The Micro and Macro Implications of Early Career Skill Mismatch*

Piero De Dominicis†

March 26, 2026

Abstract

What are the consequences of early-career skill mismatch for individual wages and aggregate productivity? To answer this question, I develop a general equilibrium dynamic Roy model in which workers have imperfect information about their multidimensional skills and accumulate task-specific human capital on the job. Estimating the model on Portuguese administrative data, I find that early-career mismatch driven by information frictions generates a 6.7% loss in aggregate productivity. Wage losses from mismatch are largest early in the career, but persist over the lifecycle because of switching costs and imperfect transferability of human capital. I then study policy counterfactuals aimed at reducing this persistent mismatch. Reducing switching costs lowers the aggregate cost of mismatch, while encouraging early-career experimentation increases wages later in life, at the cost of lower wages initially. Finally, I examine how artificial intelligence affects the cost of mismatch. I find that while AI raises aggregate output overall, it also increases the cost of mismatch, raising the aggregate productivity loss from 6.7% to 9.5%.

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

*I am particularly 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, Vasco Carvalho, Geraud Desazars, Rafael Dix-Carneiro, Jan Eckhout, Pamela Giustinelli, Joseph-Simon Görlach, Cristina Gualdani, John Grigsby, Simon Jäger, Francis Kramarz, Arnaud Maurel, Romuald Meango, Robert Miller, Andreas I. Mueller, Elena Pastorino, Nicola Pavoni, Marta Prato, Richard Rogerson, Aysegül Sahin, Karthik Sastry, and Gianluca Violante, as well as seminar participants at Bocconi, Princeton Junior Macroeconomics Workshop, 1st Aarhus Workshop on Labor Markets, European Winter Meeting of the Econometric Society 2025, CEPR Paris Symposium 2025 and BeNA Labor Economics Workshop 2025 for helpful comments. All errors are my own.

†Department of Economics, Bocconi University.

1 Introduction

Workers do not possess a single, one-dimensional skill but instead hold a portfolio of skills whose payoffs depend on the job that employs them. These latent, multidimensional skills help explain a wide range of labor-market outcomes, including earnings trajectories and occupational choice (Heckman et al. 2006; Adda and Dustmann 2023). However, workers—especially early in their careers—may have imperfect information about their skills and may therefore direct their search towards jobs that do not represent the right match given their comparative advantage (Kiss et al. 2023). A large literature proposes a process of experimentation and learning-by-doing, in which workers gradually discover where their particular skill mix is most productive while accumulating imperfectly transferable experience (Miller 1984; James 2012; Antonovics and Golan 2012; Papageorgiou 2014; Pastorino 2024). Over the life cycle, workers' comparative advantage may evolve as jobs are an important source of human capital accumulation: workers gain experience along different dimensions, especially when required to perform different types of tasks on the job (Gathmann and Schönberg 2010). Learning, imperfectly transferable human capital, and switching costs interact to create strong path dependence that prevents a substantial fraction of workers from ever sorting into their ideal career path.

In this paper, I develop and estimate a general equilibrium dynamic Roy model to assess and quantify the consequences of early-career mismatch for individual labor market outcomes and aggregate output. In the model, workers have unobserved multidimensional skills and accumulate task-specific experience over time. Skills and experience are complementary inputs in production and are jointly used to perform tasks. Although workers in each occupation perform all tasks, the productivity of each task and the degree of complementarity across tasks vary across occupations. Workers begin their careers with incomplete information about their true abilities and gradually refine their beliefs through noisy wage signals. Mismatch therefore arises endogenously, as workers sort across occupations based on beliefs rather than true comparative advantage. Task-specific experience, which accumulates at different rates across occupations, is also valued differently across jobs. Two key forces drive occupational mobility in the model: learning about unobserved multidimensional skills and the accumulation of task-specific human capital. Together, these forces shape both the timing and the direction of occupational transitions over the life cycle. I estimate the model with ten occupations and two task-specific skills[1], cognitive and manual. Using the estimated model, I show that the model replicates the main mobility patterns observed in the data. In particular, it matches two well-established facts, first documented by Groes et al. (2015): the U-shaped pattern of occupational switching within occupations, whereby workers with low and high match quality are more likely to move, and the directionality of those transitions, meaning that workers with low and high match quality within the same occupation sort into systematically different destination occupations. Learning is the key force behind this directionality and also explains why, conditional on switching, the size of the reallocation in task space depends on match quality in the origin occupation. Because workers choose occupations based on perceived comparative advantage and use wages to update beliefs about their multidimensional skills, both low and high wage realizations reveal mismatch along different dimensions and induce reallocation toward

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

occupations with different task content. Upward mobility is primarily associated with transitions toward more cognitive occupations, whereas downward mobility is mainly associated with moves toward more manual occupations.

I use the model to quantify the aggregate and individual costs of early career skills mismatch. Simulating a counterfactual stationary equilibrium where workers' beliefs coincide with their true skills from the start of their careers, I find that mismatch generate by information frictions about skills causes an aggregate productivity loss of 6.7%. These aggregate costs are largely driven by early career mismatch: on average, the wage losses are 7.5% at labor market entry, declining to 3.3% after twenty years as workers learn and change occupation accordingly. While mismatch is partially resolved over the lifecycle, the interaction of learning, imperfectly transferable human capital, and switching costs generates strong path dependence and prevents the sorting of workers into their ideal career paths. Nearly all of the aggregate costs come from improved sorting on skills rather than better experience accumulation. Under perfect information, 30% of workers would choose a different initial occupation and 50% would switch at some point, with average wage gains of 17.5% for those who reallocate. The costs are concentrated among specialists—workers with pronounced comparative advantage in one skill dimension—who have the most to gain from correct sorting. Generalists, whose skills are balanced, lose little from uncertainty. Across demographic groups, female workers experience larger wage gains than males at all education levels, with low-educated women benefiting the most.

I then explore policy counterfactuals that specifically target this path dependence. In a first policy experiment, I progressively reduce switching costs by 10% to 90% and re-solve the stationary equilibrium at each level. Two complementary findings emerge. First, the aggregate cost of mismatch—measured as the output gap between imperfect and perfect information—declines monotonically from 6.6% to 2.3%, revealing a key complementarity between information frictions and mobility frictions: when switching is cheap, workers can correct mismatches through experimentation even without knowing their true comparative advantage. Second, the direct effect on aggregate output is non-monotonic: output peaks at a 50% reduction (+2.4%) and then declines, because very low switching costs cause workers to switch more frequently in response to transitory taste shocks and amenities, destroying task-specific human capital. Switching costs thus play a dual role: they lock workers into poor matches (harmful) but also discourage unproductive churning (beneficial). In a second policy counterfactual experiment, I assign workers at labor market entry to the occupation that maximizes early information acquisition about their skills, and leave them free to choose from the second period onward. This experimentation policy generates persistent wage and utility gains with only short-run losses: workers quickly switch to better-matched occupations upon regaining free choice, and after ten years the occupational distribution mirrors the baseline while workers are substantially better matched to their true comparative advantage.

Finally, I use the model to study how artificial intelligence (AI) affects the individual and aggregate costs of early-career skills mismatch. I model AI as a task-biased technological change based on Large Language Model (LLM) exposure that shifts the task importance within occupations (Eloundou et al. 2024; Freund and Mann 2025; Althoff and Reichardt 2026). Because LLMs cannot perform physical tasks, the AI shock operates primarily through the cognitive dimension. Comparing the baseline and post-AI stationary equilibria, I find that AI overall

raises aggregate productivity but with heterogeneous wage effects: workers with strong manual comparative advantage gain as their skills become relatively more valuable, while cognitive specialists face worse outcomes as their primary skill dimension is exposed to automation. Crucially, AI amplifies the aggregate cost of mismatch: the productivity loss from information frictions rises from 6.7% to 9.5%, so workers under imperfect information capture only approximately 72% of the productivity gains that AI would deliver in an economy without information frictions. The higher wage losses are concentrated early in the career, when information frictions prevent workers from sorting into the occupations where their skills are most valuable under the new task structure, and the increase is largest for low- and medium-educated men, who have the most to gain from improved sorting when returns to manual skills rise.

2 Related Literature

This paper contributes mainly to the literature on learning, job assignment, and wage growth. Using the terminology of Pastorino (2024), learning affects wages through two distinct mechanisms. The first are the direct effects of learning[2], whereby wages respond to beliefs about ability because pay is an equilibrium outcome that depends on employers' (and/or workers') information. The second are the indirect effects of learning, whereby belief updating primarily affects wages by changing job assignment—that is, as workers and firms learn about workers' skills and comparative advantage, they reallocate workers to jobs (and, in my setting, occupations) where their skill bundle is used more productively. Pastorino (2024) argues—and finds quantitatively—that the indirect effects of learning through assignment can be the dominant force shaping wage growth and dispersion: while the direct belief-to-wage effects are relatively small, learning substantially improves sorting of workers to jobs, which then drives wage dynamics[3]. Motivated by this result, 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 choice and wages over the life cycle.

A related structural literature studies occupational sorting when workers learn about occupation-specific abilities. James (2012) uses a structural model similar to the one used in this paper, and estimates it following an algorithm described in James (2011). They investigate occupational sorting when workers learn about correlated occupation-specific abilities. Arcidiacono et al. (2025) studies how informational frictions impact schooling and work outcomes using a framework that allows for correlated learning between grades and abilities in different

[2]Much of the employer-learning literature emphasizes the direct effects channel, among others Harris and Holmström (1982), Farber and Gibbons (1996a), Altonji and Pierret (2001), Lange (2007a), 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 (1999b) and subsequent extensions (e.g. Gibbons and Waldman (2006), and the synthesis in Gibbons and Waldman (1999a)), as well as sectoral/occupational sorting frameworks in which learning affects wages both directly and through mobility/assignment (e.g., Gibbons et al. (2005a)).

[3]More recently, de Paula et al. (2025) build on Pastorino (2024) and propose a rich model with heterogeneous worker and firms, monopsony power and symmetric uncertainty and learning to reasses the importance of sorting in the labor market. They find that standard static method used to assess the importance of sorting in the labor market underestimate its importance as they ignore the dynamic effects of learning and human capital accumulation for future job matching.

occupations. Papageorgiou (2014) studies the role of search frictions in delaying optimal occupational matches in a model with learning and on-the-job search, concluding that workers sort on comparative advantage across occupations. More recently, de Melo and Papageorgiou (2025) develops a model of occupational choice with learning and partially transferable human capital, and shows that the worker's problem can be characterized and solved using Gittins indices. In these models, ability is occupation-specific, and therefore they are silent about which dimension underlies mismatch. Antonovics and Golan (2012) is the first to study the experimentation value of occupations and shows that occupational choice and wage growth in the NLSY are consistent with workers trading off optimal information acquisition and wage growth; however, their framework abstracts from important determinants of wage growth and occupational mobility, such as on-the-job human capital accumulation and non-wage amenities.

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. Motivated by this evidence, I estimate the model on administrative data and use it to quantify how learning, task content, and human capital jointly shape workers' career dynamics. The model also enriches occupational choice by allowing for non-compensating differentials, switching costs, and preference heterogeneity, which improves its quantitative fit and helps isolate the separate roles of learning and human capital accumulation in generating job mobility. Finally, I use the estimated model to perform counterfactual exercises and to analyze how technological shocks propagate through occupational mobility, wage dynamics, and the accumulation of human capital.

This paper also connects to the literatures on task-specific human capital and occupational mobility (Gibbons and Waldman 2004; Poletaev and Robinson 2008; Gathmann and Schönberg 2010). The model in this paper is similar in spirit to the task-based frameworks developed by Gathmann and Schönberg (2010) to study the transferability of human capital across occupations, and Autor and Handel (2013) who proposed a Roy (1951) model where workers sort based on comparative advantage across tasks. More recently, Freund and Mann (2025) use a model based on the canonical task-based theory (Acemoglu and Autor 2011; Acemoglu and Restrepo 2018) to study the effects of automation-driven job transformation. My framework is similar to theirs in spirit: tasks are bundled within occupations and technological change reallocates time spent across tasks within occupations. However, differently from these papers, I focus on the effects of technological change on workers' occupational allocation when there is imperfect information about skills. Some key contributions related to this paper are Guvenen et al. (2020), Lise and Postel-Vinay (2020), and Baley et al. (2022). Guvenen et al. (2020) measure multidimensional skill mismatch as the distance between workers' measured abilities (using test scores) and occupational skill requirements, but do not estimate a structural model. Lise and Postel-Vinay (2020) estimate a structural model of on-the-job search with multidimensional skill accumulation and depreciation, but workers have complete information about their skills—mismatch arises purely from search frictions. Baley et al. (2022) model learning about unobserved skills, but in their framework each job uses a single skill dimension and mismatch is defined as underqualification (the gap between a job's requirement and a worker's ability); they focus on business cycle dynamics rather than life-cycle occupational sorting.

In the model developed in this paper, a worker is "mismatched" when, given current beliefs, they choose (or remain in) an occupation that is not the one in which they are truly most productive—so mismatch reflects learning frictions and suboptimal assignment rather than a purely static gap between skill supply and task demand. The methodology I adopt does not require the econometrician to have a measure of the supply of skills (e.g., test scores); instead, I use the approach of Bunting et al. (2024) to recover the distribution of unobserved skills directly from the data. Finally, I am able to perform counterfactuals and estimate the wage cost of skill mismatch for heterogeneous workers.

A broader 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). Compared to these papers, I focus specifically on learning about multidimensional skills and task-specific human capital accumulation as the two key drivers of mobility and wage dynamics.

Finally, this paper contributes to the literature on estimating Roy-type models (Roy 1951). Building on the identification results of Heckman and Honore (1990), dynamic Roy frameworks have been applied to sectoral switching costs (Lee and Wolpin 2006), labour-market responses to trade liberalization (Dix-Carneiro 2014; Traiberman 2019), and learning about occupation-specific ability (James 2011; Hincapié 2020; Arcidiacono et al. 2025). These studies typically treat ability as a single latent factor or one scalar per occupation. I contribute by introducing a two-stage estimation strategy that accommodates multidimensional unobserved heterogeneity, building on the nonparametric identification results of Bunting et al. (2024) for what they call the *pure learning model*[4]. Two recent papers also study occupational reallocation in dynamic Roy models with task-specific human capital: Dvorkin and Monge-Naranjo (2019) and Althoff and Reichardt (2026). In both papers, workers have complete information about their skills, so occupational reallocation is always efficient. In contrast, workers in my model have imperfect information about their comparative advantage, which generates endogenous mismatch and shapes the dynamics of occupational reallocation. I use the model to perform counterfactual exercises that study how workers respond to changes in the returns to specific skills—such as those induced by artificial intelligence (AI)—when they must learn about their comparative advantage over time.

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

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

3 Data and Empirical Facts

3.1 Data

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

I use two related samples throughout the paper, each serving a distinct purpose. Detailed descriptions of the sample construction and their comparison are provided in Appendix C. 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 structurally estimating the learning and human capital parameters of the dynamic Roy model.

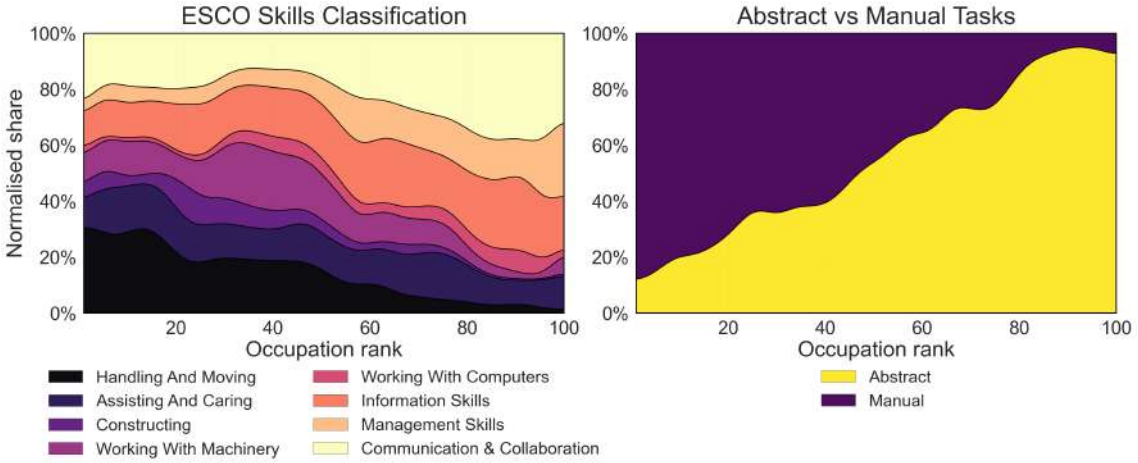
To measure the task content of occupations, I use two classifications. The first focuses on broad task categories: abstract and manual skills.[5] The second classification is ESCO, the European Skills, Competences, Qualifications and Occupations dataset, which links occupations to specific skills across European labor markets.[6] In both classifications, each occupation j receives weights ω_j^k over tasks/skills k that sum to one, $\sum_k \omega_j^k = 1$. Leveraging these measures allows for the comparison of workers within occupations and the analysis of how task content changes during occupational transitions.

To understand which skills are more important in the labor market, 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 abstract/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 abstract skills as wages rise. In the left panel, the shares of lighter-colored layers—communication & collaboration, management skills, information skills, and working with computers—increase with occupational wage rank. In contrast, the darker layers associated with more manual content—handling and moving, working with machinery, constructing, and assisting and caring—decline with occupational wage rank. The right panel confirms this pattern more starkly: abstract tasks rise with occupational wage rank while manual tasks decline.

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

[6] The ESCO skills considered include handling information, working with computers, constructing, management, and communication, among others.



(a) Task Composition by Wage Rank

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

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

3.2 The Task Content of Occupational Mobility

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

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

where $\boldsymbol{\omega}_j$ collects the skill weights of occupation j, and $\text{Sim}(j,j') \in [0,1]$ equals 1 for identical and 0 for orthogonal skill mixes.[7] The mean skills distance among actual occupation switchers is 0.52, with a standard deviation of 0.22. The closest common move is between labourers and building frame workers ($d = 0.21$), while the most distant observed switch is between mathematicians and domestic helpers ($d = 0.99$). The most popular switch for low-educated workers is between labourers and building frame workers, while for the high-educated it is between accounting clerks and bookkeepers ($d = 0.36$) or between physical science technicians and engineers ($d = 0.43$).[8] Figure 2 compares the observed distribution of skills distance

[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 5), I use the 8 aggregated ESCO skill dimensions. In addition, I use the Cognitive–Manual classification of Mihaylov and Tijdens (2019), which will serve as the task space for the structural estimation. Appendix Figure 24 shows that the patterns are robust across all granularity levels and hold for both mean and median task distance.

[8]See Appendix Table 7 for detailed summary statistics on the skills distance distribution, including the most common switches by education group.

against a random mobility counterfactual, where each switcher leaving occupation j in year t is assigned a destination drawn proportional to employment shares, following Gathmann and Schönberg (2010). Observed movers switch to occupations that are substantially closer in task space than random reallocation would predict (mean distance 0.52 vs. 0.68; Kolmogorov–Smirnov $p < 0.001$). This confirms that workers sort across occupations in a way that reflects the task content of jobs, not just their availability.

