneous switching cost $\mathcal{C}(m, d_{t-1})$ as in (25), which enters only the current deterministic index and is independent of belief updating or tenure accumulation. Conditional on choosing m at t, the realized log wage in (14) delivers the scalar signal

$$s_{mt} \;\equiv\; \log W_{mt} - p_{mt} - h_m(\boldsymbol{\tau}_t) = \boldsymbol{\lambda}'_m \boldsymbol{\xi} + \varepsilon_{mt}, \qquad \varepsilon_{mt} \sim \mathcal{N}(0, \varsigma^2_{\varepsilon m}), \quad (77)$$

and the Kalman recursion reads

$$\tilde{\boldsymbol{\xi}}_{t+1} = \tilde{\boldsymbol{\xi}}_t + \boldsymbol{K}_{mt}\, z_{mt}, \qquad z_{mt} \equiv s_{mt} - \boldsymbol{\lambda}_m' \tilde{\boldsymbol{\xi}}_t \sim \mathcal{N}\big(0,\, s_{zz}^{(m)}\big), \tag{78}$$

$$\tilde{\boldsymbol{\Sigma}}_{t+1} = \big(\boldsymbol{I}_{|\mathcal{K}|} - \boldsymbol{K}_{mt}\boldsymbol{\lambda}_m'\big)\tilde{\boldsymbol{\Sigma}}_t, \qquad \boldsymbol{K}_{mt} = \frac{\tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_m}{s_{zz}^{(m)}}, \tag{79}$$

$$s_{zz}^{(m)} \equiv \boldsymbol{\lambda}_m' \tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_m + \varsigma_{\varepsilon m}^2, \qquad \boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\omega}_m. \tag{80}$$

The next-period covariance $\tilde{\boldsymbol{\Sigma}}_{t+1}$ and the next-period tenures $\boldsymbol{\tau}_{t+1}$ are deterministic given $(\boldsymbol{x}_t, m)$; the only random object is the scalar innovation z_{mt}, with predictive variance $s_{zz}^{(m)}$. Because ε_{mt} is the only stochastic shock between two consecutive choices, the next posterior mean $\tilde{\boldsymbol{\xi}}_{t+1}$ moves only along the direction $\boldsymbol{K}_{mt} \propto \tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_m$, the rank-one image of the loadings of the chosen occupation; equivalently, its ex-ante covariance $\boldsymbol{K}_{mt}\, s_{zz}^{(m)}\, \boldsymbol{K}_{mt}'$ is rank one. Hence, although $\tilde{\boldsymbol{\xi}}_{t+1} \in \mathbb{R}^{|\mathcal{K}|}$, its conditional distribution is supported on a one-dimensional affine subspace of $\mathbb{R}^{|\mathcal{K}|}$. Continuation expectations therefore reduce to scalar integrals over z_{mt}, regardless of the number of skills $|\mathcal{K}|$.

Univariate Gauss–Hermite quadrature. For any measurable functional G of the next-period state, the change of variables $u = z/\sqrt{2\, s_{zz}^{(m)}}$ maps the Gaussian density into the Hermite weight e^{-u^2}:

$$\mathbb{E}\big[G(\tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, m) \,\big|\, \boldsymbol{x}_t, m\big] \;=\; \pi^{-1/2} \int_{\mathbb{R}} G\Big(\tilde{\boldsymbol{\xi}}_t + \sqrt{2\, s_{zz}^{(m)}}\, \boldsymbol{K}_{mt}\, u,\; \tilde{\boldsymbol{\Sigma}}_{t+1},\; \boldsymbol{\tau}_t + \boldsymbol{\omega}_m,\; m\Big)\, e^{-u^2}\, du. \tag{81}$$

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

$$\mathbb{E}\big[G(\boldsymbol{x}_{t+1}) \,\big|\, \boldsymbol{x}_t, m\big] \;\approx\; \pi^{-1/2} \sum_{r=1}^{N} w_r\, G\Big(\tilde{\boldsymbol{\xi}}_t + \sqrt{2\, s_{zz}^{(m)}}\, \boldsymbol{K}_{mt}\, u_r,\; \tilde{\boldsymbol{\Sigma}}_{t+1},\; \boldsymbol{\tau}_t + \boldsymbol{\omega}_m,\; m\Big). \tag{82}$$

The cost is therefore N function evaluations per $(\boldsymbol{x}_t, m)$ 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. Conditioning on the lagged occupation j simplifies the fixed-point computation because the switching cost factors out cleanly: only the continuous block $\boldsymbol{x} \equiv (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau})$ remains as the argument of the value function. Let

$$W_j(\boldsymbol{x}) \;\equiv\; W(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1} = j), \qquad j \in \mathcal{J},$$

denote the integrated value when the lagged occupation is j. For a prospective choice $m \in \mathcal{J}$, the deterministic index combines the expected log wage, the switching cost, and the discounted

continuation value of the integrated value function evaluated at the next-period state:[27]

$$\begin{aligned} v_{m|j}(\boldsymbol{x}) \;=\; & \underbrace{p_m + \boldsymbol{\lambda}_m' \tilde{\boldsymbol{\xi}} + h_m(\boldsymbol{\tau})}_{\text{expected log wage}} + \eta_{m,g} - \mathcal{C}(m,j) \\ & + \beta\, \mathbb{E}\big[\, W_m(\boldsymbol{x}') \;\big|\; \boldsymbol{x}, m \,\big], \end{aligned} \tag{83}$$

where the conditional expectation is taken over the scalar innovation z_{mt} and admits the exact one-dimensional integral representation in (81); the next-period covariance $\tilde{\boldsymbol{\Sigma}}'$ and tenures $\boldsymbol{\tau}' = \boldsymbol{\tau} + \boldsymbol{\omega}_m$ are deterministic given $(\boldsymbol{x}, m)$.

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 (82) for the continuation expectation in (83) and writing the deterministic flow utility as $r_m(\boldsymbol{x}) \equiv p_m + \boldsymbol{\lambda}_m' \tilde{\boldsymbol{\xi}} + h_m(\boldsymbol{\tau}) + \eta_{m,g}$, the integrated Bellman equation reads, for each $j \in \mathcal{J}$,

$$\begin{aligned} W_j(\boldsymbol{x}) \;=\; & \varrho \log \sum_{m \in \mathcal{J}} \exp\Bigg(\frac{1}{\varrho} \Bigg[r_m(\boldsymbol{x}) - \mathcal{C}(m,j) \\ & + \beta\, \pi^{-1/2} \sum_{r=1}^{N} w_r\, W_m\Big(\tilde{\boldsymbol{\xi}} + \sqrt{2\, s_{zz}^{(m)}}\, \boldsymbol{K}_m\, u_r,\; \tilde{\boldsymbol{\Sigma}}',\; \boldsymbol{\tau} + \boldsymbol{\omega}_m \Big) \Bigg] \Bigg), \qquad \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{84}$$

Equation (84) makes the fixed-point structure transparent: W_j on the left-hand side appears under the log-sum-exp aggregator on the right, with the continuation values W_m evaluated at the N Gauss–Hermite shifts of the posterior mean. Stacking $W(\boldsymbol{x}) = \big(W_1(\boldsymbol{x}), \ldots, W_{|\mathcal{J}|}(\boldsymbol{x})\big)'$, the integrated Bellman operator $\mathcal{T}$ applies (84) componentwise; by Lemma 2 it is a contraction of modulus at most β, with a unique fixed point.

