The Micro and Macro Implications of Multidimensional Skill Uncertainty*

Piero De Dominicis†

October 14, 2026

Click here for the latest version.

Abstract

What are the consequences of multidimensional skill uncertainty for workers' wages and aggregate output? I develop and estimate a general equilibrium dynamic Roy model in which workers have imperfect information about their multidimensional skills and accumulate task-specific human capital. Estimated on Portuguese administrative data, the model rationalizes key patterns of occupational mobility, with learning about comparative advantage playing a central role in occupational reallocation among young and poorly matched workers. Removing information frictions raises aggregate output by 5.1%, primarily through better skill allocation across occupations, while generating the largest wage gains early in workers' careers and among the most mismatched. A feasible information treatment about comparative advantage across tasks recovers about 19% of this output loss. Finally, job transformation induced by Large Language Model adoption raises output while widening the output gap between the imperfect- and perfect-information economies along the transition path and in the new long run.

Keywords: Multidimensional Skills, Mismatch, Information Frictions, Learning, Occupational Mobility, Human Capital, Misallocation, Wage Inequality, Dynamic Roy Model, Task-Biased Technological Change

*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 Eeckhout, Pamela Giustinelli, Cristina Gualdani, John Grigsby, Simon Jäger, Francis Kramarz, Danial Lashkari, Arnaud Maurel, Romuald Meango, Andreas I. Mueller, Theodore Papageorgiou, Elena Pastorino, Nicola Pavoni, Richard Rogerson, Aysegül Sahin, Karthik Sastry, Isaac Sorkin, Wilbur Townsend and Gianluca Violante, as well as seminar participants at Bocconi, Princeton Junior Macroeconomics Workshop, Stanford Macro Lunch, Stanford Labor/Public Economics Workshop, 1st Aarhus Workshop on Labor Markets, EWMES 2025, CEPR Paris Symposium 2025, BeNA Labor Economics Workshop 2025, 2nd Boston College PhD Conference, Theories and Methods in Macro 2026, NASMES 2026, XXIX Workshop on Dynamic Macroeconomics, CEF 2026, the NBER Summer Institute 2026 (The Micro and Macro Perspectives of the Aggregate Labor Market), EEA/ESEM 2026, EALE 2026 and the TRAILS Conference on Occupational Mismatch and Skill Shortage 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, from earnings trajectories and occupational choice (Heckman et al. 2006; Adda and Dustmann 2023) to who wins and who loses from occupational reallocation when task-biased technological shocks hit the economy (Freund and Mann 2025). However, especially early in their careers, workers may have imperfect information about their skills and may therefore direct their search toward jobs based on their perceived comparative advantage. Sorting on beliefs 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; Antonovics and Golan 2012; Papageorgiou 2014; Pastorino 2024). Mismatch along different skill dimensions predicts corrective occupational moves with implications for wages and for human capital (Guvenen et al. 2020), which accumulates along these dimensions according to the task content of jobs (Gathmann and Schönberg 2010). Learning, imperfectly transferable human capital, and other non-wage-related factors interact to create strong path dependence in occupational choice that prevents a substantial fraction of workers from ever sorting into their ideal career path. Quantifying the importance of these components in driving occupational choice and mobility is crucial for understanding the consequences of early-career mismatch generated by information frictions about skills and for designing effective policies to mitigate it.

In this paper, I develop a general equilibrium dynamic Roy model with multidimensional skill learning 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. At the same time, they accumulate task-specific experience depending on which tasks they perform on the job. The productivity of skills and experience in performing different tasks 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. In the model, two key forces drive occupational sorting: learning about unobserved multidimensional skills and the accumulation of task-specific human capital. These forces, together with utility switching costs and compensating differentials, shape both the timing and the direction of occupational transitions over the life cycle.

I estimate the model on Portuguese administrative data using a two-stage estimation procedure that allows me to control for the dynamic selection induced by belief sorting. I show that the estimated model replicates important 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, low- and high-wage workers are more likely to switch occupations 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 this narrative by looking at the change in task content for occupation switchers. I find that, conditional on switching occupations, the change in task content is also U-shaped: relative to the origin occupation, both low- and high-wage workers within an occupation make larger moves in task content than the median worker in the same occupation. On average, those at the top move up the occupational ladder to more cognitive occupations, while those at the bottom move to more manual occupations, which are on lower rungs of the ladder. I use the model to rationalize these patterns and decompose the drivers of these different occupational moves. I find that while experience accumulation mainly induces upward mobility for high-wage workers within an occupation, learning about comparative advantage is an important driver of downward mobility of poorly matched workers. In other words, low-wage workers move because they learn that they are mismatched in their current occupation and that their set of skills is more productive elsewhere, while high-wage workers move because the accumulated experience in certain tasks is more productive in other occupations. Occupation switches driven by learning are therefore wage-improving, as workers move to more productive matches.

Ideally, if workers perfectly knew their skills, they could sort into their most productive occupation right away and earn higher wages from the start of their careers. 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 the beginning of their careers. I find that aggregate output in this perfect-information economy would be 5.1% higher compared to the baseline economy with information frictions. Individual wage gains are largest at labor-market entry, where they are on average 10.5%, and fall to roughly zero after thirty years as the gap between the average wage level in the two economies closes. Almost all of the early aggregate gain comes from better sorting on skills, although some workers still benefit from sorting better immediately and accumulating more human capital along the right skill dimensions. These wage gains are concentrated at the bottom of the match-quality distribution within occupations, as these workers are the most heavily mismatched and the ones who, in the baseline economy, would switch occupations because of learning. These mismatched workers 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 skill uncertainty.

Recovering the total output and wage losses might be infeasible in practice. However, different career guidance policies exist, and I can use the estimated model to assess their effectiveness in reducing misallocation. More specifically, I study the effects of a more realistic information intervention, guided by existing experimental evidence.

This policy is an information treatment about comparative advantage in skills, which has been studied and implemented by Kiss et al. (2023). In practice, this amounts to assessing young jobseekers' skills with a computerized assessment and providing them with information about their performance in the tests relative to other candidates taking the same assessment. In my model, I simulate this intervention as a noisy signal about skills and calibrate the variance of the signal to match the average treatment effect in Kiss et al. (2023), so that the share of workers whose beliefs about their comparative advantage match their true comparative advantage rises by 13.7 percentage points. Delivered at labor-market entry, the intervention raises aggregate output by 0.97% and increases average wages by about 2.9% at entry, with the gains concentrated in the first decade and among the workers who are more mismatched in the baseline economy. The output and wage gains are highest when the assessment is delivered before labor-market entry, when workers have not yet accumulated task-specific experience: the same assessment delivered two years after entering the labor market produces gains that are only a little over half of those obtained when the signal is provided before workers start their careers. The effectiveness of career guidance policies quickly decreases over workers' careers.

The model also allows me to study how information frictions interact with standard frictions considered in the job-mobility literature. I represent these frictions as a fixed utility cost paid by workers when switching occupations. This reduced-form cost may capture search costs or broader mobility frictions that are not explicitly modeled. For example, when employment opportunities in an occupation are geographically concentrated, switching occupations may also require costly geographic relocation. Moreover, when switching occupation, workers may have to spend time unemployed while reallocating across labor markets, as in the island model of Lucas and Prescott (1974). I show that reducing this switching cost always reduces the misallocation induced by information frictions, echoing results from Papageorgiou (2014). However, decreasing this utility cost has non-monotonic effects on the level of aggregate output: small reductions in the switching cost increase the output level, but very large reductions reduce it. This non-monotonic effect is driven by two countervailing forces: on the one hand, small reductions in the switching cost allow workers to reduce their mismatch once they learn in which tasks they are most productive; on the other hand, large reductions create excessive churning that destroys accumulated experience in tasks, lowering workers' productivity. The excessive churning is driven by idiosyncratic taste shocks that make workers jump from one job to another. An economy in which a very experienced engineer decides to become a bartender because of her taste will have relatively lower aggregate output.

Finally, I study how artificial intelligence affects the misallocation induced by skill uncertainty. I focus on the consequences of Large Language Model (LLM) adoption, which changes the importance of specific tasks within occupations. The modeling of this task-biased technology shock is similar in spirit to the one recently studied by Freund and Mann (2025) and Althoff and Reichardt (2026). Following the task-exposure

framework of Eloundou et al. (2024), LLMs increase the importance of manual tasks while reducing that of cognitive tasks. In the model, this implies that time spent on cognitive tasks decreases in proportion to each occupation's exposure to LLMs, while time spent on manual tasks increases. This implies that cognitive experience is accumulated at a lower rate, and the returns to cognitive skills and experience decrease. Conversely, more time spent on manual tasks implies that experience along the manual dimension is accumulated faster, and the returns to manual skills and experience increase. The shock is modeled as an unanticipated MIT shock, and I solve for the transition path assuming that workers have perfect foresight over prices but still have imperfect information about their skills. I find that LLM adoption raises aggregate output by 2.1%. On impact, this task-biased reallocation toward manual tasks increases output by 1.0%. Gradually, as new cohorts enter the economy, output increases toward its long-run stationary equilibrium level. Output increases because new cohorts can immediately sort differently, based on the new task importance in the economy, compared to experienced workers who were already locked into certain career paths when the task-based reallocation shock hit the economy. I also show that in this long-run equilibrium, sorting on skills becomes more important, and so does learning, as cognitive experience, which was particularly important in driving wage growth, is accumulated more slowly and pays less. The wage gains from removing information frictions in the post-AI economy, compared to the pre-AI economy, are highest for high-ability workers with a manual comparative advantage. These workers are mainly highly educated individuals who enter the labor market believing that they have a comparative advantage in cognitive tasks and who benefit from reallocation toward more mixed occupations.

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 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' abilities, but do not estimate such a model. Lise and Postel-Vinay (2020) estimate a structural on-the-job search model with multidimensional skills, but workers know their skills, so mismatch is mainly driven by search frictions. Baley et al. (2022) let workers learn about unobserved skills, but each job uses a single skill dimension and their focus is on business-cycle dynamics rather than life-cycle sorting. In my model, learning is à la Jovanovic (1979) and Miller (1984): workers forecast a vector of time-invariant task-specific skills, and a worker is "mismatched" when, given current beliefs, she chooses an occupation that is not the one in which she is truly most productive. The approach does not require a measure of the supply of skills, such as test scores; I instead recover the distribution of unobserved skills directly from wages, using the approach of Bunting et al. (2024). Estimating the model allows me to quantify the micro and macro implications of early-career skill mismatch generated by imperfect information and to understand how much policies 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[1]—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 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[2]. 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.

[1]Much of the employer-learning literature emphasizes this direct channel, among others Harris and Holmström (1982), Farber and Gibbons (1996), Altonji and Pierret (2001), Lange (2007), 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).

[2]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.

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) show that occupational choice and wage growth in the NLSY are consistent with workers learning about their abilities on the job; their framework, however, abstracts from on-the-job human capital accumulation and non-wage amenities, both of which are central here.

This paper departs from this literature by embedding learning in a task-based model of occupational choice. I show empirically that mobility across tasks is informative about different types of occupational switches, and that these switches have distinct implications for wages and human capital accumulation. The model also enriches occupational choice by allowing for 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 propose a Roy (1951) model where workers sort based on comparative advantage across tasks. In my framework, as in these task-based approaches, tasks are bundled within occupations and workers' skills map into output through occupation-specific task loadings. However, unlike these papers, I focus on how workers' occupational allocation is shaped by imperfect information about multidimensional skills. The task-based approach is also the workhorse for studying technological change: routine-biased technical change and job polarization (Autor et al. 2003, 2006; Autor and Dorn 2013; Acemoglu and Autor 2011). More recently, Freund and Mann (2025) develop a general-equilibrium task-based model to study the effects of job transformation induced by Large Language Models (Eloundou et al. 2024), in which automating a task raises the importance of those that remain, and estimate the resulting individual wage effects from a distribution of task-specific skills. However, as in related general-equilibrium dynamic Roy models studying technological change (Dvorkin and Monge-Naranjo 2019; Althoff and Reichardt 2026), workers there know their own skills exactly. In this paper, I study how job transformation affects mismatch driven by information frictions about skills, and how skill uncertainty affects occupational reallocation in response to this technological 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 nonpara-

metric identification results of Bunting et al. (2024) for what they call the pure learning model[3].

The rest of the paper is organized as follows. In Section 2 I describe the data used and some of the key empirical facts on occupational mobility. In Section 3 I introduce the model, while in Section 4 I discuss identification and describe the estimation procedure. In Section 5 I use the estimated model to decompose the drivers of occupational mobility. Section 6 quantifies the costs of skill uncertainty and evaluates information and mobility policies, and Section 7 studies the AI shock. Section 8 concludes.

2 Data and Empirical Facts

2.1 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)[4]. 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 A. The large sample is designed to maximize the number of observations and representativeness for descriptive facts and baseline comparisons. It applies standard cleaning criteria to the raw data and is used to document the overarching empirical patterns of occupational mobility at the 3-digit occupation level. The estimation sample is a cohort panel of labor-market entrants, constructed by applying further restrictions to the large sample. It tracks the complete early-career trajectories of young workers entering the labor market, which is essential for proper identification and estimation of the learning and human capital parameters of the structural model.

To measure the task content of occupations, I use two classifications. The first focuses on broad task categories, cognitive and manual skills[5]; 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 clas-

[3]As I describe later in Section 4, 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 by the econometrician; while this complicates identification and estimation in general, the estimation routine can incorporate such heterogeneity by extending the expectation-maximization (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.

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

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

sifications, 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. Throughout the paper, I interpret skills and tasks as two sides of the same object: workers supply a set of skills, each of which maps into a specific task, so that skills refer to the supply side of the labor market and tasks to its demand side.

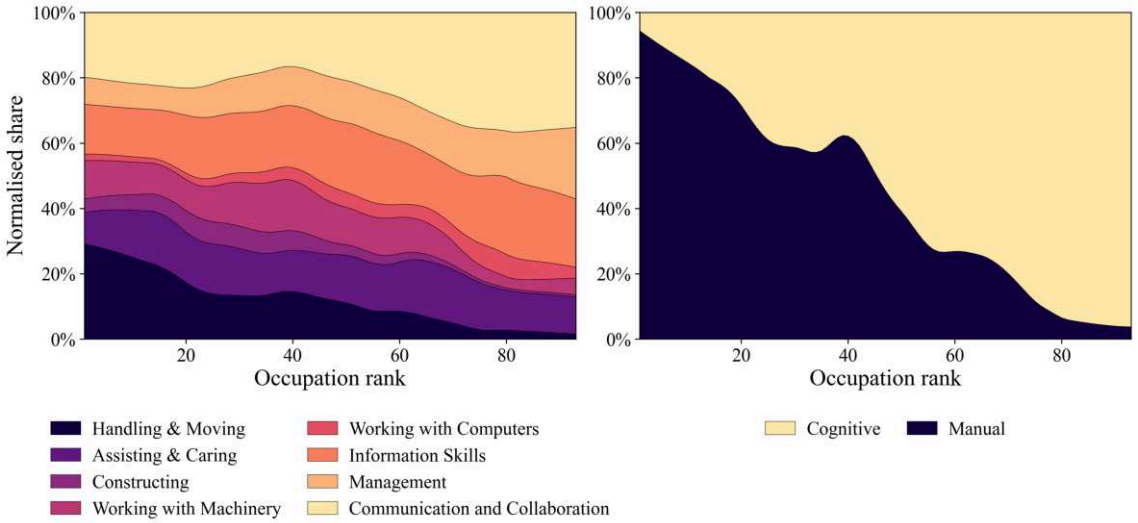


Figure 1: Task composition by occupation average wage rank. Occupations are ranked by mean hourly wage within each year (horizontal axis: occupation rank, from the lowest-paying occupation on the left to the highest-paying on the right; vertical axis: skill share in percent). *Left panel*: Each occupation's 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, and constructing—decline with occupational wage rank, while assisting and caring remains roughly constant. The right panel confirms this pattern more starkly: cognitive tasks rise with occupational wage rank while manual tasks decline.

Taken together, the figure points to a vertical job ladder in tasks: higher-paying occupations tilt increasingly toward cognitive tasks, and lower-paying occupations to-

ward manual tasks. As I show below, workers move both up and down this ladder, and the direction of the move has different implications for their wages.

2.2 Occupational Mobility, Directionality in Tasks and Wages

I now document three facts on occupational mobility. They extend the evidence of Groes et al. (2015) for Denmark in two directions: I study the task content of occupational moves as a function of the worker's position within the origin occupation, and I document the wage changes that accompany these moves.

Fact 1: Mobility and the change in task content are both U-shaped. 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:

$$w_{it} = p_j + \iota_t + f_1(\text{tenure}_{it}) + f_2(\text{experience}_{it}) + \gamma_F \text{Female}_i + \gamma_e \text{Educ}_i + m_{ij} + \nu_{it} \quad (1)$$

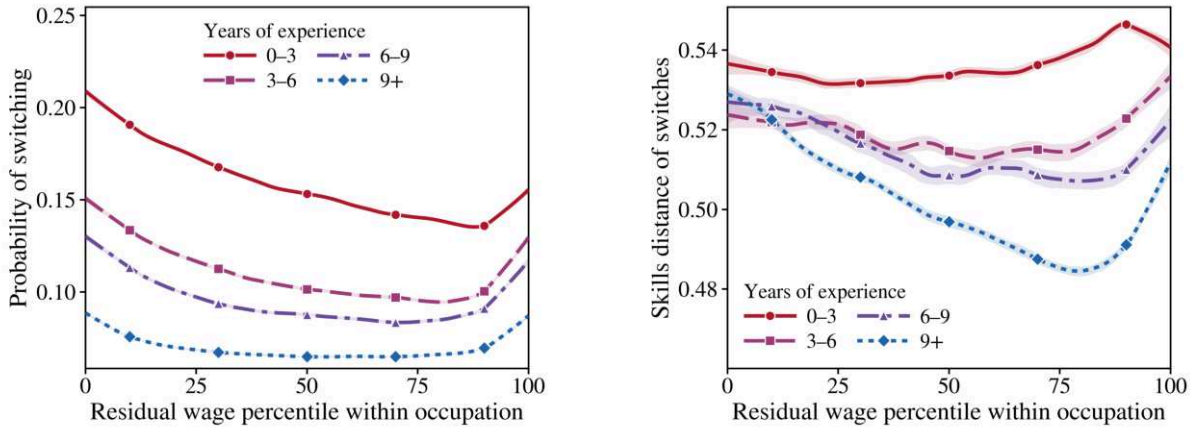
where w_{it} is the log hourly wage, p_j are occupation fixed effects (the log occupation prices), ι_t are time fixed effects, $f_1(\cdot)$ and $f_2(\cdot)$ are cubic polynomials in occupational tenure and total labor-market experience, m_{ij} is the unobserved worker-occupation match quality, and ν_{it} is mean-zero noise. The estimated residual is $\widehat{r}_{it} = \widehat{m}_{ij} + \widehat{\nu}_{it}$. Since ν_{it} is mean-zero, the percentile rank of $\widehat{r}_{it}$ within occupation j is positively correlated with m_{ij}: on average, workers with higher match quality will rank higher in the residual distribution. Therefore the "within-occupation residual percentile" used throughout this section is a noisy but informative proxy for the worker's position in the match quality distribution within occupation. Appendix A shows that all patterns documented here are robust to using raw hourly wages instead of residualized wages.

Figure 2(a)[6] 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.

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

$$d(j, j') = 1 - \text{Sim}(j, j'), \qquad \text{Sim}(j, j') = \frac{\boldsymbol{\omega}_j^\top \boldsymbol{\omega}_{j'}}{\|\boldsymbol{\omega}_j\|_2 \, \|\boldsymbol{\omega}_{j'}\|_2}, \quad (2)$$

[6] All figures for Facts 1 and 2 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 2: *Panel (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 (b)*: Skills distance between origin and destination occupations conditional on switching (vertical axis) by residualized within-occupation wage percentile (horizontal axis), computed using ESCO skill measures.

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 most common observed switch, between laborers and building frame workers, has $d = 0.21$[8]. Similarly to Gathmann and Schönberg (2010), I find that workers switch to occupations that are substantially closer in task space than a random mobility benchmark would predict (Figure 24 in the Appendix).

Figure 2(b) shows that the task distance between origin and destination occupations is related to the worker's match quality in the origin occupation. Workers at both ends of the wage distribution within occupation make larger moves in the task space, conditional on switching occupation. Task distance also falls with experience. This pattern is hard to reconcile with undirected search frictions, where the size of the move should be unrelated to the origin mismatch conditional on switching. As Guvenen et al. (2020) argue, if mismatch is induced by search frictions alone, the change in mismatch upon switching should be independent of the mismatch in the origin occupation. The relationship between match quality and the size of the task move is instead consistent with directed sorting, where workers learn about their comparative advantage and reallocate toward better-fitting occupations.

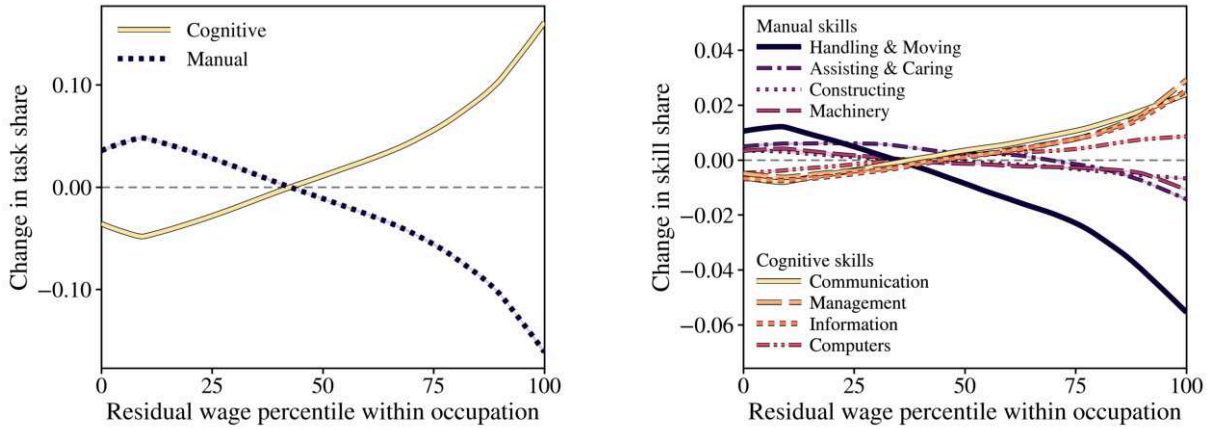
Appendix Figure 22 splits the patterns in Figure 2(a) into switches within and across

[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 3), 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.

[8]See Appendix Table 2 for detailed summary statistics on the skills distance distribution, including the most common switches by education group.

firms. The probability of switching occupation within the same firm differs from the probability of switching across firms at both ends of the within-occupation wage distribution, especially at the top, where the increase in mobility comes mainly from switches within the same firm. The directionality patterns documented next, which characterize the nature of the move, are instead consistent within and across firms (Appendix Figure 23).

Fact 2: Mobility is directional in the task space. Occupational mobility in Portugal is also directional, using the terminology introduced by Groes et al. (2015). Well-matched workers within an occupation are more likely to move to an occupation with a higher average wage compared to their current occupation, while poorly matched workers move to an occupation with a lower average wage (Appendix Figure 21). However, when moving, poorly matched workers improve their relative position in the destination occupation, while well-matched workers worsen their relative position.



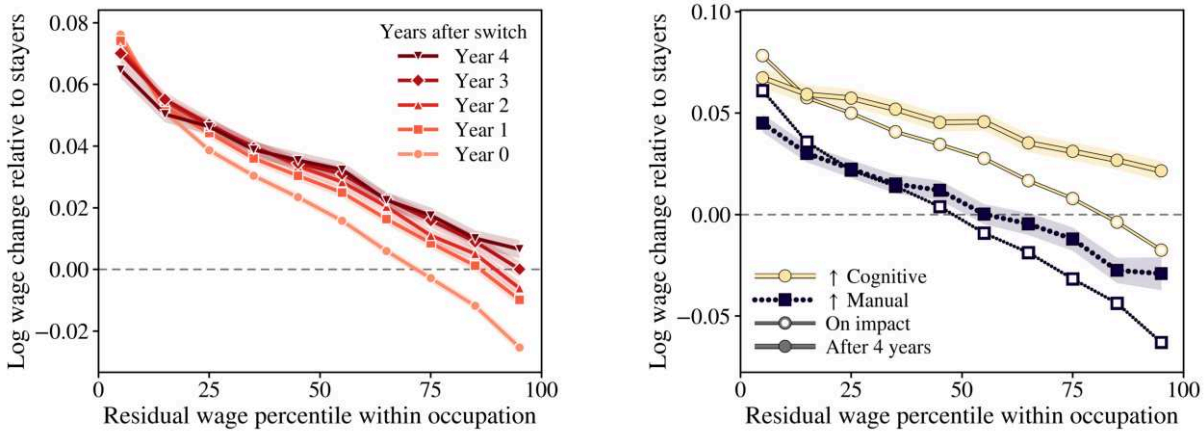
(a) Change in Cognitive and Manual Task Shares

(b) Change in ESCO Skill Shares

Figure 3: *Panel (a)*: Change in cognitive and manual task shares (vertical axis) by residualized within-occupation wage percentile (horizontal axis) for occupational movers. Change is $\Delta\text{Task-Share}_p^k = \text{LOWESS}_p\left(\omega_{j'(i)}^k - \omega_{j(i)}^k\right)$, 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 smoothed changes in shares at each percentile sum to 0.

Figure 3 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 (Panel (b)) 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.

Fact 3: Switching raises the wages of poorly matched workers. Figure 4 documents that workers realize wage gains upon switching occupations, relative to workers who start from the same position in the same occupation and stay (see Appendix A.1 for details). These gains vary systematically with a worker's initial position within their occupation: those in the bottom quintile gain 6 log points relative to stayers upon switching and 6 after four years, while workers in the top decile initially suffer wage cuts of 3 log points and close the gap by year four. Workers in the bottom quintile gain whatever the direction of the move in the task space, while workers at the top gain, after four years, only when they move toward more cognitive occupations. This pattern indicates that switching is especially beneficial for workers poorly matched to their initial occupation. The initial wage losses of high earners suggest either backloaded compensation or non-pecuniary benefits that influence job choice (Sorkin 2018).



(a) Wage change of switchers relative to stayers, by initial wage percentile

(b) Wage change of switchers relative to stayers, by initial wage percentile and task content of the switch

Figure 4: Wage changes upon occupational switches, relative to stayers in the same origin occupation, year, and decile of the initial wage percentile. Horizontal axis: residual wage percentile within the origin occupation in year $t-1$; vertical axis: cumulative change in the log real wage of switchers relative to stayers. *Panel (a)*: Years 0–4 after the switch, where Year 0 is the first year in the new occupation. *Panel (b)*: switches toward more cognitive occupations (light yellow, solid) and more manual occupations (dark blue, dotted), on impact (thin lines and open markers, Year 0) and four years after the switch (thick lines and filled markers). Shaded areas are 95% confidence bands clustered by worker.

2.3 Discussion of the Empirical Facts

Taken together, Facts 1 and 2 show that workers at both ends of the match-quality distribution are more likely to switch occupation than the median worker in the same origin occupation, and that they change tasks more when they switch. Facts 2 and 3 are instead informative about the nature of the switch: workers with low match quality improve their relative position in the destination occupation and gain relative to stayers, while workers with high match quality worsen their relative position and initially suffer wage cuts. Switches from the bottom and from the top of the distribution are

therefore likely to be driven by different forces.

The fact that, conditional on moving, the size of the change in task content depends on origin match quality is a distinctive feature of the data. If mismatch arose from undirected search, the size of the task change upon switching would be independent of the origin mismatch (Guvenen et al. 2020). The pattern documented here points instead toward directed sorting. The change in wages upon switching is also informative about the nature of the occupational move. The mobility of workers with low within-occupation match quality is consistent with learning about comparative advantage in tasks: these workers learn that they are mismatched in their current occupation and that they might be more productive elsewhere, so they are more likely to switch and, when they do, they change the task content of their job more. At the upper end of within-occupation match quality, switchers experience a wage cut on average, but their wages increase gradually afterwards. This suggests the presence of compensating differentials, which could take the form of backloaded wage compensation or of other occupation-specific amenities more closely linked to cognitive tasks. These findings call for a theory that accounts for the relationship between match quality, mobility, and the task content of transitions, while allowing for more general compensating differentials in the form of dynamic returns or non-wage amenities. The model in the next section incorporates learning about multidimensional skills, task-specific human capital, and compensating differentials to quantify the consequences of this reallocation mechanism for wages over the life cycle.

3 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.

3.1 Occupations, Tasks and Human Capital

Workers supply multidimensional human capital that is used to perform different tasks $k \in \mathcal{K}$, with $K \equiv |\mathcal{K}|$, 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:

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

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 re-

turns that may differ across occupations. The function is concave in logs, disciplining diminishing returns to learning-by-doing.[9] The primitive parameters have clean interpretations: a_{jk} is the semi-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 $\varepsilon_{ijt} \sim F_j$ 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 $(\tilde{\boldsymbol{\theta}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ it shares with the worker, and that this allocation is non-contractible. Under the Cobb-Douglas aggregation, Appendix B.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

$$Y_{ijt} = \underbrace{\exp\left(\boldsymbol{\lambda}_j' \boldsymbol{\theta}_i\right)}_{\substack{\text{unobserved} \\ \text{skills bundle}}} \quad \underbrace{H_j(\boldsymbol{\tau}_{it})}_{\substack{\text{human-capital} \\ \text{bundle}}} \quad \underbrace{\exp(\varepsilon_{ijt})}_{\substack{\text{idiosyncratic} \\ \text{shock}}}, \tag{5}$$

where the effective loadings 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 loading is the Cobb-Douglas aggregator weight ω_{jk} times the corresponding primitive task-level semi-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, I define

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

[9] In the estimation, I parameterize the log returns to be a stepwise function: $h_{jk}(\tau) = \sum_{b \in \mathcal{B}} \delta^{(b)}_{jk} \mathbf{1}\{\tau \in b\}$, 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.

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

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

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 loadings 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

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

3.2 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.[10] Firms in occupation j choose employment across worker types to maximize profits

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

subject to the occupation production function

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

[10] 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; Gibbons et al. 2005; Lange 2007; Pastorino 2024). In my framework, competitive piece-rates tie wages 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), $W_{ijt} = c_{ijt} + b_{ijt}Y_{ijt} = (1 - b_{ijt})\,\mathbb{E}\big[Y_{ijt} \mid I_{ijt}\big] + b_{ijt}Y_{ijt}$, with $b_{ijt} = 1$, so that wages depend one-for-one on realized output (up to the occupation-specific output price scaling in my model) rather than on posterior expectations.

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

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

Taking logs yields the log wage equation

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

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

3.3 Learning and Information Structure

Workers belong to an observed education group $g \in \mathcal{G}$. The objects of learning are the worker's latent skills $\boldsymbol{\theta}_i$ introduced in Section 3.1. Upon entry into the labor market,

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

Workers are Bayesian and at the beginning of period t hold a subjective belief density $\tilde{f}_{it}(\boldsymbol{\theta})$ over $\boldsymbol{\theta}_i$ 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,

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

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

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

beliefs update according to Bayes' rule

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

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 $\tilde{\boldsymbol{\theta}}_{it}$ and a variance-covariance matrix $\tilde{\boldsymbol{\Sigma}}_{it}$ such that

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

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

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

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

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.

3.4 Occupational Choice

Let the worker state at the beginning of period t be

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

where $(\tilde{\boldsymbol{\theta}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ are beliefs about latent skills, $\boldsymbol{\tau}_{it}$ is task experience, and d_{it-1} is the last period's occupation. The sequence $\{p_{js}\}_{s\geq t}$ denotes log occupation prices from period t onward. I assume workers have perfect foresight over the future path of occupational prices and take $\{p_{js}\}$ 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

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

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.[11] The switching cost is incurred only when the

[11] The taste shocks ϵ_{ijt} serve two purposes. First, they add idiosyncratic noise to the occupational choice in every period. When entering the labor market, all workers in a given education group enter with the same prior beliefs and no experience, so without the shocks all workers in a group would make the same choice. The shocks instead disperse entrants across occupations and, in later periods, help rationalize occupational choices and moves that the systematic part of the model cannot account for. Appendix E.1 illustrates the entry choice with and without the shocks. Second, they smooth the

worker changes occupation, and I parameterize it as a single utility constant:

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

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

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

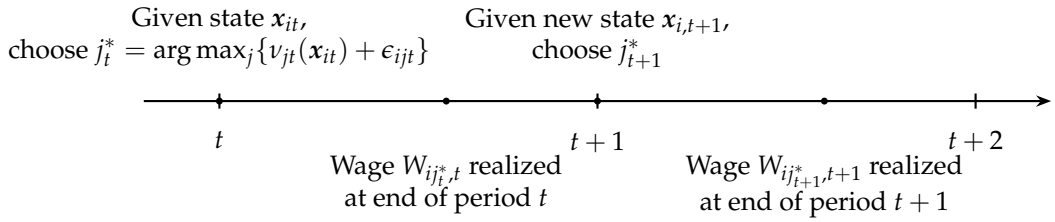


Figure 5: Timing of choices and wage realizations.

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

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

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

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

integrated value function and the conditional choice probabilities, which keeps the Bellman fixed point and the likelihood well-behaved for estimation.

continuation expectation integrates over the predictive distribution of the wage signal s_{jt}, which drives the Bayesian update of beliefs, and the next-period taste shock vector $\boldsymbol{\epsilon}_{i,t+1}$. The next-period covariance is deterministic given the chosen occupation, so the only stochastic state evolution is the posterior mean $\tilde{\boldsymbol{\theta}}_{t+1}$. Let $v_{jt}(\cdot)$ denote the choice-specific value of occupation j, the term in braces in (26) excluding the taste shock ϵ_{jt}. Integrating out ϵ yields the multinomial logit conditional choice probabilities:

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

With a continuum of workers, these conditional choice probabilities represent the mass of workers in each state choosing each occupation. The education group g is a permanent characteristic of each worker and it shifts the payoff of every occupation through the occupation-specific amenities η_{jg}. In the empirical implementation, rather than treating past occupational choice and education group as discrete states, I solve for a set of value functions, where each value function is occupation-education specific. Accordingly, the choice probabilities are also computed by education group. In practice, I approximate the integrated value function using the Smolyak method with Chebyshev polynomials (Malin et al. 2011; Judd et al. 2014), which amounts to solving for the coefficients that satisfy the fixed-point condition for the ex-ante integrated value function $EV(\cdot)$ at a set of collocation points:

$$EV(\boldsymbol{x}) = \varrho \log \sum_{j\in\mathcal{J}} \exp\left(\frac{p_j + \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\theta}} + h_j(\boldsymbol{\tau}) + \eta_{jg} - C(j, d_{-1}) + \beta\,\mathbb{E}[\,EV(\boldsymbol{x}') \mid \boldsymbol{x}, d = j]}{\varrho}\right), \tag{28}$$

where the expectation is taken over next-period state $\boldsymbol{x}'$ implied by the laws of motion under choice j. More details are provided in Appendix D.