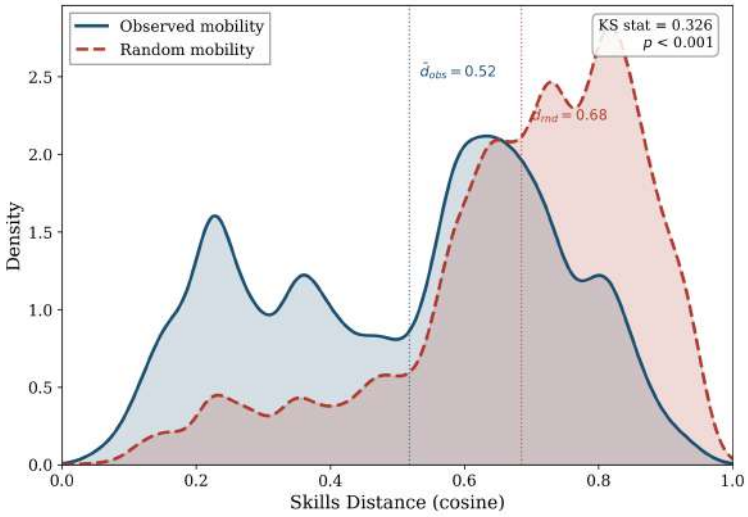
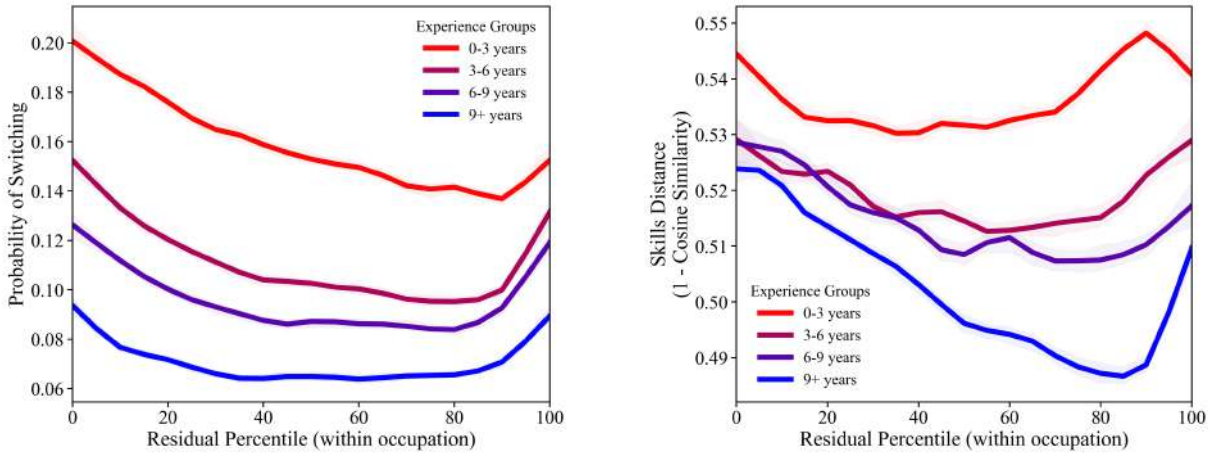


Figure 2: Distribution of skills distance for observed occupation switchers versus a random mobility counterfactual. For each switcher leaving occupation j in year t, a random destination is drawn with probability proportional to employment shares (excluding j), following Gathmann and Schönberg (2010). The random distribution is constructed from 200 Monte Carlo draws. The Kolmogorov–Smirnov test rejects equality of the two distributions ($p < 0.001$). Observed movers switch to occupations that are significantly closer in task space than what random reallocation would predict.

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

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

where α_j are occupation fixed effects, ϕ_t are time fixed effects, $f(\cdot)$ and $g(\cdot)$ are cubic polynomials in occupational tenure and total labor market experience, μ_{ij} is the unobserved worker-occupation match quality, and ν_{it} is mean-zero noise. The estimated residual is $\widehat{\epsilon}_{it} = \widehat{\mu}_{ij} + \widehat{\nu}_{it}$. Since ν_{it} is mean-zero, the percentile rank of $\widehat{\epsilon}_{it}$ within occupation j is positively correlated with μ_{ij}: workers with higher match quality tend to rank higher in the residual distribution. The "within-occupation residual percentile" used throughout this section is therefore a noisy but informative proxy for the worker's position in the match quality distribution of the current occupation. Appendix C shows that all patterns documented here are robust to using raw hourly wages instead of residualized wages.



(a) Probability of Switching by Experience

(b) Skills Distance for Occupation Switchers

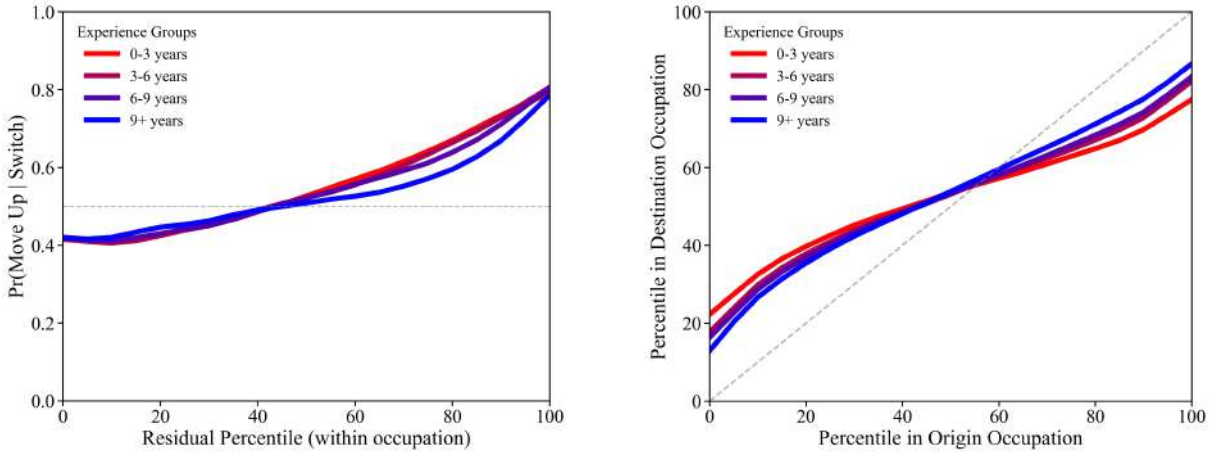
Figure 3: *Panel 3(a)*: Probability of switching occupations between t and $t+1$ (vertical axis) as a function of residualized within-occupation wage percentile (horizontal axis), for different experience groups. *Panel 3(b)*: Skills distance between origin and destination occupation conditional on switching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

Figure 3(a)[9] shows two facts about occupational mobility. First, workers at both ends of the within-occupation wage distribution switch at higher rates, resembling the "U-shape" documented by Groes et al. (2015) for Denmark. Second, mobility falls sharply with experience, from roughly 15% for workers with 0–3 years of experience to below 10% after 6 years. Mobility at the bottom is mainly driven by low-educated workers, while high-educated workers tend to rank higher in the residual wage distribution (see Appendix Figure 26). Figure 3(b) shows that the task distance between origin and destination occupations is related to the worker's match quality in the origin occupation. Workers with lower match quality make larger moves in task space, while workers with higher match quality make smaller moves. Task distance also falls with experience. This pattern is hard to reconcile with undirected search frictions, where the size of the move should be unrelated to the origin mismatch conditional on switching. As Guvenen et al. (2020) argue, if mismatch is induced by search frictions alone, the change in mismatch upon switching should be independent of the mismatch in the origin occupation. The relationship between match quality and the size of the task move is instead consistent with directed sorting, where workers learn about their comparative advantage and reallocate toward better-fitting occupations.

Occupational mobility in Portugal is also directional, using the terminology introduced by Groes et al. (2015). High-ability workers within occupation are more likely to move to an occupation with a higher average wage compared to their current occupation, while low-ability workers move to an occupation with a lower average wage (Figure 4). However, when moving, low-ability workers improve their relative position in the destination occupation, while high-ability workers worsen their relative position.

Figure 5 examines the task content of occupational transitions. Workers at different positions

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



(a) Directionality by Experience

(b) Rank-to-Rank by Experience

Figure 4: *Panel 4(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 4(b)*: Percentile position in destination occupation (vertical axis) against percentile position in origin occupation (horizontal axis) for occupational switchers, by experience group.

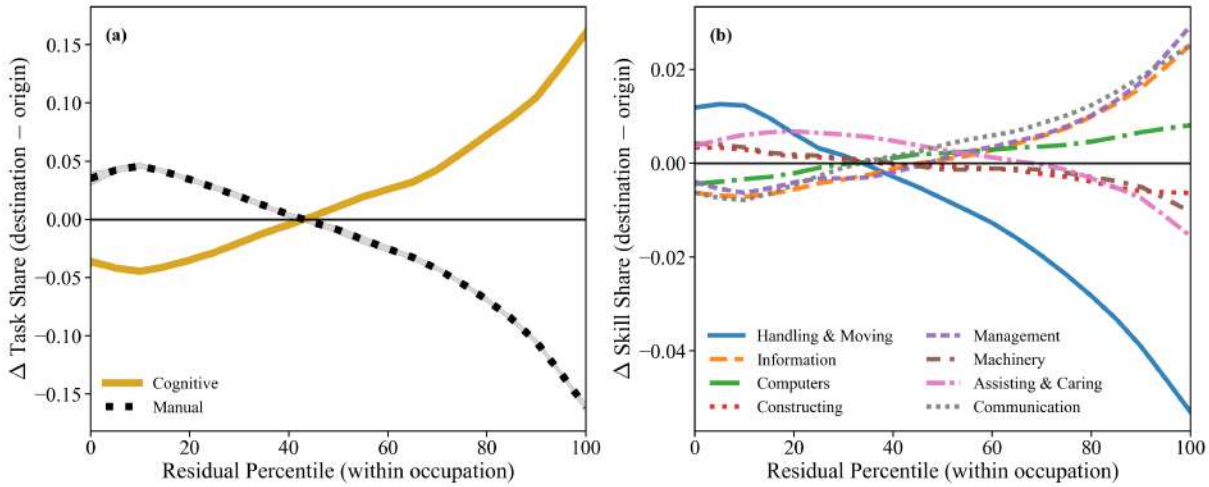


Figure 5: *Panel (a)*: Change in abstract and manual task shares (vertical axis) by residualized within-occupation wage percentile (horizontal axis) for occupational movers. Change is $\Delta\text{Task-Share}_p^k = \text{LOWESS}_p\left(\omega^k_{j'(i)} - \omega^k_{j(i)}\right)$, the smoothed difference in task intensities between destination $j'(i)$ and origin $j(i)$ occupations. *Panel (b)*: Analogous changes based on the ESCO skill classification. By construction, given that $\sum_k \omega_j^k = 1$, the non-parametric average shares per percentile sum up to 0.

in the match quality distribution move to occupations with different task content: high match quality workers move toward more abstract occupations, while low match quality workers move toward more manual ones. The ESCO classification (right panel) shows a similar pattern at a finer level: well-matched workers move to occupations requiring more communication, collaboration, creativity, and computer skills, while poorly matched workers move to occupations requiring more handling, moving, or caring skills. These patterns hold for both within-firm and across-firm switches (Appendix Figure 34).

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

3.3 Discussion of the Empirical Facts

Taken together, Figures 3(b) and 5 show that workers at both ends of the match quality distribution change tasks more when they switch, with the largest moves made by workers at the bottom. The fact that, conditional on moving, the size of the task content change depends on origin match quality is a distinctive feature of the data. Workers who discover a poor fit have stronger incentives to reallocate toward occupations with different task content, and even well-matched workers who switch do so toward occupations with somewhat different task requirements. This is consistent with learning about comparative advantage. If mismatch arises from undirected search, the size of the task change upon switching should be independent of the origin mismatch (Guvenen et al. 2020). The pattern documented here points instead toward directed sorting, where occupational choices are guided by evolving beliefs about comparative advantage across tasks. These findings call for a theory that accounts for the relationship between match quality, mobility, and the task content of transitions. The model in the next section incorporates multidimensional skills, learning about comparative advantage, and task-specific human capital to quantify the consequences of this reallocation mechanism on wages over the life cycle.

4 Quantitative Model

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

4.1 Occupations, Tasks and Human Capital

Workers supply human capital that is used to perform many tasks within each occupation. I model worker i human capital as a vector of task relevant skills $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iK})$ together with

task specific experience that evolves over time $\boldsymbol{\tau}_{it} = (\tau_{i1t}, \ldots, \tau_{iKt})$. The output of a specific task k in a given occupation j at time t is given by:

$$Y_{ijkt} = \ell_{jk}\, \theta_{ik}^{\lambda_{jk}^0}\, \tau_{ikt}^{\beta_{jk}^0}, \tag{3}$$

where λ_{jk}^0 and β_{jk}^0 are primitive task-level elasticities of output with respect to skill and experience in task k and ℓ_{jk} is the share of time allocated to task k in occupation j. When employed in occupation j at time t worker i uses this human capital to perform tasks $k = 1, \ldots, K$ and produces task output. Tasks aggregate within each occupation j according to a Cobb-Douglas technology with task weights ω_{jk}[10]:

$$Y_{ijt} = \prod_{k=1}^{K} \left(\frac{Y_{ijkt}}{\omega_{jk}} \right)^{\omega_{jk}}, \tag{4}$$

where ω_{jk} governs how intensively occupation j uses task k, with $\sum_k \omega_{jk} = 1$. I assume that the firm manages the allocation of worker time across tasks within occupation, and that this allocation is non-contractible.[11] As shown in Appendix D.1, under the normalized Cobb-Douglas aggregation, the optimal time allocation implies that $\ell_{jk}^* = \omega_{jk}$. Hence time assignment to task is independent of workers human capital but solely depends on the task weight of each occupation. Substituting this into (4) yields

$$Y_{ijt} = \prod_{k=1}^{K} \theta_{ik}^{\lambda_{jk}} \prod_{k=1}^{K} \tau_{ikt}^{\beta_{jk}}, \tag{5}$$

where the *effective* parameters are $\lambda_{jk} = \omega_{jk}\lambda_{jk}^0$ and $\beta_{jk} = \omega_{jk}\beta_{jk}^0$. The effective parameters λ_{jk} and β_{jk} determine how skills and experience contribute to output: occupations where workers spend more time on task k (higher ω_{jk}) have higher effective loadings on skill and experience in that dimension. The occupation specific vectors $\boldsymbol{\lambda}_j = (\lambda_{j1}, \ldots, \lambda_{jK})$ and $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jK})$ jointly determine how worker human capital maps into occupation j output. They govern the relative importance of skills and experience across tasks and through the multiplicative structure they pin down the degree of complementarity across tasks and between the skill and experience components. This specification implies that worker skills are bundled in the sense that once matched to an occupation a worker must perform all tasks.

4.2 Labor Demand and Wage Equation

In each period t there is a continuum of competitive firms operating in each occupation j. Firms take the occupation output price in levels P_{jt} as given and compete for workers by posting

[10] As mentioned by Freund and Mann (2025), the unitary elasticity used for tasks aggregation make ω_{jk} invariant to worker specific human capital. Most importantly, given the focus of this paper, the log-linearity of output ensure tractability for the dynamic learning problem discussed next.

[11] This assumption follows Althoff and Reichardt (2026). It simplifies the analysis by making time allocation a static firm decision that maximizes current output, rather than a strategic choice by workers. By having an effects on the continuation value, learning and experience accumulation may have interesting contracting dynamics that are beyond the scope of this paper.

output-contingent piece rate wage contracts[12] . A representative firm in occupation j chooses employment across worker types to maximize profits

$$\max_{\{n_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\}} \Pi_{jt} = P_{jt}Y_{jt} - \int W_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, n_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, dG_t(\boldsymbol{\theta},\boldsymbol{\tau}_t), \tag{6}$$

subject to the occupation production function

$$Y_{jt} = \int n_{jt}(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, Y(\boldsymbol{\theta},\boldsymbol{\tau}_t)\, dG_t(\boldsymbol{\theta},\boldsymbol{\tau}_t), \tag{7}$$

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, i.e.

$$Y(\boldsymbol{\theta},\boldsymbol{\tau}_t) \equiv \prod_{k=1}^{K} \theta_k^{\lambda_{jk}} \prod_{k=1}^{K} \tau_{kt}^{\beta_{jk}}.$$

I assume that output is subject to an idiosyncratic worker occupation time productivity shock ε_{ijt} so that realized output of worker i in occupation j at time t is $\exp(\varepsilon_{ijt})Y_{ijt}$, with $\varepsilon_{ijt} \sim H_j$ i.i.d. across workers and time. 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

$$W_{ijt} = P_{jt} \exp(\varepsilon_{ijt})\, Y_{ijt}. \tag{8}$$

Taking logs yields the log wage equation

$$w_{ijt} \equiv \log W_{ijt} = p_{jt} + \boldsymbol{\lambda}_j'\boldsymbol{\xi}_i + \boldsymbol{\beta}_j' \log \boldsymbol{\tau}_{it} + \varepsilon_{ijt}, \tag{9}$$

where $p_{jt} \equiv \log P_{jt}$ and $\boldsymbol{\xi}_i = (\log\theta_{i1}, \ldots, \log\theta_{iK})$.

4.3 Learning and Information Structure

Workers belong to a group $g \in \mathcal{G}$. To connect the production function with learning, define latent log skills $\boldsymbol{\xi}_i = (\log\theta_{i1}, \ldots, \log\theta_{iK})$. Upon entry into the labor market,

$$\boldsymbol{\xi}_i \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \tag{10}$$

so that $\boldsymbol{\theta}_i$ is lognormal conditional on g. Workers are Bayesian and at the beginning of period t hold a subjective belief density $\tilde{f}_{it}(\boldsymbol{\xi})$ over $\boldsymbol{\xi}_i$ given their information set $\mathcal{I}_{it}$, which contains

[12]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 (1996b), Gibbons et al. (2005b), Lange (2007b), Pastorino (2024)). In my framework, competitive piece rates tie pay to realized output; learning affects outcomes primarily through occupational sorting and continuation values rather than through belief-based wage-setting. Similar output-contingent wage contracts are used in Groes et al. (2015) and are consistent with empirical approaches that combine workers' choices with realized outcomes, as in Bunting et al. (2024). This wage contracting can also be viewed as a special case of the general wage contract in Camargo et al. (2022), $w_{ijt} = c_{ijt} + b_{ijt}y_{ijt} = (1 - b_{ijt})\,\mathbb{E}\left[y_{ijt} \mid I_{ijt}\right] + b_{ijt}y_{ijt}$, with $b_{ijt} = 1$, so that wages depend one-for-one on realized output (up to the occupation-specific output price scaling in my model) rather than on posterior expectations.

group membership and the history of past occupational choices, wages, and experience. I assume rational initial beliefs in the sense that the initial prior coincides with the true group distribution,

$$\tilde{f}_{i0}(\boldsymbol{\xi}) = f_{\xi|g(i)}(\boldsymbol{\xi}) = \phi\Big(\boldsymbol{\xi}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\Big), \tag{11}$$

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

$$s_{ijt} \equiv w_{ijt} - p_{jt} - \boldsymbol{\beta}_j' \log \boldsymbol{\tau}_{it} = \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}, \tag{12}$$

beliefs update according to Bayes rule

$$\tilde{f}_{it+1}(\boldsymbol{\xi}) = \frac{f_{s|\xi}\big(s_{ijt} \mid \boldsymbol{\xi}\big)\, \tilde{f}_{it}(\boldsymbol{\xi})}{\int f_{s|\xi}\big(s_{ijt} \mid \boldsymbol{\xi}'\big)\, \tilde{f}_{it}(\boldsymbol{\xi}')\, d\boldsymbol{\xi}'}. \tag{13}$$

Assume $\varepsilon_{ijt} \sim \mathcal{N}(0, \varsigma_{\varepsilon j}^2)$. Since the prior at $t = 0$ is multivariate normal and the signal in (12) is linear Gaussian, Bayes rule implies that posteriors remain multivariate normal at all dates. Therefore the relevant beliefs can be summarized by a mean vector $\tilde{\boldsymbol{\xi}}_{it}$ and a variance covariance matrix $\tilde{\boldsymbol{\Sigma}}_{it}$ such that

$$\boldsymbol{\xi}_i \mid \mathcal{I}_{it} \sim \mathcal{N}\big(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}\big), \qquad \tilde{\boldsymbol{\xi}}_{i0} = \boldsymbol{\mu}_{g(i)}, \quad \tilde{\boldsymbol{\Sigma}}_{i0} = \boldsymbol{\Sigma}_{g(i)}. \tag{14}$$

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

$$\tilde{\boldsymbol{\xi}}_{it+1} = \tilde{\boldsymbol{\xi}}_{it} + \boldsymbol{K}_{ijt}\left(s_{ijt} - \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_{it}\right), \tag{15}$$

$$\tilde{\boldsymbol{\Sigma}}_{it+1} = \left(\boldsymbol{I}_K - \boldsymbol{K}_{ijt} \boldsymbol{\lambda}_j'\right) \tilde{\boldsymbol{\Sigma}}_{it}, \tag{16}$$

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

$$\boldsymbol{K}_{ijt} \equiv \tilde{\boldsymbol{\Sigma}}_{it} \boldsymbol{\lambda}_j \left(\boldsymbol{\lambda}_j' \tilde{\boldsymbol{\Sigma}}_{it} \boldsymbol{\lambda}_j + \varsigma_{\varepsilon j}^2\right)^{-1}. \tag{17}$$

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.[13]

4.4 Occupational Choice

Let the worker state at the beginning of period t be

$$x_{it} \equiv \big(g(i),\, \tilde{\boldsymbol{\xi}}_{it},\, \tilde{\boldsymbol{\Sigma}}_{it},\, \boldsymbol{\tau}_{it},\, d_{it-1},\, \{p_{j\ell}\}_{j \in \mathcal{O}, \ell \geq t}\big), \tag{18}$$

[13] Equivalently one can write the Kalman gain as $\boldsymbol{K}_{ijt} = \tilde{\boldsymbol{\Sigma}}_{it} \boldsymbol{\lambda}_j S_{ijt}^{-1}$ where $S_{ijt} \equiv \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\Sigma}}_{it} \boldsymbol{\lambda}_j + \sigma_{\varepsilon j}^2$ is the predictive variance of the signal. Substituting this expression into the covariance recursion yields

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

where $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ are beliefs about latent log skills, $\boldsymbol{\tau}_{it}$ is task experience, and d_{it-1} is last period occupation. The sequence $\{p_{j\ell}\}_{\ell \geq t}$ denotes log occupation prices from period t onward. I assume workers have perfect foresight over the future path of occupational prices and take $\{p_{j\ell}\}$ as given.