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 $W_j(\cdot)$ on a sparse grid using the Smolyak method (Malin et al. 2011) with Chebyshev polynomials.

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). For a given multi-index $\mathbf{i}$, the tensor-product interpolant for W_j is

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

[27] The stationary equilibrium of Section 4.6 uses the same constant discount factor β; the exogenous exit probability δ is not internalized by workers and enters only the law of motion for the stationary distribution, so the algorithm below is unchanged. See Lemma 2.

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

$$\widehat{W}_j^{d,\mu}(\boldsymbol{x}) = \sum_{d \le |\mathbf{i}| \le d+\mu} (-1)^{d+\mu-|\mathbf{i}|} \begin{pmatrix} d-1 \\ d+\mu-|\mathbf{i}| \end{pmatrix} W_j^{\mathbf{i}}(\boldsymbol{x}; \boldsymbol{\phi}^{(j,\mathbf{i})}). \tag{85}$$

Fixed point in coefficient space. Substituting the Smolyak interpolant $\widehat{W}_m^{d,\mu}(\boldsymbol{x}) = \boldsymbol{\phi}^{(m)\prime}\mathbf{T}(\boldsymbol{x})$ for the unknown continuation value W_m inside (84) recasts the integrated Bellman equation as a fixed point in the finite-dimensional Smolyak coefficients $\boldsymbol{\Phi} \equiv [\boldsymbol{\phi}^{(1)}, \ldots, \boldsymbol{\phi}^{(|\mathcal{J}|)}] \in \mathbb{R}^{M \times |\mathcal{J}|}$, where M is the number of Smolyak basis functions. For each $j \in \mathcal{J}$,

$$\begin{aligned} \boldsymbol{\phi}^{(j)\prime}\mathbf{T}(\boldsymbol{x}) \;= \varrho \log \sum_{m \in \mathcal{J}} \exp\Bigg(\frac{1}{\varrho} \Bigg[& r_m(\boldsymbol{x}) - \mathcal{C}(m,j) \\ & + \beta\, \pi^{-1/2} \sum_{r=1}^{N} w_r\, \boldsymbol{\phi}^{(m)\prime}\mathbf{T}\Big(\tilde{\boldsymbol{\xi}} + \sqrt{2\, s_{zz}^{(m)}}\, \boldsymbol{K}_m\, u_r,\; \tilde{\boldsymbol{\Sigma}}',\; \boldsymbol{\tau} + \boldsymbol{\omega}_m \Big) \Bigg] \Bigg), \qquad \forall\, \boldsymbol{x} \in \mathcal{X},\; j \in \mathcal{J}. \end{aligned} \tag{86}$$

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 (86) 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 $\mathbf{T}_{m,g} = T_m(\boldsymbol{y}^{(g)})$ 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 (86) into the matrix $\widetilde{\mathbf{W}}(\boldsymbol{\Phi}) \in \mathbb{R}^{|\mathcal{J}| \times G}$ with entries $\widetilde{\mathbf{W}}_{jg}(\boldsymbol{\Phi}) = \widehat{W}_j(\boldsymbol{x}^{(g)}; \boldsymbol{\Phi})$, the fixed-point map in coefficient space is

$$\boldsymbol{\Phi} \;=\; \mathcal{F}(\boldsymbol{\Phi}) \;\equiv\; \mathbf{T}'\, \widetilde{\mathbf{W}}(\boldsymbol{\Phi})'. \tag{87}$$

We compute the fixed point $\boldsymbol{\Phi}^\star$ of (87) 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 (87) by successive substitution on the Smolyak coefficient matrix $\boldsymbol{\Phi}$. Each iteration evaluates the right-hand side of (86) at every grid point, takes a Gauss–Hermite expectation over next-period beliefs, and updates the coefficients by collocation. 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 (87).

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 $m \in \mathcal{J}$:
- Update next-period belief covariance $\tilde{\boldsymbol{\Sigma}}'(\boldsymbol{x}^{(g)}, m)$ via Kalman update.
- Build N shifted next-state nodes $\boldsymbol{x}'_r$ via the Gauss–Hermite map [(82)].
- Evaluate Smolyak interpolant $\widehat{W}_m(\boldsymbol{x}'_r; \boldsymbol{\Phi}^{(n)}) = \boldsymbol{\phi}^{(n,m)\prime}\mathbf{T}(\boldsymbol{x}'_r)$.
- Form conditional values $v_{m|j}(\boldsymbol{x}^{(g)})$ for all $j \in \mathcal{J}$ [(86)].

Aggregate via log-sum-exp: $\widetilde{W}_j(\boldsymbol{x}^{(g)}) = \varrho \log \sum_m \exp\big(v_{m|j}(\boldsymbol{x}^{(g)})/\varrho\big)$.

Collocation update: $\boldsymbol{\Phi}^{(n+1)} = \mathbf{T}'\,\widetilde{\mathbf{W}}(\boldsymbol{\Phi}^{(n)})'$ [(87)].

Convergence: if $\big\|\boldsymbol{\Phi}^{(n+1)} - \boldsymbol{\Phi}^{(n)}\big\|_F < \varepsilon$, exit.

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

Two features of the implementation are worth highlighting. First, the whole loop is compiled into a single JAX while-loop, which makes it fast and lets gradients of the outer log-likelihood flow through the converged fixed point without finite differences. Second, the construction of $\widetilde{\mathbf{W}}(\boldsymbol{\Phi}^{(n)})$ is vectorized with JAX's `vmap` over the quadrature node, the prospective choice, the previous occupation, and the grid point, so each step reduces to one batched basis evaluation, one batched log-sum-exp, and one multiplication by $\mathbf{T}'$. Convergence is fast because the Bellman operator is a contraction (Lemma 2); I have never needed Anderson acceleration or Newton–Krylov methods. Default tolerance is $\varepsilon = 10^{-8}$ and the default cap is $K_{\max} = 5000$ (overridable from the CLI to 1500).

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

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

which avoids overflow when one alternative dominates.

D.2.3 Policy Functions, CCPs, and Simulation

Given $\{\widehat{W}_j^{d,\mu}\}_{j \in \mathcal{J}}$, the CCPs at $(\boldsymbol{x}, j)$ follow from (88). To simulate a worker's path:

1. Given $(\tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, j)$, compute CCPs $\{\mathbb{P}(d_t = m \mid \boldsymbol{x}_t, j)\}_{m \in \mathcal{J}}$ and draw d_t (or use the argmax for policy).

2. Draw a wage residual $\varepsilon_{d_t,t} \sim \mathcal{N}(0, \varsigma^2_{d_t})$ 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 (77) is $s_{d_t t} = \log W_{d_t t} - p_{d_t t} - h_{d_t}(\boldsymbol{\tau}_t)$.