3.5 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 $\sigma > 0$:

$$Y_t = \left(\sum_{j\in\mathcal{J}} \alpha_j Y_{jt}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \qquad \alpha_j > 0. \tag{29}$$

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 oc-

cupation j output:

$$P_{jt} = P_t \, \alpha_j \left(\frac{Y_t}{Y_{jt}} \right)^{\frac{1}{\sigma}}. \tag{30}$$

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 3.6) gives occupation prices directly.

3.6 Stationary Competitive Equilibrium

I define a stationary competitive equilibrium in which occupation prices are constant over time. With a continuum of workers, let $\pi(\boldsymbol{x})$ denote the mass of workers in state $\boldsymbol{x} = (g, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1}) \in \mathcal{X}$ (excluding prices), where $\mathcal{X}$ is the worker state space and $g(\boldsymbol{x})$ denotes the education group of state $\boldsymbol{x}$, which augments the decision state (23) with the worker's true latent skill $\boldsymbol{\theta}$[12]. Entrants are characterized by group masses π_g (with $\sum_{g \in \mathcal{G}} \pi_g = 1$) and initial distributions $\pi_g^0(\cdot)$ that draw true skill from the group prior, $\boldsymbol{\theta} \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, set rational initial beliefs $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}) = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, and start from $\boldsymbol{\tau} = \mathbf{0}$. A stationary competitive equilibrium[13] consists of constant occupation prices $\{P_j\}_{j \in \mathcal{J}}$, a stationary mass distribution $\pi^*(\cdot)$ over worker states, aggregate output Y, occupation outputs $\{Y_j\}_{j \in \mathcal{J}}$, a value function $V(\cdot)$, and conditional choice probabilities $\mathbb{P}(d = \cdot \mid \cdot)$ satisfying the following conditions:

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

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

$$Y_j = \int \left(\int \exp(\varepsilon) \, Y_j(\boldsymbol{x}) \, dF_j(\varepsilon) \right) \mathbb{P}(d = j \mid \boldsymbol{x}) \, \pi^*(\boldsymbol{x}) \, d\boldsymbol{x}, \tag{31}$$

   where $Y_j(\boldsymbol{x}) = \exp(\boldsymbol{\lambda}_j' \boldsymbol{\theta}) \, H_j(\boldsymbol{\tau})$ is the task technology (5) net of the idiosyncratic

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

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

shock, evaluated at the worker's true skill $\boldsymbol{\theta}$ and task experience $\boldsymbol{\tau}$, F_j is the distribution of the wage shock ε, and $\mathbb{P}(d = j \mid \boldsymbol{x})\pi^*(\boldsymbol{x})$ gives the mass of workers in state $\boldsymbol{x}$ 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 (30) 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

$$P = \Big(\sum_{j\in\mathcal{J}} \alpha_j^{\sigma}\, P_j^{1-\sigma}\Big)^{\frac{1}{1-\sigma}} = 1, \tag{32}$$

the cost function dual to the aggregator (29).[14] Rather than rescaling the weights to the simplex, I choose the scale of $\{\alpha_j\}$ so that (32) holds exactly at the Stage-1 estimated prices (Section 4), 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

$$\pi^*(\boldsymbol{x}') = \sum_{j\in\mathcal{J}} \int \Big[(1-\chi)\, Q(\boldsymbol{x}' \mid \boldsymbol{x}, j) + \chi\, \pi^0_{g(\boldsymbol{x})}(\boldsymbol{x}')\Big]\, \mathbb{P}(d = j \mid \boldsymbol{x})\, \pi^*(\boldsymbol{x})\, d\boldsymbol{x}, \tag{33}$$

for all states $\boldsymbol{x}' \in \mathcal{X}$, where $Q(\cdot \mid \boldsymbol{x}, j)$ is the transition density over worker states induced by choosing occupation j: true skill $\boldsymbol{\theta}$ and group g are carried forward unchanged, task experience accumulates via (10), and the belief moments update by Bayes' rule (18) from the realized wage signal $s_j = \boldsymbol{\lambda}_j'\boldsymbol{\theta} + \varepsilon_j$, whose distribution is governed by the worker's true skill; and $\pi^0_{g(\boldsymbol{x})}(\cdot)$ is the initial mass distribution for entrants from the group $g(\boldsymbol{x})$ of the exiting worker, so that each exiting worker is replaced by an entrant from the same education group and the group masses π_g are preserved[15].

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 (30) sub-

[14] The dual index carries α_j^{σ}, not α_j; this is the unique scale-one unit cost consistent with (29) and its first-order conditions (30).

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

ject to the numeraire normalization (32), and iterate to convergence. Appendix B.2 details the algorithm.

4 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 $|\mathcal{J}| = 8$ 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 A.2. While this aggregation mechanically lowers measured mobility levels compared to the 3-digit ISCO analysis, the empirical facts documented in Section 2 remain qualitatively consistent within this clustered state space.

4.1 Identification of the Wage Equation under Learning

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

Let $d_{it} \in \mathcal{J}$ 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

$$w_{it}(j) = p_j + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \boldsymbol{\theta}_i + \varepsilon_{ijt}, \tag{34}$$

where $\boldsymbol{\theta}_i \in \mathbb{R}^K$ is a time-invariant latent skill vector and the observed log 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 $\boldsymbol{\lambda}_j' \boldsymbol{\theta}_i$.

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

occupational histories the econometrician observes,

$$d_{it} \perp\!\!\!\perp \boldsymbol{\theta}_i \mid \mathcal{H}_{it},$$

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

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

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

$$\mathbb{E}\left[w_{it}(j) \mid \mathcal{H}_{it},\, g(i),\, d_{it} = j\right] \;=\; p_j + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \,\mathbb{E}\left[\boldsymbol{\theta}_i \mid \mathcal{H}_{it},\, g(i)\right],$$

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

4.2 Identification of the Dynamic Choice Parameters

Stage 1 identifies all the parameters from the wage equation, so for every supported state $\boldsymbol{x}_{it}$ and every occupation j the econometrician can compute the expected log wage $\tilde{w}_j(\boldsymbol{x}_{it})$ 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 $\nu_j(\boldsymbol{x})$ and $\mathbb{P}_j(\boldsymbol{x})$. The state $\boldsymbol{x}$ contains the education group g, so ν_j, $\mathbb{P}_j$ and the ex-ante value EV can be defined as group-specific. With β fixed externally and ϱ calibrated, the only unknown parameters driving the dynamic choice left in the model are the amenities η_{jg} and the single switching cost κ_0 of (25). I can identify these parameters using the level and composition of mobility across destinations observed in the data. Under the Type-I extreme value assumption, the observed log-odds recover differences in the choice-specific value function,

$$\varrho \, \log \frac{\mathbb{P}_j(\boldsymbol{x})}{\mathbb{P}_m(\boldsymbol{x})} = \nu_j(\boldsymbol{x}) - \nu_m(\boldsymbol{x}),$$

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

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

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

where $\Psi_j(\boldsymbol{x})$ 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 C.2 gives its expression). Continuation values are thus not unknown parameters, but rather they are a linear function of unknown parameters with an intercept that can be computed using wage equation parameters and observed CCPs.

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

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

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

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:

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

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 ϱ is weakly identified without payoff variation that moves current utility but not the continuation value. I therefore calibrate a value of ϱ that allows the estimated model to match the average occupational mobility rate in the large sample rather than estimating it jointly by maximum likelihood. Given this calibration, the normalization $\eta_{0g} = 0$ for every g, and support conditions requiring that the occupations used are chosen with positive probability in the states used, that each occupation is observed both as an origin and as a destination, and that both stays and moves are observed, the dynamic choice parameters $\big(\{\eta_{jg}\}_{j\in\mathcal{J},g\in\mathcal{G}}, \kappa_0\big)$ are identified. Theorem 2 in Appendix C.2 states this formally.

4.3 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

$$w_{ijt} = p_j + \boldsymbol{\lambda}_j'\boldsymbol{\theta}_i + h_j(\boldsymbol{\tau}_{it}) + \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}(0, \varsigma_{\varepsilon j}^2), \tag{35}$$

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

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

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

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

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 4.2 for discussion). The observed-data likelihood for worker i factors into a wage component and a choice component:

$$L_i(w_i, d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) = L_i^w(w_i; \boldsymbol{\Gamma}) \cdot L_i^c(d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}), \tag{36}$$

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[16].

The wage likelihood integrates out latent skills,

$$L_i^w(w_i; \boldsymbol{\Gamma}) = \int \left[\prod_{t=t_0(i)}^{T(i)} \prod_{j\in\mathcal{J}} f_{\varepsilon j}\Big(w_{ijt} - p_j - \boldsymbol{\lambda}_j'\boldsymbol{\theta} - h_j(\boldsymbol{\tau}_{it}) \Big)^{d_{ijt}} \right] \phi\Big(\boldsymbol{\theta}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\Big)\, d\boldsymbol{\theta}, \tag{37}$$

where $f_{\varepsilon j}$ is the normal density with variance $\varsigma^2_{\varepsilon j}$ and $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal density. With $d_{ijt} \equiv \mathbf{1}\{d_{it} = j\}$ the indicator that worker i chooses occupation j in year t, the choice likelihood separates the first observed choice (which uses the group

[16] Starting from $L_i(w_i, d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) = \int f(w_i, d_i, \boldsymbol{\theta} \mid g(i); \boldsymbol{\Gamma}, \boldsymbol{\Delta})\, d\boldsymbol{\theta}$, the chain rule gives

$$f(w_i, d_i, \boldsymbol{\theta} \mid g(i)) = \phi(\boldsymbol{\theta}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}) \prod_t f(w_{it} \mid d_{it}, \boldsymbol{\theta}, \boldsymbol{\tau}_{it}; \boldsymbol{\Gamma}) \prod_t f(d_{it} \mid \mathcal{H}_{it}; \boldsymbol{\Gamma}, \boldsymbol{\Delta}).$$

By timing/independence, conditional on $(d_{it}, \boldsymbol{\theta}, \boldsymbol{\tau}_{it})$ wages depend only on $\boldsymbol{\Gamma}$, while choices depend on history only through $(\tilde{\boldsymbol{\theta}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}, \boldsymbol{\tau}_{it}, d_{it-1})$, hence $\prod_t f(d_{it} \mid \mathcal{H}_{it}; \boldsymbol{\Gamma}, \boldsymbol{\Delta})$ does not depend on $\boldsymbol{\theta}$. Therefore,

$$\begin{aligned} L_i(w_i, d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) &= \int \phi(\boldsymbol{\theta}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}) \prod_t f(w_{it} \mid d_{it}, \boldsymbol{\theta}, \boldsymbol{\tau}_{it}; \boldsymbol{\Gamma}) \prod_t f(d_{it} \mid \mathcal{H}_{it}; \boldsymbol{\Gamma}, \boldsymbol{\Delta})\, d\boldsymbol{\theta} \\ &= \left[\prod_t f(d_{it} \mid \mathcal{H}_{it}; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) \right] \int \phi(\boldsymbol{\theta}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}) \prod_t f(w_{it} \mid d_{it}, \boldsymbol{\theta}, \boldsymbol{\tau}_{it}; \boldsymbol{\Gamma})\, d\boldsymbol{\theta} \\ &\equiv L_i^c(d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) \cdot L_i^w(w_i; \boldsymbol{\Gamma}). \end{aligned}$$

prior) from subsequent choices:

$$L_i^c(d_i;\boldsymbol{\Gamma},\boldsymbol{\Delta}) = \underbrace{\prod_{j\in\mathcal{J}}\Big[\mathbb{P}\Big(d_{it_0(i)}=j \mid \boldsymbol{\mu}_{g(i)},\boldsymbol{\Sigma}_{g(i)},\boldsymbol{\tau}_{it_0(i)},d_{i,t_0(i)-1};\boldsymbol{\Gamma},\boldsymbol{\Delta}\Big)\Big]^{d_{ij,t_0(i)}}}_{\text{first observed choice}} \times \underbrace{\prod_{t=t_0(i)+1}^{T(i)}\prod_{j\in\mathcal{J}}\Big[\mathbb{P}\big(d_{it}=j \mid \tilde{\boldsymbol{\theta}}_{it},\tilde{\boldsymbol{\Sigma}}_{it},\boldsymbol{\tau}_{it},d_{it-1};\boldsymbol{\Gamma},\boldsymbol{\Delta}\big)\Big]^{d_{ijt}}}_{\text{subsequent choices}}, \tag{38}$$

where CCPs are given by (27) and $(\tilde{\boldsymbol{\theta}}_{it},\tilde{\boldsymbol{\Sigma}}_{it})$ 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:

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

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 $\widehat{\boldsymbol{\Delta}} = \arg\max_{\boldsymbol{\Delta}} \sum_i \log L_i^c(d_i;\widehat{\boldsymbol{\Gamma}},\boldsymbol{\Delta})$.

4.3.1 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 $\boldsymbol{\theta}_i$,
$$f(\boldsymbol{\theta}_i \mid w_i,d_i;\boldsymbol{\Gamma}^{(q)}),$$
and form the expected complete-data log-likelihood.
- **M-step.** Update $\boldsymbol{\Gamma}^{(q+1)}$ by maximizing the expected complete-data log-likelihood. In practice this step updates $(\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)$, the skill loadings $\boldsymbol{\lambda}_j$, the tenure steps $\boldsymbol{\delta}_j$, the shock variances $(\varsigma_{\varepsilon j}^2)$, and the occupation prices (p_j). The M-step explicitly operationalizes the control-function identification argument of Section 4.1: by regressing log wages on the E-step posterior mean of skills, $\hat{\boldsymbol{\zeta}}_i^{(q)} \equiv \mathbb{E}[\boldsymbol{\theta}_i \mid \boldsymbol{w}_i,\boldsymbol{d}_i;\boldsymbol{\Gamma}^{(q)}]$, 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.

4.3.2 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 $\widehat{\boldsymbol{\Gamma}}$:

$$\widehat{\boldsymbol{\Delta}} = \arg\max_{\boldsymbol{\Delta}} \log L^c(\widehat{\boldsymbol{\Gamma}},\boldsymbol{\Delta}), \tag{39}$$

where

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

The NFXP algorithm nests the solution of the worker dynamic problem inside the maximization over $\boldsymbol{\Delta}$. Fix a candidate $\boldsymbol{\Delta}$ and define the state $\boldsymbol{x} \equiv (g, \tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$, on which the value function is solved group by group. The inner step solves for the integrated ex-ante value function $EV(\cdot; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta})$ as the fixed point (28), evaluated at $(\widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta})$. Given the fixed point $EV(\cdot; \widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta})$, the CCPs used in the likelihood are computed from (27).

The outer step updates $\boldsymbol{\Delta}$ to increase (39) using a numerical optimizer. The procedure alternates between (i) solving the fixed point in (28) for EV at the current $\boldsymbol{\Delta}$ and (ii) updating $\boldsymbol{\Delta}$ given the resulting CCPs, until convergence.

4.3.3 Externally Set and Calibrated Parameters

Several parameters are set externally or calibrated to match aggregate statistics, as summarized in Table 3 in the Appendix. The discount factor is set to $\beta = 0.9$; education group shares π_g are set to the large-sample composition; and task importance weights ω_j are taken from Mihaylov and Tijdens (2019). The per-period exit hazard $\chi = 0.08$ is chosen to match the average career length in the estimation sample; as explained in Section 3.6, 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 (30) holds exactly while the final-good price index (32) equals one. Inverting the CES demand (30) leaves the CES weights proportional to $P_j Y_j^{1/\sigma}$ up to a single scalar; I fix that scalar by normalizing the aggregate price level $P(\hat{P}) = 1$ and I obtain the following closed-form calibration equation

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

where $P_j = \exp(\hat{p}_j)$ is occupation j's output price estimated in Stage 1, Y_j its output (the observed occupation wage bill deflated by P_j, because under piece-rate pricing each occupation's revenue equals its wage bill), and $s_j = P_j Y_j / \sum_m P_m Y_m$ 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.

4.4 Estimation Results

4.4.1 Stage 1 Estimation Results: Wage Equation Parameters

The EM algorithm jointly estimates occupation-specific skill and experience elasticities ($\boldsymbol{\lambda}_j$, $\boldsymbol{\delta}_j$), idiosyncratic wage variances, occupation prices, and education-specific skill distributions $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. Full estimation results are reported in Table 4 in the Appendix.

Three key findings emerge from estimation of the wage equation parameters. First, occupations differ substantially in the productivity of cognitive and manual skills, both in their skill loadings λ_{jk} and in their returns to experience[17]. Second, the estimated group skill distributions $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ (Figure 6) determine the initial comparative advantage of each education group, which is what shapes entrants' initial beliefs. Mean skills rise with education in both dimensions and the cognitive comparative advantage widens: low-educated workers sit below the origin, medium-educated workers near it, and high-educated workers above it and tilted toward cognitive skill. The estimated cross-skill correlations are weak in every education group, so workers are not strongly specialized ex ante and enter the labor market with genuine uncertainty about where their comparative advantage lies. These estimated distributions imply that more highly educated workers have (and perceive) higher skills in cognitive tasks, making them more likely to start their career in more cognitive occupations, consistent with Deming (2025). Third, occupations have heterogeneous returns to both manual and cognitive task experience, and returns to cognitive experience are generally steeper than returns to manual experience (Figure 7).

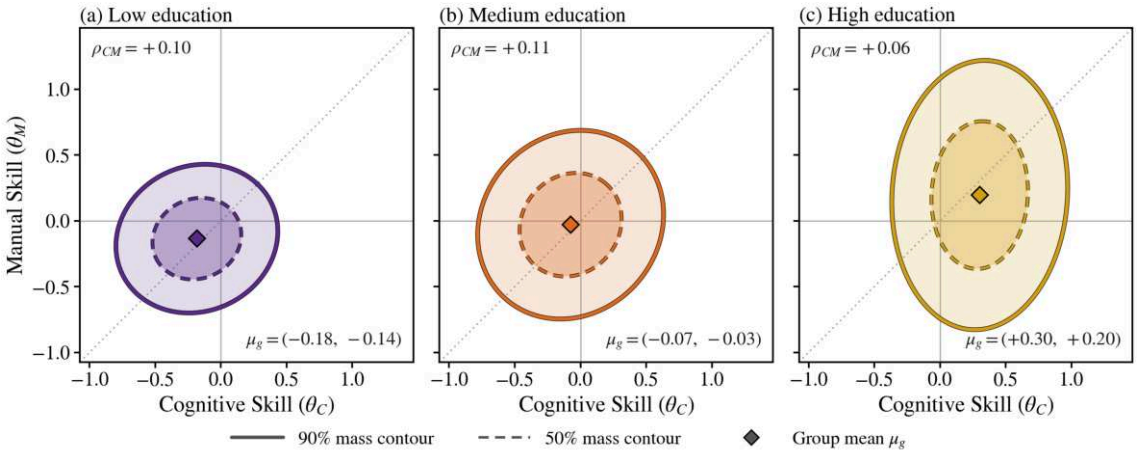


Figure 6: Estimated skill distributions by education group. Each panel (low, medium, high education) plots the joint prior distribution of cognitive skill θ_C (horizontal axis) and manual skill θ_M (vertical axis) implied by the estimated group means and covariances $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. 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 $\theta_C = \theta_M$; the within-group cognitive-manual correlation ρ_{CM} is reported in each panel.

[17] Cluster 0 is the cognitive anchor and Cluster 1 the manual anchor.

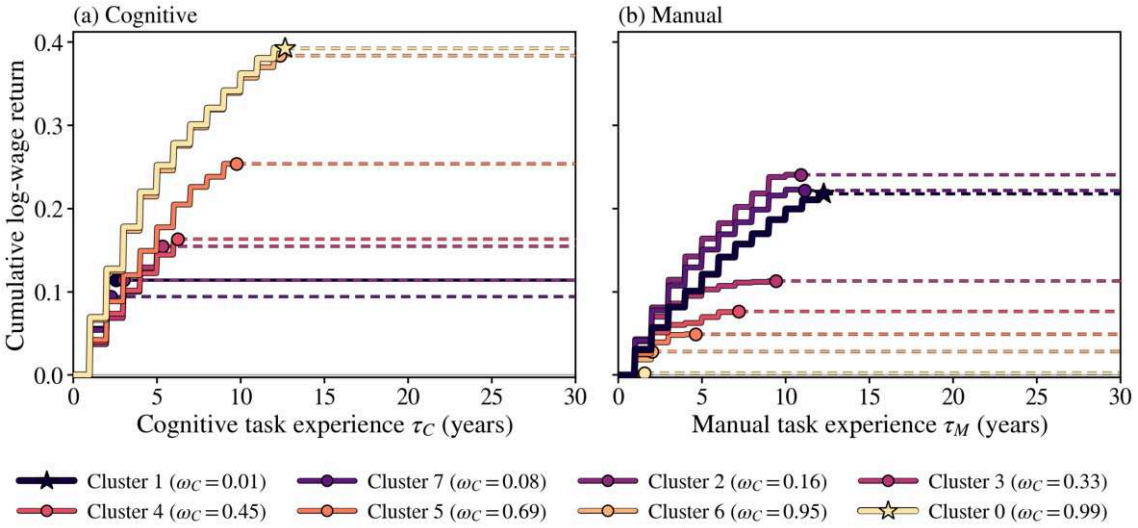


Figure 7: Estimated returns to task-specific experience by occupation. Each line plots the log-wage return to task experience, $\sum_b \delta_{jk}^{(b)} \mathbf{1}\{\tau \in b\}$, as task tenure τ accumulates from 0 to 30 years. Returns are estimated as a step function of tenure with annual bins, with the last bin being defined by the 95th percentile of observed task-specific tenure within that occupation. *Panel (a)*: returns to cognitive experience τ_C. *Panel (b)*: returns to manual experience τ_M. The marker on each curve is the occupation's 95th percentile of observed task tenure. Clusters are labeled by their cognitive task intensity ω_{jC} and shaded from navy (most manual) to cream (most cognitive); stars mark the anchor clusters (Cluster 0, cognitive; Cluster 1, manual).

4.4.2 Stage 2 Estimation Results: Dynamic Choice Parameters

The NFXP estimates are reported in Table 5 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 4.2, in dynamic learning models this parameter is weakly identified from continuation values without exogenous payoff shifters, so I pin it with this external moment while still allowing within-group heterogeneity in initial choices. The utility switching cost is moderate—a single constant of about 1.1 in log-wage units ($\kappa_0 = 1.103$)—suggesting that the life-cycle decline in occupational mobility is driven primarily by task-specific human capital accumulation and learning rather than by prohibitive switching costs. Occupation- and education-specific amenities η_{jg} display considerable heterogeneity. In the model, amenities act as compensating differentials that rationalize persistent occupational sorting patterns that cannot be explained by wage differentials alone. Potentially, these parameters capture other frictions not considered in the model. For example, if access to some occupations requires a specific educational degree, η_{jg} would capture the fact that certain groups of workers do not move to these occupations despite having attractive wages.

5 Occupational Mobility and the Role of Learning

5.1 Learning Generates Directionality in the Task Space

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

Panel (a) of Figure 8 shows the sequence of choices of a worker who starts in an occupation whose wages load on both skills (Cluster 6) and whose true skills lie in the cognitive-intensive part of the map. After period 0, this worker receives a positive signal in this occupation, so she updates her beliefs on both task dimensions and moves diagonally toward the upper-right quadrant of the skill space. She then receives a positive signal in the cognitive-intensive occupation (Cluster 0), whose wages load only on the cognitive task. Accordingly, she revises her mean cognitive belief upward and her mean manual belief downward, because the correlation between the two skills turns negative after the first signal. This sequence of signals pushes the worker toward the cognitive-intensive occupations.

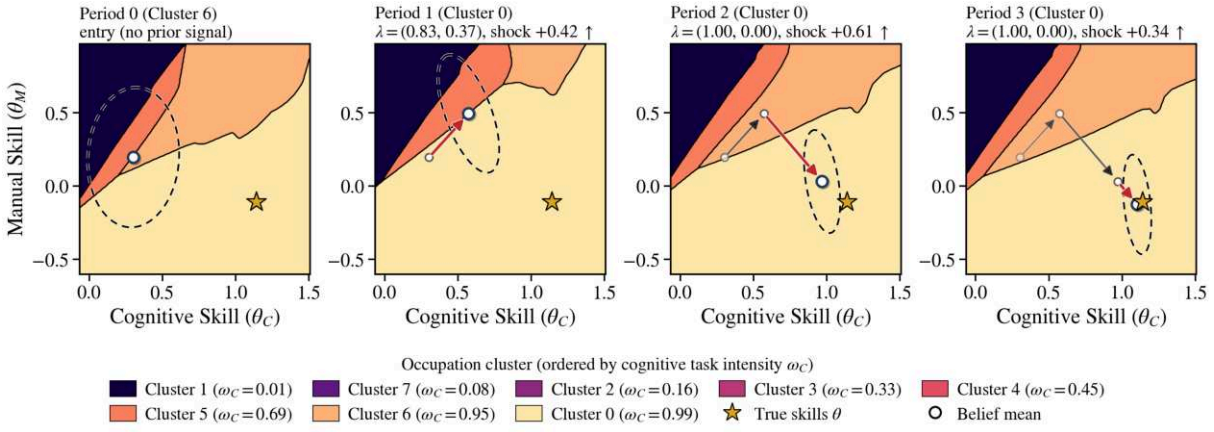
Panel (b) shows a similar pattern for a worker who starts in the cognitive-intensive occupation and ends in the manual-intensive occupation, where her comparative advantage lies. Unlike the previous case, this worker sorts into highly specialized occupations; in fact, her moves are fully horizontal or vertical, meaning that she updates along only one skill dimension. For example, after period 0 she receives a negative signal in the cognitive-intensive occupation that lowers only her cognitive belief and moves her to the manual-intensive occupation, where positive signals confirm her manual comparative advantage, and she settles there for the rest of her career.

5.2 The Drivers of Occupational Mobility

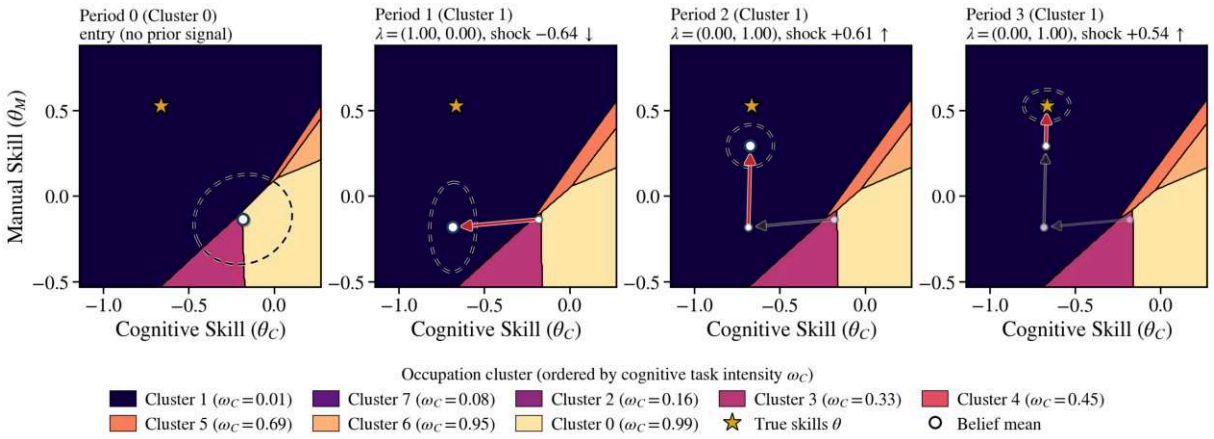
I now use the estimated model to decompose the mobility patterns documented in Section 2 into the contributions of learning about comparative advantage and of task-specific human capital accumulation.[18] To do so, I compare the estimated model with

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

$$w_{ijt} = m_{ij} + \iota_t + f_j(\text{tenure}_{ijt}) + \nu_{ijt},$$



(a) Worker entering Cluster 6 and moving to the cognitive anchor (Cluster 0)



(b) Worker entering the cognitive anchor (Cluster 0) and moving to the manual anchor (Cluster 1)

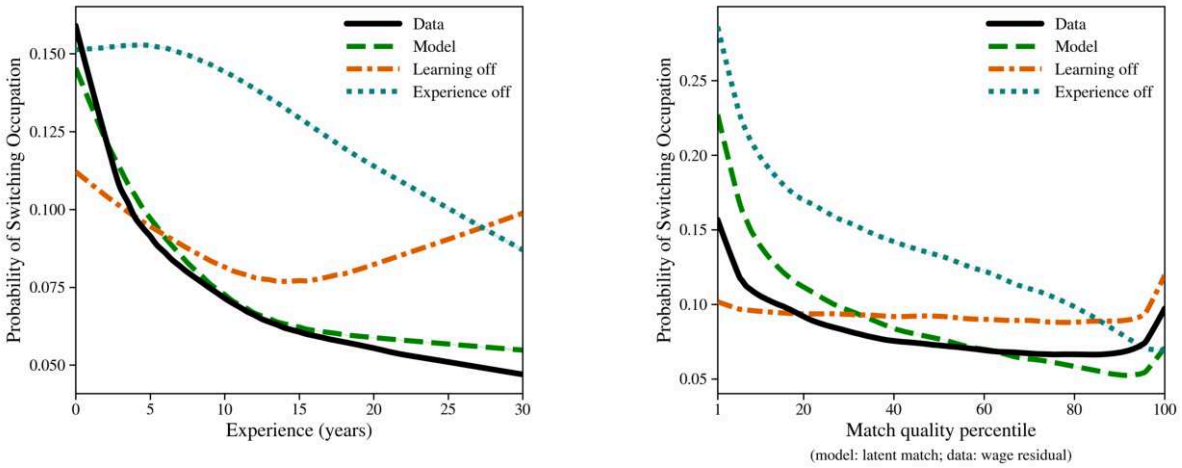
Figure 8: Belief dynamics in the space of latent skills for two simulated workers over their first four periods; each row is one worker. Horizontal axis: cognitive skill θ_C; vertical axis: manual skill θ_M. The shaded regions give the occupation chosen as a function of the belief mean with amenities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

two counterfactual choice rules, *learning off* and *experience off*.[19]

Figure 9 shows that the model reproduces both patterns of occupational mobility. Panel (a) shows the hazard of switching by experience: mobility is high at entry, around 16% in the data and 15% in the model, and falls to about 5% after two decades. The model matches both the level and the decay. It is worth stressing that the profile of this hazard rate is not targeted in the estimation routine (only the average mobility

where ι_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 $\boldsymbol{\lambda}_j'\boldsymbol{\theta}_i$: 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.

[19]Both counterfactuals change only the rule that workers use to choose their occupation, evaluated at the realized states of the simulated baseline economy, and the model and counterfactual lines plot conditional choice probabilities of switching rather than realized choices. Appendix D.6 describes the two counterfactuals in detail.



(a) Hazard of switching by experience

(b) Probability of switching by match quality

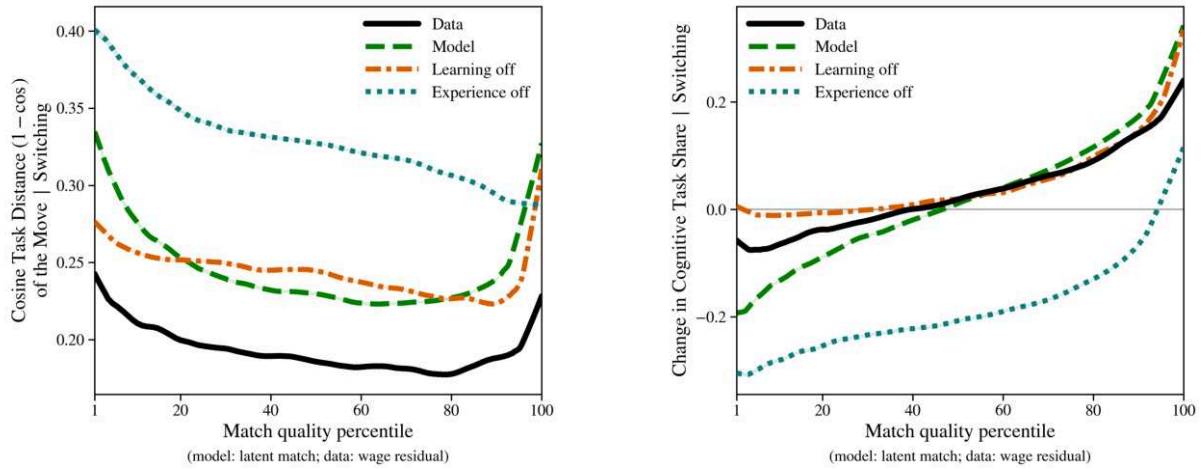
Figure 9: Occupational mobility, model versus data. Data (black), estimated model (green), *learning off* (orange), and *experience off* (teal). *Panel (a)*: probability of switching occupation by years of potential experience. *Panel (b)*: probability of switching by within-occupation match-quality percentile (model: latent match quality; data: worker–occupation fixed effect).

rate is, through ϱ), but rather naturally generated by learning and the accumulation of task-specific experience. Learning shapes this profile in two distinct ways. Early in the career, when uncertainty about comparative advantage is largest, learning generates mobility: without learning, mobility at entry would be about 11% instead of 15%. Later in the career, learning reduces mobility: once workers have learned their skills they have little reason to change occupation, whereas without learning mobility would rise again to about 10% after thirty years. Task-specific human capital instead lowers mobility throughout the career: without it, workers would keep switching at about 15% during their first five years, and mobility would decline only slowly. Appendix E.1 reports further untargeted moments: the distribution of wages at labor-market entry by education group and occupation of entry, the distribution of wages within and across occupations, the wage–tenure profile of occupational stayers, and the decline of mobility with occupational tenure.