In each period t worker i chooses an occupation $j \in \mathcal{O}$ before observing the current wage shock. I assume log utility over wages and linear switching costs in total experience $e_{it} \equiv \sum_{k=1}^{K} \tau_{ikt}$, so that $u(W) = \log W$ and $\mathcal{C}(e) = \kappa_0 + \kappa_1 e$. The realized current period utility from choosing j is

$$U_{ijt}(x_{it}) = \log W_{ijt} + \eta_j - \mathbf{1}_{\{j \neq d_{it-1}\}}\big(\kappa_0 + \kappa_1 e_{it}\big) + \epsilon_{ijt}, \tag{19}$$

where W_{ijt} is the wage paid in levels, η_j is a non pecuniary component, and ϵ_{ijt} is i.i.d. Extreme Value Type I with scale ϱ. Under log utility and the wage equation (9), expected utility is linear in expected log productivity:

$$\mathbb{E}\big[\log W_{ijt} \mid x_{it}\big] = p_{jt} + \boldsymbol{\lambda}_j' \tilde{\boldsymbol{\xi}}_{it} + \boldsymbol{\beta}_j' \log \boldsymbol{\tau}_{it}.$$

Task experience accumulation within an occupation depends on how much time a worker spend on a given task k. Given that each worker in an occupation spend the same time in a given task, then experience evolves according to

$$\boldsymbol{\tau}_{it+1} = \boldsymbol{\tau}_{it} + \boldsymbol{\omega}_j, \tag{20}$$

where $\boldsymbol{\omega}_j$ governs task specific experience accumulation when working in occupation j and $\sum_{k=1}^{K} \omega_{jk} = 1$. At the end of the period the worker observes the wage associated with the chosen occupation and updates beliefs according to Bayes rule (13).

Dropping the worker index for clarity, and emphasizing dependence on the price path, the Bellman equation is

$$\begin{aligned} V_t\big(\tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\big) = \max_{j \in \mathcal{O}} \Big\{ & \mathbb{E}\big[\log W_{jt} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t; p_{jt}\big] + \eta_j - \mathbf{1}_{\{j \neq d_{t-1}\}}\big(\kappa_0 + \kappa_1 e_t\big) + \epsilon_{jt} \\ & + \beta\,(1 - \delta(e_t))\,\mathbb{E}_t\Big[V_{t+1}\big(\tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, j; \{p_{j\ell}\}_{\ell \geq t+1}\big) \Big] \Big\}, \end{aligned} \tag{21}$$

where $\delta(e_t)$ is the experience dependent exit probability.[14] Integrating out ϵ yields the multinomial logit CCPs

$$\mathbb{P}_t\big(d_t = j \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\big) = \frac{\exp\Big(\nu_{jt}\big(\tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\big)\,/\,\varrho\Big)}{\sum_{m \in \mathcal{O}} \exp\Big(\nu_{mt}\big(\tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}; \{p_{j\ell}\}_{\ell \geq t}\big)\,/\,\varrho\Big)}. \tag{22}$$

With a continuum of workers, these conditional choice probabilities represent the mass of workers in each state choosing each occupation.

[14] In the quantitative implementation I parameterize $\delta(e)$ as a logistic function of total experience, for instance $\delta(e) = (1 + \exp\{-(\delta_0 + \delta_1 e)\})^{-1}$. This ensures a stationary environment that can be matched to the estimation sample. When a worker exits the labor market they are replaced by a new entrant drawn from the same group g with prior $\boldsymbol{\xi} \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$.

4.5 Closing the Model

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

$$Y_t = \left(\sum_{j\in\mathcal{O}} \alpha_j Y_{jt}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \qquad \alpha_j > 0, \ \sum_{j\in\mathcal{O}} \alpha_j = 1. \tag{23}$$

The aggregator weights $\{\alpha_j\}_{j\in\mathcal{O}}$ are parameters that govern the relative importance of each occupation's output in final production. Let P_t denote the price of the final good and P_{jt} the price of occupation j output. Profit maximization by the final goods producer yields the inverse demand function for occupation j output:

$$P_{jt} = \alpha_j \left(\frac{Y_t}{Y_{jt}} \right)^{\frac{1}{\sigma}}. \tag{24}$$

This expression determines relative prices across occupations as a function of relative outputs.

4.6 Stationary Competitive Equilibrium

I define a stationary competitive equilibrium in which occupation prices are constant over time. With a continuum of workers, let $\pi(x)$ denote the mass of workers in state $x = (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1})$ (excluding prices). Entrants are characterized by group masses π_g (with $\sum_{g\in\mathcal{G}} \pi_g = 1$) and initial distributions $\pi_g^0(\cdot)$ with $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}) = (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ and $\boldsymbol{\tau} = \mathbf{0}$.

A *stationary competitive equilibrium*[15] consists of constant occupation prices $\{P_j\}_{j\in\mathcal{O}}$, a stationary mass distribution $\pi^*(\cdot)$ over worker states, aggregate output Y, occupation outputs $\{Y_j\}_{j\in\mathcal{O}}$, 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 $\{P_j\}_{j\in\mathcal{O}}$ as given, workers solve the Bellman equation (21), generating the value function $V(\cdot)$ and conditional choice probabilities (22). Experience accumulates according to (20), and beliefs update via Bayes' rule (13).

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

$$Y_j = \int \left(\int \exp(\varepsilon)\, Y_j(x)\, dH_j(\varepsilon) \right) \mathbb{P}(d = j \mid x)\, \pi^*(x)\, dx, \tag{25}$$

   where $Y_j(x)$ is determined by the task technology (5), H_j is the distribution of the wage shock ε, and $\mathbb{P}(d = j \mid x)\pi^*(x)$ gives the mass of workers in state x choosing occupation j. Wages satisfy the piece-rate adjusted revenue condition (9) in each occupation.

3. **Price determination.** Occupation prices satisfy the inverse demand system (24) given aggregate output Y and occupation outputs $\{Y_j\}_{j\in\mathcal{O}}$. The aggregate price level P is defined

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

by the standard CES price index $P = \left(\sum_{j\in\mathcal{O}} \alpha_j P_j^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ for $\sigma \neq 1$ with the Cobb-Douglas limit $P = \prod_{j\in\mathcal{O}} P_j^{\alpha_j}$ when $\sigma = 1$.

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

$$\pi^*(x') = \sum_{j\in\mathcal{O}} \int \left[(1-\delta(e))\, Q(x' \mid x, j) + \delta(e) \sum_{g\in\mathcal{G}} \pi_g\, \pi_g^0(x')\right] \mathbb{P}(d = j \mid x)\, \pi^*(x)\, dx, \quad (26)$$

for all states $x' \in \mathcal{X}$, where $Q(\cdot \mid x, j)$ is the transition density over worker states induced by choosing occupation j (incorporating experience accumulation (20) and Bayesian updating (13)), e denotes total current experience, and $\pi_g^0(\cdot)$ is the initial mass distribution for entrants from group g[16].

5 Identification and Estimation

To bring the model to the data, I focus on two broad skill dimensions—abstract and manual—and aggregate occupations into ten clusters based on their task content. This coarser partition is key for the structural estimation, as it ensures that the identification assumptions are respected and keeps the dynamic programming problem computationally tractable. While this aggregation mechanically lowers measured mobility levels compared to the 3-digit ISCO analysis, the empirical facts documented in Section 3 remain qualitatively consistent within this clustered state space, as shown in the robustness checks in Appendix C.

5.1 Identification of the Wage Equation Parameters

The identification argument builds on the key insight in Bunting et al. (2024): when wages reveal productivity signals observed by both workers and the econometrician, occupational choices do not generate selection on unobserved permanent heterogeneity once one conditions on the worker's observable information set. In the model, firms pay piece rates so that log wages equal the log occupation price plus log output $w_{ijt} = p_j + y_{ijt}$.

Let the worker's observable history at a given decision point be

$$\mathcal{H}_{it} \equiv \left(g(i),\, \{d_{i\ell}\}_{\ell\le t-1},\, \{w_{i\ell}\}_{\ell\le t-1},\, \boldsymbol{\tau}_{it}\right),$$

Since $(p_j, \boldsymbol{\beta}_j)$ and $\boldsymbol{\tau}_i$ are observed, the history of wages reveals the same sequence of productivity signals to the worker and to the econometrician. Under rational beliefs and the linear–Gaussian signal structure, the worker's posterior over skills (equivalently, $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}})$) is a deterministic function of $\mathcal{H}_{it}$. Occupational choices are made as functions of $\mathcal{H}_{it}$ (and i.i.d. taste shocks), while the current-period wage shock ε_{ij} is realized after the choice. Therefore the worker's true latent skill vector is irrelevant for the current occupation decision once we condition on the

[16]Under standard regularity conditions on the induced Markov process, the death-and-rebirth mechanism makes the worker-state dynamics regenerative and causes initial conditions to wash out over time. The process therefore admits a unique invariant distribution π^*, and simulation averages converge to expectations under π^*. Appendix D.2 describes the simulation procedure used to approximate this distribution in practice.

observable history:

$$d_{it} \perp \boldsymbol{\xi}_i \mid \mathcal{H}_{it}.$$

This is the sense in which the model is a selection-on-observables environment: conditional on the observed history, which contains past productivity realizations, there is no residual selection on the unobserved permanent skill component. Because the occupation loadings $\boldsymbol{\lambda}_j$ and $\boldsymbol{\beta}_j$ are time-invariant in the stationary environment, identification of the K skill dimensions does not rely on having many calendar periods per se. Instead, it requires observing workers in occupations with sufficiently different loading vectors. In other words, identification of latent skills hinges on occupational switching (or cross-worker variation in occupational histories) that produces enough linearly independent "measurements" of the same $\boldsymbol{\xi}_i$. In the application there are $J = 10$ occupations and $K = 2$ skills: cognitive and manual.

Identification uses the following conditions.

Assumption 1 (Shocks and timing) *Workers choose occupations before the transitory productivity shock is realized. The shock ε_{ijt} is mean-zero, independent of $(\boldsymbol{\xi}_i, \mathcal{H}_{it})$, and i.i.d. across observations, with occupation-specific variance $\mathbb{V}ar(\varepsilon_{ijt} \mid d_{it} = j) = \varsigma^2_{\varepsilon j} \in (0, \infty)$.*

Assumption 2 (Switching support / repeated measurement) *There is positive probability (within each group of interest) that workers are observed (i) at least twice in the same occupation, and (ii) in at least K occupations with linearly independent loading vectors $\boldsymbol{\lambda}_j$ (not necessarily consecutively).*

Assumption 3 (Nondegenerate latent skills) *Within each group g, the latent skill distribution is nondegenerate: $\boldsymbol{\Sigma}_g \equiv \mathbb{V}ar(\boldsymbol{\xi}_i \mid g)$ is positive definite.*

Assumption 4 (Normalization: fixing the skill coordinate system) *There exists a set of K occupations $\mathcal{J}_A = \{j_1, \ldots, j_K\} \subset \mathcal{O}$ such that the $K \times K$ matrix of loading vectors $\Lambda_A \equiv \left[\boldsymbol{\lambda}'_{j_1}, \ldots, \boldsymbol{\lambda}'_{j_K}\right]$ is nonsingular.*

Assumption 5 (Location normalization) *For a reference group $g^\star$, normalize the origin of the skill space by $\mathbb{E}[\boldsymbol{\xi}_i \mid g^\star] = \mathbf{0}$.*

Assumption 1 ensures that wages are noisy but unbiased measurements of a time-invariant skill component: because the shock is realized after the occupational choice and is independent of the worker's history, conditional wage moments are not contaminated by workers selecting into occupations on unobserved transitory shocks. Assumption 2 provides the key empirical variation in a stationary environment with time-invariant loadings: observing the same worker multiple times in the same occupation separates the permanent component from transitory noise, and observing workers across occupations with linearly independent loading vectors yields multiple independent measurements of the same $\boldsymbol{\xi}_i$. Assumption 3 rules out degenerate within-group skill distributions; in particular, $\boldsymbol{\Sigma}_g \succ 0$ guarantees that both skill dimensions vary in the population and that the covariance matrix is invertible, which is necessary to recover loading vectors from wage covariances. Finally, Assumptions 4 and 5 fix the

remaining observational equivalences of the factor structure. The skill coordinates are normalized by imposing $\Lambda_A = I_K$, which fixes units, rotation, and scale of the latent skill space without restricting the group-specific covariance matrices $\mathbf{\Sigma}_g$, while $\mathbb{E}[\boldsymbol{\xi}_i \mid g^\star] = \mathbf{0}$ pins down the origin of the skill space in the presence of occupation fixed effects.

With an unrestricted group-specific covariance matrix $\mathbf{\Sigma}_g$, the factor structure has a $K \times K$ change-of-basis indeterminacy. For general K, separating the loadings from the latent distribution therefore requires K^2 normalizing restrictions. In our application $K = 2$, so $K^2 = 4$ restrictions are needed. A convenient implementation is to select two occupations with linearly independent loadings and normalize the resulting 2×2 loading matrix to the identity. In the empirical implementation, I choose these normalizing occupations based on external task measures: one cluster that is purely cognitive and one that is purely manual, which implies $\boldsymbol{\lambda}_{j_C} = (1, 0)'$ and $\boldsymbol{\lambda}_{j_M} = (0, 1)'$.

Under Assumptions 1–3 (including the normalization in Assumption 4), the wage equation parameters

$$\left\{ (p_j, \boldsymbol{\lambda}_j, \boldsymbol{\beta}_j, \varsigma^2_{\varepsilon j})_{j \in \mathcal{O}},\ (\boldsymbol{\mu}_g, \mathbf{\Sigma}_g)_{g \in \mathcal{G}} \right\}$$

are point identified.

5.2 Identification of the Dynamic Choice Parameters

Identification of the dynamic choice parameters follows a Hotz–Miller (Hotz and Miller 1993) inversion argument. Given the first–stage estimates of the wage equation, the econometrician knows the deterministic component of expected log wages $r_j(\boldsymbol{x}_t) = \mathbb{E}[\log W_{jt} \mid \boldsymbol{x}_t]$ for each occupation j and the transition kernel $\mathcal{P}(\cdot \mid \boldsymbol{x}_t, j)$ over next–period states implied by deterministic experience accumulation and Bayesian updating. The conditional choice probabilities (CCPs) in (22) are identified from the joint distribution of $(d_t, \boldsymbol{x}_t)$. Under i.i.d. Type I Extreme Value shocks with scale parameter $\varrho > 0$, CCP log–odds map into differences in the deterministic choice–specific value indices:

$$\log \frac{P(d_t = j \mid \boldsymbol{x}_t)}{P(d_t = k \mid \boldsymbol{x}_t)} = \frac{v_j(\boldsymbol{x}_t) - v_k(\boldsymbol{x}_t)}{\varrho}. \tag{27}$$

As in standard multinomial logit models, utilities are identified only up to a positive scale. In dynamic settings with learning, variation in expected wage gaps across states typically comoves with continuation values, making separate empirical identification of the taste–shock scale ϱ fragile without additional exogenous payoff shifters. Accordingly, I treat ϱ as a normalization/calibration and set $\varrho = 0.25$. Intuitively, ϱ governs how strongly idiosyncratic taste shocks move occupational choices relative to systematic components (expected wages, amenities, and switching costs); choosing a small value limits excessive occupational mobility driven purely by transitory shocks and keeps mobility primarily driven by learning and human capital accumulation. Given ϱ, equation (27) identifies $v_j(\boldsymbol{x}_t) - v_k(\boldsymbol{x}_t)$. Amenities are identified from comparisons of alternative switch destinations from a common origin, which difference out the common switching cost. Amenity levels are anchored by the normalization $\eta_0 = 0$. Switching costs are identified from move–versus–stay comparisons as total experience varies under the specification $\mathcal{C}(j, d, e) = \mathbf{1}\{j \neq d\}(\kappa_0 + \kappa_1 e)$: holding fixed the origin occupation,

differences in CCP log–odds across experience levels trace out the affine switching–cost schedule. Extensions to more general generalized extreme–value structures follow Arcidiacono and Miller (2011) and Chiong et al. (2013). A more formal argument is provided in Appendix A.1

5.3 Estimation: Two-Stage Estimation Procedure

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

$$w_{ijt} = p_j + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \boldsymbol{\beta}_j' \log \boldsymbol{\tau}_{it} + \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}(0, \varsigma_{\varepsilon j}^2), \tag{28}$$

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

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

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

$$\boldsymbol{\Delta} \equiv \Big\{ (\eta_j)_{j \in \mathcal{O}}, \kappa_0, \kappa_1 \Big\},$$

taking $\widehat{\boldsymbol{\Gamma}}$ as given and setting the taste-shock scale $\varrho = 0.25$ (see Section **??** for discussion). The complete–data likelihood for worker i factors into a wage component and a choice component:

$$L_i(w_i, d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}) = L_i^w(w_i; \boldsymbol{\Gamma}) \cdot L_i^c(d_i; \boldsymbol{\Gamma}, \boldsymbol{\Delta}), \tag{29}$$

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

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

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

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

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

The wage likelihood integrates out latent skills,

$$L_i^w(w_i;\boldsymbol{\Gamma}) = \int \left[\prod_{t=t_0(i)}^{T(i)} \prod_{j\in\mathcal{O}} f_{\varepsilon j}\Big(w_{ijt} - p_j - \boldsymbol{\lambda}_j'\boldsymbol{\xi} - \boldsymbol{\beta}_j' \log \boldsymbol{\tau}_{it} \Big)^{d_{ijt}} \right] \phi\Big(\boldsymbol{\xi}; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\Big)\, d\boldsymbol{\xi}, \tag{30}$$

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

$$\begin{aligned} L_i^c(d_i;\boldsymbol{\Gamma},\boldsymbol{\Delta}) = & \underbrace{\prod_{j\in\mathcal{O}} \Big[\mathbb{P}\Big(d_{it_0(i)} = j \mid \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}, \boldsymbol{\tau}_{it_0(i)}, d_{i,t_0(i)-1}; \boldsymbol{\Gamma}, \boldsymbol{\Delta} \Big) \Big]^{d_{ij,t_0(i)}}}_{\text{first observed choice}} \\ & \times \underbrace{\prod_{t=t_0(i)+1}^{T(i)} \prod_{j\in\mathcal{O}} \Big[\mathbb{P}\big(d_{it} = j \mid \tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it}, \boldsymbol{\tau}_{it}, d_{it-1}; \boldsymbol{\Gamma}, \boldsymbol{\Delta} \big) \Big]^{d_{ijt}}}_{\text{subsequent choices}}, \end{aligned} \tag{31}$$

where CCPs are given by (22) and $(\tilde{\boldsymbol{\xi}}_{it}, \tilde{\boldsymbol{\Sigma}}_{it})$ are the Bayesian posterior moments implied by $\boldsymbol{\Gamma}$. Because $L_i = L_i^w \cdot L_i^c$, the log–likelihood is additively separable:

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

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

5.4 Estimation Stage 1: Expectation–Maximization Algorithm

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

- **E–step.** For current $\boldsymbol{\Gamma}^{(q)}$ and observed histories (w_i, d_i), compute the posterior distribution of $\boldsymbol{\xi}_i$,
$$f(\boldsymbol{\xi}_i \mid w_i, d_i; \boldsymbol{\Gamma}^{(q)}),$$
and form the expected complete–data log–likelihood.
- **M–step.** Update $\boldsymbol{\Gamma}^{(q+1)}$ by maximizing the expected complete–data log–likelihood. In practice this step updates $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, the occupation loadings $(\boldsymbol{\lambda}_j, \boldsymbol{\beta}_j)$, the shock variances $(\sigma^2_{\varepsilon j})$, and the occupation prices (p_j).