3. Update beliefs via (78)–(80) using $m = d_t$.

4. Update tenures: $\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\omega}_{d_t}$ (cf. (10)).

5. Update the discrete state: $j \leftarrow d_t$ (next period's lagged occupation).

D.2.4 Conditional Choice Probabilities and Likelihood

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

$$\mathbb{P}\big(d_t = m \mid \boldsymbol{x}, j; \boldsymbol{\Theta}\big) \;=\; \frac{\exp\big(v_{m|j}(\boldsymbol{x})/\varrho\big)}{\sum_{\ell \in \mathcal{J}} \exp\big(v_{\ell|j}(\boldsymbol{x})/\varrho\big)}, \qquad m \in \mathcal{J}, \tag{88}$$

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

Off-grid evaluation. The fixed-point algorithm of Section D.2.2 returns the Smolyak coefficient matrix $\boldsymbol{\Phi}^\star$, but the likelihood is evaluated at the observed worker states $\boldsymbol{x}_{it}$ rather than at the sparse-grid nodes. At each $\boldsymbol{x}_{it}$ we read off $\widehat{W}_m(\boldsymbol{x}_{it}; \boldsymbol{\Phi}^\star) = \boldsymbol{\phi}^{(m)\prime} \mathbf{T}(\boldsymbol{x}_{it})$ from the Chebyshev interpolant, and the continuation expectation $\mathbb{E}[W_m(\boldsymbol{x}') \mid \boldsymbol{x}_{it}, m]$ is obtained by averaging the same interpolant across the Gauss–Hermite-shifted next-state nodes (82). No restriction to grid points is imposed at the likelihood stage.

Worker state path. For each worker i and period t the state $\boldsymbol{x}_{it}$ collects the posterior belief $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ over latent skills and the cumulative task tenures $\boldsymbol{\tau}_{it}$ (in the application $\boldsymbol{x}_{it} \in \mathbb{R}^7$: a two-dimensional posterior mean, three Cholesky entries of the posterior covariance, and two tenures). At entry the belief equals the group prior $(\hat{\boldsymbol{\mu}}_{g(i)}, \hat{\boldsymbol{\Sigma}}_{g(i)})$. At observed rows the production code approximates the per-period belief by the worker's terminal-period posterior $(\hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{V}}_i)$ from the EM E-step, broadcast to all observation rows. This sufficient-statistic approximation avoids re-running a Kalman filter inside the outer optimization at the price of using a full-information posterior in place of the period-specific one; we adopt it as a deliberate computational simplification. Tenures $\boldsymbol{\tau}_{it}$ are observed cumulative task experience and require no approximation.

Choice log-likelihood. Stacking (88) across workers and periods yields

$$\ell\big(\boldsymbol{\Theta}_{\text{dyn}}; \hat{\boldsymbol{\Theta}}_{\text{wage}}\big) \;=\; \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log \mathbb{P}\big(d_{it} \mid \boldsymbol{x}_{it}, j_{i,t-1}; \boldsymbol{\Phi}^\star(\boldsymbol{\Theta}_{\text{dyn}}), \hat{\boldsymbol{\Theta}}_{\text{wage}}\big). \tag{89}$$

The wage-block parameters $\hat{\Theta}_{\text{wage}}$ enter (89) only through the choice-specific values inside $\mathbf{\Phi}^\star$ 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 $\hat{\Theta}_{\text{wage}}$, the type-cluster means and covariances $\{\hat{\mu}_g, \hat{\Sigma}_g\}$, and the worker-specific posterior moments $\{(\hat{\zeta}_i, \hat{V}_i)\}_i$. Holding these quantities fixed, the NFXP estimator searches for the dynamic-choice parameters $\Theta_{\text{dyn}} = (\eta_j, \kappa_0, \varrho)$ that maximize the choice log-likelihood (89). For each candidate $\Theta_{\text{dyn}}^{(k)}$ proposed by the outer optimizer, the procedure executes two steps in sequence.

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

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

i.e. (87), by successive substitution on the Smolyak coefficient matrix $\mathbf{\Phi}$, returning the converged coefficients $\mathbf{\Phi}^\star\big(\Theta_{\text{dyn}}^{(k)}\big)$ via the algorithm of Section D.2.2.

Step 2 — Choice likelihood. Given $\mathbf{\Phi}^\star\big(\Theta_{\text{dyn}}^{(k)}\big)$, evaluate the conditional choice probabilities (88) at the observed worker states $\{x_{it}\}$ via the Smolyak interpolant, and form the choice log-likelihood $\ell\big(\Theta_{\text{dyn}}^{(k)}; \hat{\Theta}_{\text{wage}}\big)$ in (89).

These two steps deliver one likelihood evaluation. The outer optimizer (L-BFGS-B with JAX autodiff gradients, see Section D.2.6) uses ℓ and its gradient to propose the next candidate $\Theta_{\text{dyn}}^{(k+1)}$, and Steps 1–2 are repeated. Iteration continues until $\Theta_{\text{dyn}}^{(k)}$ converges.

D.2.6 Differentiable NFXP Implementation (JAX)

Estimation of the dynamic parameters $\Theta_{\text{dyn}} = (\eta_j, \kappa_0, \varrho)$ proceeds by NFXP (Rust 1987): for each candidate Θ_{dyn}, the inner loop solves the worker's dynamic program for $\mathbf{\Phi}^\star(\Theta_{\text{dyn}})$ via the algorithm of Section D.2.2; the outer loop evaluates the choice log-likelihood (89). The whole pipeline is implemented in JAX so that gradients of the outer objective flow through the converged inner fixed point without finite differences.

Differentiating through the inner fixed point. The Smolyak fixed point can be differentiated in two ways. The first is to backpropagate through a fixed number of Bellman iterations: $\mathbf{\Phi}^\star$ is replaced by $\mathbf{\Phi}^{(K)} = \mathcal{F}^K(\mathbf{\Phi}^{(0)})$ for a K chosen large enough that $\mathbf{\Phi}^{(K)}$ is within tolerance of the true fixed point, and reverse-mode autodiff is applied to the resulting finite composition. This is cheap to implement but introduces a small bias from truncating the iteration. The second is to apply the implicit-function theorem at the converged fixed point and propagate adjoints through the linearization of $\mathcal{F}$ at $\mathbf{\Phi}^\star$, which delivers an exact gradient. The production driver implements the second approach via a custom `jax.custom_vjp` on the inner solver, with the tolerance-based iteration compiled into a single JAX while-loop.

Outer optimization. The outer optimizer is L-BFGS-B with JAX autodiff gradients, a box bound $\varrho > 0$, and amenities anchored at $\eta_{0,g} = 0$ for each group g. We use outer tolerances ftol $= 10^{-6}$ and gtol $= 10^{-5}$, an inner Bellman tolerance of 10^{-8}, and a CLI cap of $K_{\max} = 5{,}000$ inner iterations. The full NFXP evaluation—inner fixed-point iteration plus batched likelihood on $N \simeq 2.4$ million observations—runs in approximately 31 minutes on a single NVIDIA H100 GPU node.

Standard errors. Standard errors for $\hat{\Theta}_{\text{dyn}}$ are not yet computed in the production pipeline; the outer Hessian is available via `jax.hessian`, but its inversion to a sandwich/OPG variance estimator and the Murphy–Topel correction for first-stage estimation uncertainty in $\hat{\Theta}_{\text{wage}}$ are open items left for the next revision.

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

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

Stage 1 — EM [Appendix D.1].