One of these moments deserves emphasis. In both the data and the model, the wage distributions of different occupations overlap substantially and share most of their support (Figure 29): occupation dummies explain 22% of the variance of log wages in the data and 9% in the model. The model therefore generates the U-shape and the directionality of occupational mobility even though occupations are not ranked by wage and a worker's wage does not identify her occupation. It does so because sorting operates on multidimensional skills and tasks, on compensating differentials, and on the dynamic returns to task-specific experience, rather than on a single ordering of occupations. This is an empirical improvement over the theory of Groes et al. (2015), in which workers sort on a one-dimensional ability and the wage distributions of different occupations have little or no overlap.

Panel (b) 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. The model reproduces the U-shape, although it overpredicts mobility at the very bottom of the distribution. The two counterfactuals show that the two arms of the U-shape have different drivers. Shutting down learning removes the left arm: mobility becomes flat across the match-quality distribution, so the high mobility of poorly matched workers is generated by learning about their comparative advantage. Shutting down experience accumulation instead removes the right arm: mobility falls monotonically with match quality, so the higher mobility of the best-matched workers reflects moves induced by experience accumulation.



(a) Task-content distance of the switch

(b) Change in cognitive task content

Figure 10: Task content of occupational switches by within-occupation match-quality percentile (model: latent match quality; data: worker–occupation fixed effect), in the data (black), estimated model (green), *learning off* (orange), and *experience off* (teal). *Panel (a)*: task-content distance of the switch, $1 - \cos(\boldsymbol{\omega}_{j_t}, \boldsymbol{\omega}_{j_{t+1}})$, computed on the task weights of the occupation clusters for both the data and the model. *Panel (b)*: change in the cognitive task share of the occupation held, $\Delta\omega_C$; negative values are moves toward more manual-intensive occupations.

Figure 10 turns to the task content of the moves and complements the narrative. Panel (a) shows that workers at both ends of the match-quality distribution make larger moves in task space than workers in the middle, and Panel (b) shows 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 both patterns, although it overpredicts the overall level of task moves. Learning drives the moves at the bottom: without learning, poorly matched switchers move less far in task space and no longer reallocate toward manual work. In the model, low match quality implies negative wage signals that make workers update their beliefs about their skills with a specific directionality, as shown in Figure 8. Experience accumulation drives the moves at the top: moving to more cognitive occupations has a higher continuation value, because cognitive experience becomes more important later in the career even for those workers with a manual comparative advantage. Without the experience accumulation channel, switchers would move toward more manual occupations throughout the distribution and the reallocation of

well-matched workers toward cognitive occupations would almost disappear. Learning thus generates the downward, task-reallocating moves of poorly matched workers, while experience accumulation generates the upward mobility of well-matched workers toward more cognitive occupations.

The model also allows for a more granular decomposition. Learning, for example, drives occupational choice and mobility in two ways: through passive learning and through active learning, which I define as experimentation. The first is induced by passive belief updating upon receiving a wage signal; the second instead acts through the continuation value and occurs when workers choose an occupation in order to learn. In Appendix E.2 (Figures 33 and 34), I show that in the estimated model most of the effect of learning operates through passive learning, while active experimentation plays a minor, although quantitatively relevant, role.

Appendix F.1 uses the same counterfactual choice rules to decompose life-cycle wage growth into the contributions of learning about comparative advantage and of the accumulation of task-specific human capital.

6 The Consequences of Skill Uncertainty

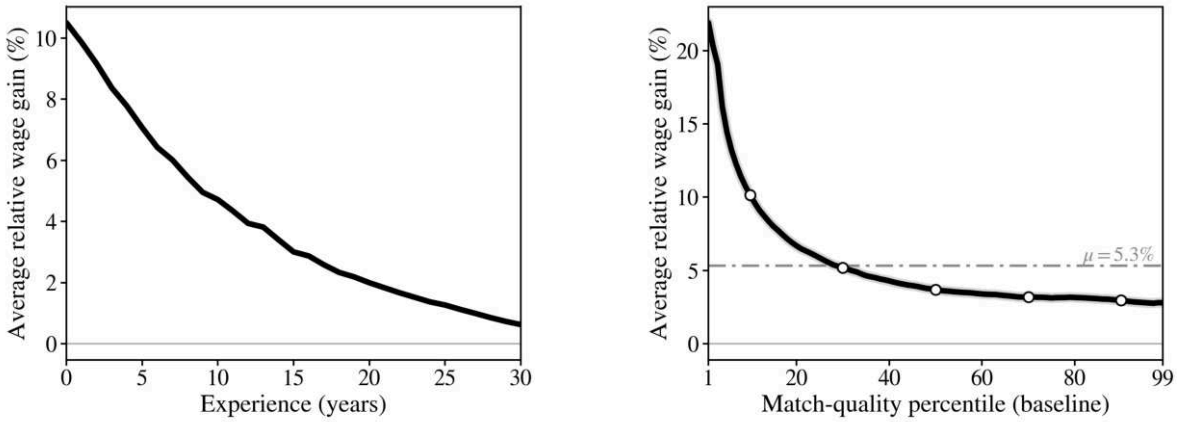
I now use the estimated model to quantify the wage cost of imperfect information about skills.[20] In the model, mismatch arises because the initial prior $\mathcal{N}(\boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)})$ is diffuse relative to each worker's realized $\boldsymbol{\theta}_i$: entrants sort on group means rather than on individual comparative advantage. Beliefs are rational in the sense that the prior beliefs that workers have when starting their careers coincide with the true within-group population distribution of $\boldsymbol{\theta}_i$. While beliefs are correct on average, a worker's own skill vector is only one draw from that distribution, and any given individual's realization can lie far from her prior. Throughout this section and the next I report the effects of each counterfactual on output and wages; Appendix F.2 reports the corresponding welfare effects, measured as consumption-equivalent variations. For all the counterfactuals that change workers' information, welfare moves in the same direction as output, with welfare gains between roughly a quarter and a half of the output gains. Welfare and output diverge when the switching cost itself is changed, as I discuss in Section 6.3, and in the forced-exploration exercise of Appendix F.4.

I construct a counterfactual general equilibrium in which workers have perfect information from the start of their careers: each worker's initial beliefs equal her true skills, and she faces no uncertainty.[21] 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

[20] All stationary counterfactual economies in this and the following sections are computed by simulating $N = 200{,}000$ workers for 400 periods and discarding the first 200 periods as burn-in, so that statistics are computed on the stationary distribution.

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

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 5.1%.[22] Figure 11 then traces the individual wage gains from removing these frictions, over the life cycle and across the match-quality distribution within occupation.



(a) Wage gains over the life cycle

(b) Wage gains by match-quality percentile

Figure 11: Wage gains from removing information frictions. *Panel (a)*: average relative wage gain over the life cycle (horizontal axis: years of potential experience; vertical axis: percent wage gain). *Panel (b)*: average relative wage gain by the worker's within-occupation match-quality percentile in the baseline economy (horizontal axis); the dashed horizontal line marks the unconditional mean gain, 5.3%. Shaded bands are bootstrap confidence intervals.

This output gain reflects the misallocation of workers to occupations that do not match their true comparative advantage under imperfect information. The wage gains are concentrated at the start of the career and flatten over the life cycle. At labor-market entry, the average worker would earn about 10.5% more if she knew her true skills. This gap narrows as workers learn and change occupation accordingly, falling to about 2% after twenty years of experience and fading toward zero (about 0.6%) after 30 years. Mismatch driven by imperfect information is largely an early-career phenomenon: the cost of imperfect information is borne mostly in the first decade, when there is more skill uncertainty and workers are still discovering their comparative advantage, and it is progressively resolved as learning and re-sorting bring imperfect-information workers closer to their perfect-information counterparts. Learning, imperfectly transferable experience, and the cost of switching interact to create a strong path dependence, so that some workers do not fully converge to their optimal career in the imperfect infor-

[22] Aggregate output in the stationary equilibrium is $Y = \left(\sum_{j\in\mathcal{J}} \alpha_j Y_j^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, where $Y_j = \int \exp(\varepsilon)\, Y_j(x)\, \mathbf{1}_{\{d=j\}}\, \pi^*(x)\, dx\, dF_j(\varepsilon)$ is occupation j's aggregate output, $Y_j(x)$ is individual output from the task technology (5) net of the idiosyncratic shock, 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.

mation scenario, but on average the residual gain closes gradually.

On average, this reallocation operates almost entirely through better sorting rather than through experience accumulation in the right dimension. Decomposing the average life-cycle 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 10.5% at entry toward zero late in the career.[23] The full decomposition is reported in Figure 35 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 5.3%. 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.

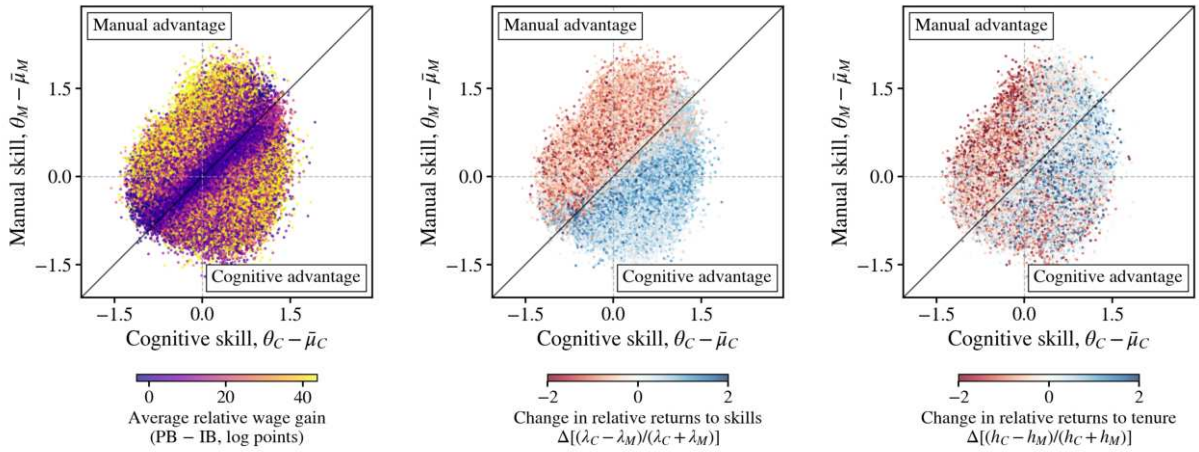
6.1 Who Benefits Most, and Why

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

[23] 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 relative returns to skills

(c) Change in relative returns to tenure

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 information, positioned by her true log cognitive skill θ_C (horizontal axis) and true log manual skill θ_M (vertical axis), both measured relative to the population mean $\bar{\mu}_k = \sum_g \pi_g \mu_{gk}$; the solid line is the 45° line $\theta_C - \bar{\mu}_C = \theta_M - \bar{\mu}_M$. *Panel (a)*: color is the worker's average relative wage gain over the life cycle, in log points. *Panel (b)*: color is the change, from the imperfect- to the perfect-information economy, in the relative returns to cognitive versus manual skills of the occupation held. *Panel (c)*: the change in the relative returns to cognitive versus manual tenure, evaluated at the worker's task tenures in each economy (periods in which both task tenures are below one year are excluded). In Panels (b) and (c) blue denotes a move toward a more cognitive occupation and red a move toward a more manual one.

ward more cognitive-intensive occupations, red workers toward more manual-intensive ones. A sharp separation along the 45° line emerges. Workers with $\theta_C - \bar{\mu}_C > \theta_M - \bar{\mu}_M$ (cognitive specialists, below the 45° line) systematically move to more cognitive occupations when given perfect information, while workers with $\theta_M - \bar{\mu}_M > \theta_C - \bar{\mu}_C$ move to more manual ones. Panel (b) shows this for the relative returns to skills $\boldsymbol{\lambda}_j$; Panel (c) shows an analogous but weaker and noisier pattern for the relative returns to tenure, driven mainly by workers with manual comparative advantage, who move toward occupations with higher returns to manual 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.

6.2 The Effects of Information Interventions

I now use the model to assess how much of the output and wage losses can be recovered by implementing realistic information interventions, and how their effectiveness changes when they are provided at different stages of the career. I take as empirical anchor the field experiment of Kiss et al. (2023). The paper studies two related information interventions; I calibrate the model intervention to the effects estimated in the first experiment, in which young jobseekers receive a short computerized assessment

covering numeracy, communication, and concept formation skills. The information treatment consists of individualized feedback about where the treated worker stands in the skill distribution of treated individuals. This information treatment is therefore informative about workers' comparative advantage in skills. Treated workers update their beliefs about their own skills and re-direct their search toward occupations that better match their assessed comparative advantage. A desirable feature of this class of intervention is that it is relatively cheap to deploy.[24] The estimated treatment effect amounts to an increase of 13.7 percentage points in the share of workers whose stated career aspiration is aligned with their measured comparative advantage.

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

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

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

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

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

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

[24] The cost of the computerized assessment and information treatment reported by Kiss et al. (2023) is \$45.74 USD PPP per worker. I convert it to 2020 EUR via the OECD USD–EUR PPP rate (\$1 = € 0.92), yielding € 42.08; deflating to 1995 real EUR via the Portuguese CPI (2020–to–1995 deflator of 0.65) gives € 27.35 per worker, which is the cost of the information treatment per worker, expressed in the model and data wage units (real hourly EUR 1995) used in this paper.

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

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

$\sigma_\zeta \approx 0.606$. I then feed this value into the estimated model and re-simulate the economy, holding all other parameters at their estimated values.

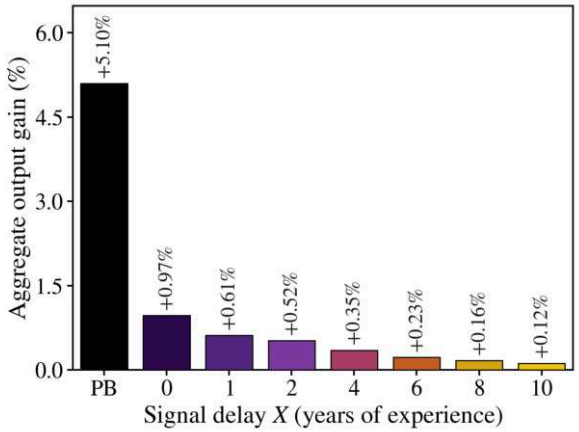
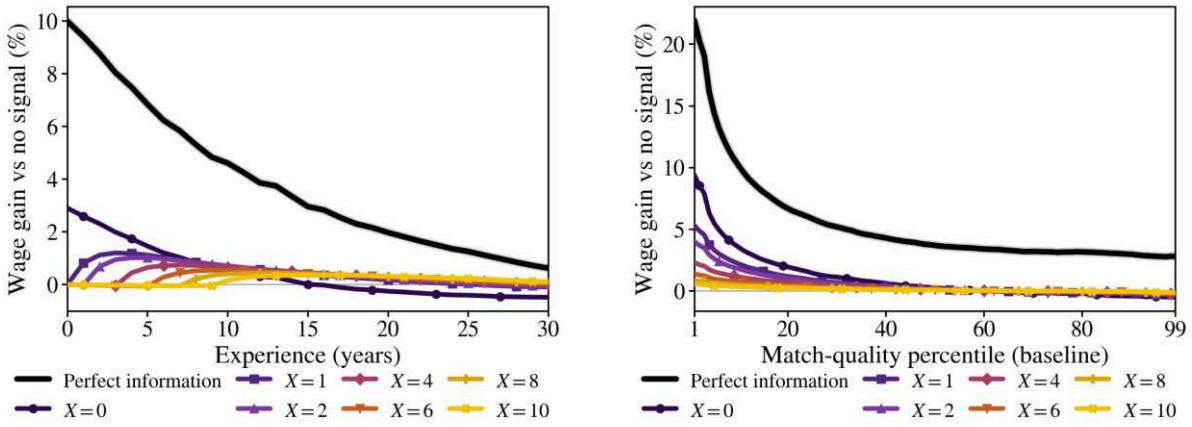


Figure 13: Aggregate output gain (percent) relative to the baseline economy when the calibrated skill assessment is delivered $X \in \{0, 1, 2, 4, 6, 8, 10\}$ years after labor-market entry, holding the calibrated signal precision σ_ζ^2 fixed at its at-entry value. Each bar is a separately solved stationary equilibrium. The leftmost bar (perfect information, PB in the figure) 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 0.97%, that is, around 19% of the total misallocation generated by information frictions about skills. The wage gains arising from the information treatment are strongly front-loaded in the career: average wages are about 2.9% above the baseline at entry, then decline as better-matched entrants accumulate experience and baseline workers catch up after updating their beliefs and sorting accordingly, crossing zero after roughly fifteen years and settling mildly negative (-0.5%) thereafter (Figure 14, Panel (a)). These gains come almost entirely from better matching on skills at the first occupational choice; the late-career reduction in the gap between the two economies reflects cross-task experience accumulation (Appendix F.1), as the policy induces early specialization and so reduces some of the exploratory experience that baseline workers accumulate under imperfect information. In other words, imperfect information may generate mismatch, but conditional on reaching their ideal career path at some point, cross-returns to different task experience may actually be beneficial.

The treatment does not help everyone equally, and Panel (b) in Figure 14 shows how the gains are distributed. The wage gains are concentrated at the bottom of the match quality distribution within occupation: workers in the lowest percentiles in the baseline economy, who are more likely to be mismatched, gain about 9%; the gain fades to roughly zero by the median. A single noisy assessment provided early in the career recovers about two fifths of the perfect-information gain for the worst-matched workers (9% against 22%), and essentially nothing for the well-matched.

A natural question is how the value of the policy depends on when it is delivered. I therefore repeat the experiment delivering the signal in (41) at $t \in \{0, 1, 2, 4, 6, 8, 10\}$



(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-occupation match-quality percentile in the baseline economy, for each delay, with the perfect-information counterfactual (solid black) as a benchmark.

periods after entry, holding its precision σ_ζ^2 fixed and keeping workers myopic about when it will arrive. The later the assessment comes, the lower are the realized output and wage gains: the output gain is largest at entry (Figure 13), drops by roughly half within the first two years, and is nearly gone by the end of the first decade. The information treatment, when provided at labor-market entry, increases wages for the first years compared to the baseline economy. A delayed assessment generates only a small hump once it arrives, mainly because task-specific experience, which is imperfectly transferable, contributes to creating a lock-in effect. Even if there is an occupation in which workers' skills might be more productive, switching would make the worker lose some task-specific experience. This implicit cost of switching is increasing over labor-market experience and directly reduces the effectiveness of the signal if delayed.

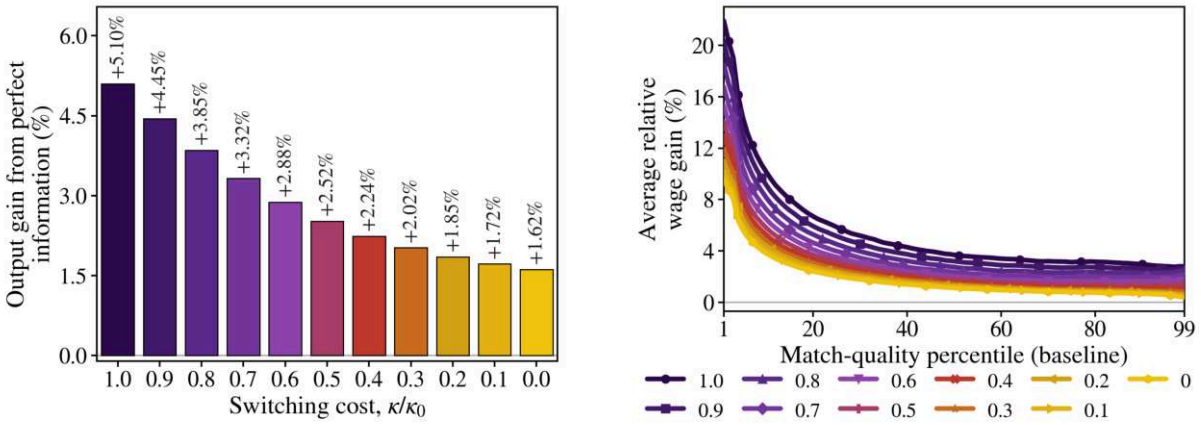
In Appendix F.4 I study an experiment that directs each labor-market entrant, for one period only, into the occupation that is most informative about her comparative advantage, after which she chooses freely. The forced exploration reduces wages early in the career but improves learning and sorting from the second period onward. The large initial reduction is driven not only by forcing workers into occupations for which they are poorly matched, but also by the downward-sloping demand curve for occupation-specific goods. Forcing everyone to move into their most informative occupation pushes prices down in the informative occupations while raising prices elsewhere. Overall, aggregate output rises by 1.11%.

6.3 Switching Costs and Information Frictions

The cost of imperfect information depends on how easily workers can switch once they learn where their best fit is. If occupational mobility were free, a worker who discovers

she is mismatched could simply switch at any point in her career, even if learning is slow. Switching costs and information frictions are thus complementary sources of misallocation, and this subsection examines how they interact. I reduce the switching cost κ from its estimated value κ_0 to zero in steps of ten percent of κ_0, 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 5.1% at the estimated switching cost to under 2% when moving is free (Panel (a)), and the gain from perfect information shrinks at every point of the match-quality distribution as the switching cost falls (Panel (b)). This result is intuitive: lower moving costs allow workers to switch once they learn their types, so misallocation generated by information frictions is lower.



(a) Output gain from perfect information

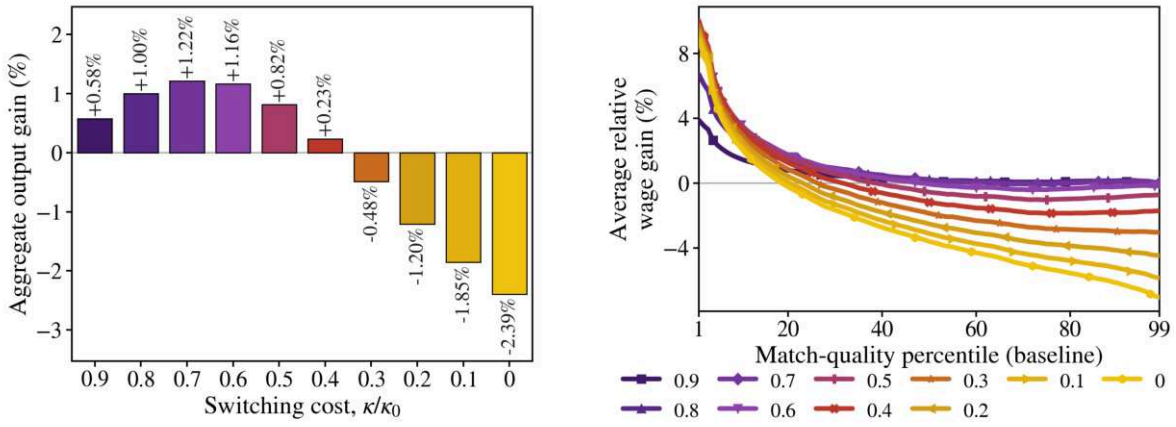
(b) Wage gain from perfect information by match quality

Figure 15: The cost of information frictions at different levels of the switching cost. At each value of κ/κ_0 both the imperfect- and the perfect-information stationary equilibria are re-solved, and the figure reports the gap between them. *Panel (a)*: aggregate output gain from moving to perfect information, $(Y^{PI} - Y^{II})/Y^{II}$, in percent, by κ/κ_0. *Panel (b)*: average relative wage gain from perfect information by the worker's within-occupation match-quality percentile, by κ/κ_0. The corresponding life-cycle wage profiles are in Appendix F.3.

The second comparison is across the imperfect-information economies themselves, and asks how aggregate output responds to the switching cost that workers actually face. Here output is not monotonically decreasing in the switching cost (Figure 16): it first rises as mobility becomes cheaper, with a peak when switching costs are 70% of the estimated value, and then falls, turning negative and reaching a loss of about 2.4% when switching is completely free[27].

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

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



(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 life-cycle wage profiles are in Appendix F.3.

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

Output is not welfare, and the two diverge in this experiment. Appendix F.2 reports, for each counterfactual, the consumption-equivalent welfare gain: the uniform and permanent wage increase that would make the average worker in the baseline economy as well off as in the counterfactual. This measure prices everything workers value in the model, that is, expected wages, amenities, the switching cost itself, and the option value of the taste shocks, whereas output counts only production. Lowering the switching cost raises welfare monotonically, by 4.2% at $\kappa/\kappa_0 = 0.7$ and by 61% when mobility is free, even where output falls (Figure 39). The switching cost is a utility cost paid by every mover, and its estimated value of 1.10 utils is equivalent to a wage

change of about 110 log points, so removing it is a direct gain to movers; in addition, workers re-sort on amenities and taste shocks that output does not count. The lower output at low switching costs therefore does not indicate that the additional mobility is inefficient from the workers' point of view: workers internalize the task-specific experience they forgo and still choose to move, because the switching-cost savings and the non-wage components of their choice are worth more to them. It indicates instead that task-specific experience is a real cost of mobility, which is borne as lower production. The welfare gain from perfect information, in contrast, falls with the switching cost like the output gain, from 2.0% at the estimated switching cost to 0.5% when moving is free.

I close this subsection by providing two practical examples of the frictions the switching cost captures. First, if occupations are concentrated in different local labor markets then workers have to change location in order to switch occupation, so part of κ_0 may reflect the monetary and psychic cost of geographic mobility (Kennan and Walker 2011). Reducing switching costs in this case can be thought of as providing moving vouchers to mismatched workers. Second, reallocating across occupations often requires passing through an unemployment spell, as in the island and sectoral-reallocation tradition in which workers move by first becoming unemployed (Lucas and Prescott 1974; Chodorow-Reich and Wieland 2020), what Pilossoph (2012) calls "move unemployment" and Alvarez and Shimer (2011) call "search unemployment". In this case, κ_0 represents the cost of being unemployed, and one can think of switching-cost reductions as policies that affect unemployment benefits (duration, overall amounts, etc.).

7 Artificial Intelligence and Skill Uncertainty

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

7.1 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 $e_{E0} = 0$, $e_{E1} = 1$, and $e_{E2} = e_{E3} = 0.5$. Manual skills are unexposed ($e = 0$),

since language models cannot perform physical tasks. Averaging skill exposure by ESCO importance within each occupation and aggregating to the $|\mathcal{J}| = 8$ clusters yields a cognitive exposure $x_{jC} \in [0,1]$ for each cluster, with $x_{jM} = 0$; 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

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

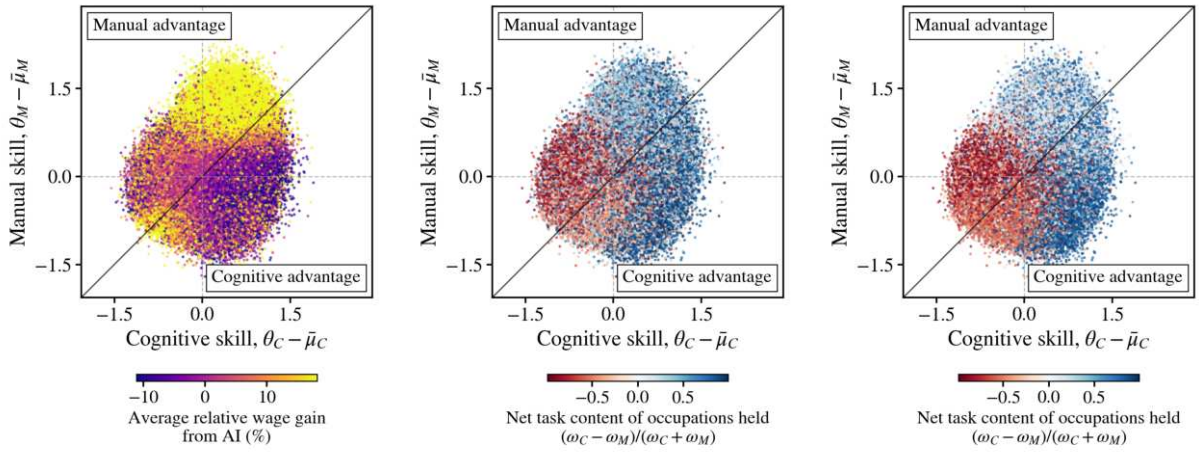
renormalized so that $\omega_{jC}^{\text{AI}} + \omega_{jM}^{\text{AI}} = 1$, so the manual weight rises mechanically. Because the effective wage parameters scale with the task weights ($\lambda_{jk} = \omega_{jk} a_{jk}$ and $h_j(\boldsymbol{\tau}) = \sum_k \omega_{jk} h_{jk}(\tau_k)$), the shock scales the skill loadings and the returns to experience 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.

7.2 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 life-cycle wage gain from AI for each worker at her true skills. Workers with strong manual comparative advantage—high-ability manual workers in the upper region, above the 45° line where $\theta_M - \bar{\mu}_M > \theta_C - \bar{\mu}_C$—gain the most, because AI raises the relative return to the manual skills in which they have a strong comparative advantage. Cognitive specialists in the lower-right region lose as cognitive tasks, in which they have a strong comparative advantage, become less important. Generalists near the 45° line are roughly unaffected.

Panels (b) and (c) trace these wage effects to occupational reallocation, showing the net task content of the occupations a worker holds over the life cycle before and after the shock. Workers, especially those with a manual comparative advantage, sort into more manual-intensive occupations after AI (the red region expands). Cognitive specialists instead remain in cognitive-intensive occupations but earn lower returns there. On the other hand, high-ability generalists benefit by moving toward more balanced occupations. These effects in the skill space aggregate into the familiar polarization pattern of the routinization literature (Autor et al. 2003): across the baseline wage distribution the wage gains are U-shaped—workers in the middle lose while those at the bottom and top gain—and employment shifts out of cognitive-intensive clusters and into manual-intensive ones, whose occupation prices rise in general equilibrium (Figure 38 in the Appendix).



(a) Average wage gain from AI

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

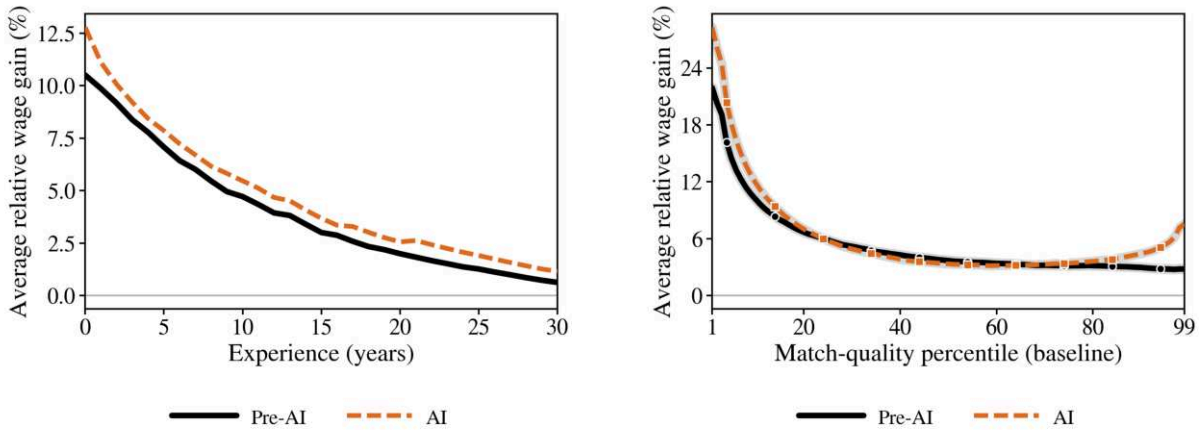
(c) Net task content of occupations held, post-AI

Figure 17: The direct effects of AI in the space of true skills. In each panel a point is a simulated worker placed at her true log cognitive skill θ_C (horizontal axis) and true log manual skill θ_M (vertical axis), both measured relative to the population mean $\bar{\mu}_k = \sum_g \pi_g \mu_{gk}$; the solid line is the 45° line $\theta_C - \bar{\mu}_C = \theta_M - \bar{\mu}_M$, 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.

7.3 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 6. 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 5.1% in the baseline to 5.8% under AI.

AI raises the cost of skill uncertainty over the whole career path. Panel (a) shows that the wage gain from perfect information is higher in the economy affected by the AI shock at every level of experience: at labor-market entry it rises from about 10.5% to about 12.8%, and the wedge between the two curves persists as experience accumulates. By raising the dispersion of returns across occupations, AI makes correct sorting more valuable, so the cost of holding the wrong occupation under imperfect beliefs is higher. Panel (b) shows who accounts for this increase. The extra cost of skill uncertainty under AI is concentrated at the two ends of the match-quality distribution. For the worst-matched workers within occupation, the gain from perfect information rises from about 22% to about 28%. At the top of the distribution AI also increases the gains from removing information frictions: well-matched workers within occupation now



(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 (orange, dashed) stationary equilibria. *Panel (a)*: by years of potential experience. *Panel (b)*: by the worker's within-occupation match-quality percentile in the corresponding baseline economy. Shaded bands are bootstrap confidence intervals.

gain 7.6%, compared with 3% in the pre-AI scenario.

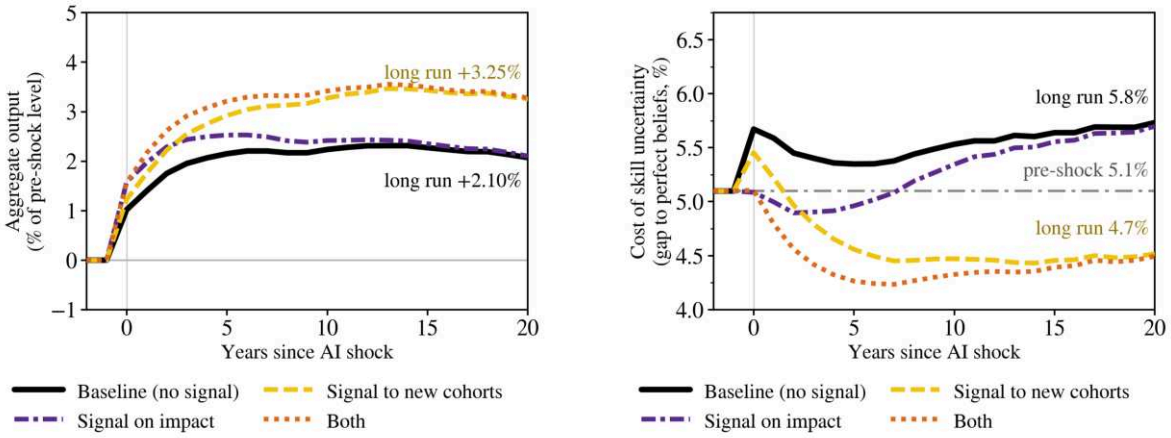
Appendix F.1 repeats the decomposition of life-cycle wage growth into learning and human capital accumulation for the post-AI economy.