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

5.5 Estimation Stage 2: Nested Fixed Point Algorithm

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

$$\widehat{\boldsymbol{\Delta}} = \arg\max_{\boldsymbol{\Delta}} \; \ell^c(\widehat{\boldsymbol{\Gamma}}, \boldsymbol{\Delta}), \tag{32}$$

where

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

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

$$W(\boldsymbol{x}) = \varrho \log \sum_{j \in \mathcal{O}} \exp\left(\frac{\tilde{w}_j(\boldsymbol{x}) + \eta_j - \mathcal{C}(j, d_{-1}, e) + \beta\, \mathbb{E}\Big[W(\boldsymbol{x}') \mid \boldsymbol{x}, d = j; \widehat{\mathbf{\Gamma}}, \mathbf{\Delta} \Big]}{\varrho} \right), \tag{33}$$

where the expectation is taken over next-period state $\boldsymbol{x}'$ implied by the law of motion under choice j (including belief updating). Given the fixed point $W(\cdot; \widehat{\mathbf{\Gamma}}, \mathbf{\Delta})$, the CCPs used in the likelihood are computed from (22).

The outer step updates $\mathbf{\Delta}$ to increase (32) using a numerical optimizer. The procedure alternates between (i) solving the fixed point in (33) for W at the current $\mathbf{\Delta}$ and (ii) updating $\mathbf{\Delta}$ given the resulting CCPs, until convergence.

5.6 Externally Set and Calibrated Parameters

Several parameters are set externally or calibrated to match aggregate statistics rather than estimated jointly with the dynamic model, as summarized in Table 2 in the Appendix. The discount factor is set to $\beta = 0.9$; exit probability parameters are calibrated to match observed panel lengths; education-gender group shares are set to the estimation sample composition; and task importance weights $\boldsymbol{\omega}_j$ are taken from Mihaylov and Tijdens (2019). The CES aggregator weights α_j are calibrated from observed wage shares and occupation prices from Stage 1, and the price index $P(p)$ is normalized to match the empirical price level.

6 Estimation Results

6.1 Stage 1 Estimation Results: Wage Equation Parameters

The EM algorithm jointly estimates occupation-specific skill and experience elasticities $(\boldsymbol{\lambda}_j, \boldsymbol{\beta}_j)$, idiosyncratic wage variances, occupation prices, and education-gender-specific skill distributions $(\boldsymbol{\mu}_g, \mathbf{\Sigma}_g)$. Full estimation results are reported in Table 3 in the Appendix.

Three patterns are central to the counterfactual analysis. First, occupations split into three natural groups—cognitive-intensive, manual-intensive, and balanced—with substantial heterogeneity in how they weight skills and experience across task dimensions. Occ_7 is normalized

to be purely cognitive and Occ_3 purely manual, following the task classification. Second, mean skills rise with education in both dimensions, but the cognitive-manual gap widens progressively: higher-educated workers are increasingly tilted toward cognitive comparative advantage, consistent with Deming (2025). Low-educated workers, especially males, have weak or negative cross-skill correlations and are more likely to be specialists; high-educated workers, especially females, have positive correlations and more balanced skills (Figure 6). These distributional differences rationalize the key empirical patterns: low-wage workers switch to exploit comparative advantage, making large task changes, while high-wage workers specialize further in cognitive tasks. Third, returns to cognitive task experience are generally steeper than returns to manual task experience (Figure 7)—an alignment with the skill elasticities that implies occupations specialized in a given task dimension reward both the initial endowment and accumulated experience in that dimension, generating substantial heterogeneity in wage profiles across worker types.

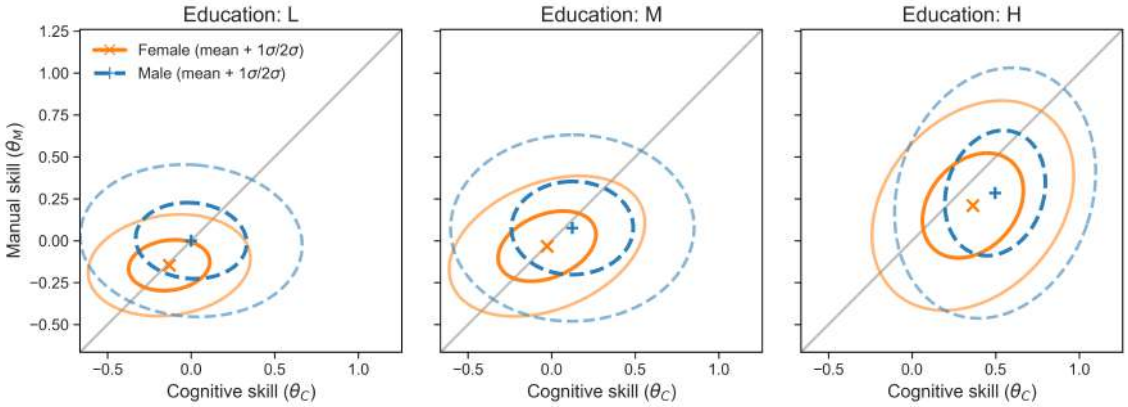


Figure 6: Simulated skill distributions by education-gender group. Each panel plots the distribution of log cognitive skill ξ_A (horizontal axis) and log manual skill ξ_M (vertical axis) using estimated parameters. Solid lines (x markers) = female workers; dashed lines (+ markers) = male workers. Ellipses show 1σ and 2σ contours. The diagonal marks $\xi_A = \xi_M$. Mean skills and the cognitive-manual gap both rise with education; cross-skill correlations are higher for women and for more educated workers.

6.2 Stage 2 Estimation Results: Dynamic Choice Parameters

The NFXP estimates are reported in Table 4 in the Appendix. The taste-shock scale is calibrated at $\varrho = 0.25$ rather than estimated, since in dynamic learning models this parameter is not separately identified from continuation values without exogenous payoff shifters; this value is conservative enough to ensure mobility is driven by economic fundamentals while still allowing within-group heterogeneity in initial choices. Switching costs are moderate—the baseline fixed cost is roughly one log point, growing only slowly with experience—suggesting that the lifecycle decline in occupational mobility is driven primarily by task-specific human capital accumulation and learning rather than increasingly prohibitive switching costs. Occupation-specific amenities η_j display considerable heterogeneity, rationalizing sorting patterns that cannot be explained by wage differentials alone.

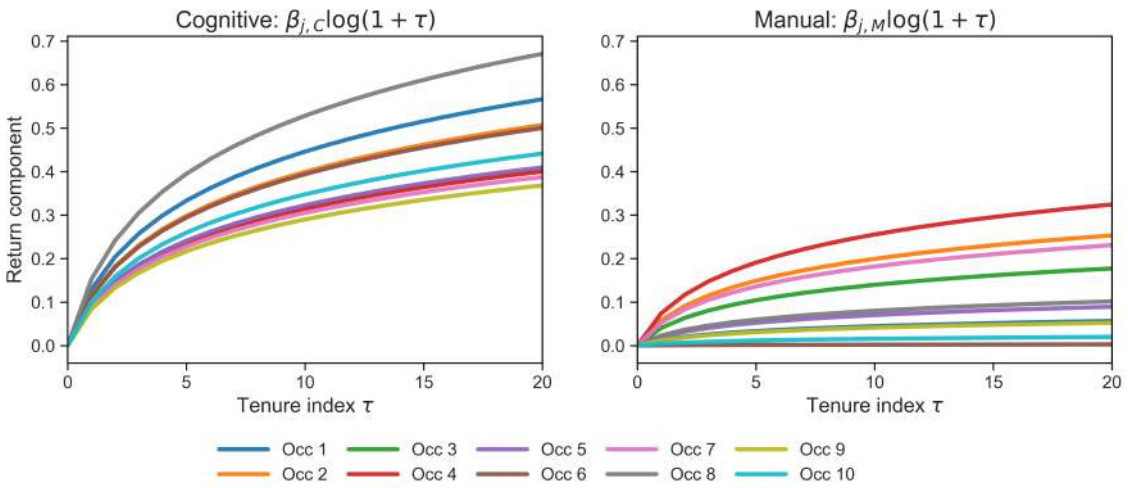
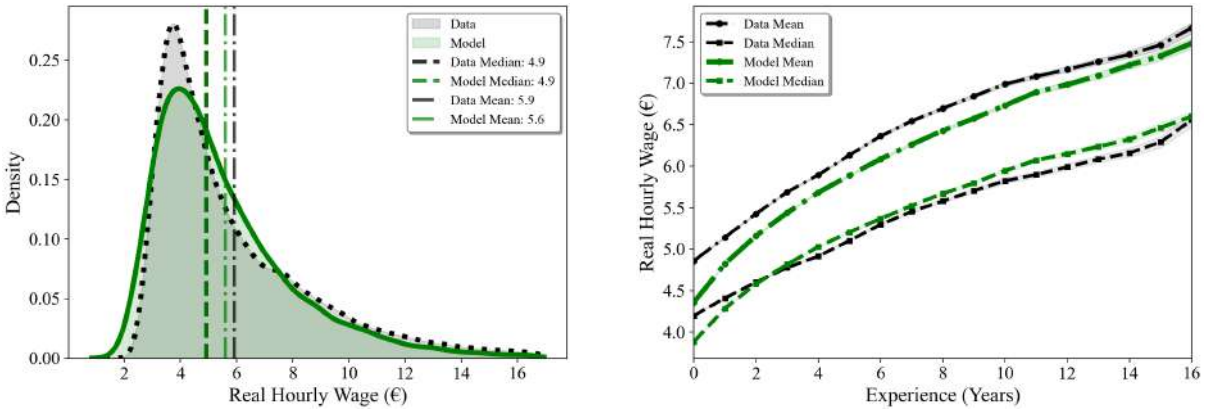


Figure 7: Returns to task-specific experience by occupation. Each line plots $\beta_{j,k}\log(\tau)$ as task tenure τ accumulates from 1 to 20 years. *Left panel*: Returns to cognitive experience. *Right panel*: Returns to manual experience. Cognitive-intensive occupations show steeper returns to cognitive experience; manual-intensive occupations show steeper returns to manual experience.

7 Quantitative Results and Counterfactuals

7.1 Empirical Fit of the Quantitative Model

I now assess how the model fit empirical moments on wages and occupational mobility patterns. Figure 8 displays the wage distribution and average wage profile over the lifecycle generated by the stationary equilibrium of the estimated model. The model generates slightly lower mean wages than the data, reflecting somewhat more mass at the lower end of the wage distribution. However, the overall fit is very good: the median wage achieves an almost perfect match, and the model captures both the general shape of the wage distribution and the increasing wage profile with experience observed in the data.

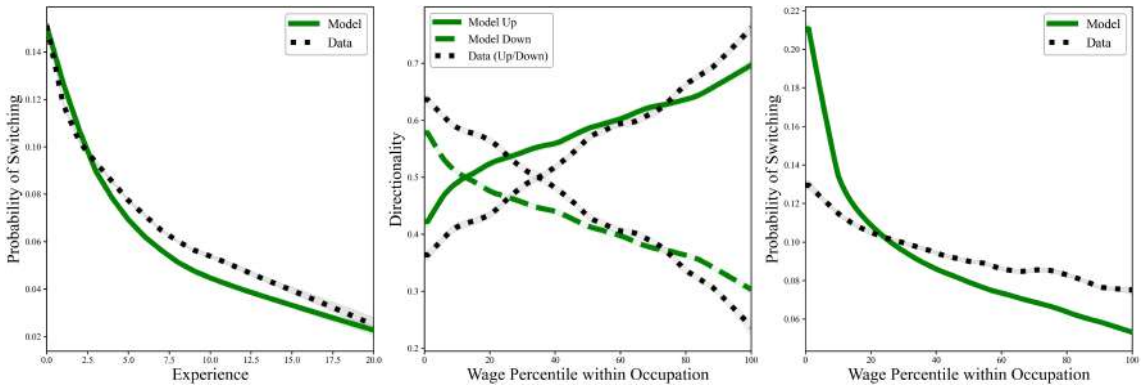


(a) Wage Distribution: Model vs Data

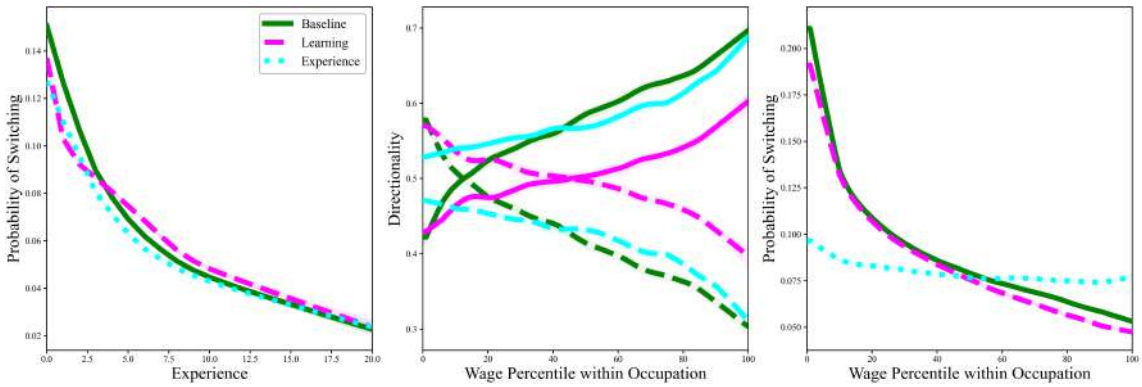
(b) Wage Profile: Model vs Data

Figure 8: Wage outcomes from the stationary equilibrium of the estimated model. *Panel 8(a)*: Distribution of real hourly wages (horizontal axis: wage; vertical axis: density). *Panel 8(b)*: Average wage profile over the lifecycle (horizontal axis: years of experience; vertical axis: average wage). Solid lines = model predictions; dashed lines = empirical patterns.

Figure 9 assesses the model's ability to reproduce the key occupational mobility patterns documented in Section 3. Overall, the model matches the empirical patterns qualitatively and achieves good quantitative performance. The model captures the declining hazard rate with experience and the negative relationship between wage percentile and switching probability. For directionality, the model reproduces the pattern that upward mobility increases with wage percentile while downward mobility decreases. Learning generates somewhat too much mobility at the bottom of the within-occupation wage distribution, but is essential for matching the directionality of occupational switches.



(a) Key Occupational Mobility Patterns: Model vs Data



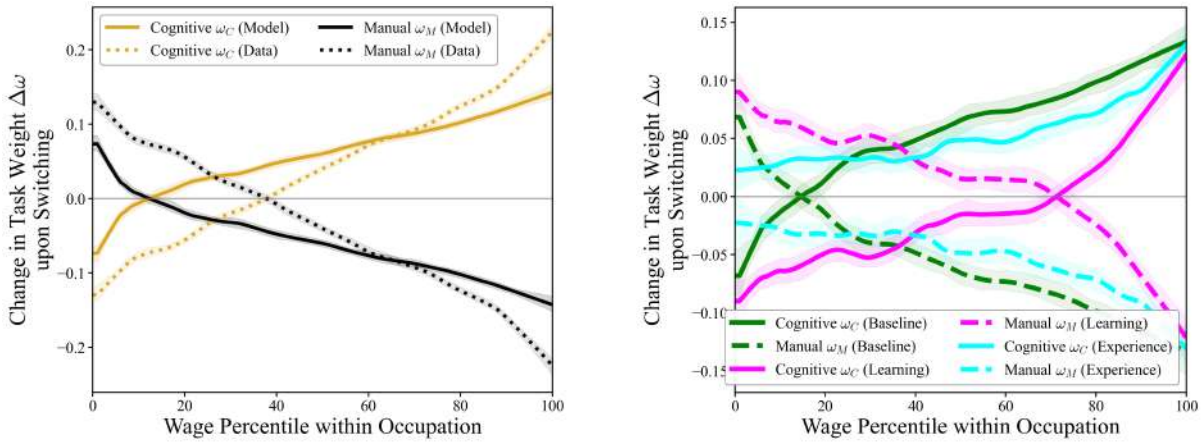
(b) Occupational Mobility Channels Decomposition in the Full Model

Figure 9: Both subfigures present three sub-panels. *Left*: Hazard rate by experience (horizontal axis: years; vertical axis: switching probability). *Middle*: Directional moves by wage percentile (horizontal axis: percentile; vertical axis: probability); solid = upward, dashed = downward. *Right*: Overall switching probability by wage percentile (horizontal axis: percentile; vertical axis: probability). *Subfigure 9(a)*: Model predictions vs empirical moments. *Subfigure 9(b)*: Counterfactual simulations under Baseline (green–solid), Learning (magenta–dashed), and Experience (cyan–dotted).

The counterfactual decomposition in Figure 9(b) reveals the distinct roles of learning and task-specific experience. All three scenarios generate similar declining hazard rates, but learning is essential for directionality: the Experience-only scenario produces much flatter patterns, failing to capture sorting based on comparative advantage. Figure 25 (Data Appendix) shows that the model also reproduces the fading "U-shape" in switching probability across wage percentiles over the life cycle, though learning overpredicts mobility for younger workers.