1. Initialize $(\boldsymbol{\Theta}_{\text{wage}}^{(0)}, \boldsymbol{\mu}_g^{(0)}, \boldsymbol{\Sigma}_g^{(0)})$.
2. Repeat for $q = 0, 1, \ldots$ until $\text{MSD}\big(\boldsymbol{\Theta}^{(q)}\big) < 10^{-9}$:
   - **E-step** [(64), (65)]: per-worker posteriors $(\hat{\boldsymbol{\zeta}}_i^{(q)}, \hat{\boldsymbol{V}}_i^{(q)})$.
   - **M-step** [(73), (74), (68)]: closed-form linear systems per occupation; group prior moments.
3. **Output:** $\hat{\Theta}_{\text{wage}},\ \{\hat{\boldsymbol{\mu}}_g, \hat{\boldsymbol{\Sigma}}_g\}_{g \in \mathcal{G}},\ \{(\hat{\boldsymbol{\zeta}}_i, \hat{\boldsymbol{V}}_i)\}_{i=1}^{N}$.

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

1. Build per-worker state path $\{\boldsymbol{x}_{it}\}$: entry belief $(\hat{\boldsymbol{\mu}}_{g(i)}, \hat{\boldsymbol{\Sigma}}_{g(i)})$; observed-row belief $(\hat{\boldsymbol{\zeta}}_i, \hat{\boldsymbol{V}}_i)$; tenures $\boldsymbol{\tau}_{it}$.
2. Initialize $\boldsymbol{\Theta}_{\text{dyn}}^{(0)}$ with anchor $\eta_{0,g} = 0$ 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 $\boldsymbol{\Phi}^{\star}\big(\boldsymbol{\Theta}_{\text{dyn}}^{(k)}\big)$ via the algorithm in Section D.2.2.
   - **CCPs** [(88)]: evaluate off-grid at $\{\boldsymbol{x}_{it}\}$ via the Smolyak interpolant.
   - **Likelihood** [(89)]: $\ell\big(\boldsymbol{\Theta}_{\text{dyn}}^{(k)}\big)$.
   - **Outer step**: L-BFGS-B update with JAX autodiff gradients (implicit-function adjoint through $\boldsymbol{\Phi}^{\star}$).
4. **Output:** $\hat{\Theta}_{\text{dyn}}$.

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

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

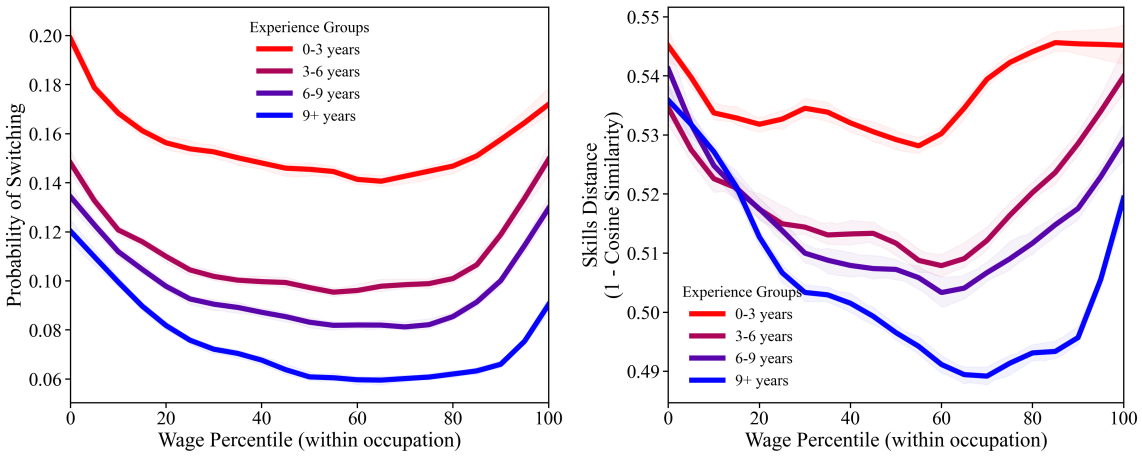
Large Sample		Estimation Sample	
Statistic	**Value**	**Statistic**	**Value**
Total observations	29,428,668	Total observations	2,316,035
Number of unique workers	4,425,614	Number of unique workers	283,765
Number of unique firms	618,009	Number of unique firms	131,626
Number of unique occupations	103	Number 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.

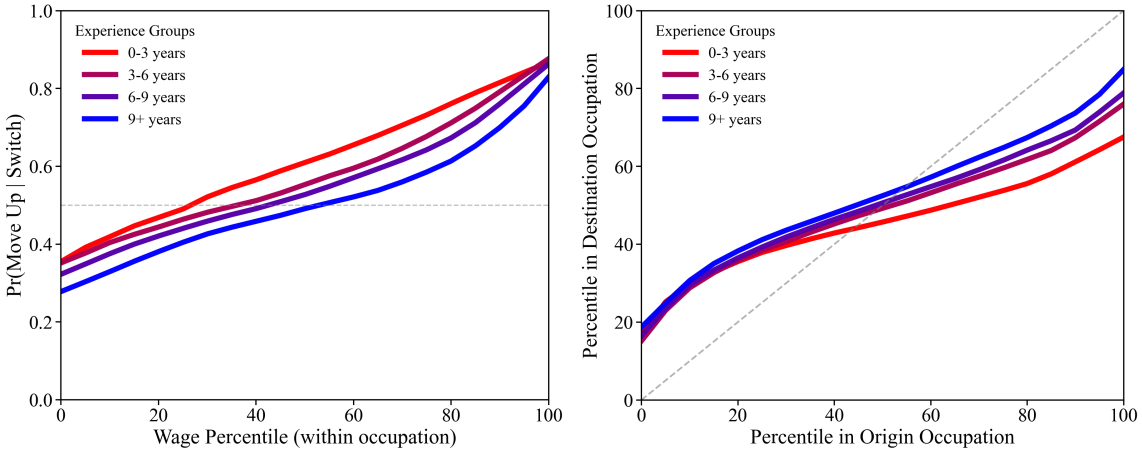
Task and Skill Classifications. The first classification, developed by Mihaylov and Tijdens (2019), classifies occupations into non-routine abstract, non-routine interactive, non-routine manual, routine cognitive, and routine manual tasks. The occupational code used in this paper is based on the ISCO88, 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 26 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

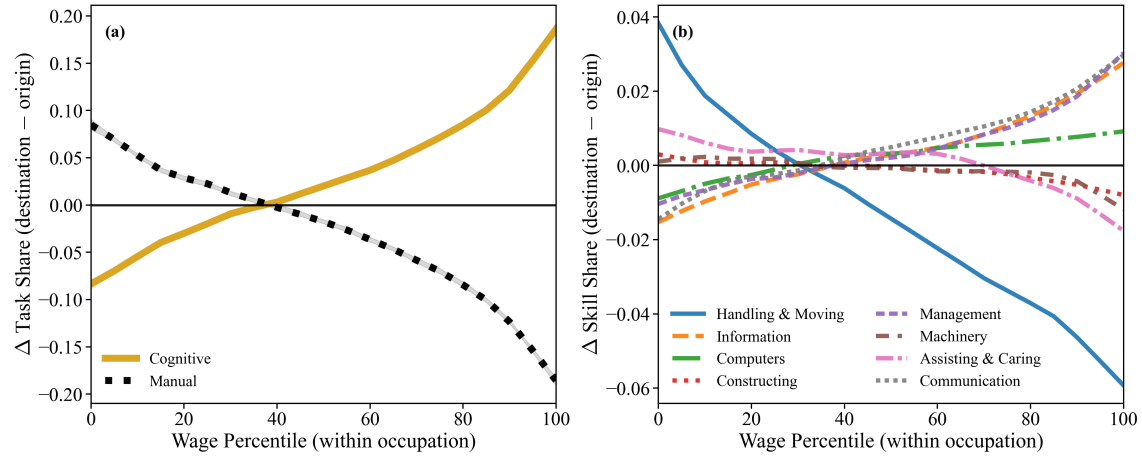
both wages and skills, and the task share changes—remain qualitatively robust when using raw wages instead of residualized wages.