7.4 Output and Misallocation along the Transition Path

The counterfactuals above compare stationary equilibria and therefore speak to the long-run incidence of the AI shock, abstracting from the adjustment dynamics that follow it. I now study the effects of the AI shock along the transition path. In particular, I compute the transition path from the pre-shock to the post-shock equilibrium[28] to understand the consequences of information frictions not only in the long run, but also on impact and along the transition path. Furthermore, I assess the effectiveness of the same calibrated information treatment in reducing misallocation over the transition path. I model the AI shock as an MIT shock: it is an unanticipated shock and workers have perfect foresight over the entire subsequent path of prices and technology. The path for equilibrium prices is solved via a backward–forward shooting algorithm, holding the CES weights and the per-period numeraire fixed exactly as in the stationary problem; the algorithm used to compute this transition path is described in Appendix B.3.

Two results stand out. First, the AI shock raises aggregate output on impact and along the entire transition (Figure 19, Panel (a)). Aggregate output jumps to about $+1.0\%$ above its pre-shock level on impact, keeps rising as workers reallocate across

[28]The transition path is simulated with the same population of $N = 200{,}000$ workers used for the stationary counterfactuals, and its pre- and post-shock endpoints correspond to the pre-AI and post-AI stationary equilibria of Section 7.3. Output along the path is measured relative to the pre-AI stationary output level.



(a) Aggregate output

(b) Cost of skill uncertainty

Figure 19: The AI transition with and without skill signals, by years since the AI shock (horizontal axis). Baseline imperfect-information economy with no signal (black, solid); skill signal delivered to all workers on impact (purple, dash-dotted); skill signal delivered to every new cohort at entry (yellow, dashed); both (orange, dotted). *Panel (a)*: aggregate output, percent deviation from the pre-shock steady state. *Panel (b)*: cost of skill uncertainty, the output gap to the perfect-information economy (percent); the dash-dotted gray line marks its pre-shock level. Labels report the pre-shock and long-run values.

occupations, peaks at roughly +2.3% after about fourteen years, and settles at its new long-run level of about +2.1%. Output does not jump to its new long-run level on impact because experienced workers are locked into career paths chosen under the old task content of occupations: part of the gain materializes only as new cohorts, who sort on the new task content from entry, replace them. Second, AI increases the misallocation generated by information frictions about skills (Figure 19, Panel (b)). The output gain from moving to perfect information jumps from about 5.1% before the shock to about 5.7% on impact, as the shock makes many existing matches worse, falls back to about 5.35% within five years as incumbents learn and reallocate, and then rises again, as cohorts enter the new economy, to roughly 5.8% in the new long run.

I then ask how effective the information treatment calibrated in Section 6.2 is along the transition, comparing two ways of delivering it. A one-shot skill signal delivered to all workers on impact raises output on impact from about +1.0% to about +1.6% and temporarily pushes the cost of skill uncertainty below its pre-shock level, but its effect is transitory: as treated cohorts retire, both output and the cost of skill uncertainty converge back to the untreated path. A skill signal delivered to every new cohort at entry instead has a small effect on impact, when only the first cohort is treated, but its effect builds up as treated cohorts replace untreated ones. In the long run it raises output to about +3.25% above the pre-shock level, 1.1 percentage points more than the untreated economy, and it lowers the cost of skill uncertainty from about 5.8% to about 4.7%, closing about a fifth of the gap. Delivering both signals combines the two profiles: the economy follows the one-shot path on impact and the entry-signal path in the long run. From a policy perspective, these results suggest that information provision is more valuable in the economy transformed by AI than in the pre-AI economy, since the long-run output gain of the entry signal is larger than the 0.97% it delivers before

the shock. The timing of the intervention matters: a one-shot assessment of incumbent workers smooths the reallocation on impact, while a permanent assessment of new entrants delivers most of the long-run gains.

8 Conclusion

In this paper, I develop and estimate a general equilibrium dynamic Roy model in which workers learn about their multidimensional skills and accumulate task-specific human capital. Using Portuguese administrative data, I enrich the occupational mobility patterns documented by Groes et al. (2015) by looking at the task content of occupational mobility and the wage change upon switching conditional on the wage position in the origin occupation. I find that changes in task content upon switching are also U-shaped in workers' positions within their origin occupations and directional in task space. Moreover, switching raises the wages of poorly matched workers the most. Through the lens of the model, I interpret the upward mobility of well-matched workers as reflecting experience accumulation, whereas the downward, task-reallocating moves of poorly matched workers are driven by learning about comparative advantage.

These findings suggest that mismatched workers could benefit from having more precise beliefs about their skills earlier in their career. Removing information frictions would raise aggregate output by 5.1%. Individual wage gains are largest at entry, fade to roughly zero over the career, and are concentrated among the worst-matched workers within occupations and among specialists. Policy can recover part of this loss. An information treatment calibrated to Kiss et al. (2023) and delivered at entry raises aggregate output by 0.97%, but its effectiveness decays quickly when delivered later in the career due to the presence of switching costs and imperfectly transferable human capital. Information frictions also interact with switching costs: modest reductions in switching costs help workers reallocate after learning about their comparative advantage, while eliminating switching costs induces excessive churning, destroys task-specific experience, and lowers output.

Finally, a task-biased technology shock induced by Large Language Model adoption raises aggregate output but increases the misallocation generated by information frictions, widening the output gap relative to the perfect-information economy. Information provision therefore becomes more valuable: delivered to every new cohort throughout the transition, the skill signal raises long-run output by about 1.1%, compared with 0.97% before the shock.

References

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

pp. 1043–1171.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Guvenen, F., Kuruscu, B., Tanaka, S. and Wiczer, D. (2020), 'Multidimensional skill mismatch', *American Economic Journal: Macroeconomics* **12**(1), 210–44.

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

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

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

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

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

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

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

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

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

Judd, K. L., Maliar, L., Maliar, S. and Valero, R. (2014), 'Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain', *Journal of Economic Dynamics and Control* **44**, 92–123.

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

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

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

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

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

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

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

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.

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

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

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

Pilossoph, L. (2012), A multisector equilibrium search model of labor reallocation. Working paper.

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

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

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

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

Topel, R. H. and Ward, M. P. (1992), 'Job mobility and the careers of young men*', *The Quarterly Journal of Economics* **107**(2), 439–479.
URL: *https://doi.org/10.2307/2118478*

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

A Appendix: 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 life-cycle 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 1 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 its mobility moments are compared with the data at the level of the occupation clusters, with workers ranked within occupation and year on the worker–occupation fixed effect of their wages in the large sample.

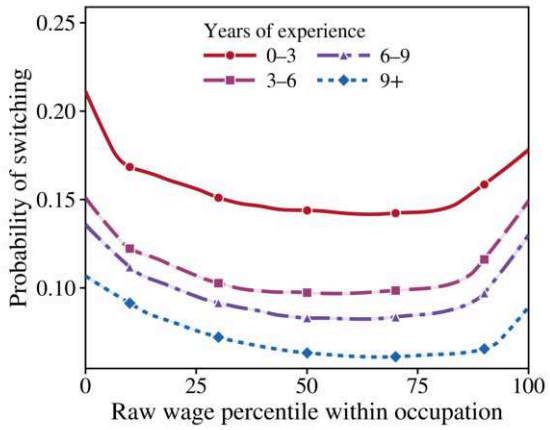
Table 1: Comparison of the Large Sample and the Estimation Sample

Large Sample		Estimation Sample	
Statistic	**Value**	**Statistic**	**Value**
Total observations	29,428,668	Total observations	2,434,141
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	1996–2019
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	57.79	Men	55.91
Women	42.21	Women	44.09
Low education (1–2)	67.90	Low education (1–2)	33.43
Medium education (3)	20.07	Medium education (3)	33.30
High education (4–5)	12.03	High education (4–5)	33.26

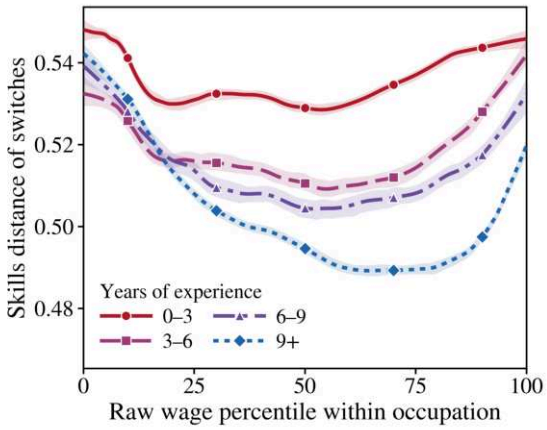
Notes: The large sample is designed to maximize representativeness and the number of observations for descriptive facts. The estimation sample is constructed from the large sample by applying the following additional restrictions: (i) entry age below 30 with zero initial experience; (ii) first consecutive 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. In the large sample, gender and education shares are computed over unique workers at their first observation; these are the education shares used in the quantitative counterfactuals.

Task and Skill Classifications. The task 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 ISCO-88, embedded in the Portuguese code CNP1994. Starting in 2010, CPP2010 was introduced in *Quadros de Pessoal*. This code is based on ISCO-08, 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.

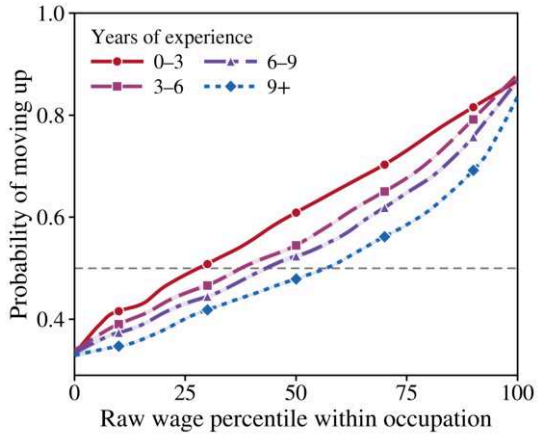
Robustness: Raw Hourly Wages. The empirical analysis in Section 2 uses residualized wages as a proxy for unobserved worker ability or match quality. Figure 20 replicates the main empirical facts from that section using raw real hourly wages to construct the within-occupation percentiles. The patterns—the 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.



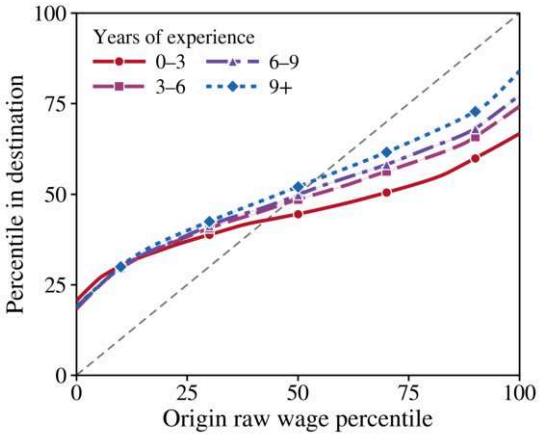
(a) Probability of switching occupations



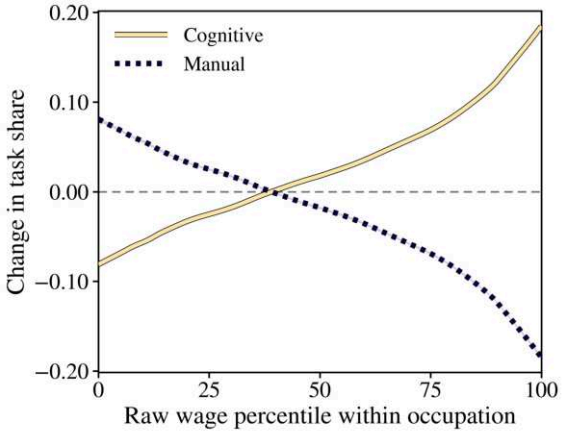
(b) Skills distance conditional on switching



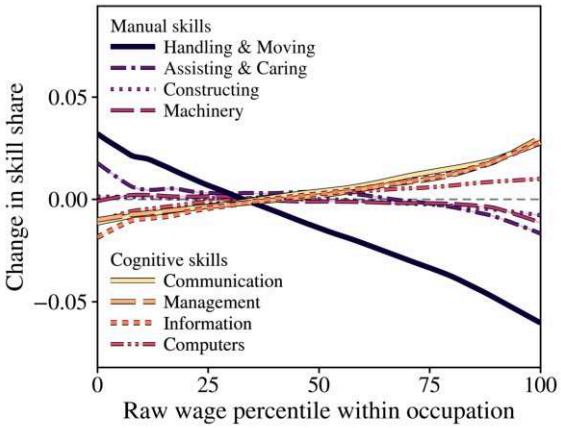
(c) Probability of moving to a higher-paying occupation



(d) Rank in the destination occupation



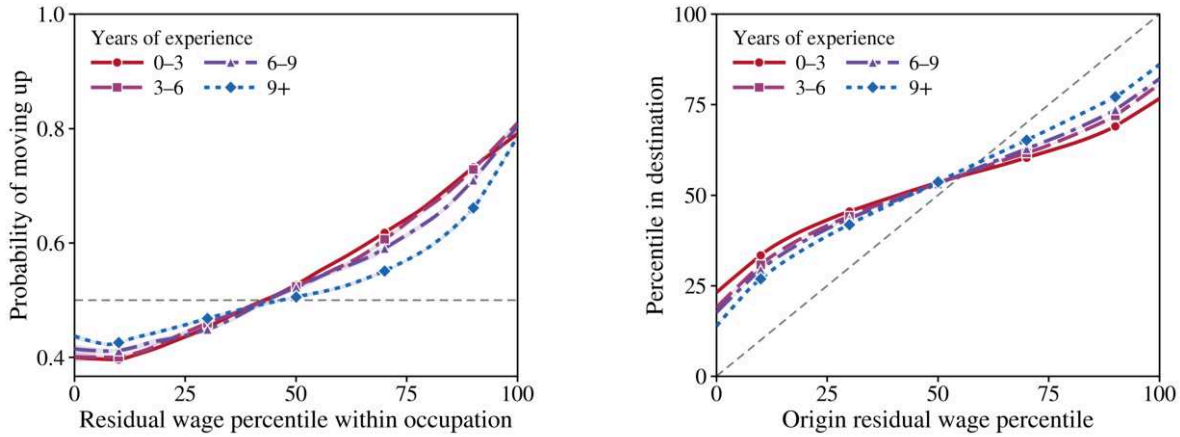
(e) Change in Cognitive–Manual task shares



(f) Change in ESCO skill shares

Figure 20: Robustness: Raw Wages. The figure replicates the main empirical facts using within-occupation percentiles of raw rather than residualized hourly wages. Panels (a) and (b): probability of switching occupations and skills distance conditional on switching, by experience group. Panels (c) and (d): probability of moving to an occupation with a higher average wage and percentile in the destination occupation, by experience group. Panels (e) and (f): change in Cognitive–Manual task shares and ESCO skill shares conditional on switching.

Directionality and Rank-to-Rank Mobility by Experience. Figure 21 reports the directionality of occupational mobility documented in Section 2. Panel 21(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 21(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 21: *Panel (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 (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 2 reports summary statistics for the skills distance measure defined in equation (2). 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), a superset of the 103 occupations observed in the large sample. 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.

Among all occupation pairs, whether or not a switch between them is observed, the most distant pair is glass and ceramics kiln operators and shoe cleaning and street services workers ($d = 0.99$), and 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 most common occupational switch by education group, by gender, and by gender $\times$ education. The most popular move for low-educated workers (education levels 1–2) is between mining and quarrying laborers and building frame workers ($d = 0.21$; 23,000 transitions), while the most common move for medium- and high-educated workers (levels 3 and 4–5) is between accounting clerks and bookkeepers ($d = 0.36$; 11,000 and 9,000 transitions). For high-educated men, the most frequent switch is between physical science technicians and architects and engineers ($d = 0.43$). The most frequent move for men is between laborers and building frame workers, while for women the most frequent move is between accounting clerks and bookkeepers.

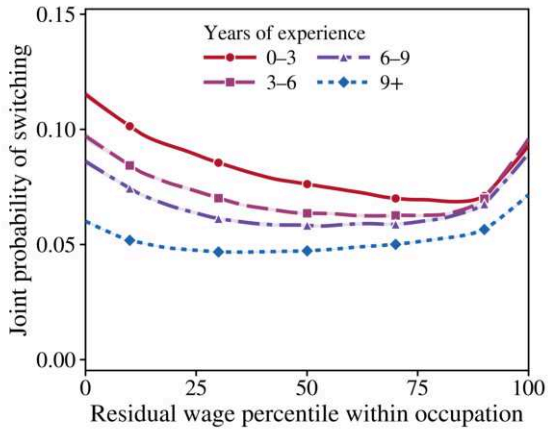
Table 2: Skills Distance Summary Statistics. Cosine distance $d(j, j') = 1 - \text{Sim}(j, j')$ computed using the 296-dimension ESCO skill classification. ISCO-88 three-digit occupation codes.

	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

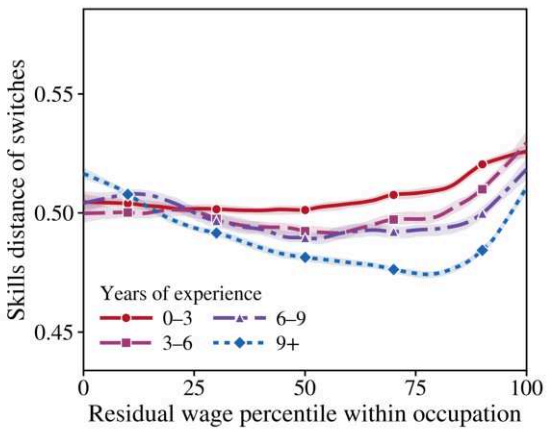
Panel C: Most common occupation switch by group

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

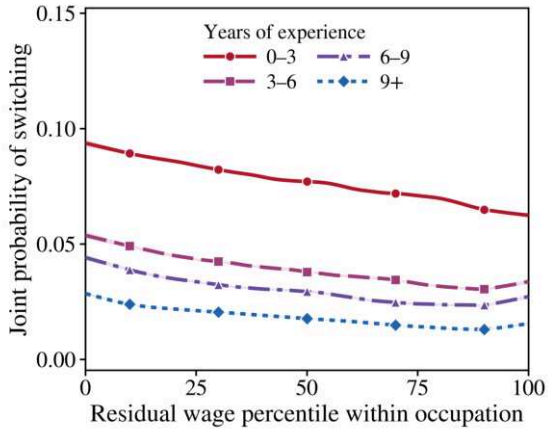
Large Sample Construction. I construct my measure of earnings by summing a worker's base salary and any regular earnings supplements and dividing the amount by her normal hours of work; this yields the same "straight-time" hourly wage used in Card et al. (2016). I obtain real hourly wages by deflating the "straight-time" wage using the CPI with 1995 as base year. From the initial dataset I eliminate entries if the worker ID is reported multiple times per year; for each worker I drop the entire history if sex changes over time, age is inconsistent, or the worker is not aged between 20 and 55. I



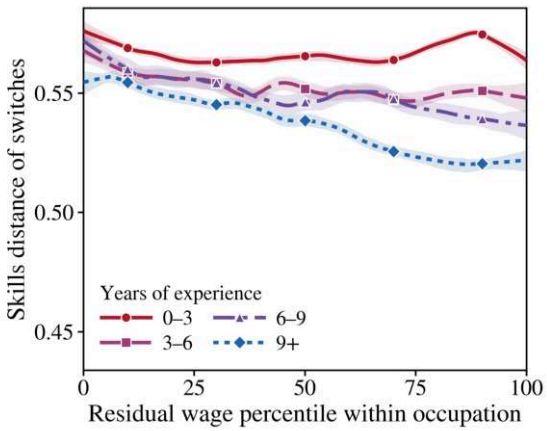
(a) Within firm: probability of switching occupation



(b) Within firm: skills distance of switches



(c) Across firms: probability of switching occupation



(d) Across firms: skills distance of switches

Figure 22: Occupational mobility within and across firms, by residualized within-occupation wage percentile (horizontal axis) and experience group. *Panels (a) and (c)*: the figures decompose the probability of switching occupation between t and $t+1$ in Figure 2(a) into the joint probability of switching occupation and staying in the same firm (Panel (a)) and the joint probability of switching occupation and changing firm (Panel (c)); the two panels sum to the total probability of switching and share a common vertical scale. *Panels (b) and (d)*: skills distance between origin and destination occupation, conditional on switching occupation within the same firm and across firms, on a common vertical scale.

also drop observations if the hiring date for any job is missing or inconsistent across observations. If the individual is observed in two consecutive years at different firms, but the hiring date for the second job is the same as the hiring date for the first job, I drop the entire history for a worker. I also drop individuals who have missing education and I drop workers with education decreasing over time. The sample is further restricted to full-time workers (at least 30 hours per week or 130 hours per month) who earn at least the minimum wage. Following Fonseca et al. (2018), I drop workers in occupations and industries associated with the primary sector, given that the survey is optional for those. I also drop workers in public administration and defense[29]

[29]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

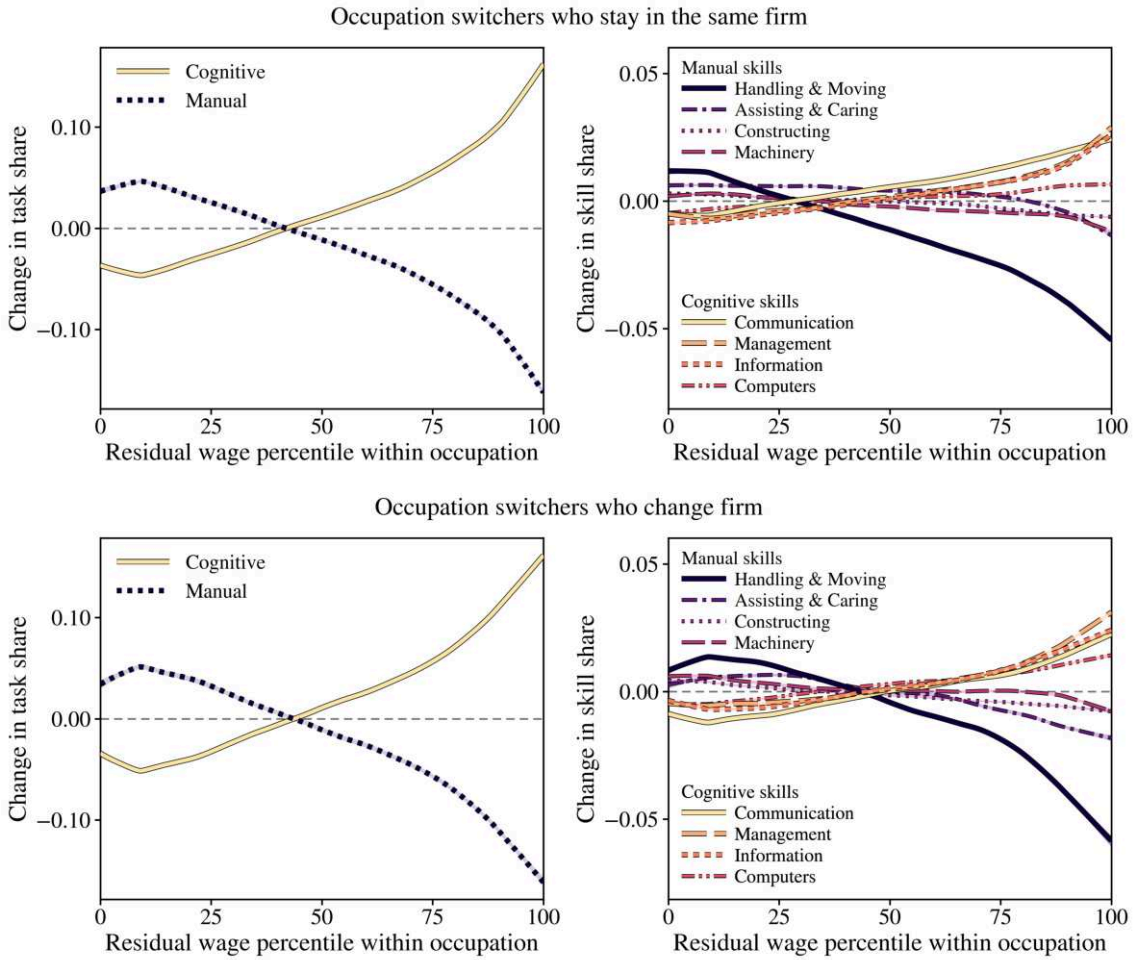


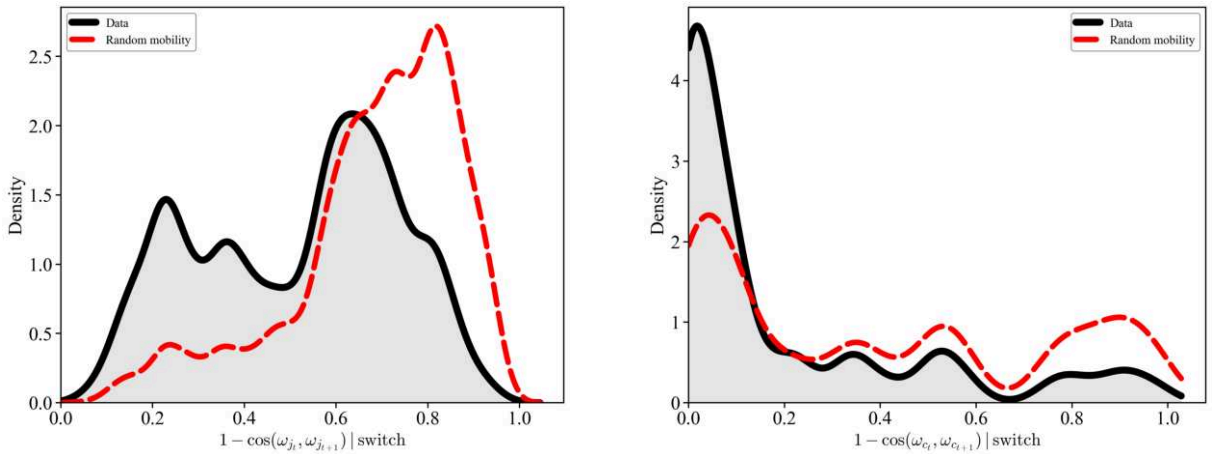
Figure 23: Change in task content for occupation switchers who stay in the same firm (top row) and who change firm (bottom row). Vertical axis: change in the task share (left, Cognitive–Manual classification) or skill share (right, ESCO skill dimensions) between destination and origin occupation; horizontal axis: residualized wage percentile within the origin occupation. Each column shares the same vertical scale across the two rows.

and drop a worker history if she has been in sectors with fewer than 100 workers reported or in occupations with fewer than 100 workers reported in a given year. As I am also interested in occupational mobility within firms, I drop workers in firms with fewer 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 2003 in place of the CAE Rev 2 used from 1995[30]. As for occupations, the National Classification of Occupations 1994 (CNP/94) was used from 1994 until 2009 and was then replaced by the Portuguese Classification of Occupations 2010 (CPP/2010) that is embedded in the ISCO-08 international classification. I develop an algorithm to minimize the num-

was introduced in Portugal, starting from 2007. Occupations dropped belong to the 2-digit groups 11, 61 and 92 for the CNP 1994 and 11, 61, 62, 63 and 92 for the CPP 2010.

[30]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 the 2-digit 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.

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



(a) ESCO skill distance (296 dimensions)

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

Figure 24: 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.

A.1 Wage Changes upon Occupational Switches

This appendix documents what happens to the wages of workers who switch occupation in the large sample, comparing them with stayers who start from the same position in the same occupation. Let w_{it} denote the log real hourly wage of worker i in year t. As in the rest of the empirical analysis, a switcher is a worker observed in occupation j_t in year t and in a different three-digit occupation j_{t+1} in year $t+1$; stayers are workers observed in the same occupation in t and $t+1$. The initial position of a worker is the percentile of the residual wage $\widehat{r}_{i,t-1}$ of equation (1), as in Facts 1 and 2, measured in the year before the switch and ranked within the origin occupation and year. Ranking on year $t-1$ keeps w_{it}, which enters the wage change, out of the ranking, so that a transitory wage shock in t does not generate mean reversion mechanically. The sample is therefore restricted to workers observed in the origin occupation in both $t-1$ and t.

Wage changes by initial position within occupation. Panel (a) of Figure 4 reports, for each horizon $h \in \{0,\dots,4\}$, the coefficients β_{hq} of the regression

$$w_{i,t+1+h} - w_{it} = \sum_{q=1}^{10} \beta_{hq}\, S_{it}\, \mathbf{1}\{q_{it}=q\} + v^{h}_{c(i,t)} + \xi_h' X_{it} + e_{ith}, \tag{45}$$

where S_{it} indicates a switch between t and $t+1$, q_{it} is the decile of the initial position, v_c^h are fixed effects for origin occupation $\times$ year $\times$ decile cells, and X_{it} includes education fixed effects, experience and experience^2. Standard errors are clustered by worker. Horizon $h = 0$ is the first year in the new occupation. The coefficient β_{hq} is the cumulative wage change of switchers relative to stayers who start from the same decile of the same occupation in the same year, so that it nets out the mean reversion in wages that is common to the two groups. The wage change upon switching declines monotonically with the initial position. Relative to stayers, switchers in the bottom quintile gain 6 log points on impact and 6 after four years, switchers around the median gain 2 and 3, and switchers in the top decile lose 3 on impact and close the gap after four years.

Wage changes by initial position and task content of the switch. Panel (b) of Figure 4 groups switches by the change in the cognitive task share of the occupation, $\Delta\omega_C = \omega_{C,j_{t+1}} - \omega_{C,j_t}$, measured with the Cognitive–Manual classification of Mihaylov and Tijdens (2019): ↑Cognitive if $\Delta\omega_C > 0.10$ and ↑Manual if $\Delta\omega_C < -0.10$. Switches with a smaller change in the cognitive task share are included in the regression as a third group and are not reported in the figure. Panel (b) reports the coefficients of equation (45) on impact ($h = 0$) and after four years ($h = 4$) when the switch indicator is interacted with both the decile of the initial position and the three groups, so that each group of switchers is compared with the stayers in the same origin occupation, year, and decile. Workers in the bottom quintile gain relative to stayers whatever the direction of the move. On impact they gain 7 log points when they move to more cognitive occupations and 5 when they move to more manual occupations; after four years the gains are 6 and 4. The return to moving toward more manual occupations falls with the initial position: around the median it is 0 log points on impact and 1 after four years, and in the top decile it is -6 on impact and -3 after four years. In the top decile, workers moving to more cognitive occupations lose 2 log points on impact and gain 2 after four years. In both groups the wage change is more dispersed across initial positions on impact than after four years. Workers at the bottom of the wage distribution of their occupation are also those who move more often toward manual occupations (34% of the switches in the bottom decile against 15% in the top decile).

Event studies. Figure 25 reports the dynamics around the switch. Let $\Delta w_{it} \equiv w_{it} - w_{it-1}$ and let $q \in \{\text{bottom quintile}, \text{middle}, \text{top quintile}\}$ index the initial position; in this paragraph q_{it} denotes the position group of worker i in year t rather than the decile. For each relative time $\tau \in \{-3, \ldots, 4\}$ I run a separate OLS regression

$$\Delta w_{i,t+1+\tau} = \sum_q \beta_{\tau q}\, S_{it}\, \mathbf{1}\{q_{it} = q\} + v^{\tau}_{c(i,t)} + \xi'_\tau\, X_{it} + e_{it\tau}, \tag{46}$$

where v_c^τ are fixed effects for origin occupation $\times$ year $\times$ position group cells. The coefficients $\beta_{\tau q}$ measure the difference in annual wage growth between switchers and stayers with the same initial position, and $\tau = 0$ is the wage change spanning the

switch, $w_{i,t+1} - w_{it}$. To present wage levels around the event I cumulate the estimated growth effects and normalize at $\tau = -1$, similar to Heise and Porzio (2022):

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

with 95% confidence bands computed from the joint worker-clustered covariance of the estimates across τ. In Panel (a), switchers in the bottom quintile gain 6 log points relative to stayers at the switch, and the gain is persistent: it is 6 after four years. Switchers in the middle of the distribution gain 2 on impact and 4 after four years, while switchers in the top quintile lose 2 on impact and recover the loss within four years.

Before the switch, wage growth of switchers and stayers with the same initial position is similar: the cumulative difference is below 1 log point in absolute value in all groups. Ranking workers on the wage in t instead selects switchers whose last wage in the origin occupation contains a transitory component. Under that ranking, the wage of switchers in the bottom quintile falls by 2 log points relative to stayers in the three years before the switch, and the wage of switchers in the top quintile rises by 3. Under that ranking, the gain of switchers in the bottom quintile would be 7 log points on impact and 6 after four years, and the loss of switchers in the top quintile would be 3 on impact and 1 after four years.

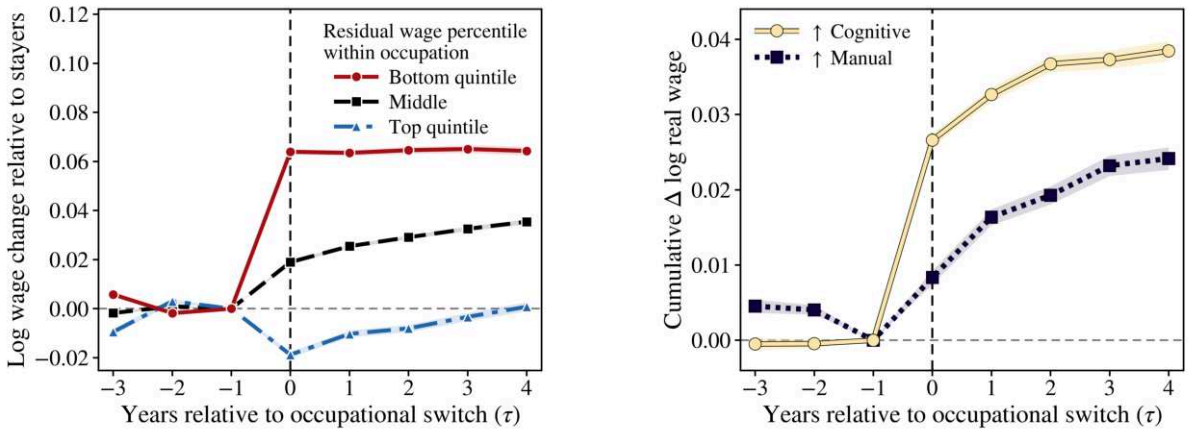
Panel (b) groups switchers by the task content of the switch, with the groups defined above. The specification is the one in equation (46), with the switch indicator interacted with the groups, all stayers as the omitted category, and the dependent variable and the controls demeaned by calendar year in place of the cell fixed effects. Workers moving to more cognitive occupations gain 3 log points on impact and 4 after four years, while workers moving toward more manual occupations gain 1 on impact and 2 after four years.