Figure 10 shows that the model captures the task content of occupational switches. Panel 10(a) confirms that the model reproduces the task directionality of occupational mobility: low-wage

workers within occupation move toward more manual-intensive occupations, while high-wage workers move toward more cognitive-intensive ones. This pattern is consistent with the interpretation that low-wage switchers sort on comparative advantage—making large task adjustments to find better-matched occupations—while high-wage switchers specialize further within a similar task bundle. The counterfactual decomposition of the task directionality in Panel 10(b) confirms that learning is essential for generating these task-based sorting patterns—without it, the Experience-only scenario fails to capture heterogeneity in task directionality across the wage distribution.



(a) Task Content of Switches: Model vs Data

(b) Counterfactual Decomposition

Figure 10: *Panel 10(a)*: Comparison of model predictions (solid) vs empirical patterns (dashed) for the probability of moving toward more cognitive-intensive or manual-intensive occupations, by wage percentile within occupation. *Panel 10(b)*: Counterfactual decomposition of the task content of occupational switches, showing how different model channels (Baseline, Learning, Experience) contribute to the task directionality of occupational mobility by wage percentile within occupation.

To further validate the role of learning, I simulate a simplified model without switching costs, amenities, or idiosyncratic shocks (see Figure 27 in the Data Appendix). The results confirm that learning is crucial for generating sorting based on comparative advantage: without it, workers do not effectively use their wage position as a signal about their skills.

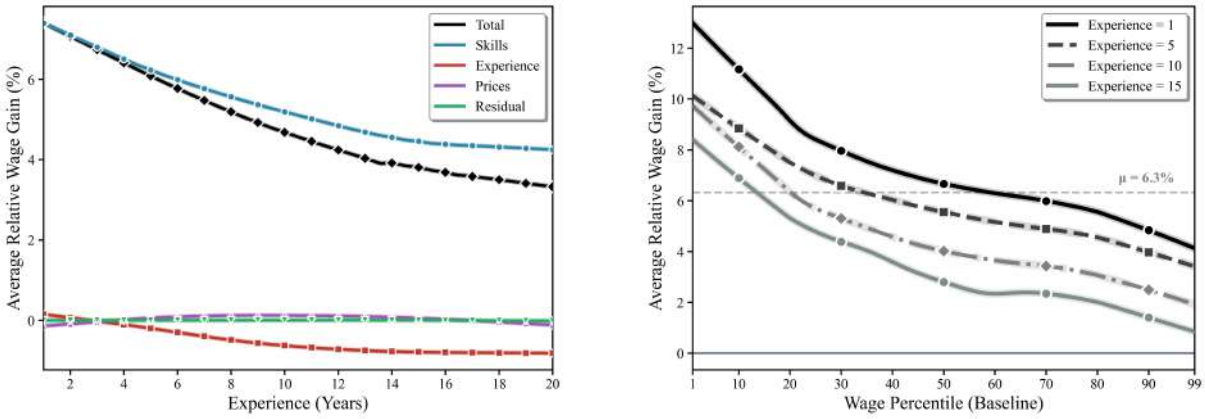
Summary: The model matches the magnitude, scale, and shape of key occupational mobility patterns and of the wage distribution and wage profile over the lifecycle. Learning generates somewhat too much mobility at the bottom of the wage distribution but remains essential for matching directionality and task-based sorting. Switching costs, non-compensating differentials, and idiosyncratic shocks help calibrate overall mobility levels.

7.2 The Aggregate Implications of Early Career Skills Mismatch

I now use the estimated model to quantify the wage cost of imperfect information about skills. I construct a counterfactual general equilibrium in which workers have perfect beliefs from the start of their careers: each worker's initial beliefs equal their true skills, and they face no uncertainty.[18] In both the baseline and counterfactual economies, workers face identical switch-

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

ing costs, idiosyncratic taste shocks, and wage noise. The only difference is the information structure at labor market entry. This exercise quantifies the wage cost of imperfect information about skills, that is the wage loss associated with sorting based on imperfect rather than perfect knowledge of one's comparative advantage. Figure 11 presents the average relative wage gain from removing information frictions over the life cycle and by wage percentile. Four findings emerge.



(a) Wage Gains from Perfect Information over the Life Cycle

(b) Average Wage Gains by Wage Percentile

Figure 11: *Panel (a)*: Average relative wage gain over the life cycle (horizontal axis: years of experience; vertical axis: percent wage gain) from removing information frictions. *Panel (b)*: Wage gains by position in the overall wage distribution (horizontal axis: wage percentile; vertical axis: percent wage gain), with separate lines for different experience levels. Workers at lower percentiles gain more from removing information frictions. The decomposition of wage gains into channels conditional on different demographic groups is reported in Figure 28 in the Appendix.

First, information frictions generate an aggregate productivity loss of 6.7%.[19] Comparing the two stationary equilibria, the economy under imperfect information produces 6.7% less output than under perfect information. This reflects the misallocation of workers to occupations that do not match their true comparative advantage: despite the slightly negative price channel—perfect information induces reallocation that compresses some occupation prices—the productivity gain from better sorting dominates.

Second, these aggregate costs are driven by early career mismatch that is only partially resolved over the life cycle. At labor market entry, the average worker would earn 7.5% more if they knew their true skills. This gap narrows over time as workers learn and re-sort, falling to 3.3% after twenty years. However, mismatch is never fully eliminated. Learning, imperfectly transferable experience, and switching costs interact to create strong path dependence: workers who sort into the wrong occupation early accumulate experience in the wrong task dimension, raising the opportunity cost of switching later in the career. Early mismatch thus compounds over time, and a substantial fraction of workers never converge to their optimal career path.

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

Third, improved sorting on skills accounts for nearly all of the wage gains. I decompose the wage gains into four channels (see Figure 28 in the Appendix): (i) *Skills*, gains from sorting into occupations that better match true comparative advantage holding experience fixed; (ii) *Experience*, additional gains from accumulating task-specific human capital in better-matched occupations; (iii) *Prices*, gains from general equilibrium adjustments in occupation prices; and finally (iv) *Residual*[20]. The Skills channel almost perfectly tracks the total wage gain throughout the lifecycle. The Experience channel is small and only slightly positive by year 20: while perfect information workers accumulate experience in better-matched occupations from the start, the gain is modest because experience accumulated in suboptimal occupations is not entirely wasted. Given the complementary, multiplicative production structure, workers who spent time in a different occupation carry cross-dimensional experience that retains some productivity value in their eventual optimal occupation, partially offsetting the experience loss from early mismatch. This cross-task complementarity becomes central in the switching costs counterfactual (Section 7.3), where it drives a reversal of the wage gap late in the career. The Price channel is slightly negative, as the reallocation induced by perfect information shifts relative labor supply across occupations.

Fourth, removing information frictions reduces wage inequality. Panel (b) of Figure 11 plots average wage gains by workers' position in the overall wage distribution, with separate lines for different experience levels. Workers at lower percentiles experience substantially larger wage gains from perfect information: at labor market entry, those in the 10th percentile gain approximately 13%, while those in the 90th percentile gain only about 5%. This pattern persists, though attenuated, across all experience levels. The finding reflects the model's core mechanism: low-wage workers are more likely to be mismatched, sorting into jobs that poorly fit their true comparative advantage. When provided with perfect information, these workers reallocate to better-matched occupations, generating the largest wage improvements. By disproportionately benefiting low-wage workers, removing information frictions compresses the overall wage distribution.

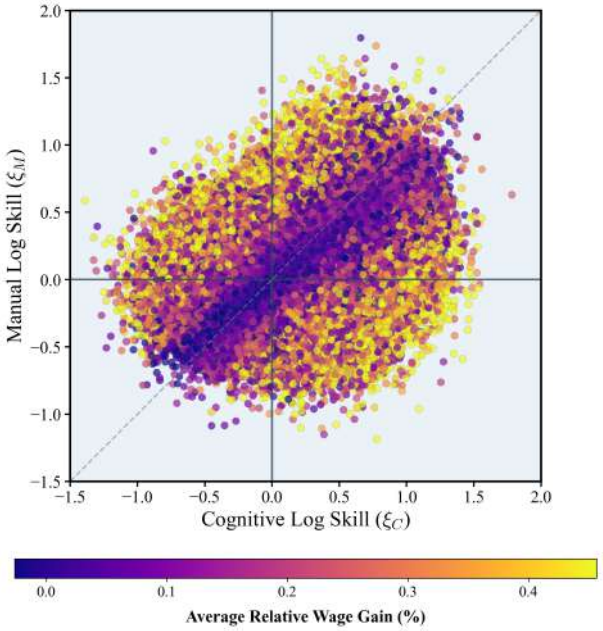
7.2.1 Who Benefits Most? Specialists versus Generalists

Importantly, not all workers reallocate when provided with perfect information. The model includes other important determinants of occupational choice, such as idiosyncratic taste shocks, non-compensating differentials, and switching costs, that have been shown to be important in determining labor mobility. Under perfect information, 30% of workers would choose a different initial occupation, and 50% would switch occupations at some point during their career, depending on education, gender, and experience. Among those who do reallocate, the wage gains are substantial, with average gains of 17.5%.

Figure 12 reveals striking heterogeneity in who benefits from removing information frictions, focusing on reallocated workers only. Each point represents a worker who switches occupations, positioned according to their true cognitive skill (θ_C, horizontal axis) and manual skill (θ_M, vertical axis), with color indicating their average wage gain over the lifecycle. A clear pattern emerges: *specialists*, workers with a strong comparative advantage in one dimension, gain most from perfect information, while *generalists*, workers near the 45-degree line with similar

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

cognitive and manual skills, gain relatively little.



(a) Wage gains for reallocated workers by true skill

Figure 12: Each point represents a reallocated worker (those who choose different occupations under perfect versus imperfect beliefs), positioned by their true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis). Color indicates lifecycle-average wage gain (color scale: yellow/light = high gains, purple/dark = low gains). The dashed 45-degree line separates cognitive specialists (below) from manual specialists (above). Specialists with pronounced comparative advantage gain most from removing information frictions.

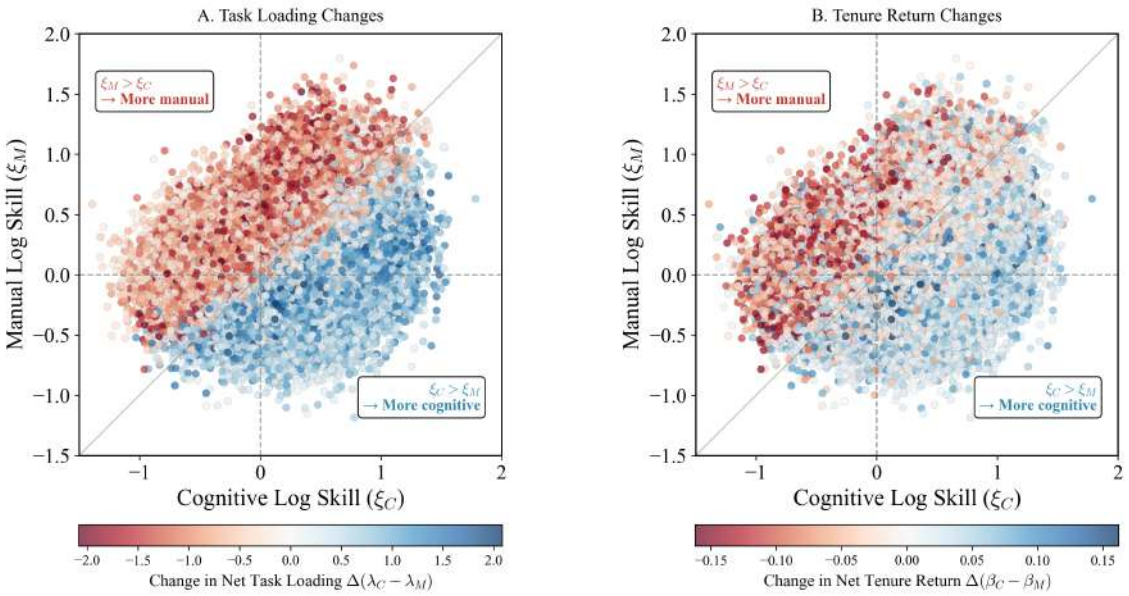
This pattern has an intuitive economic interpretation. Specialists have high returns to correct sorting: a worker with $\theta_C \gg \theta_M$ loses substantially if they mistakenly enter a manual-intensive occupation. When beliefs are imprecise, such workers may fail to recognize their comparative advantage and sort into the wrong occupation for several years. In contrast, generalists have the lowest returns to improving information because they are approximately equally productive across occupations, so the cost of sorting "incorrectly" is small.

This finding has implications for the design of career guidance policies. Interventions that provide better information about skills, such as aptitude testing, structured internships, or early career mentoring, may yield the largest returns for workers with pronounced but unrecognized comparative advantage.

7.2.2 Wage Gains are Driven by Improved Sorting on Skills

Figure 13 illustrates the mechanism through which perfect information improves sorting. Each panel plots occupational switchers, workers who choose different occupations under perfect versus imperfect beliefs, positioned by their true skills. The color indicates the *direction* of reallocation: blue workers move toward more cognitive-intensive occupations (positive $\Delta(\lambda_C - \lambda_M)$), while red workers move toward more manual-intensive occupations.

The horizontal axis sorts workers by their true cognitive skill θ_C, and the vertical axis by their true manual skill θ_M. A sharp diagonal pattern emerges: workers with $\theta_C > \theta_M$ (cognitive



(a) The sorting mechanism: reallocation by skill type

Figure 13: Each point represents an occupational switcher, positioned by true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis). Color indicates direction of reallocation: blue = move to more cognitive-intensive occupations, red = move to more manual-intensive occupations. *Panel A*: Reallocation measured by change in skill loadings $\Delta(\lambda_C - \lambda_M)$. *Panel B*: Reallocation measured by change in experience returns $\Delta(\beta_C - \beta_M)$. The diagonal pattern confirms that specialists reallocate toward occupations matching their true comparative advantage.

specialists, below the 45-degree line) systematically move to more cognitive occupations when given perfect information. Conversely, workers with $\theta_M > \theta_C$ (manual specialists, above the 45-degree line) move to more manual occupations. Panel A shows this pattern for task loadings λ; Panel B shows an analogous pattern for experience returns β, where workers with cognitive comparative advantage move to occupations with higher returns to cognitive experience. This suggests that while the contribution of experience to wage gains across the two scenarios is small, there might be a role for task-specific experience in generating wage costs that prevents workers from switching to the right occupation as they discover their true comparative advantage.

This evidence confirms the model's core mechanism: imperfect beliefs lead workers to sort based on perceptions rather than true comparative advantage, and removing this friction allows them to specialize in their strongest dimension.

7.2.3 Heterogeneity of Wage Gains across Demographic Groups

The aggregate wage gains mask substantial heterogeneity across demographic groups. Figure 28 in Appendix C presents the decomposition separately by education and gender. Two patterns emerge. First, female workers experience larger wage gains than males across all education levels, with gains reaching 5–6% at labor market entry for low- and medium-educated women. This reflects both the initial sorting patterns, as women may face greater early-career uncertainty about their comparative advantage, and the occupational distribution, as female-dominated occupations may offer different return structures. Second, the Experience channel becomes more important relative to Skills for high-educated workers, particularly women, sug-

gesting that information frictions distort not only initial occupational choice but also the entire trajectory of human capital accumulation for workers who invest intensively in task-specific skills.

Table 8 in Appendix C provides detailed statistics for the subset of workers who reallocate, those who choose different occupations under perfect versus imperfect beliefs. Under perfect information, 30% of workers would choose a different initial occupation, and 50% would switch at some point during their career. Among reallocators, median wage gains range from 5% to 16%, with substantial right-skew: the 75th percentile exceeds 13–30%, and some workers experience gains above 40%. Not all reallocators benefit: the 5th percentile shows losses of 2–12%, reflecting workers for whom the general equilibrium price adjustments or suboptimal experience accumulation under perfect beliefs (due to different switching patterns) generate lower wages. Low-educated women exhibit the highest reallocation rates (35–41%) and substantial median gains (6–16%), while high-educated workers show similar reallocation rates (35–38%) with median gains of 5–11%.

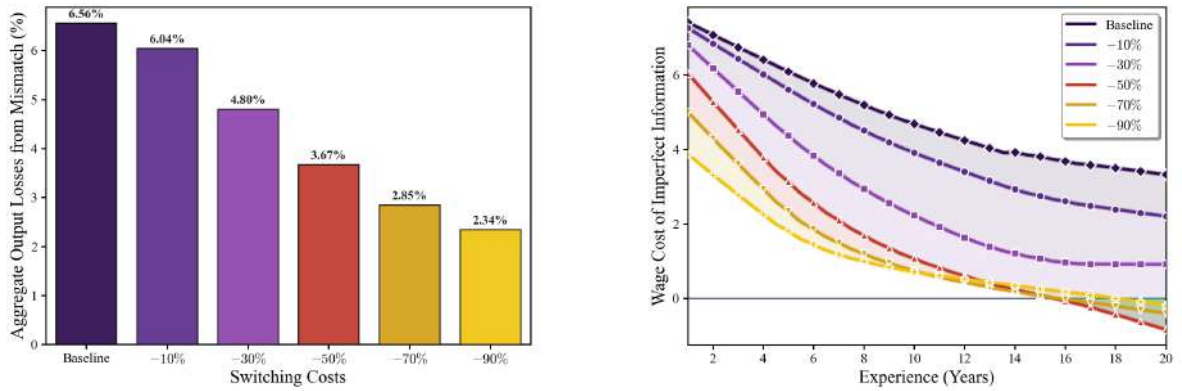
Summary. Information frictions about skills generate economically meaningful wage costs: 7.5% at labor market entry, declining to 3.3% after twenty years. Nearly all of these costs reflect misallocation rather than suboptimal experience accumulation. Crucially, the costs are concentrated among specialists, workers with pronounced comparative advantage who have the most to gain from correct sorting. Generalists, whose skills are balanced across dimensions, lose little from uncertainty and may actually benefit from the experimentation that imperfect beliefs induce. These findings suggest that policies aimed at improving early-career information about skills should target workers most likely to possess (but not recognize) strong comparative advantage.

7.3 Policy Counterfactuals

Reducing Switching Costs. The lifecycle dynamics documented above point to switching costs, imperfectly transferable human capital and learning as potential key drivers of persistent late-career misallocation. Even after workers learn their true comparative advantage through accumulated wage signals, high switching costs may prevent them from acting on this information: mismatch that could in principle be resolved through occupational mobility instead persists because the cost of switching is too large relative to the remaining career horizon gains. I now examine whether policies that reduce switching costs—for example, reduced search frictions or subsidized moving costs for changing occupations—can meaningfully lower the aggregate cost of early-career mismatch.

I study a series of counterfactuals in which the switching cost parameters κ_0 and κ_1 are jointly reduced by 10%, 30%, 50%, 70%, and 90% relative to their estimated values, holding all other parameters at their baseline. For each reduction level, I re-solve the full stationary equilibrium under both imperfect information (the baseline learning environment) and perfect information (workers know their true skills at entry). Comparing these two equilibria at each switching cost level yields the *cost of mismatch*: the aggregate output loss attributable to imperfect information at that level of occupational mobility frictions. I also compare the imperfect-information equilibria across switching cost levels to measure the *direct effect* of reducing switching costs on aggregate output.

Switching costs and the cost of mismatch. Figure 14 reports the results from the within-scenario comparison—imperfect versus perfect information at each switching cost level. The aggregate output loss from mismatch declines monotonically as switching costs fall: from 6.6% under baseline costs, to 6.0% with a 10% reduction, 4.8% with a 30% reduction, 3.7% with a 50% reduction, 2.8% with a 70% reduction, and 2.3% with a 90% reduction (Panel (a)). This monotonic decline reveals a key complementarity between information frictions and mobility frictions: switching costs amplify the damage of imperfect information by locking workers into poor matches. When switching is cheap, workers can correct misallocations through experimentation—trying occupations, observing wage signals, and switching again—even without knowing their true comparative advantage ex ante. The cost of not knowing thus falls, because the consequences of being wrong are less persistent.