(a) Probability of Switching & Skills Distance



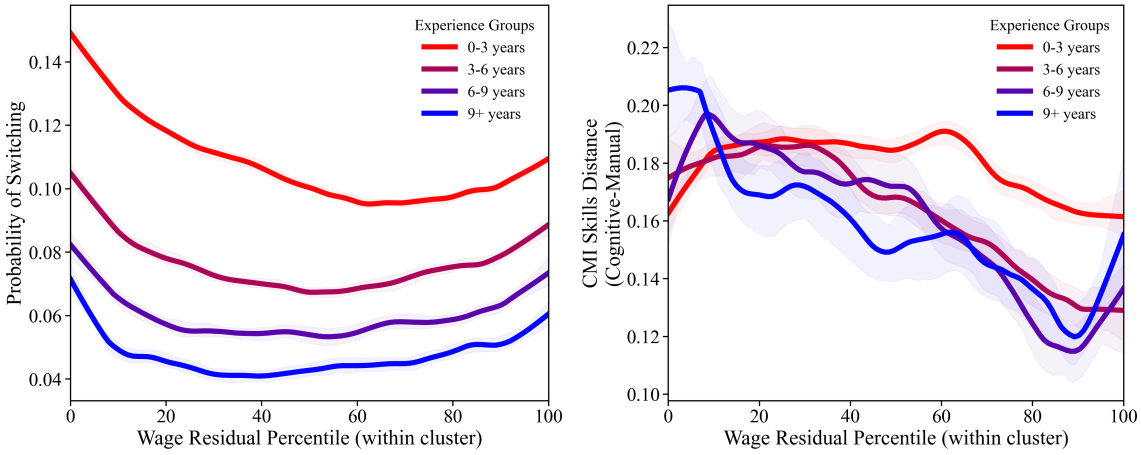
(b) Directionality by Experience



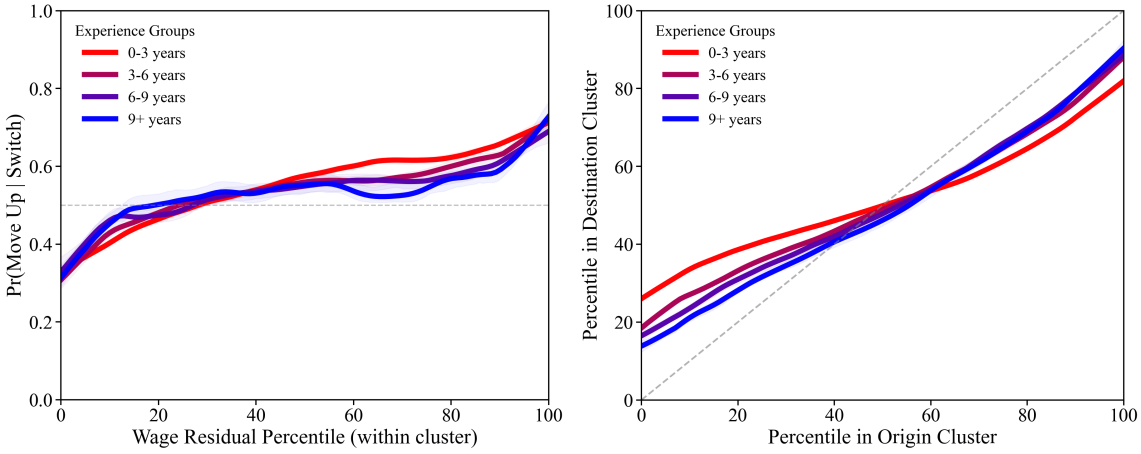
(c) Task Content Directionality (Cognitive/Manual and ESCO)

Figure 26: 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.

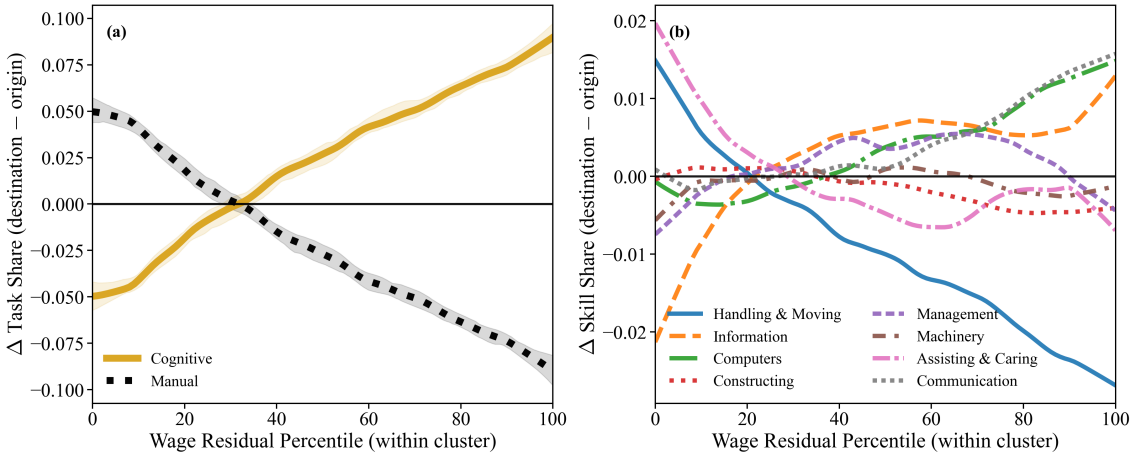
Robustness: Estimation Sample (Clusters). Furthermore, Figure 27 reproduces the main empirical analysis using the smaller estimation sample, where occupations are aggregated into ten task-based clusters. As discussed in Section 3, this aggregation mechanically lowers mobility levels, and the smaller 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 27: 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 28 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.