A.2 Occupation Clustering

Let $o = 1, \ldots, n$ index three-digit ISCO occupations and let $c \in \{0, \ldots, J-1\}^n$ denote a partition into J clusters. Each occupation has a standardized two-dimensional vector defining the task content $x_o \in \mathbb{R}^2$ summarizing its cognitive and manual content, a mass weight s_o equal to its share of person-periods in the panel, and pairwise mobility flows $F_{oo'}$ between different occupations measured in the large sample. The partition minimizes

$$L(c) \;=\; \underbrace{\sum_{o} s_o \, \| x_o - \bar{x}_{c_o} \|^2}_{\text{task-content dispersion}} \;+\; \lambda_F \underbrace{\sum_{o,o'} F_{oo'} \, \mathbf{1}\{c_o = c_{o'}\}}_{\text{absorbed flow}},$$

where $\bar{x}_c$ is the mass-weighted centroid of cluster c. The first term ensures that occupations with similar task content are clustered together; the second term penalizes observed worker flows that are absorbed into a cluster, so that aggregated cluster-



(a) Event study of wages around the switch, by initial position in the origin occupation

(b) Event study of wages around the switch, by task content of the switch

Figure 25: Event studies of wages around occupational switches. Horizontal axis: event time relative to the switch, where $\tau = -1$ is the last year in the origin occupation (year t) and $\tau = 0$ is the first year in the new occupation (year $t+1$); vertical axis: cumulative change in the log real wage of switchers relative to stayers, normalized at $\tau = -1$. *Panel (a)*: workers in the bottom quintile (red, solid), the middle (black, dashed), and the top quintile (blue, dash-dot) of the residual wage distribution of the origin occupation in year $t-1$, relative to stayers with the same initial position. *Panel (b)*: switches toward more cognitive occupations (light yellow, solid) and more manual occupations (dark blue, dotted), relative to all stayers. Shaded areas are 95% confidence bands clustered by worker.

to-cluster transitions remain informative with respect to the original occupation-to-occupation mobility graph. In choosing the relative importance of the two measures for this loss function, I choose $\lambda_F = 10$ considering the trade-off between within-cluster dispersion and cross-cluster mobility share. Importantly, there are two feasibility constraints that I impose while minimizing the above loss function: the cluster-level mobility graph must remain connected, and every non-seed cluster must retain at least 5% of total mass. To make sure that two clusters have a clear cognitive and manual interpretation, I fix two seed sets of occupations, each of which must be contained in a single cluster. These occupations are chosen based on their extreme task content, and are either fully manual or fully cognitive. The procedure is run from several starting points (mass-weighted k-means initializations) and the best partition is retained. Results with respect to the two metrics used in the loss function are plotted in Figure 26. As displayed by the vertical dashed line, $J = 8$ is selected balancing parsimony in the number of clusters considered, within-cluster skill/task content dispersion and the cross-cluster mobility share. For larger J the measure improves but the number of observations included in each cluster falls below the 5% threshold, with potential implications for the identification of some of the cluster-specific components of the model. The clustering is implemented using the two different skill requirement and task content classifications already described in the main body of the paper: an ESCO variant in which each granular ESCO skill is labeled as either cognitive or manual, and (C_o, M_o) is occupation o's share of cognitive vs. manual skills aggregated from those labels using its ESCO skill mix, and a task-based variant in which $(C_o^{\text{MT}}, M_o^{\text{MT}})$ is computed from the routine/non-routine task shares of Mihaylov and Tijdens (2019).

Cluster labels are deterministic: the cluster containing the cognitive seed set is Cluster 0, the cluster containing the manual seed set is Cluster $J-1$, and the remaining clusters, Clusters 1 to $J-2$, are ordered by ascending mean cognitive share. The seed sets only serve the clustering; the anchor occupations used to identify the wage equation are chosen separately as the clusters with the highest cognitive and manual task shares, Clusters 0 and 1. The manual anchor is therefore the non-seed cluster with the lowest cognitive share, not the cluster containing the manual seed set.

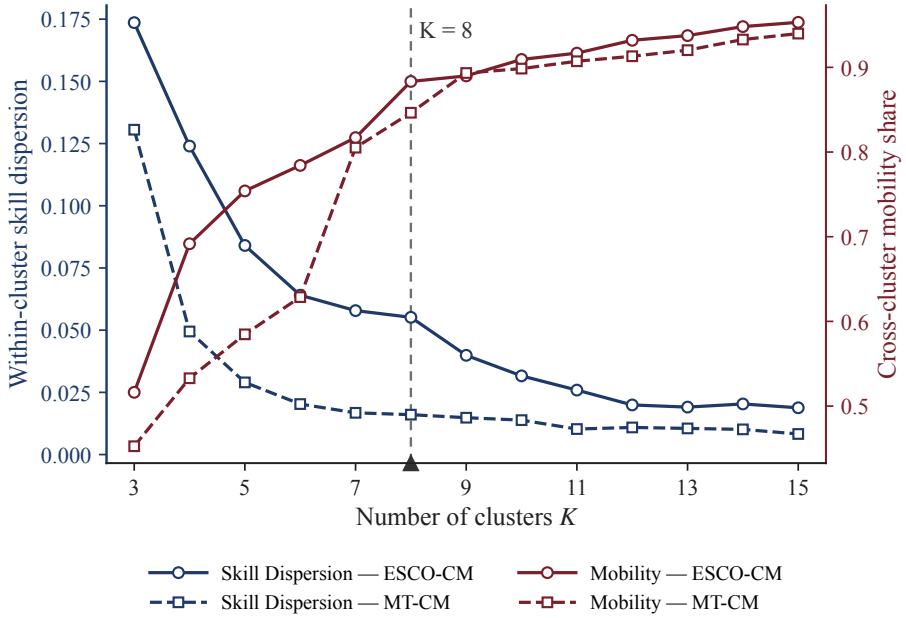


Figure 26: Headline diagnostic for the J-selection step of the anchored skill+mobility-penalty clustering. The horizontal axis is the number of clusters J on the 3–15 grid (with $\lambda_F = 10$ fixed at the empirical knee). Curves report the mass-weighted within-cluster skill dispersion and the cross-cluster mobility share at the L-minimizing partition, separately for the ESCO-based cognitive/manual shares (C_o, M_o) and the task-based cognitive/manual shares $(C_o^{\text{MT}}, M_o^{\text{MT}})$.

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, 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.

B Appendix: Model

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

B.1 Firm's Time Allocation Problem

This subsection 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 $(\tilde{\boldsymbol{\theta}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ shared with the worker, where the expectation is taken under the learning block's posterior $\boldsymbol{\theta}_i \mid \mathcal{I}_{it} \sim \mathcal{N}(\tilde{\boldsymbol{\theta}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$.

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

$$y_{jk} = \ell_{jk} \exp\big(a_{jk}\,\theta_{ik}\big)\, H_{jk}(\tau_{ikt}), \qquad H_{jk}(\tau) \;\equiv\; \exp\big(h_{jk}(\tau)\big), \tag{47}$$

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

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

Taking logs,

$$\log Y_{ijt} = \sum_{k=1}^{K} \omega_{jk} \log \ell_{jk} + \sum_{k=1}^{K} \omega_{jk}\, a_{jk}\, \theta_{ik} + \sum_{k=1}^{K} \omega_{jk}\, h_{jk}(\tau_{ikt}) - \sum_{k=1}^{K} \omega_{jk} \log \omega_{jk}. \tag{49}$$

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

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

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

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

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

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

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

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

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

B.2 Stationary Equilibrium Algorithm

A stationary equilibrium is a constant vector of occupation log prices $\tilde{\boldsymbol{p}} = (\log P_j)_{j\in\mathcal{J}}$ at which every occupation's goods market clears. The CES aggregator weights $\{\alpha_j\}$ are held fixed at their pre-shock calibration and any candidate price vector is renormalized so that the aggregate price index equals one, $P(\tilde{\boldsymbol{p}}) = \left(\sum_j \alpha_j^\sigma P_j^{1-\sigma}\right)^{1/(1-\sigma)} = 1$ (Section 3.6). I compute the stationary equilibrium of the model by iterating on the time-independent price vector: at each iteration, I solve the worker's dynamic problem given the current prices, simulate a synthetic panel, and compute the update for the price vector using the labor demand equation. When simulating, workers exit at rate χ[31] and are replaced by entrants drawn from the group priors. After discarding an initial burn-in, this death-and-rebirth process makes initial conditions irrelevant and the simulated panel approximates the ergodic cross-section $\pi^\star(\tilde{\boldsymbol{p}}^\star)$ resulting from the equilibrium price vector $\tilde{\boldsymbol{p}}^\star$.

In the simulation, occupation-specific output is aggregated using choice probabilities as weights. To reduce the noise induced by numerical simulations, I do not assign each worker to a unique occupation as this would generate discrete jumps for small moves in the price vector. Rather, I let her contribute to every occupation in proportion to her conditional choice probability ($\mathbb{P}_{ij}(\tilde{\boldsymbol{p}})$), which is given by

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

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

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

weighted sum of these wages deflated by the price in that occupation,

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

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

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

Algorithm: Stationary Equilibrium (price tâtonnement)

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

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

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

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

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

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

B.3 Transition Path Algorithm

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

tionary algorithm (solve, simulate, aggregate, update), but now applied to the whole price path by a backward–forward shooting algorithm.

The technology shock considered in the experiment changes 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, $\lambda_{jk}^{\mathrm{AI}} = \lambda_{jk}\,\omega_{jk}^{\mathrm{AI}}/\omega_{jk}$, and the log human-capital bundle h_j inherits the new task weights $\boldsymbol{\omega}_j^{\mathrm{AI}}$, while noise variances, amenities, and switching costs are unchanged. Let Λ collect the parameters of this technology. I compute an "instantaneous" technology adoption, where the production function technology mapping skills and experience to output jumps from Λ_{pre} to Λ_{post} at $t = 0$.

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

$$\boldsymbol{\Phi}_t = \mathcal{S}^{-1}\Big[\varrho \log \sum_{j\in\mathcal{J}} \exp\Big(\big\{r_j(\boldsymbol{x};\tilde{\boldsymbol{p}}_t,\Lambda_t) + \beta\,\mathbb{E}\big[EV_{t+1}(\boldsymbol{x}') \mid \boldsymbol{x}, j\big]\big\}/\varrho\Big)\Big], \qquad t = T-2,\ldots,0, \tag{56}$$

where $\mathcal{S}^{-1}$ is the Chebyshev/Smolyak inversion[32], $r_j(\boldsymbol{x};\tilde{\boldsymbol{p}}_t,\Lambda_t)$ is the flow payoff of occupation j in state $\boldsymbol{x}$ (the expected log wage at the period-t prices and technology, plus the amenity η_{jg}, minus the switching cost $\mathcal{C}(j, d_{-1})$), so that the flow value uses the period-t prices and technology, and the expectation integrates the Gaussian belief update by Gauss–Hermite quadrature. I then forward simulate a single panel where workers face the time-varying value functions (so the value function is now not stationary), prices, and technology, and occupation outputs Y_{jt} are formed by exactly the price-dependent choice-probability aggregation described before. This is now applied period by period as I am solving for a sequence of competitive equilibria. Finally, I use the per-period CES inverse demand to get the updating price sequence, which I renormalize to the numeraire in each period. The path is iterated until the largest price change across iterations and occupations falls below a certain tolerance level. The procedure is summarized in the algorithm box below.

[32]The model is solved via collocation, hence I obtain the value function on the collocation nodes, and the Chebyshev polynomial coefficients are obtained by inverting the interpolation (basis) matrix at the collocation nodes.

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 χ, taste-shock scale ϱ, and the estimated parameters $(\widehat{\mathbf{\Gamma}}, \widehat{\mathbf{\Delta}})$; horizon T, damping ζ, tolerance *tol*.

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

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

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

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

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

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

C Appendix: Proofs

C.1 Identification of the Wage Equation Parameters under Learning

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

The wage equation is

$$w_{it}(j) = p_{jt} + h_j(\boldsymbol{\tau}_{it}) + \boldsymbol{\lambda}_j' \boldsymbol{\theta}_i + \varepsilon_{ijt}, \tag{57}$$

where p_{jt} is the occupation-time log price, $h_j(\boldsymbol{\tau}_{it})$ is the human-capital component, $\boldsymbol{\lambda}_j$ is the occupation loading on the latent skill vector, $\boldsymbol{\theta}_i$ 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, $\boldsymbol{\theta}_i \mid g(i) = g \sim N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ 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 $\boldsymbol{\theta}_i$, and realized after the occupational choice is made.*

Assumption 2 (Pure-learning selection) *Let*

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

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

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

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

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

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

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

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

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

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:

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

Stack the signals observed before period t into

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

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

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

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

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

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

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

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

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

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 $\boldsymbol{\theta}_i$ or about future wage shocks (Assumption 2). Let $q_t(j \mid \mathcal{H}_{it}) \equiv \Pr(d_{it} = j \mid \mathcal{H}_{it})$ be the conditional choice probability, which is identified from the joint distribution of histories and choices. To see how it works as a weight, write the joint density sequentially, using the timing of the model: the worker enters period t with history $\mathcal{H}_{it}$, chooses her occupation, and only then observes her wage. Hence, I obtain

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

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

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

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

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

Everything on the right is in the data, so the distribution of potential log wages along every supported path is identified. This weighting scheme reweights each worker by the inverse of how likely her path was; without it, workers who believe that they have high cognitive skills, and who are therefore more likely to self-select into occupations with a high cognitive task content, would be over-represented in those occupations. Every moment I use below is a moment of this reweighted distribution.

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

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

is jointly Gaussian, because it is a linear map of the Gaussian skill vector $\boldsymbol{\theta}_i$ and the Gaussian wage shocks. To characterize the path of potential wages, the mean and

covariance are therefore all I need, and repeated measurement ensures I can identify them: for a worker observed multiple times, the observed wages share her permanent skill $\boldsymbol{\theta}_i$ while shocks are independent over time. What is left after reweighting is thus the same latent skill vector, with different loadings if the worker switches occupation, and independent noise. I write the three moments below, after netting out the returns to task experience.

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

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

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

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

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

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

Then

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

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. Each of them looks at the workers observed in a given occupation at a given date, so it is enough to reweight each worker by the probability of the choice she made at that date, $q_t(j \mid \mathcal{H}_{it})$. The logic is the same as above: conditional on her history, the choice of a worker reveals nothing further about her skills or her future wage shocks (Assumption 2), so the workers who choose j at t, weighted by the inverse of the probability of doing so, are representative of all the workers with the same history (Assumption 3 ensures that this probability is positive). For these workers, the observed adjusted wage Z_{it} is the potential one,

$Z_{it}(j)$.

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

$$m_{jgt} \equiv \mathbb{E}\left[\frac{\mathbf{1}\{d_{it}=j\}\, Z_{it}}{q_t(j \mid \mathcal{H}_{it})} \;\middle|\; g(i)=g\right] = \mathbb{E}[Z_{it}(j) \mid g(i)=g] = p_{jt} + \boldsymbol{\lambda}_j'\boldsymbol{\mu}_g.$$

Only the workers observed in j at t enter, each weighted by the inverse of the probability that, given her history, she chose j at t; after reweighting, their average adjusted wage is the average across all the workers of group g.

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, each date reweighted by its own choice probability,

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

which contains only permanent skill, since the transitory shocks are independent across periods. The two dates are reweighted one after the other: the wage at the earlier date s is already part of the history at t, so reweighting by the choice at t leaves it unchanged, and reweighting by the choice at s then removes the selection into ℓ. The third is the same-period variance in occupation j,

$$V_{jg} \equiv \mathbb{E}\left[\frac{\mathbf{1}\{d_{it}=j\}\,(Z_{it}-m_{jgt})^2}{q_t(j \mid \mathcal{H}_{it})} \;\middle|\; g(i)=g\right] = \boldsymbol{\lambda}_j'\boldsymbol{\Sigma}_g\boldsymbol{\lambda}_j + \varsigma_{\epsilon j}^2,$$

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

Identification of the skill distribution: the means μ_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

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

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

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

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

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

I obtain

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

Then

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

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

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

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

Using the anchor normalization,

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

Therefore,

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

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

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

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

From the covariance formula,

$$C'_{jA,g} = \lambda'_j \Sigma_g.$$

Since Σ_g has already been identified,

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

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 λ_j must be recovered from each of them—a further overidentifying restriction.

Identification of the occupation prices p_{jt}. Once λ_j and μ_g are known, the mean adjusted wage equation

$$m_{jgt} = p_{jt} + \lambda'_j \mu_g$$

identifies the occupation-time price:

$$p_{jt} = m_{jgt} - \lambda'_j \mu_g.$$

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

$$V_{jg} = \lambda'_j \Sigma_g \lambda_j + \varsigma^2_{\varepsilon j}.$$

The first term is now known, since λ_j and Σ_g have been identified. Subtracting it,

$$\varsigma^2_{\varepsilon j} = V_{jg} - \lambda'_j \Sigma_g \lambda_j.$$

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

$$\left\{ (p_{jt}, \lambda_j, \delta_j, \varsigma^2_{\varepsilon j})_{j \in \mathcal{J}, t},\ (\mu_g, \Sigma_g)_{g \in \mathcal{G}} \right\}.$$

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 man-

ual 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.

C.2 Identification of the Dynamic Choice Parameters

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

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

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

$$(Q_j f)(\boldsymbol{x}) \equiv \int f(\boldsymbol{x}')\, Q(d\boldsymbol{x}' \mid \boldsymbol{x}, j).$$

Then, I denote by Q_j^s the same operator when the assignment to j is repeated for s consecutive periods. Therefore $Q_j^s f(\boldsymbol{x})$ propagates the state forward for s periods under repeated choice of occupation j. Notice that this choice is defined exogenously, that is, I am forcing $d_{im} = j$ for any m from period t to period $t+s-1$. The conditional value function can be written using this operator as

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

The conditional choice probabilities are

$$\mathbb{P}_j(\boldsymbol{x}) = \frac{\exp(\nu_j(\boldsymbol{x})/\varrho)}{\sum_{m\in\mathcal{J}}\exp(\nu_m(\boldsymbol{x})/\varrho)}.$$

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

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

Define the choice-specific value correction

$$\psi_j(\boldsymbol{x}) \equiv EV(\boldsymbol{x}) - \nu_j(\boldsymbol{x}).$$

This term measures the value lost if the worker is required to commit to j in the current period, before observing the taste shocks, rather than being free to choose optimally, while future choices remain optimal. This is the probability correction term of Section 4.2. Below, I use the Hotz and Miller (1993) inversion argument to show that under the multinomial logit assumption I can recover these terms directly from the choice probabilities, that is $\psi_j(\boldsymbol{x}) = -\varrho \log \mathbb{P}_j(\boldsymbol{x})$. The correction $\psi_j = -\varrho \log \mathbb{P}_j$ can be understood as a penalty for using j as the reference action. From the CCP representation this term has an intuitive revealed-preference interpretation: it is small when j is chosen often (close to optimal) and large when j is rarely chosen.

Assumption 6 (Stationarity and known primitives) *The data are generated by the stationary equilibrium of Section 3.6. The wage equation and wage-signal distribution are identified, so $\tilde{w}_j$ and Q_j are known for every occupation j. The discount factor $\beta \in (0,1)$ and taste-shock scale $\varrho > 0$ are fixed. The conditions of Lemma 2 hold, so the Bellman equation has a unique integrated value function EV.*

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

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

I write $\boldsymbol{x}_{dg} = (\cdot, d, g)$ for the state of a worker whose last-period occupation was d and who belongs to group g, with $\cdot$ collecting the remaining state variables (mean and variance-covariance of the belief distribution, and task-specific tenure).

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

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

I now show how using a standard inversion argument à la Hotz and Miller (1993) allows me to recover continuation values.

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

$$\psi_j(\boldsymbol{x}) \equiv EV(\boldsymbol{x}) - \nu_j(\boldsymbol{x}) = -\varrho \log \mathbb{P}_j(\boldsymbol{x}).$$

Moreover,

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

where

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

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

Proof. Taking logs of the logit CCP gives

$$\varrho \log \mathbb{P}_j(\boldsymbol{x}) = \nu_j(\boldsymbol{x}) - \varrho \log \sum_{m \in \mathcal{J}} \exp\left(\frac{\nu_m(\boldsymbol{x})}{\varrho} \right) = \nu_j(\boldsymbol{x}) - EV(\boldsymbol{x}),$$

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

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

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

times yields

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

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

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

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

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

□

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

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

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

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

is identified.

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

$$\varrho \log \frac{\mathbb{P}_j(\boldsymbol{x})}{\mathbb{P}_m(\boldsymbol{x})} - \left[\left(\tilde{w}_j(\boldsymbol{x}) + \beta \Psi_j(\boldsymbol{x}) \right) - \left(\tilde{w}_m(\boldsymbol{x}) + \beta \Psi_m(\boldsymbol{x}) \right) \right] = \frac{\eta_{jg} - \eta_{mg}}{1-\beta} - \kappa_0 \left(\mathbf{1}_{\{j \neq d\}} - \mathbf{1}_{\{m \neq d\}} \right). \tag{58}$$

Amenities. Fix g and $j \neq 0$, and evaluate equation (58) at a supported state $\boldsymbol{x}_{dg}$ with $d \notin \{j, 0\}$, as guaranteed by Assumption 8. Because both j and 0 are then moves, their switching costs cancel. Comparing occupation j with the baseline normalized

occupation where $\eta_{0g} = 0$ yields

$$\eta_{jg} = (1-\beta)\left\{\varrho \log \frac{\mathbb{P}_j(\boldsymbol{x}_{dg})}{\mathbb{P}_0(\boldsymbol{x}_{dg})} - \left[\left(\tilde{w}_j(\boldsymbol{x}_{dg}) + \beta\Psi_j(\boldsymbol{x}_{dg})\right) - \left(\tilde{w}_0(\boldsymbol{x}_{dg}) + \beta\Psi_0(\boldsymbol{x}_{dg})\right)\right]\right\}.$$

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

Switching cost. At a supported state $\boldsymbol{x}_{dg}$ whose occupation d admits a destination $j \neq d$, compare moving to j with staying in d. Equation (58) gives

$$\kappa_0 = \frac{\eta_{jg} - \eta_{dg}}{1-\beta} - \varrho \log \frac{\mathbb{P}_j(\boldsymbol{x}_{dg})}{\mathbb{P}_d(\boldsymbol{x}_{dg})} + \left[\left(\tilde{w}_j(\boldsymbol{x}_{dg}) + \beta\Psi_j(\boldsymbol{x}_{dg})\right) - \left(\tilde{w}_d(\boldsymbol{x}_{dg}) + \beta\Psi_d(\boldsymbol{x}_{dg})\right)\right].$$

All terms on the right-hand side are now known, so κ_0 is identified. Intuitively, wages, amenities and continuation values imply a certain amount of switching; if workers move less than that, the gap is attributed to the switching cost. □

Remark. A just-identifying subset contains

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

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

The simple comparison argument relies on the switching cost being common across destinations and paid only when the worker moves. If switching costs varied with destination, direction, or experience, they would no longer cancel in the first step and would have to be identified jointly with amenities.

C.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.

Throughout, $\boldsymbol{x}_t$ is the state defined in equation (23), dropping the worker index. The price path is part of the state, so the future evolution of prices is known from the worker's point of view.

Let $f(\boldsymbol{x}' \mid \boldsymbol{x}, j)$ denote the density of next period's state when the worker chooses occupation j today; it includes the update of beliefs after the wage signal. The deter-

ministic value of choosing occupation j is

$$v_{jt}(\boldsymbol{x}_t; EV) = \mathbb{E}\left[\log W_{jt} \mid \boldsymbol{x}_t\right] + \eta_{jg} - \mathcal{C}(j, d_{t-1}) + \beta \int EV(\boldsymbol{x}')\, f(\boldsymbol{x}' \mid \boldsymbol{x}_t, j)\, \mathrm{d}\boldsymbol{x}',$$

where the current taste shocks have been integrated out, and the Bellman operator is

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

I normalize the taste shocks to have mean zero. Under the other common normalization, the value function would only shift by the constant $\varrho\gamma_E/(1-\beta)$, where γ_E is Euler's constant, and the CCPs would not change.

The difficulty in this setting is that beliefs are not bounded. A worker's belief about her skills, $\tilde{\boldsymbol{\theta}}$, can in principle take any value, and so can her expected log wage, which is linear in $\tilde{\boldsymbol{\theta}}$, and hence her value function. The standard contraction argument, which works with bounded value functions, therefore does not apply directly. I instead use the weighted contraction mapping theorem of Boyd (1990), and show that $\mathcal{T}$ is a contraction once value functions are compared relative to the size of beliefs. This makes existence, uniqueness and convergence of value iteration hold from any starting value. The next lemma states the result.

Lemma 2 (Bellman contraction in a weighted norm) *Suppose $\beta \in (0,1)$, $\varrho > 0$, $\varsigma_{\varepsilon j}^2 > 0$ for every j, and prices, amenities, switching costs and returns to experience are bounded. Then the Bellman equation has a unique solution EV among value functions that are continuous in beliefs and grow at most linearly in beliefs, and iterating the Bellman operator from any such starting value converges to it at rate β.*

Proof.

The value function cannot grow faster than beliefs do, so I define a weight function w, which measures the size of beliefs, and measure the distance between two value functions relative to it:

$$w(\boldsymbol{x}) = \left(1 + \|\tilde{\boldsymbol{\theta}}\|^2 + \operatorname{tr}\tilde{\boldsymbol{\Sigma}}\right)^{1/2}, \qquad \|EV - \widetilde{EV}\|_w = \sup_{\boldsymbol{x}} \frac{|EV(\boldsymbol{x}) - \widetilde{EV}(\boldsymbol{x})|}{w(\boldsymbol{x})}.$$

Each term of w has a simple interpretation. The constant 1 accounts for the bounded parts of the payoff, such as prices, amenities and switching costs, and keeps w away from zero. The term $\|\tilde{\boldsymbol{\theta}}\|^2$ measures how large the worker's beliefs are, which is what makes her expected log wage large. The term $\operatorname{tr}\tilde{\boldsymbol{\Sigma}}$ measures how much she still does not know about her skills. Together, the last two terms are the expected squared size of her true skill, $\mathbb{E}[\|\boldsymbol{\theta}\|^2 \mid \boldsymbol{x}] = \|\tilde{\boldsymbol{\theta}}\|^2 + \operatorname{tr}\tilde{\boldsymbol{\Sigma}}$: what she believes, plus what she is still uncertain about. A value function has finite w-norm exactly when it grows at most linearly in beliefs. Throughout, I consider value functions that are continuous in beliefs and grow at most linearly in beliefs. The Bellman operator preserves continuity: the expected continuation value changes smoothly with beliefs, because beliefs are updated

by a normally distributed wage signal, and the log-sum of finitely many continuous choice-specific values is itself continuous.

In economic terms, Boyd's theorem states that $\mathcal{T}$ is a contraction of modulus β on this set of value functions if three conditions hold:

1. *Monotonicity:* a higher continuation value can never make the worker worse off today, that is, $EV \leq \widetilde{EV}$ implies $\mathcal{T}EV \leq \mathcal{T}\widetilde{EV}$.

2. *Linear growth of the myopic value:* the value of a worker who cares only about today's payoff, $\mathcal{T}0$, grows at most linearly in beliefs.

3. *Discounting:* raising the continuation value by $c\,w(\boldsymbol{x}')$ raises today's value by at most $\beta c\,w(\boldsymbol{x})$, that is, $\mathcal{T}(EV + c\,w) \leq \mathcal{T}EV + \beta c\,w$ for every constant $c \geq 0$.

I now verify the three conditions in turn.

Condition 1. Only the continuation term of $v_{jt}(\boldsymbol{x}_t; EV)$ depends on EV. If $EV \leq \widetilde{EV}$ at every state, then the expected continuation value is weakly higher under $\widetilde{EV}$, so $v_{jt}(\boldsymbol{x}_t; EV) \leq v_{jt}(\boldsymbol{x}_t; \widetilde{EV})$ for every occupation j. The log-sum is increasing in each choice-specific value, since its derivatives are the choice probabilities, which are positive, and therefore

$$(\mathcal{T}EV)(\boldsymbol{x}_t) = \varrho \log\Big[\sum_{j\in\mathcal{J}} \exp\Big(\frac{v_{jt}(\boldsymbol{x}_t; EV)}{\varrho}\Big)\Big] \leq \varrho \log\Big[\sum_{j\in\mathcal{J}} \exp\Big(\frac{v_{jt}(\boldsymbol{x}_t; \widetilde{EV})}{\varrho}\Big)\Big] = (\mathcal{T}\widetilde{EV})(\boldsymbol{x}_t).$$

Condition 2. With no continuation value, $\mathcal{T}0$ is the log-sum of the flow payoffs. It lies between the highest flow payoff and the highest flow payoff plus $\varrho \log|\mathcal{J}|$, the option value of the taste shocks with $|\mathcal{J}|$ choices. The only unbounded term of the flow payoff is the expected skill component $\boldsymbol{\lambda}_j'\tilde{\boldsymbol{\theta}}$, which is at most a constant times $\|\tilde{\boldsymbol{\theta}}\| \leq w(\boldsymbol{x})$ in absolute value; prices, amenities, switching costs and returns to experience are bounded by assumption. Hence $\mathcal{T}0$ is bounded by a constant times w and grows at most linearly in beliefs. This is where the constant 1 in w is used.

Condition 3. The key property of w is that it does not grow on average from one period to the next. A wage signal moves the worker's beliefs up or down, but it does not push them in any particular direction: on average, her belief next period equals her belief today. Learning does, however, shift weight between the two terms of w. As the worker learns, her beliefs move away from the prior mean, so $\|\tilde{\boldsymbol{\theta}}\|^2$ rises on average, while her uncertainty $\operatorname{tr}\tilde{\boldsymbol{\Sigma}}$ falls by the same amount: what she learns moves from her uncertainty into her beliefs. Because priors are rational, the sum is her own forecast of the size of her true skill, $w(\boldsymbol{x})^2 = \mathbb{E}[1 + \|\boldsymbol{\theta}\|^2 \mid \boldsymbol{x}]$, and since her true skill does not change, she cannot expect this forecast to change either. Next period's w^2 therefore equals this period's on average and, since the square root is concave,

$$\int w(\boldsymbol{x}')\, f(\boldsymbol{x}' \mid \boldsymbol{x}, j)\, \mathrm{d}\boldsymbol{x}' \leq w(\boldsymbol{x}).$$

This is why the uncertainty term is needed: with beliefs alone, w would grow on average as the worker learns. Now raise the continuation value from EV to $EV + c\,w$. Each choice-specific value rises by $\beta c \int w(\boldsymbol{x}')\, f(\boldsymbol{x}' \mid \boldsymbol{x}, j)\, \mathrm{d}\boldsymbol{x}' \leq \beta c\, w(\boldsymbol{x})$. The log-sum of values that each rise by at most δ rises by at most δ: its derivatives are the choice probabilities, which are nonnegative and sum to one, so a common increase in all choice values leaves the choice probabilities unchanged and passes through one-for-one, and an uneven increase passes through by less than its largest component. Hence $\mathcal{T}(EV + c\,w) \leq \mathcal{T}EV + \beta c\, w$. The worker discounts the future and w does not grow on average, so a uniform increase in the continuation value is worth less today than it is tomorrow.

All three conditions hold, so $\mathcal{T}$ is a contraction of modulus β, and the contraction mapping theorem gives existence, uniqueness and convergence of the iterations from any starting value that is continuous in beliefs and grows at most linearly in beliefs. This fixed point is the integrated value function used in the occupational choice problem. □

In computation, the state space is a bounded grid, the expectation is replaced by Gauss–Hermite weights, which are positive and sum to one, and points outside the grid are projected back onto it. Beliefs are then bounded, the same argument applies in the ordinary sup norm, and the computed Bellman operator is also a contraction. The Smolyak coefficient iteration additionally interpolates between grid points, so this result does not cover it; its convergence is checked numerically (Appendix D.5.1).

D Appendix: Estimation and Numerical Methods

D.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 $\chi = 0.08$ (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 3.6). Education shares are set to match the large sample: low-educated workers comprise 67.9%, medium-educated 20.1%, and high-educated 12.0%. 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 as $\alpha_j = P_j\, Y_j^{1/\sigma} / (\sum_m P_m Y_m)^{1/\sigma}$ (equivalently $\alpha_j^\sigma = s_j P_j^{\sigma-1}$ with wage-bill share s_j), where $P_j = \exp(p_j)$ is the occupation price in levels and 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 3: Externally Set and Calibrated Parameters

Parameter	Description	Empirical Moment	Value
Preferences & Demographics			
β	Discount factor		0.900
χ	Per-period exit hazard	Avg. career length (est. sample)	0.080
Education Distribution			
Low education (L)	Population share	Empirical shares	67.9%
Medium education (M)	Population share	Empirical shares	20.1%
High education (H)	Population share	Empirical shares	12.0%
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: Parameters are either set externally, to standard values or to match aggregate statistics, or calibrated in closed form. The per-period exit hazard χ 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 $\alpha_j = P_j Y_j^{1/\sigma} / (\sum_m P_m Y_m)^{1/\sigma}$ (equivalently $\alpha_j^\sigma = s_j P_j^{\sigma-1}$, with wage-bill share s_j), where $P_j = \exp(p_j)$ is the Stage-1 occupation price in levels 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.