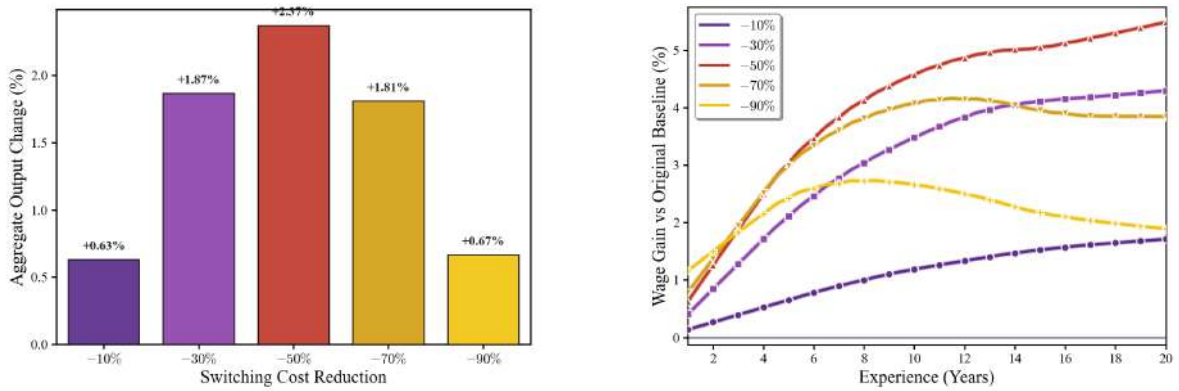
(a) Aggregate output losses from mismatch

(b) Lifecycle wage cost of imperfect information

Figure 14: *Panel (a)*: Bars show the aggregate output loss from mismatch (vertical axis: percent; gap between perfect and imperfect information stationary equilibria) for each switching cost level. The cost of mismatch declines monotonically as switching costs are reduced. *Panel (b)*: Lifecycle profile of the wage cost of imperfect information (vertical axis: percent; horizontal axis: years of potential experience) for each switching cost level.

The lifecycle profile reveals where the reduction is concentrated, and delivers a striking additional result (Figure 14, Panel (b)). Under baseline switching costs the wage gap declines slowly. Under reduced switching costs it falls much faster—and then turns slightly negative after approximately fourteen years of experience. The wage decomposition under reduced switching costs (Figure 32 in the Appendix) clarifies why. The Skills channel remains positive throughout: workers with low switching costs progressively sort into better-matched occupations as they learn, so the sorting gain relative to perfect-information workers keeps shrinking but never reverses. The Experience channel, however, turns negative and deepens over time. Workers under imperfect information who spent their early career in a different occupation have, by the time they switch, accumulated experience in that occupation's task dimension. Because output depends on experience in both skill dimensions, this cross-task experience turns out to be productive: it generates returns that workers who sorted immediately into their optimal occupation—and accumulated experience almost exclusively along one dimension—do not have. After approximately fourteen years, this experience advantage outweighs the remaining sorting disadvantage, and the wage gap turns negative.

The direct effect of reducing switching costs on output. The within-scenario comparison measures how the *cost of imperfect information* varies with switching costs, but a policymaker also cares about the total effect on output. Figure 15 reports aggregate output in the imperfect-information economy (the empirically relevant scenario) as a function of the switching cost reduction. Output increases with moderate reductions, peaking at approximately +2.4% for a 50% reduction, but then declines with further reductions.



(a) Aggregate output change

(b) Lifecycle wage gains vs. original baseline

Figure 15: *Panel (a)*: Aggregate output change (percent, relative to the baseline imperfect-information economy) for each switching cost reduction level. Output increases non-monotonically, peaking at a 50% reduction. *Panel (b)*: Lifecycle wage gains (vertical axis: percent; horizontal axis: years of experience) for each reduced switching cost scenario relative to the original imperfect-information baseline.

This non-monotonicity reflects two opposing forces. On one hand, lower switching costs allow workers to reallocate toward better-matched occupations, improving sorting and raising output. On the other hand, when switching costs become very low, non-productivity-related factors—idiosyncratic taste shocks and occupation-specific amenities—become relatively more important in occupational choice. Workers switch more frequently in response to transitory preference shocks, destroying task-specific human capital accumulated in their previous occupation. At large reductions, this churning effect dominates the sorting gain, and aggregate output falls.

These results carry a nuanced policy implication. Moderate reductions in the costs of occupational mobility—reduced search frictions or subsidized career transitions—yield substantial returns: both the cost of mismatch and aggregate output improve. However, the non-monotonicity implies that eliminating switching costs entirely would be counterproductive. The optimal policy balances the benefit of allowing workers to correct mismatches against the cost of excessive mobility driven by non-productivity factors. The monotonic decline in the cost of mismatch, combined with the non-monotonic output response, suggests that switching costs serve a dual role in the economy: they lock workers into poor matches but also discourage unproductive churning.

Forced Exploration at the Beginning of the Career. The switching costs counterfactual shows that frictions to mobility are a key driver of persistent mismatch, but even with low switching costs the baseline information structure limits how quickly workers can sort opti-

mally. A complementary policy question is whether a targeted intervention at the very start of the career—before any experience or wage signal has been observed—can accelerate learning and improve lifetime sorting. I study a one-period forced exploration policy: in their first period, workers are assigned to the occupation that maximizes the information gain about their latent skills. I measure the learning value of occupation j for entrants in group g as the mutual information between skills and the first wage signal,

$$IG_{jg}^{\text{entry}} \equiv I(\xi_i; s_{ij1} \mid g) = \frac{1}{2} \log\left(1 + \frac{\boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j}{\varsigma_{\varepsilon j}^2}\right), \tag{34}$$

where $\boldsymbol{\lambda}_j$ are the skill loadings of occupation j, $\boldsymbol{\Sigma}_g$ is the prior skill covariance of group g, and $\varsigma_{\varepsilon j}^2$ is the occupation-specific wage noise variance. This has a natural signal-to-noise interpretation: an occupation is more informative when it loads strongly on dimensions about which entrants are ex ante uncertain and when wages are a precise signal of skills[21]. Each worker is assigned to $\arg\max_j IG_{jg}^{\text{entry}}$ in their first period. From period 2 onward they choose freely under their updated beliefs. The assignment is welfare-neutral in the sense that workers receive the going market wage in the assigned occupation; the only constraint is on occupational choice in the first period.

The forced occupations are those that best reveal a worker's comparative advantage: manual-intensive occupations (Occ_6) for high-educated workers and medium-educated males, and cognitive-intensive occupations (Occ_7 and Occ_8) for low-educated groups and medium-educated females. These are the occupations that load most strongly on one skill dimension relative to the prior uncertainty, maximizing the signal-to-noise ratio of the first wage observation. Figure 17 shows the occupational shares at different career stages under forced exploration relative to the baseline.

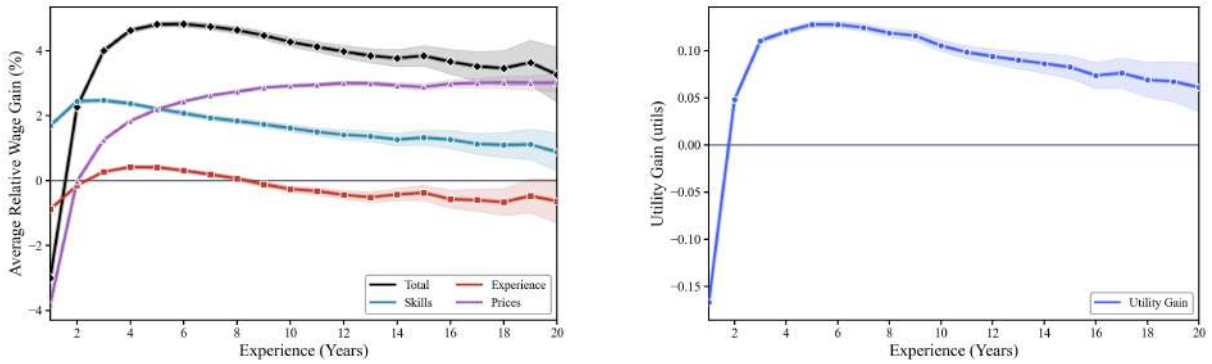
Figures 16 and 17 together tell a clear story. In period 1, workers are concentrated in the few informative occupations (Figure 17), a sharp departure from the baseline sorting. This forced bunching generates a GE price response: excess labor supply in the assigned occupations drives their prices down, while prices in other occupations rise, which explains the wage and utility loss in period 1 visible in Figure 16. Already in period 2, workers are free to choose and quickly switch to better-matched occupations: armed with much more precise beliefs about their comparative advantage, a large fraction immediately transitions into the occupation that truly fits their skills. The occupation shares quickly converge back toward the baseline—by around period 10 the aggregate distribution looks very similar—but now workers are much

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

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

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

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



(a) Wage decomposition: Forced exploration vs Baseline

(b) Utility gains: Forced exploration vs Baseline

Figure 16: *Panel (a)*: Lifecycle wage decomposition of the gain from forced exploration relative to the baseline, by channel (Skills, Experience, Prices, and Residual). *Panel (b)*: Lifecycle utility gains from forced exploration relative to the baseline. Horizontal axis: years of potential experience; vertical axis: percent gain. Both panels show an initial cost in period 1 followed by persistent gains over the rest of the career.

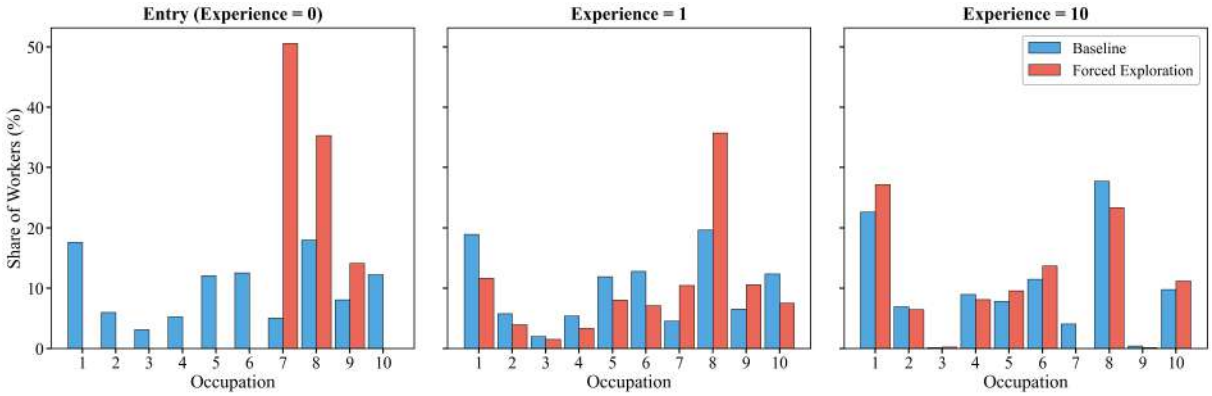


Figure 17: Occupational shares in the forced exploration counterfactual versus the baseline at different career stages.

better matched within each occupation. This rapid reallocation generates a persistent Skills gain that dominates throughout the career. The Price channel, initially negative due to the first-period distortion, normalizes over a few periods as the economy adjusts to the new sorting pattern. The net result is that the one-period nudge more than pays for itself: workers bear a brief first-period cost but recover it quickly through better sorting, with both wages and utility persistently above baseline.

These results highlight a key asymmetry in the value of early information. Because the first wage signal is the most valuable—it arrives when prior uncertainty is highest—even a single period of optimally chosen exploration substantially improves the quality of all subsequent sorting decisions. A policy that temporarily redirects workers toward their most informative occupation at labor market entry can thus be both productivity and welfare enhancing, at the cost of a one-period disruption that the general equilibrium and individual wage gains quickly offset.

7.4 Artificial Intelligence Counterfactuals

The estimated model provides the ideal framework to study how workers respond to changes in the returns to specific skills—such as those induced by artificial intelligence (AI)—when they must learn about their comparative advantage over time. In this section, I use the model to analyze the labor market effects of a task-biased technological shock that differentially affects the returns to cognitive and manual skills. The analysis compares stationary equilibria with imperfect information about skills to quantify how AI adoption reshapes occupational sorting, wage distributions, and the value of information.

7.4.1 AI Shock Design

The production function introduced in Section 4 provides a natural framework for analyzing the effects of AI-driven job transformation. Recall that tasks aggregate according to a Cobb-Douglas technology with importance weights ω_{jk} as exponents, which govern the relative importance of each task type for occupation j. When AI automates or augments tasks requiring certain skills, the importance of those skills for production changes. This shift in task importance has two effects. First, under optimal firm behavior, time allocation adjusts to match the new task importance weights ($\ell^*_{jk} = \omega_{jk}$), so workers spend different amounts of time on tasks requiring each skill type. Second, and more directly, the effective skill loadings $\lambda_{jk} = \omega_{jk}\lambda^0_{jk}$ and experience returns $\beta_{jk} = \omega_{jk}\beta^0_{jk}$ change proportionally with the task importance weights.

As previously mentioned, the baseline task importance weights ω_{jk} is calibrated using occupation-specific task importance measures from Mihaylov and Tijdens (2019), who provide detailed measures of cognitive and manual task requirements for ISCO-08 occupations.

AI (LLM) Exposure Framework. To construct occupation-specific AI shocks, I adopt the exposure framework developed by Eloundou et al. (2024). Their approach defines AI exposure based on whether access to a large language model (LLM) would reduce the time required to complete a task by at least 50% while maintaining quality. They propose four exposure categories:

- *E0 (No Exposure)*: The task cannot be meaningfully accelerated by LLMs.
- *E1 (Direct LLM Exposure)*: The task can be directly performed or substantially accelerated by a text-based LLM (e.g., writing, coding, summarization).
- *E2 (LLM+ Exposure)*: The task requires LLM capabilities combined with additional software tools (e.g., data analysis with code execution, document processing).
- *E3 (Image Exposure)*: The task requires AI systems with image recognition or generation capabilities.

To have a more granular classification of how LLM can affect different occupation, I apply this framework to the 296 Level 3 skills in the ESCO (European Skills, Competences, Qualifications and Occupations) classification, which provides granular skill requirements for 125 ISCO-08

3-digit occupations[22]. To classify each skill into one of the four exposure categories, I use the large language model Claude Opus 4.5 (Anthropic), prompting it with the Eloundou et al. (2024) criteria and the skill description. The model evaluates whether the skill involves text generation (*E1*), data processing with tools (*E2*), visual analysis (*E3*), or physical/interpersonal tasks that LLMs cannot perform (*E0*).

A key finding from this classification is that AI exposure is concentrated in cognitive skills. Of the 296 ESCO skills, 138 are classified as manual and all receive *E0* (no exposure)—LLMs cannot perform physical tasks such as construction, manufacturing, or patient care. Among the 158 cognitive skills, the distribution is: 37 skills (23%) with *E0*, 41 skills (26%) with *E1*, 66 skills (42%) with *E2*, and 14 skills (9%) with *E3*. This asymmetry implies that AI shocks operate primarily through the cognitive dimension of the task space, leaving manual task requirements largely unaffected.

From Exposure to Skill Importance Shocks. Following Eloundou et al. (2024), I assign exposure weights: $\phi_{E0} = 0$, $\phi_{E1} = 1$, $\phi_{E2} = 0.5$, and $\phi_{E3} = 0.5$. For each occupation j, I compute a weighted average exposure for the cognitive dimension:

$$\alpha_{jC} = \frac{\sum_{s \in \text{Cognitive}} \phi_{E(s)} \cdot \text{importance}_{sj}}{\sum_{s \in \text{Cognitive}} \text{importance}_{sj}}, \tag{35}$$

where $E(s)$ is the exposure category of skill s and importance_{sj} is the importance of skill s in occupation j from the ESCO skills-occupations matrix. The manual exposure $\alpha_{jM} = 0$ for all occupations since all manual skills have E0 exposure.

Therefore, the AI shock reduces the cognitive task importance proportionally to exposure:

$$\omega_{jC}^{\text{new}} = \omega_{jC} \cdot (1 - \alpha_{jC}), \qquad \omega_{jM}^{\text{new}} = \omega_{jM}, \tag{36}$$

where $\alpha_{jC} \in [0, 1]$ is the calibrated shock magnitude. The task importance weights are then renormalized so that $\omega_{jC}^{\text{new}} + \omega_{jM}^{\text{new}} = 1$. This renormalization implies that while the cognitive skill importance decreases in absolute terms, the manual skill importance increases mechanically to maintain the constraint that task importance weights sum to one within each occupation. The change in task importance weights ω_{jk} propagates through the model as follows. Since the effective parameters are $\lambda_{jk} = \omega_{jk}\lambda_{jk}^0$ and $\beta_{jk} = \omega_{jk}\beta_{jk}^0$, where λ_{jk}^0 and β_{jk}^0 are primitive occupation-task-level elasticities, both the skill loadings and experience returns scale proportionally with the task importance weights: $\lambda_{jk}^{\text{new}} = \lambda_{jk} \cdot (\omega_{jk}^{\text{new}}/\omega_{jk})$ and $\beta_{jk}^{\text{new}} = \beta_{jk} \cdot (\omega_{jk}^{\text{new}}/\omega_{jk})$. When AI reduces the importance of cognitive skills (lower ω_{jC}), cognitive skills and experience contribute less to output and wages. Additionally, task-specific experience accumulation rates equal the task importance weights, so workers accumulate less cognitive experience because they spend less time performing tasks requiring cognitive skills. Table 9 in the Appendix reports the resulting parameter changes for each occupation cluster, and Figure 33 provides examples of the LLM classification output.

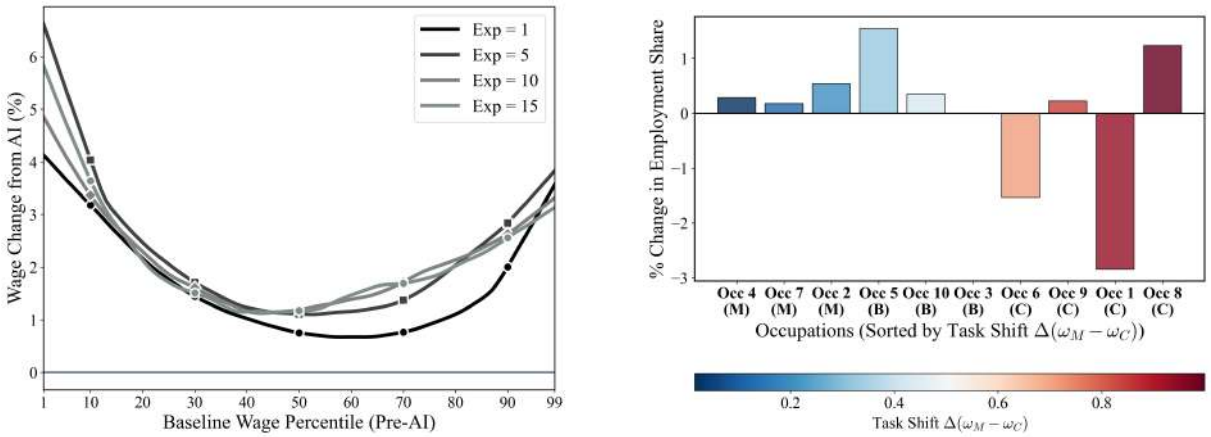
This methodology generates heterogeneous shocks across occupations based on their AI expo-

[22]Skill requirements is not exactly the same as task content. The underlying assumptions here is that there is a one-to-one mapping between skills and tasks, so that "cognitive" skills are used to perform "cognitive" tasks, for example.

sure. Occupations with high cognitive exposure—such as clerical work, data processing, and routine information tasks—experience larger reductions in cognitive skill importance, while occupations intensive in interpersonal interaction, management, and creative tasks have lower exposure and smaller shocks. Crucially, manual-intensive occupations are largely unaffected by LLM-based AI, as physical tasks remain outside the capability of current language models.