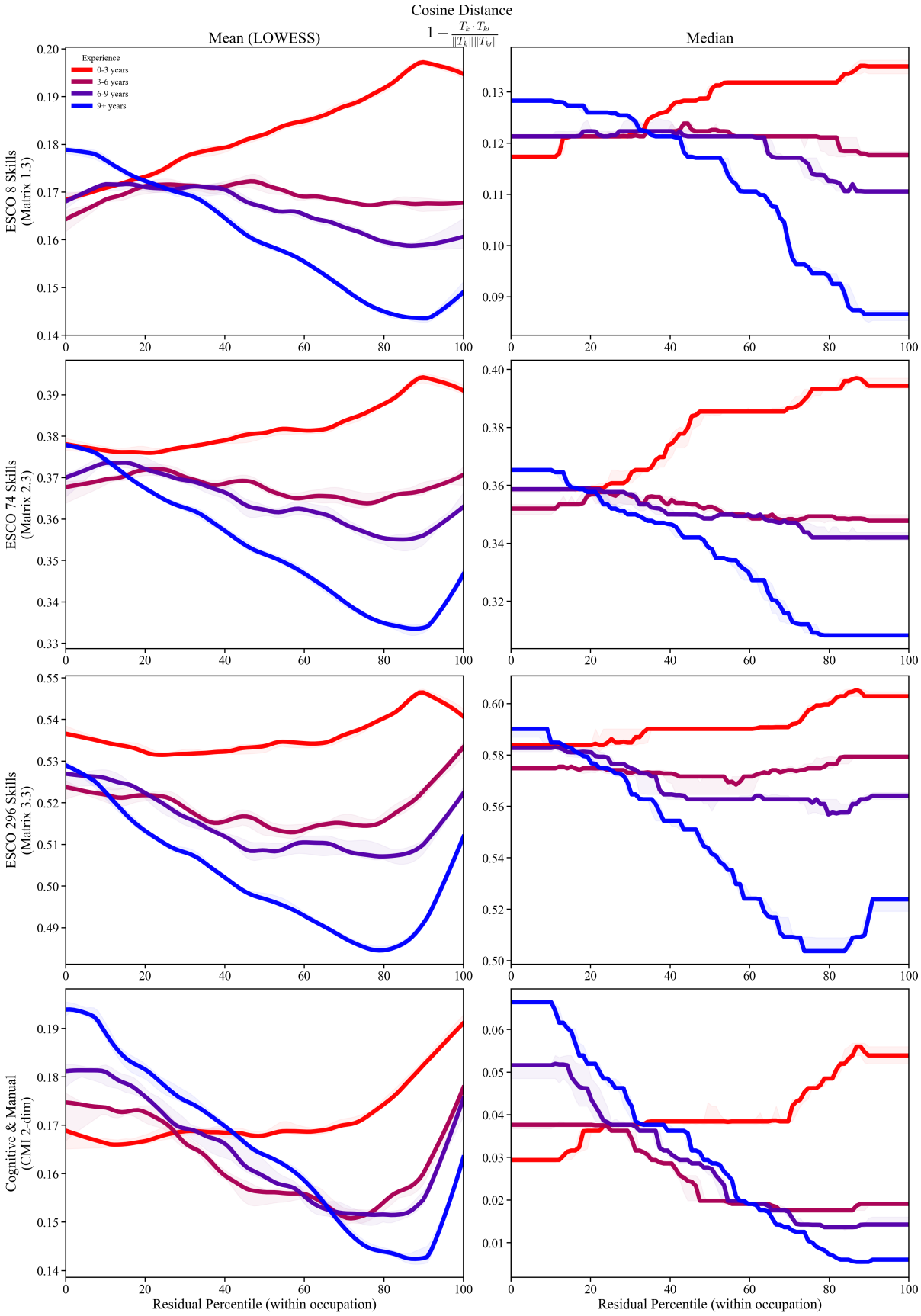
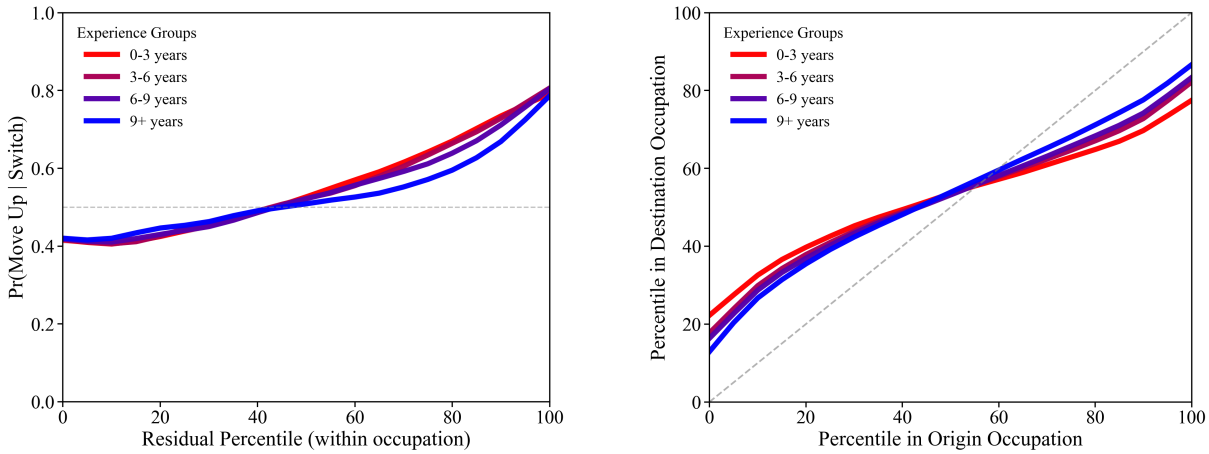


Figure 28: Robustness: Skill Granularity. Task distance (cosine distance, $1 - \text{Sim}(j, j')$, as in Gathmann and Schönberg 2010) by residualized within-occupation wage percentile and experience group. Left column: mean (LOWESS). Right column: median. Rows from top to bottom: ESCO 8 dimensions, ESCO 74 dimensions, ESCO 296 dimensions, Cognitive–Manual (2 dimensions).

Directionality and Rank-to-Rank Mobility by Experience. Figure 29 reports the directionality of occupational mobility documented in Section 3. Panel 29(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 29(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 29: *Panel 29(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 29(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 labourers 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 bookkeep-

ers ($d = 0.36$; 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. Men's most frequent move is between labourers and building frame workers—a short-distance switch within manual occupations—while women's most frequent move is between accounting clerks and bookkeepers. These gender differences are largely driven by occupational sorting: low-educated men concentrate in construction and manufacturing, while low-educated women concentrate in domestic and care services.

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

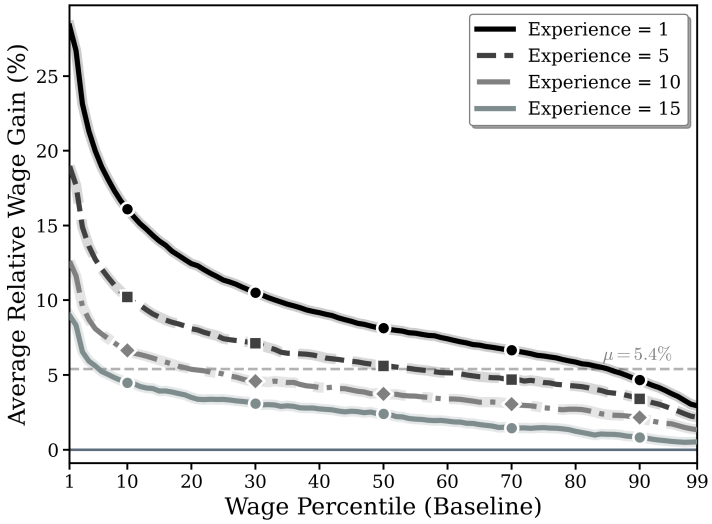


Figure 30: Average wage gain from removing information frictions by position in the overall wage distribution (horizontal axis: wage percentile; vertical axis: percent wage gain), with separate lines for different experience levels. Workers at lower percentiles gain substantially more—those near the 10th percentile gain about 21% one year into the career against roughly 4% near the 90th—so removing information frictions compresses the overall wage distribution.