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

Table 4 reports the full estimation results. Panel A shows the skill loadings λ_{jk}, the idiosyncratic wage variances $\varsigma_{\varepsilon j}^2$, and the log occupation prices p_j. For identification, Cluster 0 is the cognitive anchor ($\lambda_0 = (1, 0)$) and Cluster 1 the manual anchor ($\lambda_1 = (0, 1)$). Occupations differ sharply in their loadings: some load mainly on cognitive skills (Clusters 0, 3, and 6), others mainly on manual skills (Clusters 1, 2, and 5), and Clusters 4 and 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 $|\mathcal{J}| = 8$ 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 4: Stage 1: Expectation-Maximization Estimates of Wage Equation Parameters

Panel A: Loadings, Shock Variances, and Occupational Prices

Occupation	$\lambda_{j,C}$	$\lambda_{j,M}$	$\varsigma^2_{\varepsilon j}$	p_j
Cluster 0	1.000	0.000	0.0235	1.538
Cluster 1	0.000	1.000	0.0200	1.443
Cluster 2	0.202	0.826	0.0197	1.442
Cluster 3	0.875	0.000	0.0271	1.520
Cluster 4	0.527	0.567	0.0210	1.481
Cluster 5	0.350	0.737	0.0211	1.496
Cluster 6	0.827	0.375	0.0264	1.522
Cluster 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)
Cluster 0	0.988	0.012
Cluster 1	0.009	0.991
Cluster 2	0.162	0.838
Cluster 3	0.327	0.673
Cluster 4	0.448	0.552
Cluster 5	0.692	0.308
Cluster 6	0.946	0.054
Cluster 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 $\varsigma^2_{\varepsilon j}$, and the log occupation prices p_j (the occupation intercept in the wage equation). By normalization *Cluster 0* is the cognitive anchor ($\lambda_{0,C} = 1$, $\lambda_{0,M} = 0$) and *Cluster 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 $|\mathcal{J}| = 8$ 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 (θ_C) and manual (θ_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 5: 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)
Cluster 0 (base)	0.000	0.000	0.000
Cluster 1	0.169	0.088	−0.219
Cluster 2	0.188	0.119	−0.034
Cluster 3	0.176	0.123	0.116
Cluster 4	0.242	0.211	0.198
Cluster 5	0.191	0.143	0.133
Cluster 6	0.087	0.047	0.017
Cluster 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, Cluster 0, has $\eta = 0$ within each education group (21 free amenities). The taste-shock scale ϱ is weakly identified from continuation values without exogenous payoff shifters and is therefore set to match the average occupational mobility rate in the large sample.

D.3 Stage 2 Estimation Results: Dynamic Choice Parameters

D.4 EM Algorithm: Computational Details

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

For worker i employed in occupation j at date t,

$$w_{ijt} \;=\; p_j \;+\; \boldsymbol{\lambda}_j'\boldsymbol{\theta}_i \;+\; h_j(\boldsymbol{\tau}_{it}) \;+\; \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}\big(0, \varsigma^2_{\varepsilon j}\big), \tag{59}$$

where p_j is the occupation log price, $\boldsymbol{\theta}_i \in \mathbb{R}^K$ collects latent skills, $\boldsymbol{\lambda}_j \in \mathbb{R}^K_+$ are skill loadings, and $\boldsymbol{\tau}_{it} \in \mathbb{R}^K_+$ 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 $g(i) \in \mathcal{G}$ defined by education. The skill prior is conditionally Gaussian within group,

$$\boldsymbol{\theta}_i \mid g(i) = g \;\sim\; \mathcal{N}\big(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g\big). \tag{60}$$

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

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

with q indexing the EM iteration.

D.4.1 E-step

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

$$r_{it}^q \;:=\; w_{ijt} \;-\; p_{j(i,t)}^q \;-\; h_{j(i,t)}^q(\boldsymbol{\tau}_{it}). \tag{61}$$

Completing the square in $\boldsymbol{\theta}_i$ in the product of the wage likelihood implied by (59) and the group prior (60) delivers

$$\boldsymbol{V}_i^q = \left((\boldsymbol{\Sigma}_{g(i)}^q)^{-1} + \sum_{t=t_0(i)}^{T(i)} \frac{1}{\varsigma_{\varepsilon j(i,t)}^{2,q}} \boldsymbol{\lambda}_{j(i,t)}^q (\boldsymbol{\lambda}_{j(i,t)}^q)' \right)^{-1}, \tag{62}$$

$$\boldsymbol{\zeta}_i^q = \boldsymbol{V}_i^q \left((\boldsymbol{\Sigma}_{g(i)}^q)^{-1} \boldsymbol{\mu}_{g(i)}^q + \sum_{t=t_0(i)}^{T(i)} \frac{\boldsymbol{\lambda}_{j(i,t)}^q \, r_{it}^q}{\varsigma_{\varepsilon j(i,t)}^{2,q}} \right), \tag{63}$$

$$\boldsymbol{\theta}_i \mid \text{data}, \boldsymbol{\Gamma}^q, g(i) \;\sim\; \mathcal{N}\!\left(\boldsymbol{\zeta}_i^q,\, \boldsymbol{V}_i^q\right). \tag{64}$$

I store the worker-specific pair $(\boldsymbol{\zeta}_i^q, \boldsymbol{V}_i^q)$ at the end of every E-step iteration. Equations (62)–(63) are themselves a K-dimensional linear system per worker weighted by precision of the signals, solved in closed form by inverting the precision matrix in (62).

D.4.2 M-step

The M-step maximizes the expected complete-data log-likelihood over $\boldsymbol{\Gamma}$ while taking as given the posteriors $(\boldsymbol{\zeta}_i^q, \boldsymbol{V}_i^q)$ from the E-step. The complete-data log-likelihood sums, across workers, the log prior density of $\boldsymbol{\theta}_i$ in (60) and the log density of each wage observation in (59). Taking its expectation under the E-step posterior $\boldsymbol{\theta}_i \mid w_i, d_i, \boldsymbol{\Gamma}^q \sim \mathcal{N}(\boldsymbol{\zeta}_i^q, \boldsymbol{V}_i^q)$ gives

$$\begin{aligned} Q(\boldsymbol{\Gamma} \mid \boldsymbol{\Gamma}^q) = \; & -\tfrac{1}{2} \sum_i \Big[\log\left|2\pi \boldsymbol{\Sigma}_{g(i)}\right| + \operatorname{tr}\!\left(\boldsymbol{\Sigma}_{g(i)}^{-1} \boldsymbol{V}_i^q\right) + \left(\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_{g(i)}\right)' \boldsymbol{\Sigma}_{g(i)}^{-1} \left(\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_{g(i)}\right) \Big] \\ & -\tfrac{1}{2} \sum_i \sum_t \Big[\log\left(2\pi \varsigma_{\varepsilon j}^2\right) + \frac{\left(w_{ijt} - p_j - \boldsymbol{\lambda}_j' \boldsymbol{\zeta}_i^q - h_j(\boldsymbol{\tau}_{it})\right)^2 + \boldsymbol{\lambda}_j' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j}{\varsigma_{\varepsilon j}^2} \Big], \end{aligned} \tag{65}$$

with $j = j(i,t)$ being the occupation chosen by individual i in period t. This Q-function is additively separable into a prior block, separable across groups g, and a wage block, separable across occupations j; the M-step maximizes each block in closed form. Below I take the first-order conditions of (65) block by block: first for the prior moments, then for each occupation's wage parameters.

Prior moments optimization. Maximizing the prior block of (65) over $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ gives

the standard updates

$$\boldsymbol{\mu}_g^{q+1} = \frac{1}{N_g} \sum_{i:\, g(i)=g} \boldsymbol{\zeta}_i^q, \tag{66}$$

$$\boldsymbol{\Sigma}_g^{q+1} = \frac{1}{N_g} \sum_{i:\, g(i)=g} \Big[\boldsymbol{V}_i^q + (\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_g^{q+1})(\boldsymbol{\zeta}_i^q - \boldsymbol{\mu}_g^{q+1})' \Big]. \tag{67}$$

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

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

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

$$\boldsymbol{W}_j \;=\; \big[\, \mathbf{1},\;\; \boldsymbol{\zeta}^q,\;\; \boldsymbol{D}(\boldsymbol{\tau}) \,\big] \;\in\; \mathbb{R}^{N_j \times (1+K+|\mathcal{B}|)}, \tag{68}$$

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

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

The wage block of (65) is quadratic in $\boldsymbol{\vartheta}_j$, so updating the wage parameters is just a per-occupation least-squares regression of the stacked log wages $\boldsymbol{y}_j$ on $\boldsymbol{W}_j$. Its first-order conditions are given by the following closed-form equations

$$\boldsymbol{\vartheta}_j^{q+1} \;=\; \boldsymbol{S}_j^{-1}\boldsymbol{b}_j, \qquad \boldsymbol{S}_j \;=\; \boldsymbol{W}_j'\boldsymbol{W}_j \;+\; \boldsymbol{\Delta}_j^q, \qquad \boldsymbol{b}_j \;=\; \boldsymbol{W}_j'\boldsymbol{y}_j. \tag{69}$$

The only departure from textbook OLS is the correction $\boldsymbol{\Delta}_j^q = \sum_{t \in j} \boldsymbol{V}_{i(t)}^q$, which adds the summed posterior variances $\sum_{t \in j} \boldsymbol{V}_i^q$ to the $K \times K$ loading block of $\boldsymbol{S}_j$ and is zero elsewhere. This term is the curvature of the $\boldsymbol{\lambda}_j' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j$ correction in (65); it corrects the loadings $\boldsymbol{\lambda}_j$ for the residual posterior uncertainty about $\boldsymbol{\theta}_i$. Intuitively, residual wages are noisy measurements of multidimensional skills, because of the iid shocks and because skills enter wages only through the loadings. If I had observed every worker for many periods moving across different occupations, then this uncertainty would be close to zero for everyone. This system of equations has dimension $1 + K + |\mathcal{B}|$ and is solved separately for each occupation, delivering $(p_j^{q+1}, \boldsymbol{\lambda}_j^{q+1}, \boldsymbol{\delta}_j^{q+1})$ jointly.

The inclusion of the E-step posterior $\boldsymbol{\zeta}^q$ as a regressor in $\boldsymbol{W}_j$ represents the practical

implementation of the correction for selection on beliefs that the identification argument of Section 4.1 describes: regressing log wages on $(\mathbf{1}, \boldsymbol{\zeta}^q, \boldsymbol{D}(\boldsymbol{\tau}))$ per occupation absorbs endogenous occupational choice through the observed history, and so controls for belief-based selection into occupation, while the augmentation $\boldsymbol{\Delta}_j^q$ corrects the loading block for residual posterior uncertainty.

Anchored normalization. The factor model is identified up to a $K \times K$ invertible transformation of $(\boldsymbol{\theta}, \boldsymbol{\lambda})$. To eliminate this rotation and scale indeterminacy, I pin the full loading vector at two anchor occupations: a cognitive anchor a_C and a manual anchor a_M. In the default implementation of the EM algorithm I set the anchor loadings $\boldsymbol{\lambda}_{a_C} = (1,0)'$ and $\boldsymbol{\lambda}_{a_M} = (0,1)'$. This choice imposes four scalar restrictions (two scale, two rotation), which is exactly what is required to identify the orientation of the latent skill space. At the two anchor occupations I substitute $\boldsymbol{\lambda}_j$ out of (69) and solve the reduced system for $(p_j, \boldsymbol{\delta}_j)$ only.

Additionally, and to make sure that the parameters of the model have a meaningful economic content, I constrain the skill loadings to be non-negative, i.e., $\boldsymbol{\lambda}_j \geq \mathbf{0}$. This assumption ensures a meaningful ordinal ranking of skills: if $\lambda_{jk} < 0$ for any occupation j in task k then a very low skill worker in dimension k might have an unreasonably high productivity in occupation j.

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

$$\varsigma_{\varepsilon j}^{2,q+1} \;=\; \frac{1}{N_j} \sum_{(i,t):\, j(i,t)=j} \Big[\big(r_{it}^{q+1} - (\boldsymbol{\lambda}_j^{q+1})' \boldsymbol{\zeta}_i^q\big)^2 + (\boldsymbol{\lambda}_j^{q+1})' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j^{q+1} \Big], \tag{70}$$

where r_{it}^{q+1} is the residual from (61) evaluated at the new wage parameters. The second term inside the bracket is the variance-of-the-projection correction $\boldsymbol{\lambda}_j' \boldsymbol{V}_i^q \boldsymbol{\lambda}_j$ from (65) and ensures that posterior uncertainty about $\boldsymbol{\theta}_i$ is correctly transmitted into $\varsigma_{\varepsilon j}^2$.

The algorithm iterates (62)–(64) and (66)–(70) until convergence. Two stopping rules are available. The default monitors the mean squared change in the flattened parameter vector across consecutive iterations and stops once this falls below 10^{-9}; an alternative monitors the absolute change in the Q-function with tolerance 10^{-4}.

D.4.3 Likelihood Evaluation

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

$$\tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \;=\; \int_{\mathbb{R}^K} \left\{ \prod_{t=t_0(i)}^{T(i)} \phi\Big(w_{ij(i,t)t};\; p_{j(i,t)} + \boldsymbol{\lambda}_{j(i,t)}' \boldsymbol{\theta}_i + h_{j(i,t)}(\boldsymbol{\tau}_{it}),\; \varsigma_{\varepsilon j(i,t)}^2 \Big) \right\} \phi\big(\boldsymbol{\theta}_i; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\big)\, d\boldsymbol{\theta}_i, \tag{71}$$

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

I approximate (71) by Monte Carlo. For each worker i I fix a sequence of standard-normal draws $\boldsymbol{Z}_i^{(b)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and map them through the current prior,

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

and form the worker-level time-summed log contribution

$$S_i^{(b)}(\boldsymbol{\Gamma}) = \sum_{t=t_0(i)}^{T(i)} \log \phi\Big(w_{ij(i,t)t};\; p_{j(i,t)} + \boldsymbol{\lambda}_{j(i,t)}' \boldsymbol{\theta}_i^{(b)} + h_{j(i,t)}(\boldsymbol{\tau}_{it}),\; \varsigma^2_{\varepsilon j(i,t)} \Big).$$

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

$$\log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \;\approx\; \text{logsumexp}\big(\{S_i^{(b)}(\boldsymbol{\Gamma})\}_{b=1}^B\big) - \log B, \tag{72}$$

which is numerically stable (it avoids underflow) at extreme parameter values.[33]

D.5 NFXP Algorithm: Computational Details

D.5.1 Group–Occupation–Specific Integrated Value Functions and Fixed Point

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

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

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

$$s_{jt} \;\equiv\; \log W_{jt} - p_{jt} - h_j(\boldsymbol{\tau}_t) = \boldsymbol{\lambda}_j' \boldsymbol{\theta} + \varepsilon_{jt}, \qquad \varepsilon_{jt} \sim \mathcal{N}(0, \varsigma^2_{\varepsilon j}), \tag{73}$$

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

$$\tilde{\boldsymbol{\theta}}_{t+1} = \tilde{\boldsymbol{\theta}}_t + \boldsymbol{K}_{jt}\, z_{jt}, \qquad z_{jt} \equiv s_{jt} - \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\theta}}_t \sim \mathcal{N}\big(0,\, s_{zz}^{(j)}\big), \tag{74}$$

$$\tilde{\boldsymbol{\Sigma}}_{t+1} = \big(\boldsymbol{I}_K - \boldsymbol{K}_{jt} \boldsymbol{\lambda}_j'\big) \tilde{\boldsymbol{\Sigma}}_t, \qquad \boldsymbol{K}_{jt} = \frac{\tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_j}{s_{zz}^{(j)}}, \tag{75}$$

$$s_{zz}^{(j)} \equiv \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\Sigma}}_t \boldsymbol{\lambda}_j + \varsigma^2_{\varepsilon j}, \qquad \boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\omega}_j. \tag{76}$$

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

The next-period covariance $\tilde{\boldsymbol{\Sigma}}_{t+1}$ and the next-period tenures $\boldsymbol{\tau}_{t+1}$ are both deterministic transitions given $(\boldsymbol{x}_t, j)$; the only random object is the scalar innovation z_{jt}, with predictive variance $s_{zz}^{(j)}$. Because ε_{jt} is the only stochastic shock between two consecutive choices, the next posterior mean $\tilde{\boldsymbol{\theta}}_{t+1}$ moves only along the direction defined by the Kalman gain $\boldsymbol{K}_{jt}$. Although $\tilde{\boldsymbol{\theta}}_{t+1} \in \mathbb{R}^K$, its conditional distribution is supported on a one-dimensional subspace of $\mathbb{R}^K$. Therefore the expectation over the continuation value reduces to a one-dimensional integral over z_{jt}, regardless of the number of skills K.

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

$$\mathbb{E}\big[G(\tilde{\boldsymbol{\theta}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, j) \,\big|\, \boldsymbol{x}_t, j\big] \;=\; \pi^{-1/2} \int_{\mathbb{R}} G\Big(\tilde{\boldsymbol{\theta}}_t + \sqrt{2\, s_{zz}^{(j)}}\, \boldsymbol{K}_{jt}\, u,\; \tilde{\boldsymbol{\Sigma}}_{t+1},\; \boldsymbol{\tau}_t + \boldsymbol{\omega}_j,\; j\Big)\, e^{-u^2}\, du. \tag{77}$$

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

$$\mathbb{E}\big[G(\boldsymbol{x}_{t+1}) \,\big|\, \boldsymbol{x}_t, j\big] \;\approx\; \pi^{-1/2} \sum_{r=1}^{R} \varpi_r\, G\Big(\tilde{\boldsymbol{\theta}}_t + \sqrt{2\, s_{zz}^{(j)}}\, \boldsymbol{K}_{jt}\, u_r,\; \tilde{\boldsymbol{\Sigma}}_{t+1},\; \boldsymbol{\tau}_t + \boldsymbol{\omega}_j,\; j\Big). \tag{78}$$

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

Deterministic index and integrated value. In practice, rather than including the discrete states in the state space, I define group- and occupation-specific value functions, with only the continuous variables $\boldsymbol{x} \equiv (\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau})$ remaining as arguments. From here on in this appendix, $\boldsymbol{x}$ denotes this continuous part of the state, $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau})$, rather than the full state $\boldsymbol{x}_t$ above. Let

$$EV_{mg}(\boldsymbol{x}) \;\equiv\; EV\big(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1} = m,\, g\big), \qquad m \in \mathcal{J},\; g \in \mathcal{G},$$

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

$$\begin{aligned} \nu_{j|m,g}(\boldsymbol{x}) \;=\; & \underbrace{p_j + \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\theta}} + h_j(\boldsymbol{\tau})}_{\text{expected log wage}} + \eta_{jg} - \mathcal{C}(j, m) \\ & + \beta\, \mathbb{E}\big[\, EV_{jg}(\boldsymbol{x}') \;\big|\; \boldsymbol{x}, j \,\big], \end{aligned} \tag{79}$$

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

With Type-I extreme value taste shocks of scale ϱ under the mean-zero normalization (so the Euler–Mascheroni constant is absorbed into the location parameter and no additive term in the Euler–Mascheroni constant times ϱ remains), the occupation-specific integrated values satisfy a log-sum-exp fixed point. Substituting the univariate Gauss–Hermite approximation (78) for the continuation expectation in (79) and writing the deterministic flow utility as $u_{jg}(\boldsymbol{x}) \equiv p_j + \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\theta}} + h_j(\boldsymbol{\tau}) + \eta_{jg}$, the integrated Bellman equation reads, for each $m \in \mathcal{J}$ and $g \in \mathcal{G}$,

$$EV_{mg}(\boldsymbol{x}) = \varrho \log \sum_{j\in\mathcal{J}} \exp\left(\frac{1}{\varrho}\Big[\, u_{jg}(\boldsymbol{x}) - \mathcal{C}(j,m) + \beta\, \pi^{-1/2} \sum_{r=1}^{R} \varpi_r\, EV_{jg}\Big(\tilde{\boldsymbol{\theta}} + \sqrt{2\, s_{zz}^{(j)}}\, \boldsymbol{K}_j\, u_r,\ \tilde{\boldsymbol{\Sigma}}',\ \boldsymbol{\tau} + \boldsymbol{\omega}_j\Big) \,\Big]\right), \qquad \boldsymbol{x} \in \mathcal{X}. \tag{80}$$

For each group g, collect the integrated values across incumbent occupations into the vector $\boldsymbol{EV}_g(\boldsymbol{x}) = \big(EV_{mg}(\boldsymbol{x})\big)_{m\in\mathcal{J}}$. The right-hand side of (80) for a given incumbent m aggregates over all possible occupation choices $j \in \mathcal{J}$, and each choice carries its own continuation value EV_{jg} evaluated with the transitions implied by choosing j, so every component of $\boldsymbol{EV}_g$ enters the update of every other component. The $|\mathcal{J}|$ equations therefore constitute a single vector fixed point,

$$\boldsymbol{EV}_g = \mathcal{T}(\boldsymbol{EV}_g), \qquad \big[\mathcal{T}(\boldsymbol{EV}_g)\big]_m(\boldsymbol{x}) = \varrho \log \sum_{j\in\mathcal{J}} \exp\Big(\tfrac{1}{\varrho}\big[\, u_{jg}(\boldsymbol{x}) - \mathcal{C}(j,m) + \beta\, \mathbb{E}\big(EV_{jg}(\boldsymbol{x}') \mid \boldsymbol{x}, j\big)\,\big]\Big), \tag{81}$$

which must be solved jointly for all $|\mathcal{J}|$ components rather than one occupation at a time. By Lemma 2, $\mathcal{T}$ is a contraction of modulus at most β, so (81) has a unique fixed point $\boldsymbol{EV}_g^{\star}$. The fixed point is solved separately for each $g \in \mathcal{G}$. In what follows I suppress the dependence on g to simplify the notation.

Smolyak approximation of the integrated value function on a sparse grid. The continuous state $\boldsymbol{x} = (\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau})$, which collects the posterior means, the distinct elements of the posterior variance-covariance matrix, and task experience, can be high-dimensional. I approximate each value function $EV_{mg}(\cdot)$ on a sparse grid using the Smolyak method (Malin et al. 2011; Judd et al. 2014) with Chebyshev polynomials. Because the construction is identical for every incumbent occupation m and group g, I describe it for a generic value function $EV(\boldsymbol{x})$ and suppress the (m, g) index.

Fix lower and upper bounds for each component of $\boldsymbol{x}$. Let $D = 2K + K(K+1)/2$ be the continuous-state dimension ($D = 7$ with two skills) and $\bar{\ell}$ the Smolyak level. The

sparse grid is

$$\mathcal{H}^{D,\bar{\ell}} = \bigcup_{\mathbf{i}:\,|\mathbf{i}|=D+\bar{\ell}} \left(\mathcal{S}^{n(i_1)} \times \cdots \times \mathcal{S}^{n(i_D)}\right), \qquad |\mathbf{i}| := i_1 + \cdots + i_D,$$

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

$$EV^{\mathbf{i}}(\boldsymbol{x}; \boldsymbol{\phi}^{(\mathbf{i})}) = \sum_{l_1=1}^{n(i_1)} \cdots \sum_{l_D=1}^{n(i_D)} \phi^{(\mathbf{i})}_{l_1 \cdots l_D}\, T_{l_1}(x_1) \cdots T_{l_D}(x_D) = \boldsymbol{\phi}^{(\mathbf{i})\prime} \mathbf{T}(\boldsymbol{x}),$$

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

$$\widehat{EV}^{D,\bar{\ell}}(\boldsymbol{x}) = \sum_{D \le |\mathbf{i}| \le D+\bar{\ell}} (-1)^{D+\bar{\ell}-|\mathbf{i}|} \binom{D-1}{D+\bar{\ell}-|\mathbf{i}|}\, EV^{\mathbf{i}}(\boldsymbol{x}; \boldsymbol{\phi}^{(\mathbf{i})}). \tag{82}$$

Fixed point in coefficient space. Substituting the Smolyak interpolant $\widehat{EV}_j(\boldsymbol{x}) = \boldsymbol{\phi}^{(j)\prime}\mathbf{T}(\boldsymbol{x})$ for the unknown continuation value EV_j inside (80) recasts the integrated Bellman equation as a fixed point in the finite-dimensional Smolyak coefficients $\boldsymbol{\Phi} \equiv [\boldsymbol{\phi}^{(m)}]_{m \in \mathcal{J}} \in \mathbb{R}^{M \times |\mathcal{J}|}$, whose columns index the incumbent occupation and where M is the number of Smolyak basis functions. For each incumbent $m \in \mathcal{J}$,

$$\begin{aligned} \boldsymbol{\phi}^{(m)\prime}\mathbf{T}(\boldsymbol{x}) \;= \varrho \, \log \sum_{j \in \mathcal{J}} \exp\Bigg(\frac{1}{\varrho} \Bigg[& u_j(\boldsymbol{x}) - \mathcal{C}(j,m) \\ & + \beta\, \pi^{-1/2} \sum_{r=1}^{R} \varpi_r\, \boldsymbol{\phi}^{(j)\prime} \mathbf{T}\Big(\tilde{\boldsymbol{\theta}} + \sqrt{2\, s_{zz}^{(j)}}\, \boldsymbol{K}_j\, u_r, \\ & \tilde{\boldsymbol{\Sigma}}',\; \boldsymbol{\tau} + \boldsymbol{\omega}_j \Big) \Bigg] \Bigg), \qquad \forall\, \boldsymbol{x} \in \mathcal{X},\; m \in \mathcal{J}. \end{aligned} \tag{83}$$

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 $EV_m(\cdot)$.

To obtain the matrix-stacked map solved in practice, evaluate (83) at every node of the sparse grid. Let $\{\boldsymbol{x}^{(n)}\}_{n=1}^{N_G}$ denote the N_G grid points in $\mathcal{H}^{D,\bar{\ell}}$, and let $\boldsymbol{y}^{(n)} \in [-1,1]^D$ be $\boldsymbol{x}^{(n)}$ rescaled to the Chebyshev domain. Define the basis matrix $\mathbf{T} \in \mathbb{R}^{M \times N_G}$ by $\mathbf{T}_{\ell,n} = T_\ell(\boldsymbol{y}^{(n)})$ (basis function $\ell = 1, \ldots, M$, grid point n) and the collocation pseudo-inverse

$$\mathbf{T}^{+} \;\equiv\; (\mathbf{T}\mathbf{T}^{\top})^{-1}\mathbf{T} \;\in\; \mathbb{R}^{M \times N_G};$$

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 (83) into the matrix $\widetilde{\mathbf{EV}}(\boldsymbol{\Phi}) \in \mathbb{R}^{|\mathcal{J}| \times N_G}$ with entries $\widetilde{\mathbf{EV}}_{mn}(\boldsymbol{\Phi}) = \widehat{EV}_m(\boldsymbol{x}^{(n)}; \boldsymbol{\Phi})$, the fixed-point map

in coefficient space is

$$\mathbf{\Phi} \;=\; \mathcal{F}(\mathbf{\Phi}) \;\equiv\; \mathbf{T}^{+}\,\widetilde{\mathbf{EV}}(\mathbf{\Phi})'. \tag{84}$$

I compute the fixed point $\mathbf{\Phi}^{\star}$ of (84) by successive substitution, compiled into a single JAX while-loop; gradients of the outer likelihood with respect to $\mathbf{\Theta}_{\text{dyn}}$ are obtained by implicit differentiation of (84) at the converged coefficients (Section D.5.6), not by differentiating through the iterations.

D.5.2 Fixed-Point Algorithm for Integrated Value Function Smolyak Approximation

The inner loop of the NFXP procedure solves (84) by successive iteration on the Smolyak coefficient matrix $\mathbf{\Phi}$. Each iteration evaluates the right-hand side of (83) at every grid point, takes a Gauss–Hermite expectation over next-period beliefs, and updates the coefficients. The algorithm box below gives the steps.

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

---

Inputs.

- Structural parameters $\mathbf{\Theta}$.
- Sparse-grid nodes $\{\boldsymbol{x}^{(n)}\}_{n=1}^{N_G} \subset \mathcal{H}^{D,\bar{\ell}}$.
- Chebyshev basis $\mathbf{T} \in \mathbb{R}^{M\times N_G}$ and pseudo-inverse $\mathbf{T}^{+} = (\mathbf{T}\mathbf{T}^{\top})^{-1}\mathbf{T}$.
- Univariate Gauss–Hermite nodes and weights $\{(u_r, \varpi_r)\}_{r=1}^{R}$.
- Tolerance $\epsilon_{\text{tol}} > 0$; maximum iterations $n_{\max}$.

Output. Fixed-point Smolyak coefficient matrix $\mathbf{\Phi}^{\star} \in \mathbb{R}^{M\times|\mathcal{J}|}$ solving (84).

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

Main loop. Repeat for $s = 0, 1, \ldots, n_{\max} - 1$ until convergence:

For each grid point n and prospective choice $j \in \mathcal{J}$:

- Update next-period belief covariance $\tilde{\mathbf{\Sigma}}'(\boldsymbol{x}^{(n)}, j)$ via Kalman update.
- Build R shifted next-state nodes $\boldsymbol{x}'_r$ via the Gauss–Hermite map [(78)].
- Evaluate Smolyak interpolant $\widehat{EV}_j(\boldsymbol{x}'_r; \mathbf{\Phi}^{(s)}) = \boldsymbol{\phi}^{(s,j)\prime}\mathbf{T}(\boldsymbol{x}'_r)$.
- Form conditional values $v_{j|m}(\boldsymbol{x}^{(n)})$ for all incumbents $m \in \mathcal{J}$ [(83)].

Aggregate via log-sum-exp: $\widehat{EV}_m(\boldsymbol{x}^{(n)}) = \varrho \log \sum_j \exp\big(v_{j|m}(\boldsymbol{x}^{(n)})/\varrho\big)$.

Collocation update: $\mathbf{\Phi}^{(s+1)} = \mathbf{T}^{+}\,\widetilde{\mathbf{EV}}(\mathbf{\Phi}^{(s)})'$ [(84)].

Convergence: if $\big\|\mathbf{\Phi}^{(s+1)} - \mathbf{\Phi}^{(s)}\big\|_F < \epsilon_{\text{tol}}$, exit.

Return $\mathbf{\Phi}^{\star} = \mathbf{\Phi}^{(s+1)}$.

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

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

which avoids overflow when one alternative dominates.

D.5.3 Policy Functions, CCPs, and Simulation

Given $\{\widehat{EV}_m\}_{m\in\mathcal{J}}$, the CCPs at $(\boldsymbol{x}, m)$ follow from (85). To simulate a worker's path:

1. Given $(\tilde{\boldsymbol{\theta}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, m)$, compute CCPs $\{\mathbb{P}(d_t = j \mid \boldsymbol{x}_t, m)\}_{j\in\mathcal{J}}$ and draw d_t (or use the argmax for policy).
2. Draw a wage residual $\varepsilon_{d_t,t} \sim \mathcal{N}(0, \varsigma^2_{\varepsilon d_t})$ and set

$$\log W_{d_t t} = p_{d_t} + \boldsymbol{\lambda}'_{d_t}\boldsymbol{\theta} + h_{d_t}(\boldsymbol{\tau}_t) + \varepsilon_{d_t,t}.$$

   Equivalently, the signal in (73) is $s_{d_t t} = \log W_{d_t t} - p_{d_t} - h_{d_t}(\boldsymbol{\tau}_t)$.
3. Update beliefs via (74)–(76) using $j = d_t$.
4. Update tenures: $\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{\omega}_{d_t}$ (cf. (10)).
5. Update the discrete state: $m \leftarrow d_t$ (next period's lagged occupation).

D.5.4 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 $\{\widehat{EV}_j(\boldsymbol{x}; \boldsymbol{\Phi}^\star)\}_{j\in\mathcal{J}}$ and a worker arriving at state $\boldsymbol{x}$ with lagged occupation m, the multinomial-logit CCPs implied by the Type-I extreme value taste shocks are

$$\mathbb{P}\big(d_t = j \mid \boldsymbol{x}, m; \boldsymbol{\Theta}\big) \;=\; \frac{\exp\big(v_{j|m}(\boldsymbol{x})/\varrho\big)}{\sum_{\ell\in\mathcal{J}} \exp\big(v_{\ell|m}(\boldsymbol{x})/\varrho\big)}, \qquad j \in \mathcal{J}, \tag{85}$$

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

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

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

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

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

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

Kalman recursion (74)–(76) evaluated at the observed choice d_{it},

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

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

The fixed-point algorithm of Section D.5.2 returns the Smolyak coefficient matrix $\boldsymbol{\Phi}^\star$ that approximates the value function at the current parameter estimates. To evaluate the choice likelihood, I must compute CCPs for each observation in the estimation sample. Because I approximate the value function using Chebyshev polynomials, I can simply compute the value function at every observed state.

Stacking (85) across workers and periods yields

$$\log L^c\big(\boldsymbol{\Theta}_{\text{dyn}}; \hat{\boldsymbol{\Theta}}_{\text{wage}}\big) \;=\; \sum_{i=1}^{N}\sum_{t=t_0(i)}^{T(i)} \log \mathbb{P}\big(d_{it} \mid \boldsymbol{x}_{it}, d_{i,t-1}; \boldsymbol{\Phi}^\star(\boldsymbol{\Theta}_{\text{dyn}}), \hat{\boldsymbol{\Theta}}_{\text{wage}}\big). \tag{86}$$

The wage-block parameters $\hat{\boldsymbol{\Theta}}_{\text{wage}}$ enter (86) only through the choice-specific values inside $\boldsymbol{\Phi}^\star$ and are held fixed at their Stage 1 EM estimates. Entry rows contribute via $d_{i,-1} = \varnothing$ (coded as -1), in which case no switching cost is incurred.

D.5.5 NFXP Procedure

Stage 1, the EM routine, delivers the wage-block estimates $\hat{\boldsymbol{\Theta}}_{\text{wage}}$ and the means and covariances $\{\hat{\boldsymbol{\mu}}_g, \hat{\boldsymbol{\Sigma}}_g\}$ for each group g. Taking these parameters as given, and using them to build the path of beliefs for each worker as described above, the NFXP estimator searches for the dynamic-choice parameters $\boldsymbol{\Theta}_{\text{dyn}} = \big(\{\eta_{jg}\}_{j\neq 0,\,g}, \kappa_0\big)$ that maximize the choice log-likelihood (86), holding the discount factor β and the taste-shock scale ϱ fixed. For each candidate $\boldsymbol{\Theta}^{(k)}_{\text{dyn}}$ proposed by the outer optimizer, the procedure executes two steps in sequence.

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

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

i.e. (84), by successive substitution on the Smolyak coefficient matrix $\boldsymbol{\Phi}$, returning the converged coefficients $\boldsymbol{\Phi}^\star\big(\boldsymbol{\Theta}^{(k)}_{\text{dyn}}\big)$ via the algorithm of Section D.5.2.

Step 2 — Choice likelihood. Given $\boldsymbol{\Phi}^\star\big(\boldsymbol{\Theta}^{(k)}_{\text{dyn}}\big)$, evaluate the CCPs (85) at the observed worker states $\{\boldsymbol{x}_{it}\}$ using the approximated value function, and compute the choice log-likelihood $\log L^c\big(\boldsymbol{\Theta}^{(k)}_{\text{dyn}}; \hat{\boldsymbol{\Theta}}_{\text{wage}}\big)$ in (86).