7.4.2 Stationary Equilibrium Comparison under Imperfect Information

I compare the pre-shock and post-shock stationary equilibria when workers have imperfect information about their skills. The AI shock changes the allocation of time across tasks within occupations, thereby affecting not only the returns to skills and experience, but also the accumulation rate of task-specific experience.



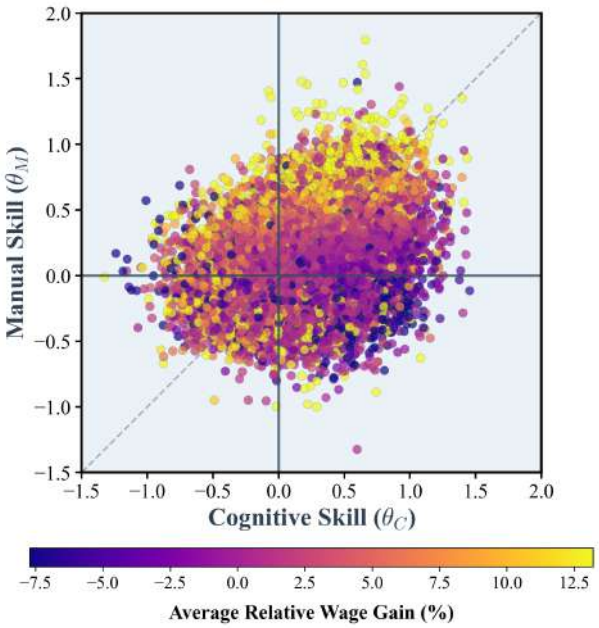
(a) Wage gains from AI by wage percentile

(b) Detailed reallocation by task change across occupations

Figure 18: *Panel (a)*: Wage gains from AI adoption by baseline wage percentile (horizontal axis) for workers with different years of experience (separate lines); vertical axis shows percent wage change. *Panel (b)*: Occupation-specific task reallocation patterns following AI adoption, showing how workers in each occupation redistribute across occupations with different task content.

AI generates overall wage increases, but with substantial heterogeneity across the wage distribution (Figure 18). Panel (a) reveals a striking U-shaped pattern: gains are larger at both the bottom and top of the cross-sectional wage distribution across all experience levels. This pattern is consistent with the job polarization documented in the routinization literature (Autor et al. 2003), where technological change hollows out middle-skill, routine-intensive occupations while benefiting both high-skill and low-skill workers in non-routine tasks. In my framework, polarization arises mechanically: LLM-based AI disproportionately affects cognitive-routine tasks concentrated in the middle of the occupational wage ladder, while leaving manual-intensive occupations at the bottom and pushing more balanced occupations at the top. Across demographic groups, high-educated workers experience the largest average wage gains, with high-educated men benefiting the most, followed by high-educated women (see Figure 30 in the Appendix for the full demographic breakdown). This pattern reflects skill composition: high-educated workers tend to have higher absolute skill levels in both dimensions, including manual skills that become relatively more valuable under AI. Notably, among low-educated workers, women gain more than men, consistent with their stronger manual

comparative advantage within this education group. Panel (b) shows that AI generates substantial occupational reallocation, with employment shifting away from cognitive-intensive occupations toward more manual-intensive or balanced occupations. As the AI shock reduces cognitive task importance, some previously cognitive-heavy occupations become more balanced or even manual-oriented, attracting workers whose manual comparative advantage is now better rewarded. Despite the net employment inflow into manual-intensive occupations, general equilibrium effects lead occupation prices in these occupations to increase. This price increase reflects the productivity gains from AI adoption: even with more workers sorting into these occupations, the higher effective productivity raises the marginal product of labor and thus equilibrium wages.



(a) Wage gains from AI by worker skill type

Figure 19: Each point represents a simulated worker positioned according to their true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis). Color indicates lifecycle-average wage gain from AI adoption (color scale: yellow/light = positive gains, purple/dark = negative gains). The dashed 45-degree line separates cognitive specialists (below the line, where $\xi_C > \xi_M$) from manual specialists (above the line, where $\xi_M > \xi_C$).

The distribution of AI gains across the skill space reveals the underlying mechanism driving these aggregate patterns (Figure 19). Cognitive specialists—workers positioned in the lower-right region of the skill space where $\xi_C > \xi_M$—are particularly harmed by AI adoption as their primary skill dimension is directly exposed to LLM automation. In contrast, workers with strong manual comparative advantage—those in the upper-left region where $\xi_M > \xi_C$—experience systematically positive wage gains, with high-skilled manual workers (predominantly men) benefiting the most. These workers gain because the LLM-based AI shock reduces the returns to cognitive-intensive tasks, making their manual skill advantage relatively more valuable. Generalists near the 45-degree line also benefit from AI by sorting into occupations that remain balanced but have become relatively more manual-intensive following the shift in task weights. This heterogeneity reflects the occupation-specific nature of the AI shock: since LLMs cannot perform physical tasks, manual-intensive occupations are largely insulated from

automation while cognitive-intensive occupations face the largest productivity shifts.

The reallocation patterns following AI adoption reveal important heterogeneity in how workers adjust their occupational choices (see Figure 29 in the Appendix). Low-wage workers—predominantly those with manual comparative advantage—sort directly into more manual-intensive occupations, as these jobs become relatively more attractive when cognitive tasks are automated. In contrast, high-wage workers maintain their sorting into cognitive or balanced occupations even after the AI shock. Among high-wage workers, those who benefit most are the highly educated individuals with manual comparative advantage: these workers can leverage their high absolute skill levels while their manual skills become relatively more valuable. Cognitive specialists continue sorting into cognitive-intensive occupations, and experience lower returns as the task content of their jobs shifts toward automation-exposed activities.

7.4.3 AI Increases the Wage Cost of Early Career Skills Mismatch

I now perform the same counterfactual analysis as in Section 7.2 to quantify how AI adoption affects the value of removing information frictions. I simulate stationary equilibria with perfect information about skills using the post-AI task importance. The value of information is then calculated as the percentage difference in average lifetime wages between the perfect information and imperfect information equilibria. A key finding from these counterfactuals is that AI adoption increases the value of removing information frictions—the wage gain workers would obtain if they knew their true skills from labor market entry.

Table 1: Aggregate Output and Mismatch Cost: Baseline and Post-AI Stationary Equilibria

	Aggregate Output		Mismatch Cost
	Imperfect Info	Perfect Info	(PI − II) / II
Baseline	856.6	913.9	6.7%
Post-AI	918.6	1005.6	9.5%
AI Output Gain	+7.2%	+10.0%	

Notes: Aggregate output in efficiency units. The mismatch cost is the percentage output gap between the perfect (PI) and imperfect (II) information stationary equilibria. The AI output gain rows report the percentage change in output from baseline to post-AI under each information regime. The difference in AI gains (10.0% − 7.2% = 2.8 pp) is the share of AI's potential productivity gains lost to amplified mismatch under imperfect information.

Table 1 reports aggregate output across all four stationary equilibria. The mismatch cost—the aggregate output losses from mismatch between the perfect and imperfect information economies—rises from 6.7% in the baseline to 9.5% post-AI. To see what this implies for AI's aggregate impact, compare the AI output gain across information regimes: under perfect information, AI raises aggregate output by 10.0%; under imperfect information, it raises output by only 7.2%. The difference of 2.8 percentage points represents the share of AI's potential productivity gains that is lost to amplified mismatch. Equivalently, workers under imperfect information capture approximately 72% of the aggregate gains that AI would deliver in a frictionless economy. At the individual level, the wage cost of information frictions rises from 6.3% to 8.6%

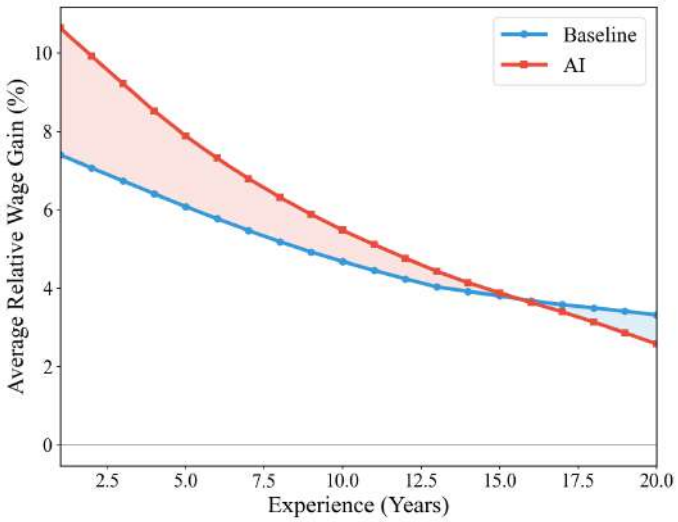
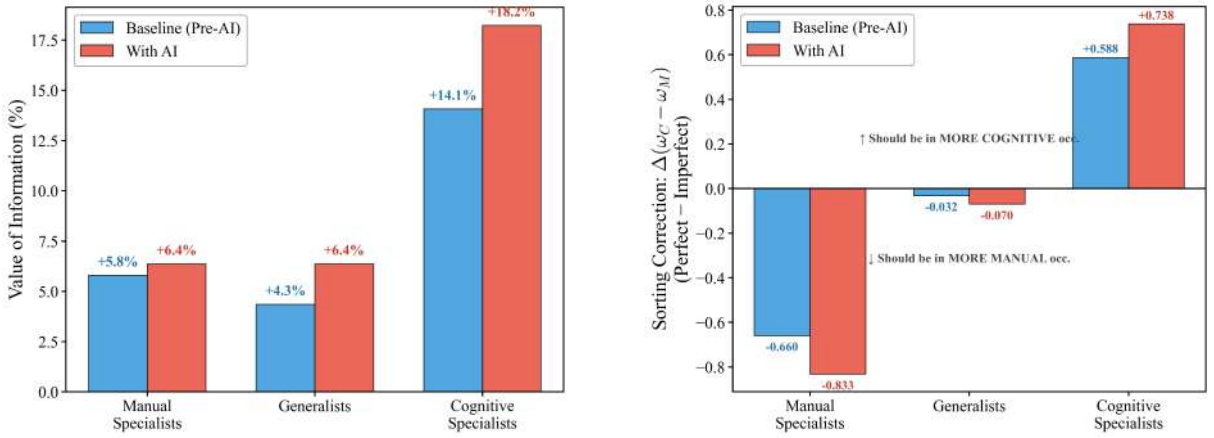


Figure 20: Value of information over the life cycle under baseline and post-AI task structure. Horizontal axis: years of potential experience; vertical axis: percent wage gain from removing information frictions. Separate lines for baseline and AI economies. The breakdown by education-gender group is reported in Figure 31 in the Appendix.



(a) Value of information by skill group

(b) Direction of misallocation

Figure 21: *Panel (a)*: Value of information decomposed by skill group (cognitive specialists, manual specialists, generalists), comparing baseline and post-AI economies. *Panel (b)*: Direction of misallocation under imperfect information, showing how workers sort relative to their true comparative advantage in the baseline versus post-AI economy.

(Figures 20 and 21). While in the baseline economy the misallocation due to information frictions about skills is larger for women than for men, under AI the value of information increases mainly for men. This asymmetry arises because the LLM-based AI shock disproportionately increases the returns to manual skills, in which men tend to have overall higher skill levels at all education levels. More broadly, the increase in the value of information is heterogeneous across demographic groups: low- and medium-educated men experience the largest increase in the wage cost of imperfect information, while high-educated workers of both genders show the smallest increase (see Figure 31 in the Appendix for the full demographic breakdown). This pattern reflects the fact that low- and medium-educated men have the most to gain from improved sorting when the returns to manual skills rise, as their manual comparative advantage is least recognized under imperfect information.

The mechanism connects to the reallocation patterns documented above. Because AI shifts the optimal occupational assignment, workers who were well-matched in the baseline may now be mismatched: cognitive specialists face larger losses from misallocation when their skills are devalued, while workers with unrecognized manual comparative advantage miss larger gains. The value of knowing one's true skills rises because the stakes of correct sorting are higher under the post-AI task structure.

This effect is heterogeneous across skill groups. Cognitive specialists bear the largest increase in the wage cost of information frictions (from 14.1% to 18.2%), as AI directly devalues their primary skill dimension and amplifies the penalty from incorrect sorting. Generalists bear the smallest increase, as their balanced skill profiles make them approximately equally productive across occupations regardless of the shift in task weights. AI thus amplifies the specialist/-generalist asymmetry: specialists have more to lose from misallocation when the returns to their comparative advantage shift, making accurate self-knowledge more valuable precisely for those workers whose skills are most affected by technological change.

Conclusion

I develop and estimate a general equilibrium dynamic Roy model to quantify the consequences of early-career mismatch on individual labor market outcomes and aggregate output. In the model, workers have unobserved multidimensional skills, accumulate task-specific experience while working, and learn about their comparative advantage over time through noisy wage signals. Estimating the model on Portuguese administrative data with ten occupations and two skill dimensions—cognitive and manual—I show that it successfully matches key occupational mobility patterns: low-wage workers within an occupation are more likely to move to occupations with substantially different task content, sorting mainly on comparative advantage, while high-wage workers specialize and move upward along the occupational job ladder.

Using the estimated model, I find that early career mismatch generates a 6.7% aggregate productivity loss, driven by workers sorting into occupations that do not match their true comparative advantage. At the individual level, the wage cost is 7.5% at labor market entry and declines to 3.3% after twenty years as workers learn and change occupation accordingly, but is never fully eliminated. The costs are heterogeneous: under perfect information, 30% of workers would choose a different initial occupation and 50% would switch at some point over the career, with average wage gains of 17.5% for those who reallocate. The losses are concentrated among specialists—workers with pronounced comparative advantage in one skill dimension—who have the most to gain from correct sorting; generalists, whose skills are balanced, lose little from uncertainty. Across demographic groups, female workers and low-educated workers benefit the most from improved sorting, reflecting greater initial uncertainty about comparative advantage in these groups.

The interaction of learning, imperfectly transferable experience, and switching costs generates strong path dependence in occupational choices, preventing a substantial fraction of workers from sorting into their optimal occupation even as they learn. Switching costs are the key friction sustaining this residual mismatch. Progressively reducing switching costs from 10% to 90% reveals that the aggregate cost of mismatch declines monotonically—from 6.6%

to 2.3%—confirming that information frictions per se resolve through learning, but their wage consequences persist because workers cannot cheaply act on what they learn. Lower switching costs allow workers to correct mismatches through occupational experimentation even under imperfect information. However, the direct effect on aggregate output is non-monotonic: output peaks at a 50% reduction and then declines, as very low switching costs induce excessive mobility driven by transitory preference shocks that destroys task-specific human capital. Switching costs thus play a dual role in the economy: they lock workers into poor matches (harmful) but also discourage unproductive churning (beneficial), implying that moderate—not maximal—reductions in mobility frictions are optimal. A complementary experimentation policy—assigning workers at labor market entry to the most informationally rich occupation for one period before they freely choose—generates persistent wage and utility gains with only short-run losses: workers quickly switch to better-matched occupations in the second period, and after a decade the occupational distribution mirrors the baseline while workers are substantially better matched to their true comparative advantage.

I then study how AI amplifies the costs of mismatch. Modeling AI as a task-biased shock that reduces the importance of cognitive tasks, I find that AI raises aggregate productivity but generates heterogeneous wage effects: workers with manual comparative advantage gain while cognitive specialists face worse outcomes. More importantly, AI raises the aggregate cost of early career skills mismatch: the aggregate productivity loss from information frictions rises from 6.7% to 9.5%, and workers under imperfect information capture only approximately 72% of the productivity gains that AI would deliver in a frictionless economy. The higher wage losses are concentrated early in the career, when information frictions prevent workers from sorting into the occupations where their skills are most valuable under the new task structure, and the increase is largest for low- and medium-educated men, who have the most to gain from improved sorting when returns to manual skills rise.

In ongoing work, I extend the analysis to study the transition dynamics following an AI shock. While the comparison of stationary equilibria quantifies the long-run effects of AI under imperfect information, the transition path captures how the labor market adjusts over time. I compute the perfect foresight equilibrium transition path using a standard backward-forward iteration algorithm (see Appendix D.3 for technical details). This extension will focus on two key objects: the speed of convergence to the new stationary equilibrium and welfare along the transition path. This analysis will provide insights into the dynamic costs of skill uncertainty when the economy faces technological change, and help evaluate policies that could accelerate efficient reallocation in response to AI.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review*, 108(6):1488–1542.

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.

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

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

Arcidiacono, P., Aucejo, E., Maurel, A., and Ransom, T. (2025). College attrition and the dynamics of information revelation. *Journal of Political Economy*, 133(1):53–110.

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

Autor, D. H. and 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., 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.

Chiong, K., Galichon, A., and Shum, M. (2013). Duality in dynamic discrete choice models. *Quantitative Economics*, 4(3):569–594.

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.

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. (1996a). Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047.

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

Fonseca, T., Lima, F., and Pereira, S. C. (2018). Job polarization, technological change and routinization: Evidence for portugal. *Labour Economics*, 51:317–339.

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

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. (2005a). Comparative advantage, learning, and sectoral wage determination. *Journal of Labor Economics*, 23(4):681–724.

Gibbons, R., Katz, L. F., Lemieux, T., and Parent, D. (2005b). Comparative advantage, learning, and sectoral wage determination. *Journal of labor economics*, 23(4):681–724.

Gibbons, R. and Waldman, M. (1999a). Careers in organizations: Theory and evidence. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, pages 2373–2437. Elsevier.

Gibbons, R. and Waldman, M. (1999b). 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.

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*, pages 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.

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.

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

Kiss, A., Garlick, R., Orkin, K., and Hensel, L. (2023). Jobseekers' beliefs about comparative advantage and (mis)directed search. Technical Report 23-388, W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

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

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

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

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

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

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

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

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

Papageorgiou, T. (2014). Learning your comparative advantages. *Review of Economic Studies*, 81(3):1263–1295.