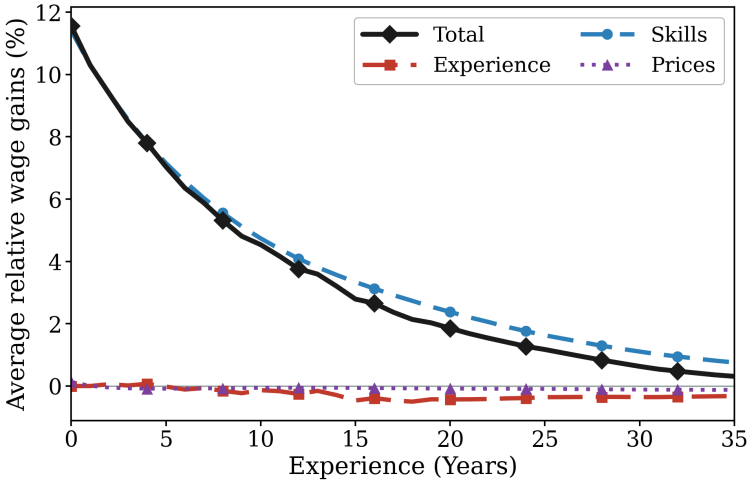
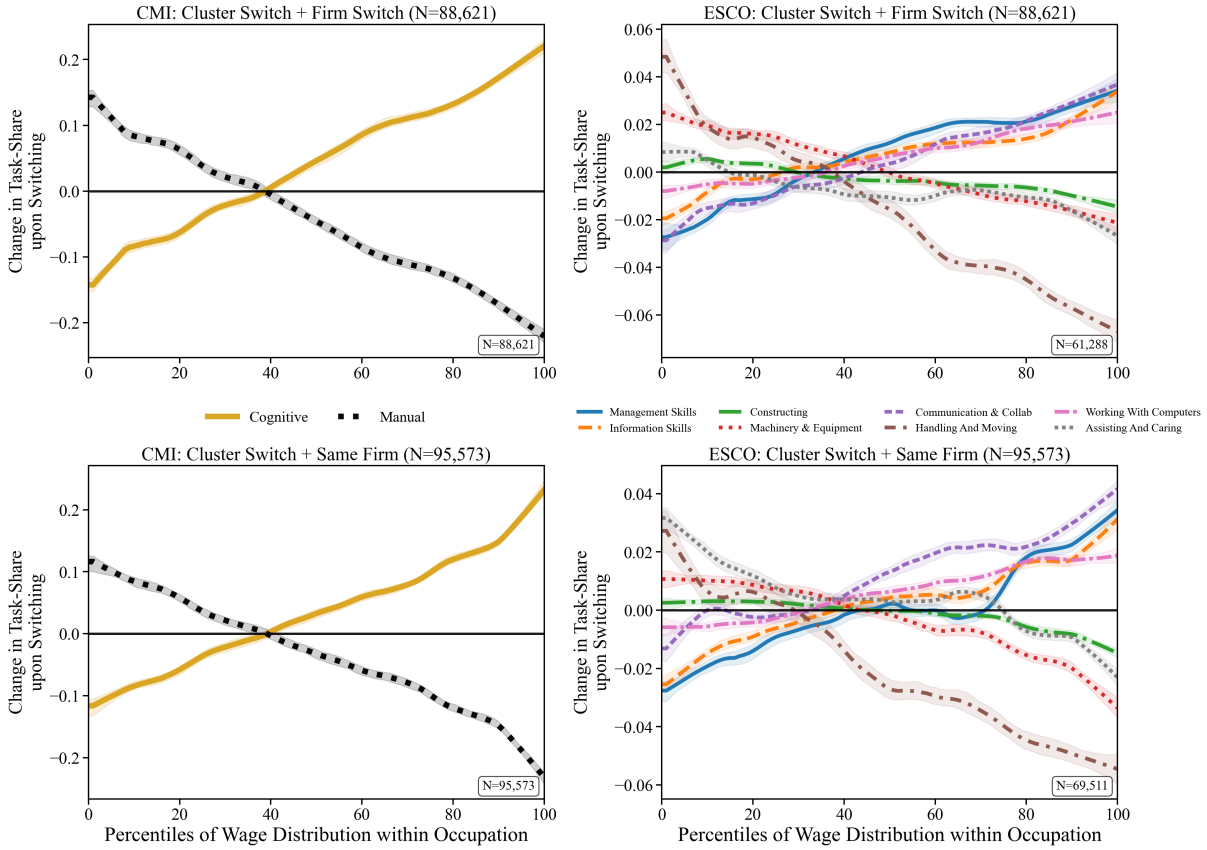


Figure 31: Decomposition of the lifecycle wage gain from removing information frictions (Figure 11, Panel (a)) into channels: Total (black), Skills (blue), Experience (red), and Prices (purple). Horizontal axis: years of potential experience; vertical axis: average relative wage gain (percent). The Skills channel holds experience fixed and isolates the gain from sorting into better-matched occupations; the Experience channel adds the effect of accumulating task-specific human capital in those occupations; the Prices channel captures the general-equilibrium adjustment of occupation prices.

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



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

Figure 32: 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.

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 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[28] 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[29]. As for occupations, the

[28]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.

[29]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.

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 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 skill vector $x_i \in \mathbb{R}^2$ summarizing its cognitive and manual content, a mass weight ω_i equal to its share of person-periods in the panel, and a symmetrized pairwise mobility flow F_{ij} measured from observed consecutive employer-to-employer transitions in the large sample. The partition minimizes

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

where μ_c is the ω-weighted centroid of cluster c. The first term keeps occupations with similar skill content together; the second term penalizes absorbing observed worker flows into within-cluster diagonals, so that aggregated cluster-to-cluster transitions remain informative—the opposite of what standard community-detection objectives would do. The weight $\lambda = 10$ is chosen at the knee of the trade-off between within-cluster dispersion and cross-cluster mobility share. Two feasibility constraints discipline the partition: the cluster-level mobility graph must remain connected, and every non-anchor cluster must retain at least 5% of total mass. Two anchor groups are pinned to fixed labels—a cognitive anchor (ISCO 241, 412, 431) at cluster 1 and a manual anchor at cluster $K = 8$—and anchored occupations are forbidden from leaving their pinned cluster. The procedure is run from several starting points (ω-weighted k-means initializations) and the best partition is retained; $K = 8$ is selected by a standard model-selection criterion balancing fit and parsimony on a 3–15 grid. The clustering is implemented in two variants of the underlying skill measure: 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 granularly, 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,