These two steps must be computed for each likelihood evaluation. In the practical

implementation, L-BFGS-B (with JAX automatic differentiation; see Section D.5.6) uses $\log L^c$ and its gradient to propose the next candidate $\boldsymbol{\Theta}_{\text{dyn}}^{(k+1)}$, and Steps 1–2 are repeated. Iteration continues until $\boldsymbol{\Theta}_{\text{dyn}}^{(k)}$ converges.

D.5.6 Differentiable NFXP Implementation (JAX)

The estimation routine is implemented in JAX and runs on the GPU. For a given vector of dynamic choice parameters $\boldsymbol{\Theta}_{\text{dyn}}$, I solve the inner fixed point $\boldsymbol{\Phi}^\star$ by iterating the Bellman map $\mathcal{F}$ until successive iterations change by less than 10^{-8}. I then take the gradient of the likelihood by implicitly differentiating the fixed-point condition $\boldsymbol{\Phi}^\star = \mathcal{F}(\boldsymbol{\Phi}^\star)$; this keeps the gradient exact while remaining fast and light on memory. The dynamic choice parameters are updated with L-BFGS-B using these gradients. A full evaluation, the inner solve plus the choice likelihood over $N \simeq 2.4$ million observations, takes about 31 minutes on a single NVIDIA H200 GPU node.

D.5.7 Two-Step Estimation Summary

The complete estimation procedure pairs the EM algorithm of Appendix D.4 (Stage 1, wage and skill block) with the NFXP procedure described above (Stage 2, dynamic-choice block). The algorithm box below summarizes the full pipeline.

Algorithm: Two-Step Estimation (EM + NFXP)

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

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

Stage 1 — EM [Appendix D.4].

1. Initialize $(\Theta^{(0)}_{\text{wage}}, \mu_g^{(0)}, \Sigma_g^{(0)})$.
2. Repeat for $q = 0, 1, \ldots$ until $\text{MSD}(\Theta^{(q)}) < 10^{-9}$:
   - **E-step** [(62), (63)]: per-worker posteriors $(\hat{\zeta}_i^{(q)}, \hat{V}_i^{(q)})$.
   - **M-step** [(69), (70), (66)]: closed-form linear systems per occupation; group prior moments.
3. **Output:** $\hat{\Theta}_{\text{wage}}$, $\{\hat{\mu}_g, \hat{\Sigma}_g\}_{g\in\mathcal{G}}$, $\{(\hat{\zeta}_i, \hat{V}_i)\}_{i=1}^N$.

Stage 2 — NFXP on $\Theta_{\text{dyn}} = (\{\eta_{jg}\}, \kappa_0)$, with β and ϱ fixed [Sections D.5.2, D.5.4].

1. Build per-worker state path $\{x_{it}\}$: entry belief $(\hat{\mu}_{g(i)}, \hat{\Sigma}_{g(i)})$; observed-row belief from the forward Kalman recursion on wage residuals at the first-stage estimates; tenures $\boldsymbol{\tau}_{it}$.
2. Initialize $\Theta^{(0)}_{\text{dyn}}$ with normalization $\eta_{0g} = 0$ for each $g \in \mathcal{G}$.
3. Repeat for $k = 0, 1, \ldots$ until L-BFGS-B convergence (ftol $< 10^{-6}$, gtol $< 10^{-5}$):
   - **Inner solve**: fixed point $\Phi^\star(\Theta^{(k)}_{\text{dyn}})$ via the algorithm in Section D.5.2.
   - **CCPs** [(85)]: evaluate CCPs at panel state variables $\{x_{it}\}$ using the approximated value function.
   - **Likelihood** [(86)]: $\log L^c(\Theta^{(k)}_{\text{dyn}})$.
   - **Outer step**: L-BFGS-B update with JAX autodiff gradients.
4. **Output:** $\hat{\Theta}_{\text{dyn}}$.

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

D.6 Counterfactual Choice Rules: Learning and Experience Off

The decomposition of occupational mobility in Figures 9 and 10 compares the estimated model with two counterfactual choice rules. Both counterfactuals keep the simulated baseline economy fixed and change only the rule that workers use to choose their occupation. For each counterfactual $c \in \{\text{lo}, \text{eo}\}$, I first re-solve the fixed point in coefficient space (84) under a modified decision problem, obtaining the Smolyak coefficients Φ^{lo} (learning off) and Φ^{eo} (experience off), while occupation prices, amenities, the switching cost, and the taste-shock scale are held at their baseline values. Let $x_{it} = (\tilde{\theta}_{it}, \tilde{\Sigma}_{it}, \boldsymbol{\tau}_{it})$ and $m = d_{i,t-1}$ denote the realized state and incumbent occupation of worker i of group g in period t of the simulated baseline economy. In the estimated model, the choice-specific value of occupation j is computed from the baseline coefficients $\Phi^\star$ as in (80),

$$v^\star_{j|m}(x_{it}) = u_{jg}(x_{it}) - \mathcal{C}(j, m) + \beta\, \pi^{-1/2} \sum_{r=1}^{R} \varpi_r\, \boldsymbol{\phi}^{\star(j)\prime} \mathbf{T}\Big(\tilde{\theta}_{it} + \sqrt{2\, s_{zz}^{(j)}}\, K_j\, u_r,\ \tilde{\Sigma}'_{it},\ \boldsymbol{\tau}_{it} + \boldsymbol{\omega}_j\Big).$$

Learning off ($\boldsymbol{\Phi}^{\text{lo}}$). The belief law of motion is replaced by the identity: workers never update their beliefs, so the continuation value of choosing occupation j is evaluated at unchanged beliefs and task experience $\boldsymbol{\tau} + \boldsymbol{\omega}_j$, with no expectation over the wage signal. When the choice rule is evaluated, each worker's beliefs are set to the prior of her group, $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, while her task experience and incumbent occupation are the realized ones:

$$v^{\text{lo}}_{j|m}(\boldsymbol{x}_{it}) = p_j + \boldsymbol{\lambda}_j'\boldsymbol{\mu}_g + h_j(\boldsymbol{\tau}_{it}) + \eta_{jg} - \mathcal{C}(j,m) + \beta\, \boldsymbol{\phi}^{\text{lo}(j)\prime}\mathbf{T}\big(\boldsymbol{\mu}_g,\ \boldsymbol{\Sigma}_g,\ \boldsymbol{\tau}_{it} + \boldsymbol{\omega}_j\big).$$

Experience off ($\boldsymbol{\Phi}^{\text{eo}}$). The returns to task-specific experience are set to zero, $h_j(\boldsymbol{\tau}) = 0$, and experience is not accumulated, both in the current expected wage and in the continuation value. Learning is unchanged, and the choice rule is evaluated at the worker's realized posterior beliefs:

$$\begin{aligned} v^{\text{eo}}_{j|m}(\boldsymbol{x}_{it}) = {} & p_j + \boldsymbol{\lambda}_j'\tilde{\boldsymbol{\theta}}_{it} + \eta_{jg} - \mathcal{C}(j,m) \\ & + \beta\, \pi^{-1/2} \sum_{r=1}^{R} \varpi_r\, \boldsymbol{\phi}^{\text{eo}(j)\prime}\mathbf{T}\Big(\tilde{\boldsymbol{\theta}}_{it} + \sqrt{2\, s^{(j)}_{zz}}\, \boldsymbol{K}_j\, u_r,\ \tilde{\boldsymbol{\Sigma}}'_{it},\ \boldsymbol{\tau}_{it}\Big). \end{aligned}$$

Counterfactual choice probabilities. For each rule $c \in \{\star, \text{lo}, \text{eo}\}$, the conditional choice probabilities are the multinomial-logit probabilities (85) evaluated at the choice-specific values of that rule,

$$\mathbb{P}^c\big(d_t = j \mid \boldsymbol{x}_{it}, m\big) = \frac{\exp\big(v^c_{j|m}(\boldsymbol{x}_{it})/\varrho\big)}{\sum_{\ell \in \mathcal{J}} \exp\big(v^c_{\ell|m}(\boldsymbol{x}_{it})/\varrho\big)}, \qquad j \in \mathcal{J}.$$

The counterfactual therefore answers the question: given everything that happened to the worker up to period t, which occupation would she have chosen if her beliefs were fixed at her prior, or if task-specific human capital neither paid off nor accumulated? The probability of switching is

$$\mathbb{P}^c(\text{switch}_{it}) = 1 - \mathbb{P}^c\big(d_t = m \mid \boldsymbol{x}_{it}, m\big),$$

and the characteristics of the move, such as the task distance or the change in the cognitive task share between m and the destination j', are averaged with weights $\mathbb{P}^c(d_t = j' \mid \boldsymbol{x}_{it}, m)/\mathbb{P}^c(\text{switch}_{it})$ for $j' \neq m$. Using choice probabilities rather than realized choices removes the noise generated by the taste shocks, and evaluating all rules at the same realized states ensures that the differences across lines reflect only differences in the choice rule. Model match quality is the within-(occupation, year) percentile of the latent match $\boldsymbol{\lambda}_j'\boldsymbol{\theta}_i$. As a robustness check, for the learning-off counterfactual I also fully re-simulate the synthetic panel under $\boldsymbol{\Phi}^{\text{lo}}$, so that the changed choices feed back into later states; the resulting mobility dynamics are nearly identical.

E Appendix: Model Fit

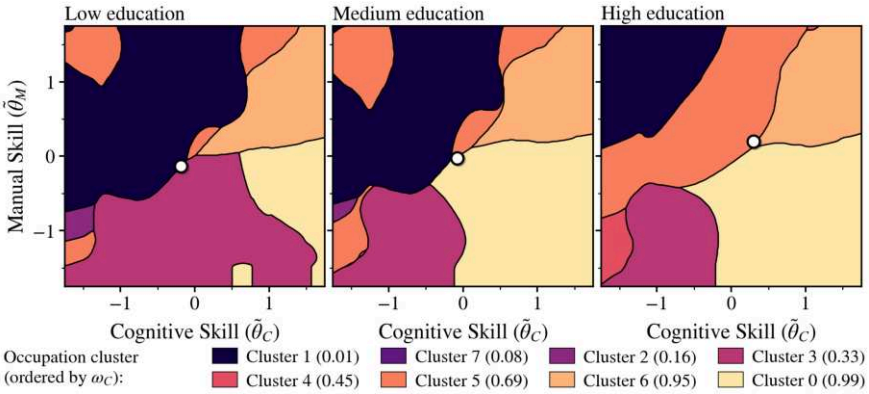
E.1 Entry into the Labor Market and Additional Model Fit

This appendix illustrates how workers enter the labor market in the model and compares the model with moments of the large sample that are not targeted in estimation: the distribution of wages at labor-market entry, the distribution of wages within and across occupations, the wage–tenure profile of occupational stayers, and the decline of mobility with occupational tenure. In all the comparisons the data are the log real hourly wage in the large sample, and the model is the baseline stationary simulation of Section 6, with the large-sample education shares.

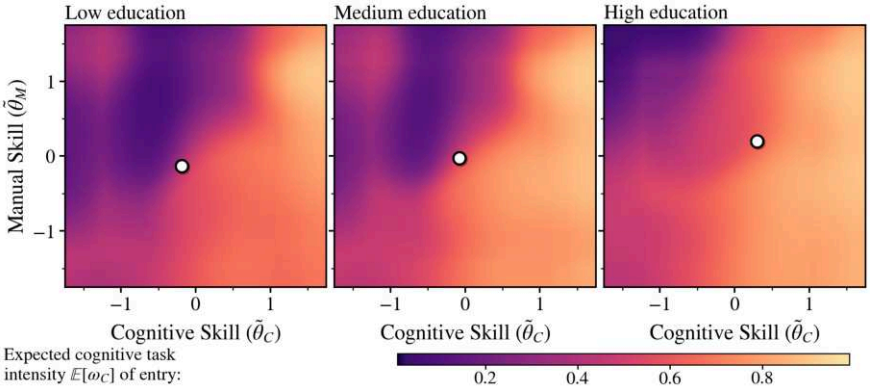
Taste shocks and the entry occupational choice. At the beginning of the career, every worker in education group g shares the same prior beliefs $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ and has no experience, so the systematic part of the choice value is identical within the group. In the deterministic limit $\varrho \to 0$ the entry choice is therefore degenerate: every worker of a specific group would enter the same occupation and workers' paths would diverge only later, through the wage signals they happen to draw and the experience they accumulate. Panel (a) of Figure 27 maps this deterministic assignment as a function of entry beliefs $(\tilde{\theta}_C, \tilde{\theta}_M)$; the white marker locates each group's prior mean $\boldsymbol{\mu}_g$.

The idiosyncratic taste shocks ϵ_{ijt} break this degenerate choice at labor-market entry. With taste shocks the entry choice is noisy around the occupation that would have been chosen in the degenerate case. Workers from the same group, who have identical beliefs at entry, spread across neighboring occupations. Panel (b) of Figure 27 plots the expected cognitive task intensity $\mathbb{E}[\omega_C]$ of the entry occupation, in expectation over the taste shocks, as a function of entry beliefs. The dispersion in the occupational choice still depends on the initial prior: Panel (c) reports the probability of the most likely entry occupation, $\max_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\theta}})$, which falls along the indifference lines where a worker is nearly indifferent between occupations, and rises toward one where comparative advantage is strong. For example, with the estimated priors, low-educated workers would always enter Cluster 1, which has a very high manual content. However, with taste shocks, they may end up in any occupation but are more likely to end up in either Cluster 3 with $\omega_M = 0.67$ or Cluster 1 with $\omega_M = 0.99$. Over the life cycle, taste shocks also allow the model to rationalize occupational choices and moves that its systematic component alone would not predict.

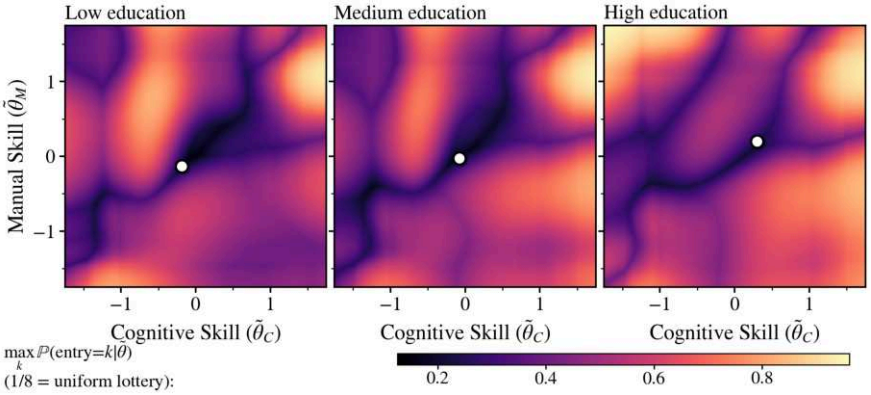
Wages at labor-market entry. The entry wage distribution is an informative test of the model because, conditional on education, all entrants share the same prior beliefs and have no experience, so the systematic part of their occupational choice is identical. Entrants are assigned to occupations by the prior and by the taste shocks alone, as illustrated above, and their wages reveal skills that neither they nor the econometrician observe at that point. The dispersion of entry wages within an education group therefore



(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, $\max_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\theta}})$

Figure 27: 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 ($\tilde{\theta}_C$, horizontal axis) and manual skill ($\tilde{\theta}_M$, 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, $\mathbb{E}[\omega_C] = \sum_k \mathbb{P}(\text{entry} = k \mid \tilde{\boldsymbol{\theta}})\, \omega_{kC}$. *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 calibrated taste-shock scale $\varrho = 0.271$.

reflects only the estimated skill distribution of the group, the loadings of the occupations entrants select into, and the wage noise. Figure 28 compares the quantiles of the entry wage distribution in the model with those in the data, by education group. The model matches the mean and the median of each group (mean log wage 1.31 against 1.32 for low-educated, 1.43 against 1.44 for medium-educated, and 1.84 against 1.77 for high-educated workers) and the dispersion of the two larger groups (standard deviation 0.29 against 0.29 and 0.33 against 0.34), while it under-disperses the entry wages of high-educated workers (0.45 against 0.37). The quantile–quantile plots show where the differences lie: the model compresses both tails of the low- and medium-education distributions slightly (the gap at the 10th percentile is -0.05 to -0.08 log points and at the 90th $+0.04$) and misses the upper tail of the high-education distribution (the gap at the 90th percentile is -0.13). Within education $\times$ entry-occupation cells, the correlation between the mean entry wage in the data and in the model is 0.85.

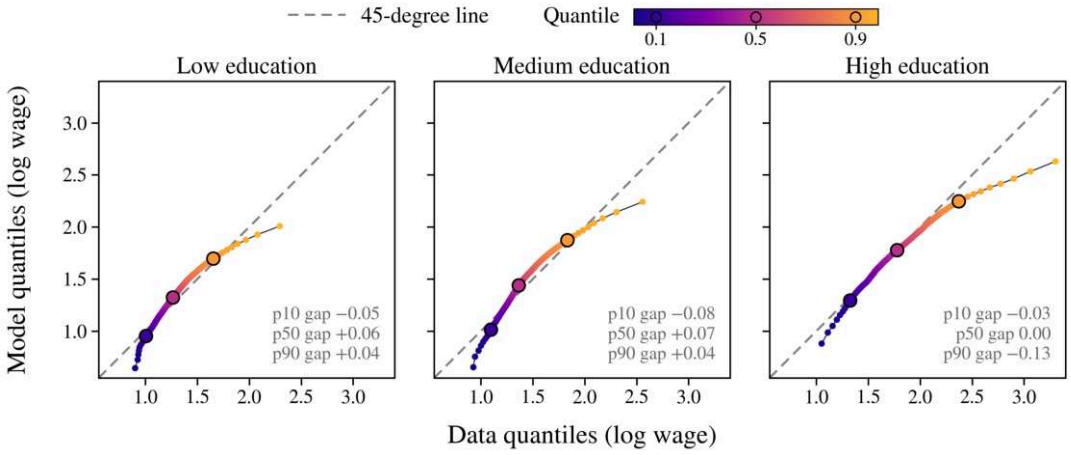


Figure 28: Quantile–quantile plots of the log real hourly wage at labor-market entry, model against data, by education group. Entrants are workers in their first year in the labor market. Each point is a quantile of the entry wage distribution, from the 1st to the 99th percentile; the dashed line is the 45° line, and the gaps between the model and the data at the 10th, 50th, and 90th percentiles are reported in each panel.

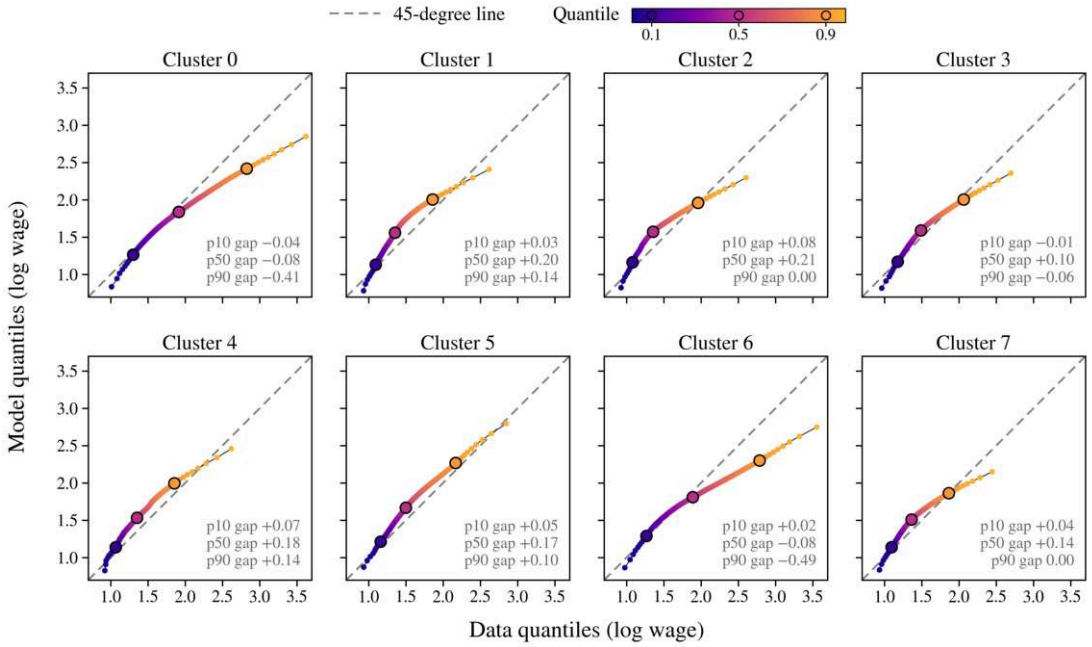
Wage distributions across occupations. Figure 29 compares the distribution of log wages in each of the eight occupations over all worker-years. The quantile–quantile plots in Panel (a) show that the model reproduces the within-occupation dispersion of the manual- and mixed-task occupations (the standard deviation of the log wage is 0.34 against 0.34 in Cluster 1, 0.36 against 0.33 in Cluster 3, 0.35 against 0.34 in Cluster 4, and 0.42 against 0.41 in Cluster 5), with the model distribution shifted up by 0.1 to 0.2 log points at the median in these occupations, and that it under-disperses the two most cognitive-intensive occupations, Clusters 0 and 6: their medians are matched within 0.08 log points, but the long right tails of high earners are missing in the model (the gap at the 90th percentile is -0.41 and -0.49 log points), so that the standard deviation is 0.60 in the data against 0.44 and 0.40 in the model. Panel (b) overlays the eight distributions: in both the data and the model the wage distributions of different occupations overlap substantially, so a worker's occupation is far from determining her

wage, in contrast with models of occupational sorting in which the wage distributions of different occupations are disjoint. The model generates, if anything, more overlap than the data: occupation dummies explain 22% of the variance of log wages in the data and 9% in the model, because the estimated occupation prices and loadings are closer to each other than the raw occupation means, which also reflect the selection of workers across occupations.

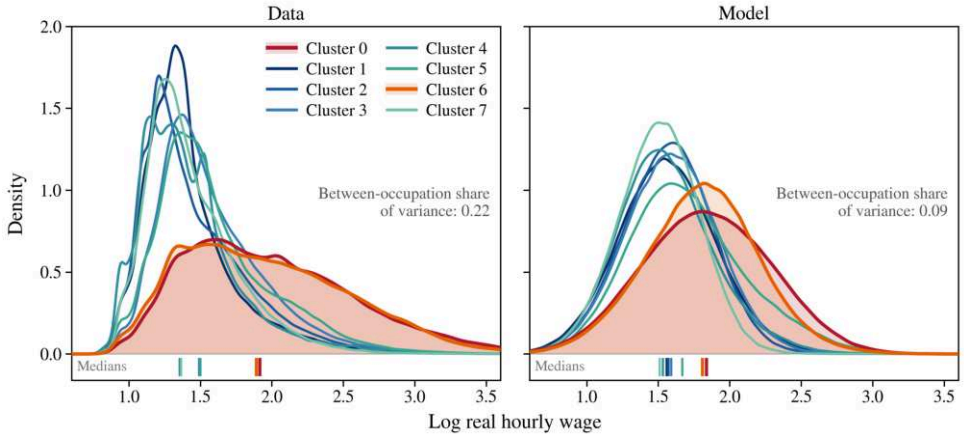
Wage–tenure profiles of stayers and mobility by tenure. Because wages reflect realized rather than expected output, within-occupation wage growth in the model comes from the accumulation of task-specific experience and from the selection of workers who stay. Panel (a) of Figure 30 follows workers observed from entry who remain without interruption in their entry occupation, and plots their mean log wage relative to the entry mean of all entrants. The model tracks the data closely over the first ten years and overshoots thereafter ($+0.49$ against $+0.43$ log points after fifteen years). The dotted line isolates the structural return to task experience of the same workers in the model, $+0.31$ log points after fifteen years; the difference from the model's profile is the selection of stayers, who are increasingly those whose wage signals confirmed a good match. The within-worker wage growth of stayers, which nets out the entry wage, is $+0.29$ log points after fifteen years in both the data and the model. Panel (b) reports the probability of switching occupation by years of occupational tenure, which complements the hazard by experience of Figure 9. The model reproduces the decline of mobility with tenure, but the decline is flatter than in the data: the model underpredicts switching in the first year of a spell (13% against 19%) and over-predicts it after ten years of tenure (5% against 3%), since the single switching cost κ_0 does not depend on tenure.

E.2 Additional Model-versus-Data Figures

Figures 31 and 32 report additional model-versus-data moments of occupational mobility by within-occupation match quality, and Figures 33 and 34 decompose the effect of learning on mobility into experimentation and passive learning.

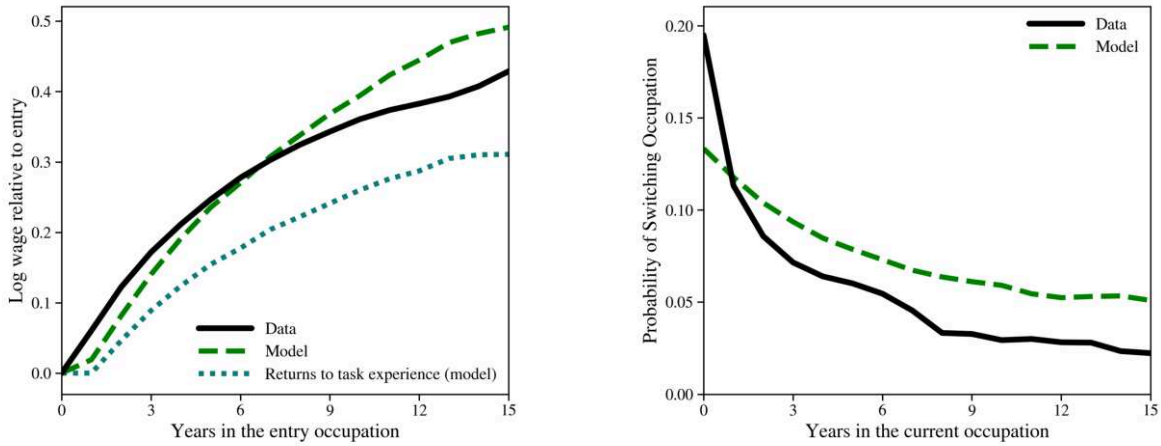


(a) Quantile–quantile plots by occupation



(b) Overlap of the occupation wage distributions

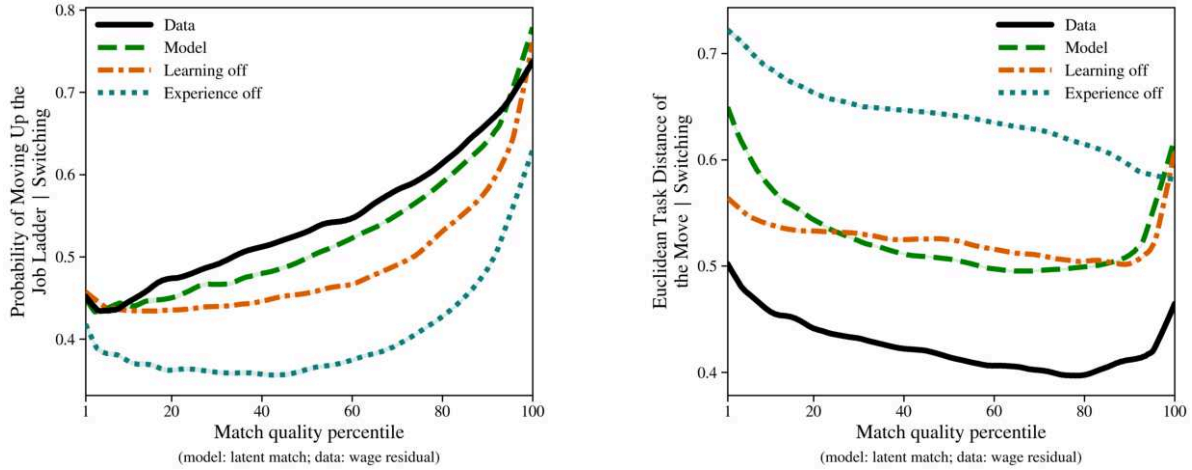
Figure 29: Wage distributions by occupation over all worker-years, data and model. *Panel (a)*: model quantiles against data quantiles of the log real hourly wage in each occupation cluster, from the 1st to the 99th percentile; the dashed line is the 45° line, and the gaps between the model and the data at the 10th, 50th, and 90th percentiles are reported in each panel. *Panel (b)*: densities of the log real hourly wage of the eight occupation clusters overlaid, in the data (left) and in the model (right); the two most cognitive-intensive occupations, Clusters 0 and 6, are shaded, the ticks below the horizontal axis mark the median of each occupation, and each panel reports the share of the variance of log wages explained by occupation dummies.



(a) Wage–tenure profile of occupational stayers

(b) Probability of switching by occupational tenure

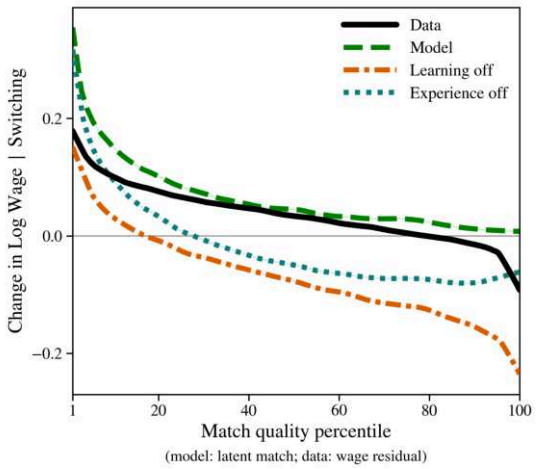
Figure 30: Wage growth of stayers and mobility by tenure, data (black, solid) and model (green, dashed). *Panel (a)*: mean log wage, relative to the mean entry wage of all entrants, of workers observed from entry who are still in their entry occupation after the number of years on the horizontal axis; the dotted line is the structural return to task experience of the same workers in the model. The model's stayers are reweighted to the education mix of the entrants in the data. *Panel (b)*: probability of switching occupation between t and $t+1$ by years of tenure in the current occupation.



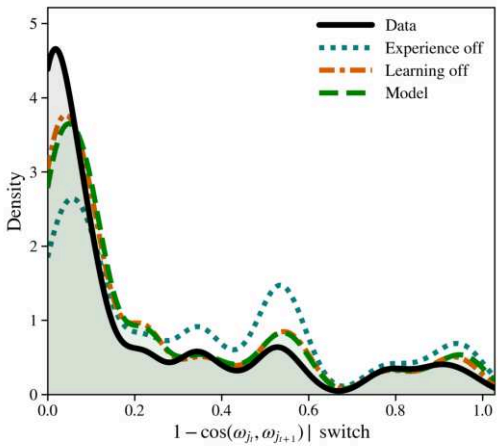
(a) Probability of moving up the job ladder

(b) Euclidean task distance of the switch

Figure 31: Additional model-versus-data moments of occupational mobility by within-occupation match-quality percentile (model: latent match quality; data: worker–occupation fixed effect), in the data (black), estimated model (green), *learning off* (orange), and *experience off* (teal). *Panel (a)*: probability that a switch is to an occupation with a higher mean wage rank (moving up the job ladder). *Panel (b)*: Euclidean task distance of the switch, $\|\boldsymbol{\omega}_{j_{t+1}} - \boldsymbol{\omega}_{j_t}\|_2$.

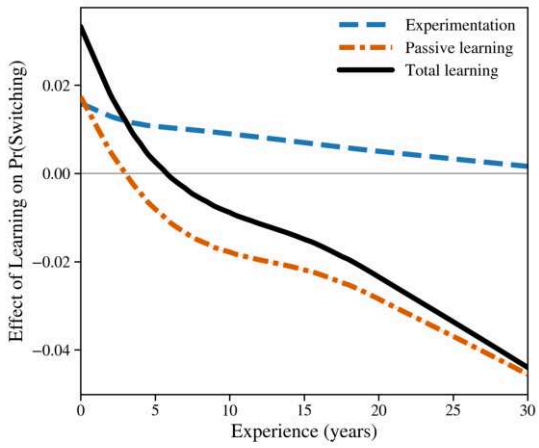


(a) Change in log wage upon switching

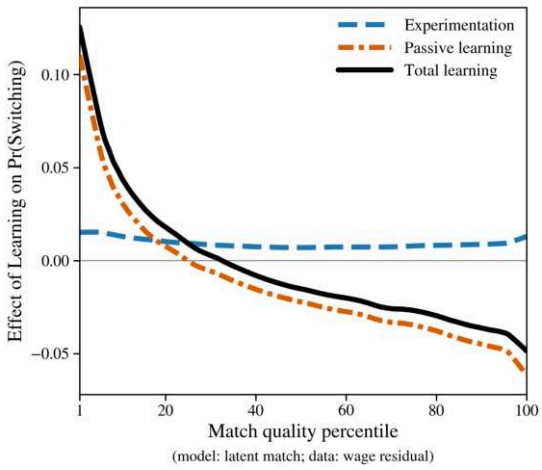


(b) Distribution of the task-content distance of switches

Figure 32: *Panel (a)*: change in log wage upon switching occupation, by within-occupation match-quality percentile (model: latent match quality; data: worker–occupation fixed effect). *Panel (b)*: distribution of the task-content distance of switches, $1 - \cos(\omega_{j_t}, \omega_{j_{t+1}})$, computed on the task weights of the occupation clusters for both the data and the model, conditional on switching. In both, the data (black), estimated model (green), *learning off* (orange), and *experience off* (teal).

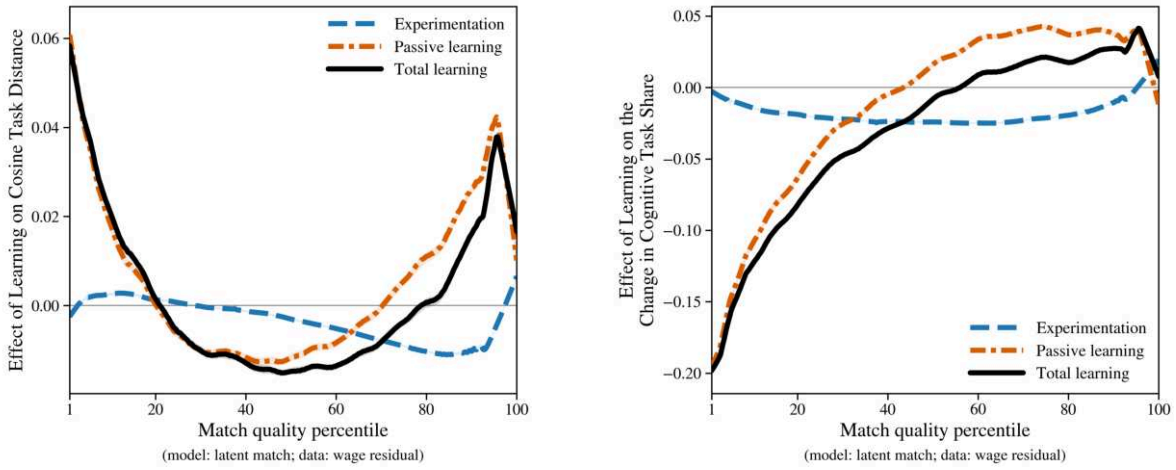


(a) Probability of switching by experience



(b) Probability of switching by match quality

Figure 33: Decomposition of the total effect of learning into experimentation and passive learning. In each panel the total learning effect (black, the gap between the estimated model and the *learning off* counterfactual) is split into the contribution of *experimentation* (blue, the option value of trying occupations to acquire information about comparative advantage) and *passive learning* (orange, the updating of beliefs from realized wage signals). *Panel (a)*: probability of switching by years of experience. *Panel (b)*: probability of switching by within-occupation match-quality percentile.



(a) Task-content distance of the switch

(b) Change in cognitive task content

Figure 34: Decomposition of the total effect of learning into experimentation and passive learning, for the task content of switches by within-occupation match-quality percentile, as in Figure 33: total learning effect (black), *experimentation* (blue), and *passive learning* (orange). *Panel (a)*: task-content distance of the switch, $1 - \cos(\boldsymbol{\omega}_{j_t}, \boldsymbol{\omega}_{j_{t+1}})$, computed on the task weights of the occupation clusters for both the data and the model. *Panel (b)*: change in the cognitive task share, $\Delta\omega_C$.

F Appendix: Additional Results and Counterfactuals

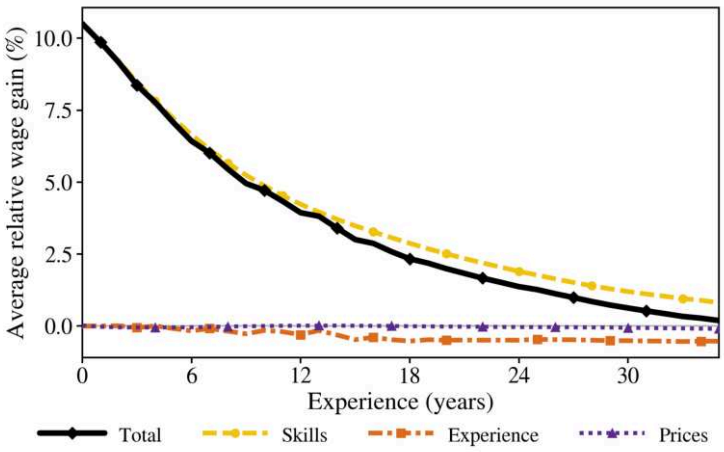


Figure 35: Decomposition of the life-cycle wage gain from removing information frictions (Figure 11, Panel (a)) into channels: Total (black), Skills (yellow, dashed), Experience (orange, dash-dotted), and Prices (purple, dotted). Horizontal axis: years of potential experience; vertical axis: average relative wage gain (percent).

F.1 Learning, Human Capital Accumulation and Wage Growth

This appendix uses the model to assess the importance of different channels in generating wage growth over the life cycle, using the counterfactual choice rules of Appendix D.6. I simulate the model in partial equilibrium, sequentially shutting down learning and/or human capital accumulation, to quantify the importance of each component in generating wage growth. In the model, skills are permanent, but learn-

ing allows workers to improve their occupational sorting; therefore, higher wages are achieved by switching occupations over the life cycle. Let $G_{h\ell}(e)$ denote cumulative log wage growth from labor-market entry to experience e in the economy in which the human-capital channel is active if $h = 1$ and the learning channel if $\ell = 1$; thus G_{11} is the estimated model, G_{00} the baseline in which both channels are shut down, G_{10} the human-capital-only economy, and G_{01} the learning-only economy. I can decompose cumulative wage growth as

$$G_{11}(e) = \underbrace{G_{00}(e)}_{\text{no HC, no learning}} + \underbrace{\big[G_{10}(e) - G_{00}(e)\big]}_{\text{human capital}} + \underbrace{\big[G_{01}(e) - G_{00}(e)\big]}_{\text{learning}} + \underbrace{\big[G_{11}(e) - G_{10}(e) - G_{01}(e) + G_{00}(e)\big]}_{\text{HC–learning interaction}}. \tag{87}$$

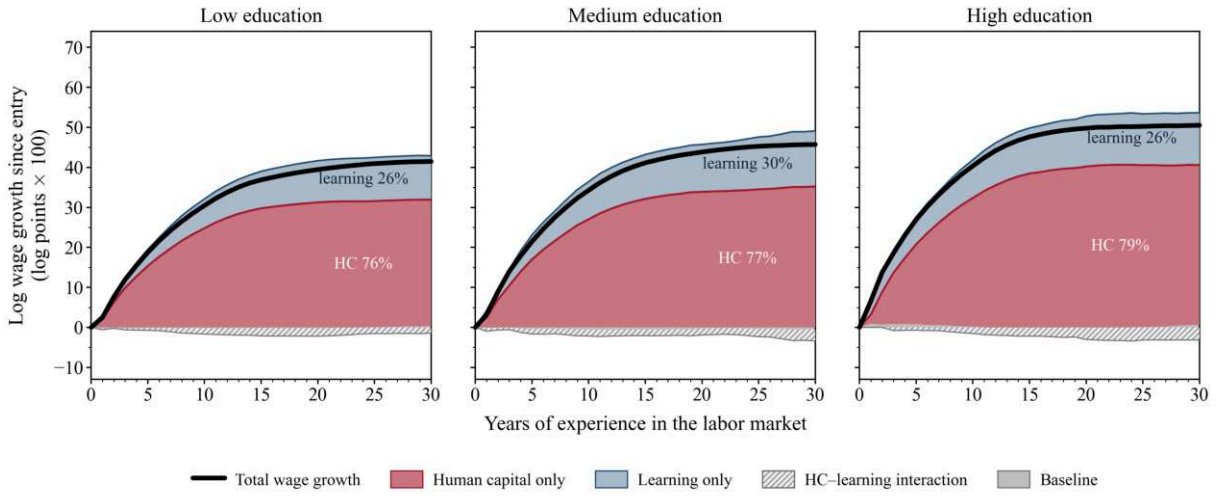


Figure 36: Decomposition of cumulative log wage growth since labor-market entry (vertical axis, log points ×100) by years of experience (horizontal axis), separately for low, medium, and high education. The total (black) is decomposed into the contribution of task-specific human capital (red), of learning about comparative advantage (blue), and their interaction (hatched).

In Figure 36, I plot the wage decomposition separately for each education group. Across groups, learning drives approximately one fourth of total wage growth, and experience accumulation drives the remaining three quarters. Besides starting their careers with higher wage levels, more educated workers see their wages grow more and faster over the life cycle. More educated individuals start their careers in more cognitive-intensive jobs, which have higher and steeper returns to tenure (Deming 2025). This decomposition exercise also shows how learning and human capital accumulation interact in the baseline model. As suggested by the growing interaction term in the decomposition plots, average wage growth in the estimated economy is lower than the sum of the wage growth achieved in the economies with only learning and with only human capital accumulation. This implies that human capital accumulation, while generating wage growth within occupations, in fact prevents some workers from eventually achieving the higher wages that they would achieve in an economy

without experience accumulation. This accumulation channel implicitly generates a cost and makes workers less willing to switch occupations later in their careers.

I compute the decomposition in (87) for the economy with the change in tasks induced by AI, and I report the results in Figure 37. AI increases the importance of learning, particularly for high-educated workers, and decreases the importance of human capital accumulation for low- and medium-educated workers: on average, the share of life-cycle wage growth due to human capital accumulation falls from about 79% to about 66%, that is, by about 13 percentage points, while for highly educated workers the negative interaction between learning and human capital widens. In the pre-AI economy, cognitive tasks were an important driver of the returns to experience, both through direct returns and through the accumulation function. In the post-AI economy, cognitive experience is accumulated more slowly and cognitive tasks are less important. At the same time, manual skills become more important and, particularly for highly educated workers, so does learning: given the estimated skill distribution, highly educated workers enter the labor market believing that they have a comparative advantage in performing cognitive tasks. These initial beliefs are particularly penalizing for workers with more mixed skills or higher manual skills, who benefit substantially from learning their comparative advantage and reallocating accordingly.

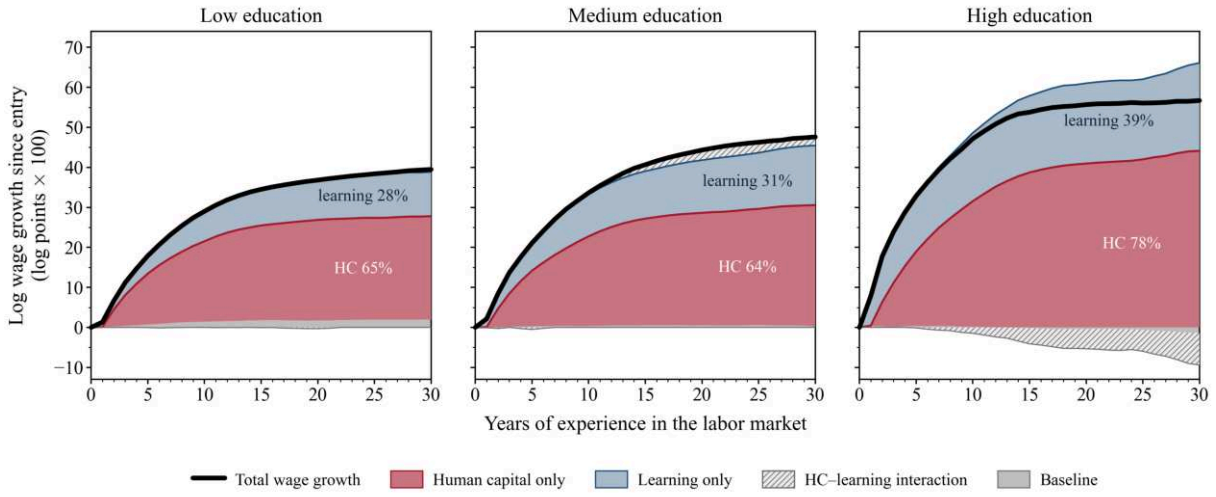
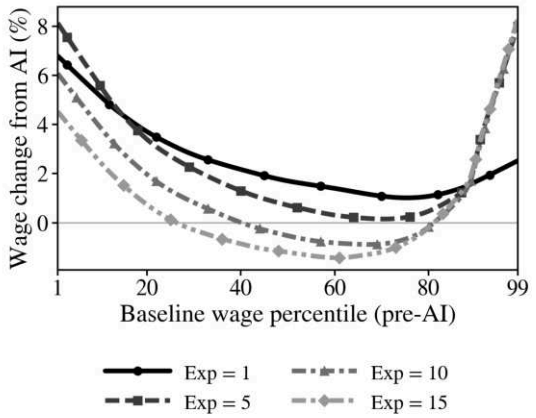


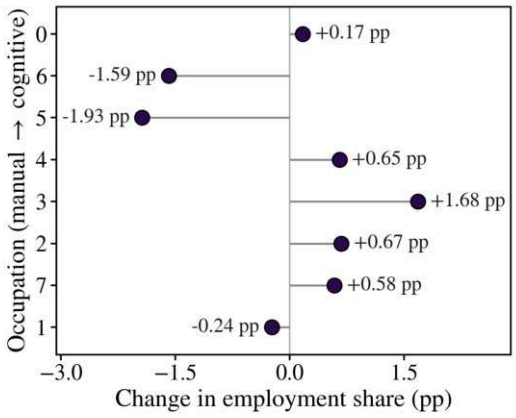
Figure 37: Decomposition of cumulative log wage growth since labor-market entry in the post-AI stationary equilibrium (vertical axis, log points $\times 100$) by years of experience (horizontal axis), separately for low, medium, and high education. The total (black) is decomposed into the contribution of task-specific human capital (red), of learning about comparative advantage (blue), and their interaction (hatched).

F.2 Welfare Measure

This appendix defines the welfare measure used in Sections 6 and 7 and reports it for every counterfactual. For each simulated worker-period (i, t) the simulation stores the ex-ante value $EV(\boldsymbol{x}_{it})$ of equation (28): the expected discounted sum of the worker's remaining flow utilities, log wages plus amenities minus switching costs, integrated



(a) Wage gains from AI by baseline wage percentile



(b) Reallocation by task content of the occupation

Figure 38: *Panel (a)*: percent wage change from AI by baseline (pre-AI) wage percentile, for workers with different years of experience (separate lines). *Panel (b)*: change in each occupation cluster's employment share between the pre-AI and post-AI stationary equilibria, in percentage points; clusters are ordered from the most manual-intensive (bottom) to the most cognitive-intensive (top). Employment moves out of the cognitive-intensive Clusters 5 and 6 and into the mixed and manual-intensive clusters, mostly Cluster 3.

over future wage signals, taste shocks, and the worker's own future choices. Because the switching cost enters the inclusive value, EV already nets out the utility cost of every future move. Aggregate welfare in an economy is the average of this value over the stationary cross-section of alive workers,

$$\mathcal{W} = \frac{1}{|\mathcal{S}|} \sum_{(i,t)\in\mathcal{S}} EV(\boldsymbol{x}_{it}), \tag{88}$$

where $\mathcal{S}$ collects the worker-periods of the stationary simulation after the burn-in. Since the cross-section contains every stage of the career in its stationary proportions, $\mathcal{W}$ is the expected remaining lifetime value of a randomly drawn worker.

Utility is logarithmic in the wage, so multiplying every wage in every period by $1 + \Delta^{\text{CEV}}$ adds $\log(1 + \Delta^{\text{CEV}})/(1 - \beta)$ to $\mathcal{W}$. The consumption-equivalent variation of a counterfactual economy relative to a baseline is the Δ^{CEV} that equalizes the two,

$$\Delta^{\text{CEV}} = \exp\left((\mathcal{W}^{\text{cf}} - \mathcal{W}^{\text{base}})(1 - \beta)\right) - 1, \tag{89}$$

that is, the uniform and permanent wage increase that would make the average worker in the baseline economy as well off as the average worker in the counterfactual. The measure prices everything that enters workers' utility: expected wages, and therefore skills, task-specific experience, and prices; amenities; the switching cost; and the option value of the taste shocks. It does not price the exit hazard, which workers do not internalize. For the information interventions delivered after entry, the signal is unanticipated, so the value of the entering cohort is unchanged by construction; I therefore report the population measure throughout.

Table 6 reports the output and welfare gains of each counterfactual. For every counterfactual that changes workers' information, welfare moves in the same direction as output, with welfare gains between roughly a quarter and a half of the output gains. The welfare gain is smaller than the output gain partly because the consumption-equivalent variation nets out the switching costs and amenity losses that re-sorting workers incur, which output does not count, and partly because output is a level of production whereas the consumption-equivalent variation is a wage-equivalent averaged across workers. Figures 40 and 41 report the welfare counterparts of the information-provision, life-cycle, and transition figures of the main text. Along the AI transition, welfare rises by 0.8% on impact and by 1.1% in the long run, and the skill signal delivered to every new cohort raises the long-run gain to 1.5% and closes 17% of the long-run welfare gap to the perfect-information economy, against 20% of the output gap. The two measures diverge in the switching-cost experiment (Figure 39), because the switching cost is a utility cost that the counterfactual removes directly (Section 6.3), and in the forced-exploration exercise of Appendix F.4, in which output rises but the entering cohort bears the forced assignment.

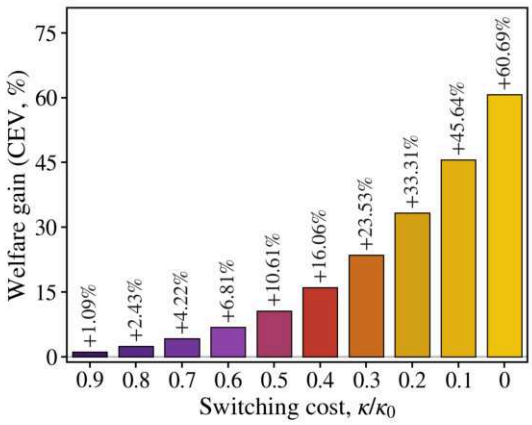
Table 6: Output and Welfare Effects of the Counterfactuals

Counterfactual	Comparison	Output gain (%)	Welfare gain (%)
Perfect information	vs. baseline	5.10	2.03
Skill signal at entry	vs. baseline	0.97	0.32
Skill signal after 2 years	vs. baseline	0.52	0.22
Skill signal after 6 years	vs. baseline	0.23	0.11
Skill signal after 10 years	vs. baseline	0.12	0.06
Perfect information at $\kappa/\kappa_0 = 0.7$	vs. same κ	3.32	1.07
Perfect information at $\kappa/\kappa_0 = 0.5$	vs. same κ	2.52	0.77
Perfect information at $\kappa/\kappa_0 = 0$	vs. same κ	1.62	0.45
Switching cost $\kappa/\kappa_0 = 0.7$	vs. baseline	1.22	4.22
Switching cost $\kappa/\kappa_0 = 0.5$	vs. baseline	0.82	10.61
Switching cost $\kappa/\kappa_0 = 0.3$	vs. baseline	−0.48	23.53
Switching cost $\kappa/\kappa_0 = 0$	vs. baseline	−2.39	60.69
Forced exploration	vs. baseline	1.11	−0.03
AI shock	vs. pre-AI baseline	2.08	1.05
Perfect information, post-AI	vs. post-AI baseline	5.84	2.48
Skill signal at entry, post-AI	vs. post-AI baseline	1.11	0.40

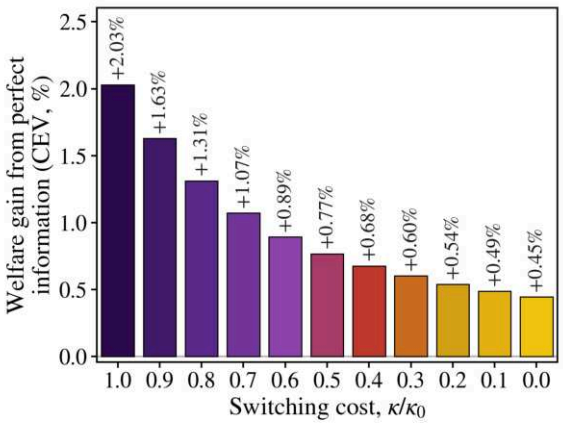
Notes: Output gain is the percent change in CES aggregate output between the two stationary equilibria. Welfare gain is the consumption-equivalent variation of the counterfactual relative to the comparison economy, in percent. All economies are simulated with $N = 200{,}000$ workers for 400 periods, discarding the first 200 periods.

F.3 Switching Costs: Life-Cycle Wage Profiles

This appendix reports the life-cycle wage profiles behind the switching-cost counterfactuals of Section 6.3, which the main text summarizes by match-quality percentile.

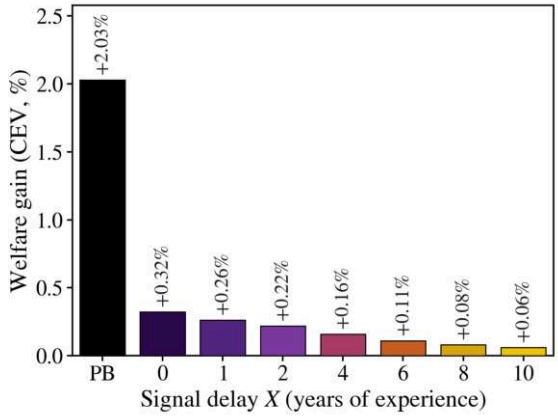


(a) Welfare gain from reducing the switching cost

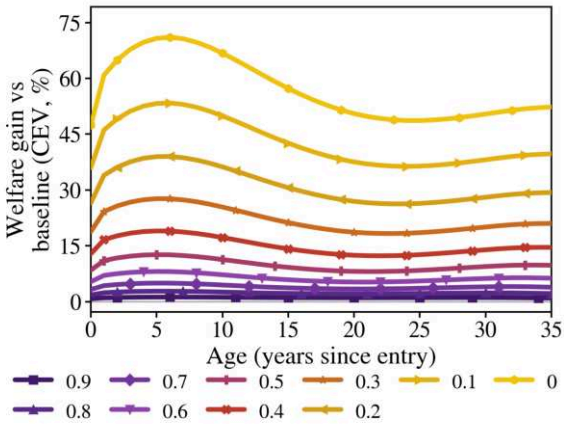


(b) Welfare gain from perfect information, by κ/κ_0

Figure 39: Welfare effects of switching costs and information frictions, the welfare counterparts of Panel (a) of Figures 16 and 15. *Panel (a)*: consumption-equivalent welfare gain of each imperfect-information economy relative to the baseline, by the switching cost as a fraction of its estimated value. *Panel (b)*: consumption-equivalent welfare gain from moving to perfect information at each switching-cost level, both economies solved at the same κ/κ_0.

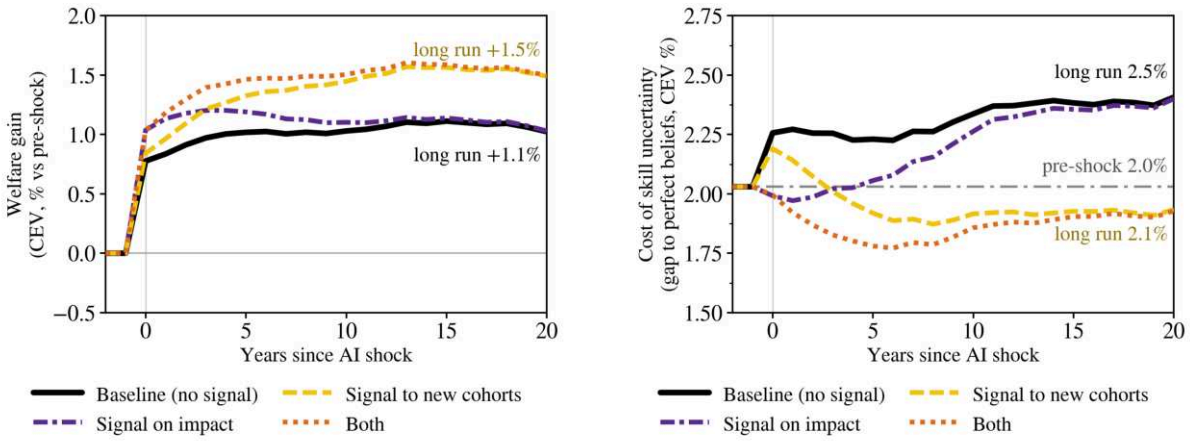


(a) Welfare gain from the skill signal, by delay



(b) Welfare gain from reducing the switching cost, by age

Figure 40: Welfare counterparts of Figure 13 and of Panel (b) of Figure 42. *Panel (a)*: consumption-equivalent welfare gain relative to the baseline economy when the calibrated skill signal is delivered X years after labor-market entry, and under perfect information. *Panel (b)*: consumption-equivalent welfare gain relative to the baseline economy by years since entry, for each switching cost κ/κ_0; the value at age zero is the welfare gain of the entering cohort.



(a) Welfare along the transition

(b) Welfare gap to the perfect-information economy

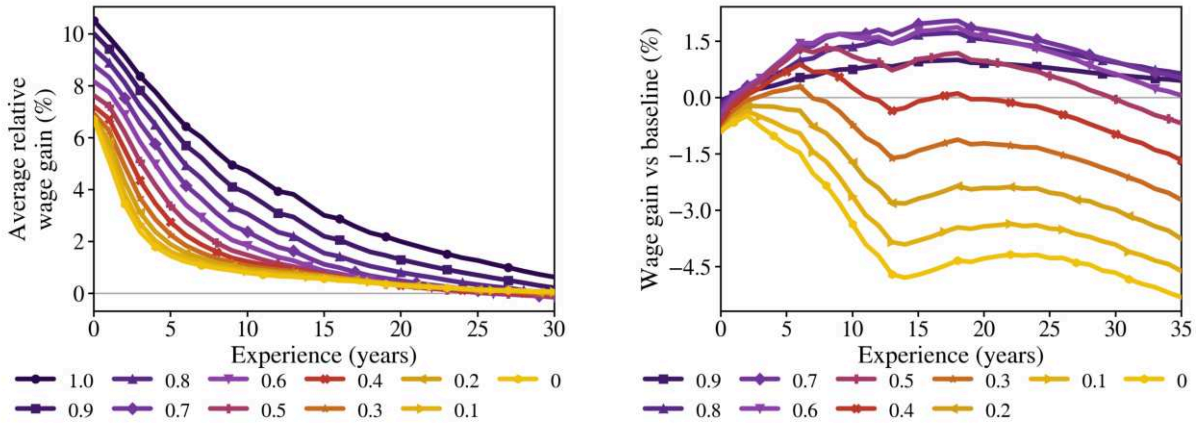
Figure 41: Welfare counterparts of Figure 19, by years since the AI shock (horizontal axis). Baseline imperfect-information economy with no signal (black, solid); skill signal delivered to all workers on impact (purple, dash-dotted); skill signal delivered to every new cohort at entry (yellow, dashed); both (orange, dotted). *Panel (a)*: consumption-equivalent welfare gain of the cross-section alive in each period relative to the pre-shock stationary economy. *Panel (b)*: consumption-equivalent variation that the cross-section alive in each period would need to be as well off as in the perfect-information economy in the same period; the dash-dotted gray line marks its pre-shock level. Labels report the pre-shock and long-run values.

Figure 42 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 life-cycle counterpart of Figure 15), and Panel (b) is the gain relative to the estimated baseline economy (the life-cycle counterpart of Figure 16). In Panel (b), the wage gain of reducing switching costs gradually becomes more negative over the life cycle: with lower switching costs, taste-shock-driven churning makes workers repeatedly restart their task-specific experience, and these losses compound over the career; when the reduction is extreme, wage losses materialize almost immediately.

F.4 Forced Exploration at the Beginning of the Career

The switching-cost counterfactual (Section 6.3) shows that frictions to mobility are a key driver of persistent mismatch, but even with low switching costs the baseline information structure limits how quickly workers can sort optimally. A complementary policy question is whether a targeted intervention at the very start of the career can accelerate learning and improve lifetime sorting and wages. I study a one-period "forced exploration" policy: in their first period, workers are assigned to the occupation that maximizes the information gain about their latent skills. I measure the learning value of occupation j for entrants in group g as the mutual information between skills and the first wage signal,

$$IG_{jg}^{\text{entry}} \equiv I(\boldsymbol{\theta}_i; s_{ij1} \mid g) = \frac{1}{2}\log\left(1 + \frac{\boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j}{\varsigma_{\varepsilon j}^2}\right), \tag{90}$$



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

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

Figure 42: Life-cycle 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.

where $\boldsymbol{\lambda}_j$ are the skill loadings of occupation j, $\boldsymbol{\Sigma}_g$ is the prior skill covariance of group g, and $\varsigma^2_{\varepsilon j}$ 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[34]. Each worker is assigned to $\arg\max_j IG^{\text{entry}}_{jg}$ in their first period. From period 2 onward they choose freely under their updated beliefs. The assignment involves no transfers: 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 cognitive anchor (Cluster 0) and, for a smaller share of workers, the manual anchor (Cluster 1). This is intuitive: the anchors load on a single skill dimension, so their wage is a noisy signal of a single skill and therefore more informative than the wage of an occupation that mixes skills. Figure 44 shows the occupational shares at entry, after two years, and after ten years, under forced exploration and in the baseline.

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

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

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

where $h(Y) = \mathbb{E}[-\log f_Y(Y)]$ denotes differential entropy. Under the model's linear-Gaussian signal structure, $s_{ij1} = \boldsymbol{\lambda}_j'\boldsymbol{\theta}_i + \varepsilon_{ij1}$, $\boldsymbol{\theta}_i \mid g \sim N(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, $\varepsilon_{ij1} \sim N(0, \varsigma^2_{\varepsilon j})$, the signal is scalar Gaussian with unconditional variance $\boldsymbol{\lambda}_j'\boldsymbol{\Sigma}_g\boldsymbol{\lambda}_j + \varsigma^2_{\varepsilon j}$ and conditional variance $\varsigma^2_{\varepsilon j}$. Since the entropy of a scalar Gaussian $N(m, v)$ is $h(Y) = \frac{1}{2}\log(2\pi e\, v)$, it follows that

$$I(\boldsymbol{\theta}_i; s_{ij1} \mid g) = \frac{1}{2}\log\Big(2\pi e\left[\boldsymbol{\lambda}_j'\boldsymbol{\Sigma}_g\boldsymbol{\lambda}_j + \varsigma^2_{\varepsilon j}\right]\Big) - \frac{1}{2}\log\Big(2\pi e\,\varsigma^2_{\varepsilon j}\Big) = \frac{1}{2}\log\left(1 + \frac{\boldsymbol{\lambda}_j'\boldsymbol{\Sigma}_g\boldsymbol{\lambda}_j}{\varsigma^2_{\varepsilon j}}\right).$$

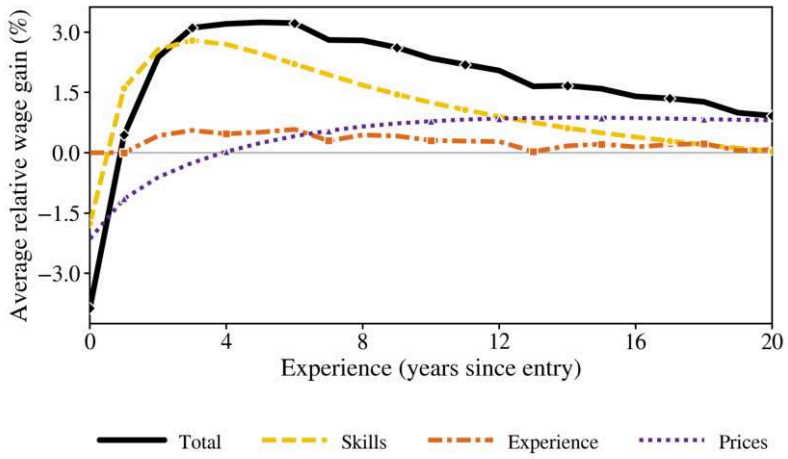


Figure 43: Life-cycle wage gain from forced exploration relative to the baseline economy, decomposed into channels: Total (black), Skills (yellow, dashed), Experience (orange, dash-dotted), and Prices (purple, dotted). Horizontal axis: years of potential experience since entry; vertical axis: average relative wage gain (percent).

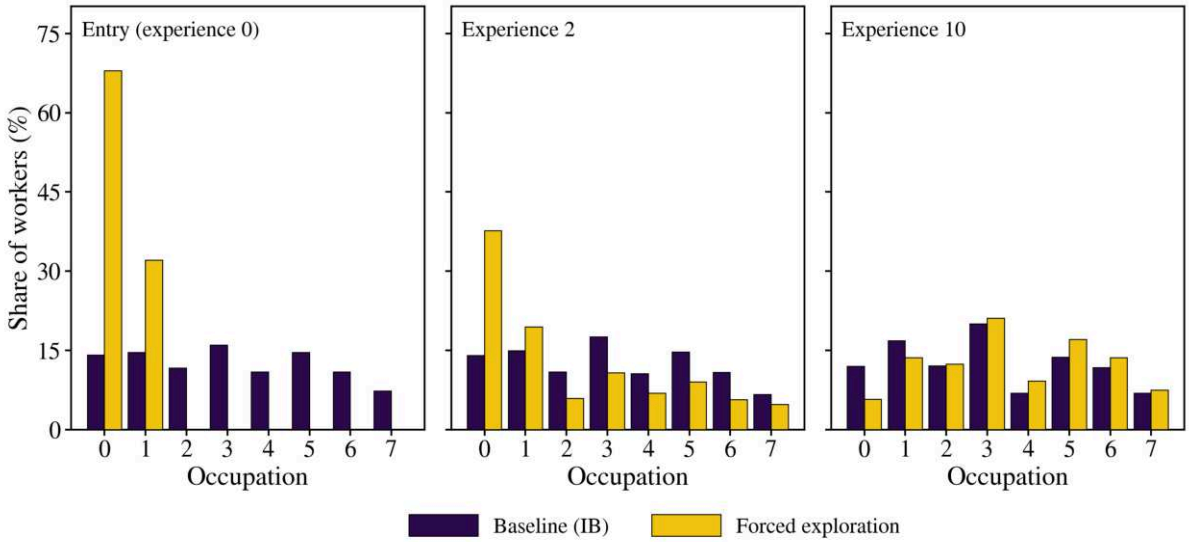


Figure 44: Share of workers in each occupation cluster (vertical axis) under forced exploration (yellow) and in the baseline imperfect-information economy (dark purple), at entry, after two years, and after ten years of experience.

line, where entrants are spread across all eight occupations (Figure 44). This forced bunching triggers a strong general-equilibrium price response: excess labor supply in the informative occupations drives their prices down while prices in the other occupations rise. This drop in prices produces a sharp wage loss on impact, about 3.9% at entry (Figure 43). From the second period workers choose freely, and they reallocate quickly to occupations that better match their comparative advantage in tasks. In fact, the Skills channel jumps to about +2.6% within two years and stays above the baseline for most of the career, suggesting that in this counterfactual workers are indeed better matched on skills. The total wage gain turns positive after the first year, reaching about +3.2% by the third year and staying close to that level until about the sixth year. As

shown in Figure 44, the occupational distribution converges back toward the baseline distribution but workers are now better matched within each occupation. The net result is that the later gains more than offset the short-run wage loss: aggregate output in the exploration economy is 1.11% higher than in the baseline. The welfare effect is instead negligible: the population consumption-equivalent gain is -0.03%, and the entering cohort, which bears the forced assignment, loses 6.4%, because the assignment overrides the occupation these workers would have chosen (Appendix F.2). These results highlight a key asymmetry in the value of early information. This counterfactual experiment suggests that a policy that temporarily redirects workers toward their most informative occupation upon entering the labor market can be productivity-enhancing later in the career, at the cost of short-run wage losses.

Appendix References

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

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

Boyd, III, J. H. (1990), 'Recursive utility and the Ramsey problem', *Journal of Economic Theory* **50**(2), 326–345.

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

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

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

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

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

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

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

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

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

Judd, K. L., Maliar, L., Maliar, S. and Valero, R. (2014), 'Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain', *Journal of Economic Dynamics and Control* **44**, 92–123.

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

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

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