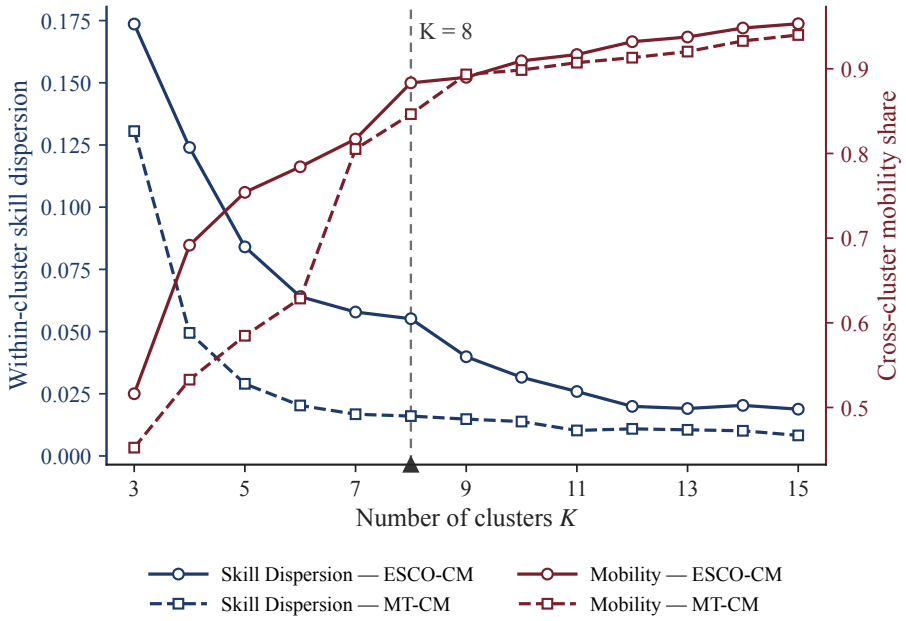


Figure 33: Headline diagnostic for the K-selection step of the anchored skill+mobility-penalty clustering. The horizontal axis is the number of clusters K on the 3–15 grid (with $\lambda_{\text{user}} = 10$ fixed at the empirical knee). Curves report the ω-weighted within-cluster skill dispersion and 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 $(C_i^{\text{MT}}, M_i^{\text{MT}})$, in both cases with the cognitive anchor $\{241, 412, 431\}$ and the manual anchor pinned. The chosen $K = 8$ sits inside the dispersion plateau and one notch below the feasibility ceiling under both feature sets.

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.

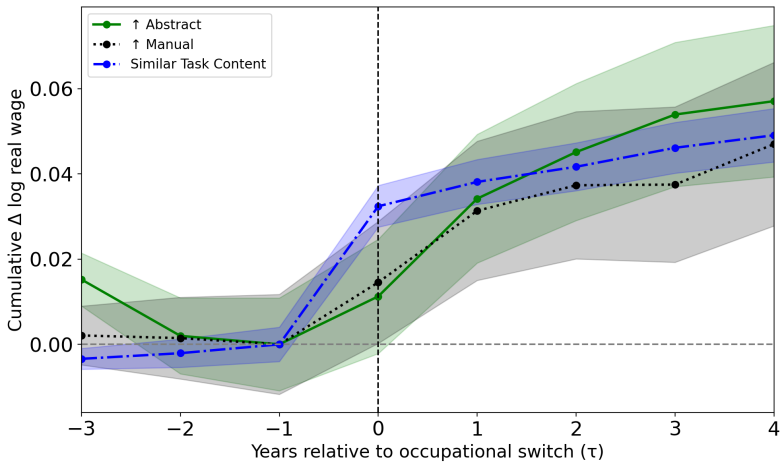
E.2 Empirical evidence on wage changes upon occupational switches

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

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

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 $g \in \{\text{Cog}, \text{Man}, \text{Other}\}$.



(a) Cumulative wage changes around occupational switches

Figure 34: 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.

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:

$$\underbrace{\left(\Delta \log w_{it+\tau}\right)^{\circ}}_{\text{demeaned by year } t} = \alpha_\tau + \sum_g \beta_{\tau g} D_{it}^g + \theta_\tau' X_{it}^{\circ} + \varepsilon_{it\tau}, \tag{90}$$

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—matching the figure—we cumulate the estimated growth effects across relative time and normalize at $\tau = -1$:

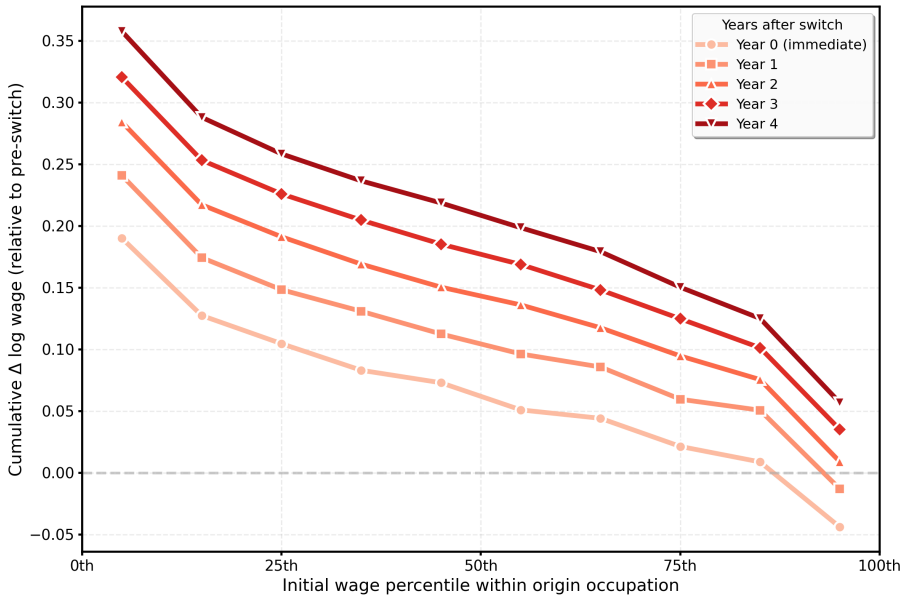
$$C_g(k) \;=\; \sum_{\tau=-3}^{k} \hat{\beta}_{\tau g} \;-\; \sum_{\tau=-3}^{-1} \hat{\beta}_{\tau g}, \qquad k \in \{-3, \ldots, 4\},$$

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

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 35 shows cumulative wage changes for switchers across the initial wage percentile distribution.

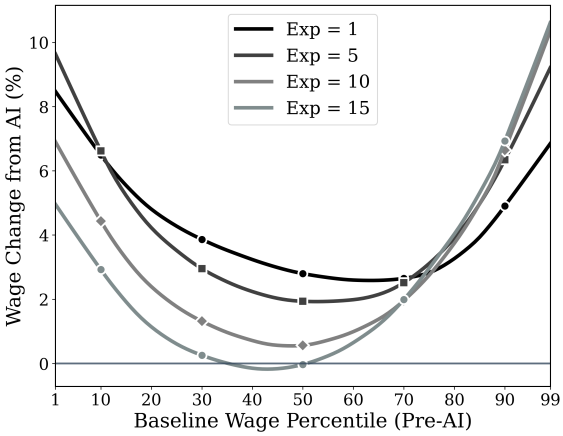
For each switcher, we calculate their wage percentile within their origin occupation (in the year before switching) and track cumulative log wage changes over the following four years. The



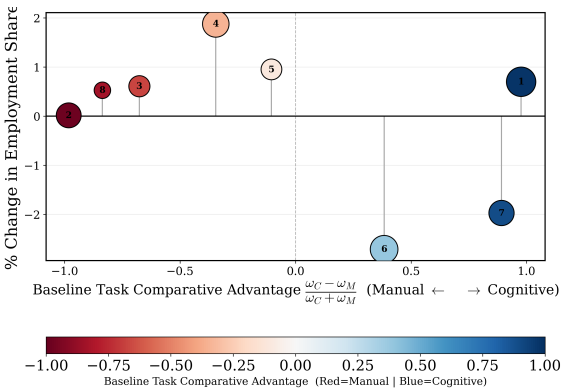
(a) Cumulative wage changes by initial wage percentile

Figure 35: 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.

plot reveals a strong negative relationship: workers initially at the bottom of their occupation's wage distribution experience cumulative gains exceeding 30%, while those at the top initially lose wages but recover to modest gains of approximately 10% by year four. This pattern persists across all time horizons 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 36: *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.