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

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

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

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

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

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

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

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

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

A Appendix: Proofs

A.1 Wage Equation Parameters Identification Proof

This proof establishes point identification of

$$\left\{(p_j, \boldsymbol{\lambda}_j, \boldsymbol{\beta}_j, \varsigma^2_{\varepsilon j})_{j\in\mathcal{O}},\ (\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g\in\mathcal{G}}\right\}$$

under Assumptions 1–5. Because the environment is stationary, $(p_j, \boldsymbol{\lambda}_j, \boldsymbol{\beta}_j, \varsigma^2_{\varepsilon j})$ are constant across observations. Let ℓ index observations along a worker's history (suppressing calendar time). In observation ℓ, worker i is in occupation $d_{i\ell} \in \mathcal{O}$ with wage $w_{i\ell}$ and experience vector $\boldsymbol{\tau}_{i\ell}$. The wage equation is

$$w_{i\ell} = p_{d_{i\ell}} + \boldsymbol{\lambda}'_{d_{i\ell}}\boldsymbol{\xi}_i + \boldsymbol{\beta}'_{d_{i\ell}} \log \boldsymbol{\tau}_{i\ell} + \varepsilon_{i\ell}, \qquad \varepsilon_{i\ell} \mid (d_{i\ell} = j) \sim \mathcal{N}(0, \varsigma^2_{\varepsilon j}),$$

with $\varepsilon_{i\ell}$ independent across observations and independent of $(\boldsymbol{\xi}_i, \mathcal{H}_{i\ell})$. Fix an occupation j. Consider any worker i observed in occupation j in two observations ℓ and ℓ' (Assumption 2). Subtracting the wage equation across the two observations eliminates the time-invariant term $p_j + \boldsymbol{\lambda}'_j\boldsymbol{\xi}_i$:

$$w_{i\ell} - w_{i\ell'} = \boldsymbol{\beta}'_j(\log \boldsymbol{\tau}_{i\ell} - \log \boldsymbol{\tau}_{i\ell'}) + (\varepsilon_{i\ell} - \varepsilon_{i\ell'}).$$

Because $\varepsilon_{i\ell}$ is mean-zero and independent of $\boldsymbol{\tau}_{i\ell}$, the coefficient vector $\boldsymbol{\beta}_j$ is identified from the conditional linear projection of $(w_{i\ell} - w_{i\ell'})$ on $(\log \boldsymbol{\tau}_{i\ell} - \log \boldsymbol{\tau}_{i\ell'})$ within the subpopulation of stayers in j, provided the differences in $\log \boldsymbol{\tau}$ have full rank on a set of positive probability (Assumption 1(iii) and 2). Repeating for all j identifies $\{\boldsymbol{\beta}_j\}_{j\in\mathcal{O}}$. Using the same within-occupation pairs, define the differenced residual

$$u_{i,\ell\ell'}(j) \equiv (w_{i\ell} - w_{i\ell'}) - \boldsymbol{\beta}'_j(\log \boldsymbol{\tau}_{i\ell} - \log \boldsymbol{\tau}_{i\ell'}) = \varepsilon_{i\ell} - \varepsilon_{i\ell'}.$$

By independence across observations,

$$\mathbb{V}ar\big(u_{i,\ell\ell'}(j) \mid d_{i\ell} = d_{i\ell'} = j\big) = \mathbb{V}ar(\varepsilon_{i\ell} - \varepsilon_{i\ell'} \mid d = j) = 2\varsigma^2_{\varepsilon j},$$

so $\varsigma^2_{\varepsilon j}$ is identified for each occupation j. With $\boldsymbol{\beta}_j$ identified, define the residualized wage

$$r_{i\ell} \equiv w_{i\ell} - \boldsymbol{\beta}'_{d_{i\ell}} \log \boldsymbol{\tau}_{i\ell} = p_{d_{i\ell}} + \boldsymbol{\lambda}'_{d_{i\ell}}\boldsymbol{\xi}_i + \varepsilon_{i\ell}.$$

Thus, conditional on the observed occupation sequence, $\{r_{i\ell}\}$ is a linear measurement system for the fixed latent vector $\boldsymbol{\xi}_i$, with occupation-specific loadings and i.i.d. transitory noise. Without normalization, the mapping $(\boldsymbol{\lambda}_j, \boldsymbol{\xi}_i) \mapsto (A'\boldsymbol{\lambda}_j, A^{-1}\boldsymbol{\xi}_i)$ leaves $\boldsymbol{\lambda}'_j\boldsymbol{\xi}_i$ invariant for any invertible A. With an unrestricted covariance matrix, this indeterminacy has K^2 degrees of freedom; when $K = 2$, this is 4 degrees of freedom. Fixing two loading vectors $\boldsymbol{\lambda}_{j_C} = (1, 0)'$ and $\boldsymbol{\lambda}_{j_M} = (0, 1)'$ imposes exactly four restrictions and forces $A = I_2$, delivering uniqueness of the recovered loadings and the full (unnormalized) covariance matrix $\boldsymbol{\Sigma}_g$. Let j_C and j_M denote

the pure cognitive and pure manual anchor occupations, normalized as in Assumption **??**:

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

For any observation in the cognitive anchor occupation j_C,

$$r_{i\ell} = p_{j_C} + \xi_{iC} + \varepsilon_{i\ell},$$

and for any observation in the manual anchor occupation j_M,

$$r_{i\ell} = p_{j_M} + \xi_{iM} + \varepsilon_{i\ell}.$$

Taking expectations conditional on group g yields

$$\mathbb{E}[r \mid g, d = j_C] = p_{j_C} + \mu_{gC}, \qquad \mathbb{E}[r \mid g, d = j_M] = p_{j_M} + \mu_{gM}.$$

Impose the location normalization $\boldsymbol{\mu}_{g^\star} = \mathbf{0}$ (Assumption 5). Then

$$p_{j_C} = \mathbb{E}[r \mid g^\star, d = j_C], \qquad p_{j_M} = \mathbb{E}[r \mid g^\star, d = j_M],$$

and for any group g,

$$\mu_{gC} = \mathbb{E}[r \mid g, d = j_C] - \mathbb{E}[r \mid g^\star, d = j_C], \qquad \mu_{gM} = \mathbb{E}[r \mid g, d = j_M] - \mathbb{E}[r \mid g^\star, d = j_M].$$

Hence (p_{j_C}, p_{j_M}) and all group mean vectors $\{\boldsymbol{\mu}_g\}$ are identified. Consider workers in group g observed at least twice in the cognitive anchor (Assumption 2). For two such observations $\ell \neq \ell'$ with $d_{i\ell} = d_{i\ell'} = j_C$,

$$\mathbb{C}ov(r_{i\ell}, r_{i\ell'} \mid g) = \mathbb{C}ov(\xi_{iC} + \varepsilon_{i\ell}, \xi_{iC} + \varepsilon_{i\ell'} \mid g) = \mathbb{V}ar(\xi_{iC} \mid g) = \Sigma_{g,CC},$$

because ε is independent across observations. Similarly, for two observations in the manual anchor,

$$\mathbb{C}ov(r_{i\ell}, r_{i\ell'} \mid g,\, d_{i\ell} = d_{i\ell'} = j_M) = \Sigma_{g,MM}.$$

Finally, for workers observed once in each anchor (in any order), say $d_{i\ell} = j_C$ and $d_{i\ell'} = j_M$,

$$\mathbb{C}ov(r_{i\ell}, r_{i\ell'} \mid g) = \mathbb{C}ov(\xi_{iC} + \varepsilon_{i\ell}, \xi_{iM} + \varepsilon_{i\ell'} \mid g) = \mathbb{C}ov(\xi_{iC}, \xi_{iM} \mid g) = \Sigma_{g,CM}.$$

Thus all elements of the 2×2 covariance matrix $\boldsymbol{\Sigma}_g$ are identified for each group g. Fix any non-anchor occupation j. Consider workers in a given group g who are observed at least once in occupation j and at least once in each anchor (Assumption 2). Let ℓ be an observation with $d_{i\ell} = j$, and let ℓ_C and ℓ_M denote (possibly different) observations with $d_{i\ell_C} = j_C$ and $d_{i\ell_M} = j_M$. Centering is not required for covariance calculations because constants drop out. Using the residualized equation,

$$r_{i\ell} = p_j + \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{i\ell}.$$

Then

$$\mathbb{C}ov(r_{i\ell}, r_{i\ell_C} \mid g) = \mathbb{C}ov(\boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{i\ell},\ \xi_{iC} + \varepsilon_{i\ell_C} \mid g) = \boldsymbol{\lambda}_j' \mathbb{C}ov(\boldsymbol{\xi}_i, \xi_{iC} \mid g) = \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{e}_C,$$

and similarly

$$\mathbb{C}ov(r_{i\ell}, r_{i\ell_M} \mid g) = \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{e}_M,$$

where $\boldsymbol{e}_C = (1,0)'$ and $\boldsymbol{e}_M = (0,1)'$. Stack these two covariances in a vector:

$$\boldsymbol{c}_{jg} \equiv \begin{bmatrix} \mathbb{C}ov(r_j, r_{j_C} \mid g) \\ \mathbb{C}ov(r_j, r_{j_M} \mid g) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{e}_C \\ \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{e}_M \end{bmatrix} = \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j,$$

where the last equality uses symmetry of $\boldsymbol{\Sigma}_g$. Since $\boldsymbol{\Sigma}_g$ is positive definite (hence invertible), we obtain the explicit inversion formula

$$\boldsymbol{\lambda}_j = \boldsymbol{\Sigma}_g^{-1} \boldsymbol{c}_{jg}.$$

Therefore all non-anchor loading vectors $\{\boldsymbol{\lambda}_j\}$ are identified. Because $\boldsymbol{\lambda}_j$ is common across groups, different groups provide overidentifying restrictions. For any occupation j and group g, taking expectations of the residualized equation gives

$$\mathbb{E}[r \mid g, d = j] = p_j + \boldsymbol{\lambda}_j' \boldsymbol{\mu}_g.$$

Using the reference group normalization $\boldsymbol{\mu}_{g^\star} = \mathbf{0}$ and the now-identified $\boldsymbol{\lambda}_j$,

$$p_j = \mathbb{E}[r \mid g^\star, d = j].$$

Hence all occupation prices $\{p_j\}$ are identified.

The above arguments provide a constructive identification mapping from observable moments of wages to the parameter vector

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

Because $\boldsymbol{\lambda}_j$ and $\boldsymbol{\beta}_j$ are constant in the stationary environment, identification relies on occupational switching that generates covariances with the anchor occupations rather than on a large number of calendar periods.

A.2 Identification of Dynamic Choice Parameters

From the first stage, the parameter vector Γ is identified. Hence the econometrician knows (i) expected log wages $r_j(\boldsymbol{x})$ for each j and state $\boldsymbol{x}$ and (ii) the transition kernel $\mathcal{P}(\cdot \mid \boldsymbol{x}, j)$ induced by experience accumulation and Bayesian updating. CCPs $P(d = j \mid \boldsymbol{x})$ are identified from the joint distribution of $(d, \boldsymbol{x})$. With i.i.d. EV–I$(0, \varrho)$ shocks, only utility differences measured in units of ϱ are directly revealed by CCPs. In dynamic learning models, state variation that changes current expected wage gaps generally also changes continuation values, so ϱ is difficult to pin down empirically without additional exogenous payoff shifters that move current utility in known units but do not affect state transitions. I therefore normalize/calibrate the taste–shock scale and set a value for ϱ. Given ϱ, the deterministic index for occupation j in state $\boldsymbol{x}_t$ is

$$v_j(\boldsymbol{x}_t) = r_j(\boldsymbol{x}_t) + \eta_j - \mathbf{1}\{j \neq d_{t-1}\}(\kappa_0 + \kappa_1 e_t) + \beta(1 - \delta(e_t))\, EV_j(\boldsymbol{x}_t),$$

where $EV_j(\boldsymbol{x}_t) \equiv \mathbb{E}[W(\boldsymbol{x}_{t+1}) \mid \boldsymbol{x}_t, d_t = j]$ and $W(\cdot)$ is the integrated ex–ante value function solving (33). Under EV–I shocks, Hotz–Miller inversion implies that CCP log–odds identify index differences:

$$\ell_{j,k}(\boldsymbol{x}_t) \equiv \log \frac{P(d_t = j \mid \boldsymbol{x}_t)}{P(d_t = k \mid \boldsymbol{x}_t)} = \frac{\nu_j(\boldsymbol{x}_t) - \nu_k(\boldsymbol{x}_t)}{\varrho}, \qquad \Rightarrow \qquad \nu_j(\boldsymbol{x}_t) - \nu_k(\boldsymbol{x}_t) = \varrho\, \ell_{j,k}(\boldsymbol{x}_t).$$

Since CCPs are identified, the left-hand side is identified for all j, k and $\boldsymbol{x}_t$. Fix an origin occupation d and consider two destination occupations $j \neq d$ and $j' \neq d$. For any state $\boldsymbol{x}$ with lagged occupation $d_{t-1} = d$, both moves incur the same switching cost $(\kappa_0 + \kappa_1 e)$, so the switching–cost term cancels in the difference:

$$\nu_j(\boldsymbol{x}) - \nu_{j'}(\boldsymbol{x}) = \big[r_j(\boldsymbol{x}) - r_{j'}(\boldsymbol{x})\big] + \big(\eta_j - \eta_{j'}\big) + \beta(1 - \delta(e))\big[EV_j(\boldsymbol{x}) - EV_{j'}(\boldsymbol{x})\big].$$

The left-hand side is identified from CCP log–odds via Hotz–Miller, and $r_j(\boldsymbol{x})$ is known from Stage 1. For any candidate vector of amenities $\{\eta_j\}$ and switching–cost parameters (κ_0, κ_1), the continuation values $EV_j(\boldsymbol{x})$ are uniquely determined by the Bellman fixed point (33) given the known transition kernel. Thus the system of log–odds restrictions across states and occupation pairs identifies amenity differences. Imposing the normalization $\eta_0 = 0$ pins down amenity levels. Next compare a move $j \neq d$ to staying in the origin d at the same state $\boldsymbol{x}$ (so $d_{t-1} = d$). Only the move pays the switching cost:

$$\nu_j(\boldsymbol{x}) - \nu_d(\boldsymbol{x}) = \big[r_j(\boldsymbol{x}) - r_d(\boldsymbol{x})\big] + \big(\eta_j - \eta_d\big) - \big(\kappa_0 + \kappa_1 e\big) + \beta(1 - \delta(e))\big[EV_j(\boldsymbol{x}) - EV_d(\boldsymbol{x})\big].$$

Since $\eta_j - \eta_d$ is identified from the previous step (and $r_j(\boldsymbol{x})$ is known), the remaining term $(\kappa_0 + \kappa_1 e)$ is identified from the way the move–vs–stay index difference varies with experience e. In particular, holding fixed the origin occupation d and varying e traces out the affine schedule $\kappa_0 + \kappa_1 e$, identifying κ_1 from the slope and κ_0 from the intercept. Given the normalization $\varrho = 0.25$ and $\eta_0 = 0$, and under the regularity condition that CCPs are strictly positive on the relevant support and that total experience varies within origin occupations, the dynamic choice parameters $(\{\eta_j\}_{j \in \mathcal{O}}, \kappa_0, \kappa_1)$ are point identified.

A.3 Contraction of the Integrated–Value Bellman Operator

Let the state be $\boldsymbol{x} \equiv (\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}, d_{-1}) \in \mathcal{X}$, where $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\Sigma}})$ are the posterior mean and covariance of log skills (13), $\boldsymbol{\tau}$ are task-specific tenures, and $d_{-1} \in \mathcal{O}$ is last period's occupation. Define total experience $e \equiv \sum_{k=1}^{K} \tau_k$. For each occupation $m \in \mathcal{O}$, define the deterministic utility index

$$r_m(\boldsymbol{x}) \;\equiv\; p_m + \boldsymbol{\lambda}_m' \tilde{\boldsymbol{\xi}} + \boldsymbol{\beta}_m' \log \boldsymbol{\tau} + \eta_m,$$

which is the expected log wage in occupation m given beliefs $\tilde{\boldsymbol{\xi}}$ plus the non-pecuniary amenity η_m (see (9) and (19)). Define the survival probability $s(\boldsymbol{x}) \equiv 1 - \delta(e) \in (0, 1)$ (with $\delta(e) > 0$ for all e by the logistic hazard) and the switching cost $\mathcal{C}(m, d_{-1}, e) = \mathbf{1}\{m \neq d_{-1}\}(\kappa_0 + \kappa_1 e) \geq 0$. For any bounded continuous $W : \mathcal{X} \to \mathbb{R}$, define the integrated-value Bellman operator $\mathcal{T}$:

$C_b(\mathcal{X}) \to C_b(\mathcal{X})$ by

$$(\mathcal{T}W)(\boldsymbol{x}) \;\equiv\; \varrho \log \sum_{m \in \mathcal{O}} \exp\left(\frac{1}{\varrho} \Big[r_m(\boldsymbol{x}) - \mathcal{C}(m, d_{-1}, e) + \beta\, s(\boldsymbol{x})\, \mathbb{E}\big[\, W(\tilde{\boldsymbol{\xi}}', \tilde{\boldsymbol{\Sigma}}', \boldsymbol{\tau}', m) \;\big|\; \boldsymbol{x}, m \,\big] \Big] \right), \tag{37}$$

where $\varrho > 0$ is the scale of the EV-I taste shocks ϵ_{jt}. Following the standard normalization in which the EV-I shocks have mean $-\varrho\gamma$ (i.e. the location parameter absorbs the Euler-Mascheroni constant γ), integrating out ϵ yields the log-sum-exp formula above without the $\varrho\gamma$ additive term; this is consistent with the fixed point (33). The expectation is over $(\tilde{\boldsymbol{\xi}}', \tilde{\boldsymbol{\Sigma}}', \boldsymbol{\tau}')$: beliefs update via Kalman filtering (13) and experience accumulates according to (20), both conditional on $(\boldsymbol{x}, m)$. We endow $C_b(\mathcal{X})$ with the sup norm $\|W\|_\infty = \sup_{\boldsymbol{x} \in \mathcal{X}} |W(\boldsymbol{x})|$.

Claim. $\mathcal{T}$ is a contraction on $(C_b(\mathcal{X}), \|\cdot\|_\infty)$ with modulus at most $\beta \in (0,1)$. Therefore there exists a unique fixed point $W^\star = \mathcal{T}W^\star$.

Proof. We verify Blackwell's sufficient conditions.

(i) Monotonicity. Let $W_1, W_2 \in C_b(\mathcal{X})$ with $W_1 \le W_2$ pointwise. Since beliefs and tenures evolve deterministically given $(\boldsymbol{x}, m)$ via (13) and (20), the conditional expectations satisfy $\mathbb{E}[W_1(\tilde{\boldsymbol{\xi}}', \tilde{\boldsymbol{\Sigma}}', \boldsymbol{\tau}', m) \mid \boldsymbol{x}, m] \le \mathbb{E}[W_2(\tilde{\boldsymbol{\xi}}', \tilde{\boldsymbol{\Sigma}}', \boldsymbol{\tau}', m) \mid \boldsymbol{x}, m]$ for all $(\boldsymbol{x}, m)$. The log-sum-exp function $v \mapsto \varrho \log \sum_m \exp(v_m/\varrho)$ is increasing in each component, and $r_m(\boldsymbol{x})$ and $\mathcal{C}(m, d_{-1}, e)$ do not depend on W. Hence the arguments in (37) increase componentwise when W_1 is replaced by W_2, so $(\mathcal{T}W_1)(\boldsymbol{x}) \le (\mathcal{T}W_2)(\boldsymbol{x})$ for all $\boldsymbol{x}$.

(ii) Discounting. Fix $c \ge 0$ and $W \in C_b(\mathcal{X})$. By linearity of expectation, $\mathbb{E}[W(\cdot) + c] = \mathbb{E}[W(\cdot)] + c$. Therefore,

$$\begin{aligned}
(\mathcal{T}(W + c))(\boldsymbol{x}) &= \varrho \log \sum_{m \in \mathcal{O}} \exp\left(\frac{1}{\varrho} \Big[r_m(\boldsymbol{x}) - \mathcal{C}(m, d_{-1}, e) + \beta\, s(\boldsymbol{x}) \big(\mathbb{E}[W(\cdot)] + c \big) \Big] \right) \\
&= \varrho \log \left(e^{\beta s(\boldsymbol{x}) c/\varrho} \sum_{m \in \mathcal{O}} \exp\left(\frac{1}{\varrho} \Big[r_m(\boldsymbol{x}) - \mathcal{C}(m, d_{-1}, e) + \beta\, s(\boldsymbol{x})\, \mathbb{E}[W(\cdot)] \Big] \right) \right) \\
&= (\mathcal{T}W)(\boldsymbol{x}) + \beta\, s(\boldsymbol{x})\, c.
\end{aligned}$$

Since $s(\boldsymbol{x}) \in (0,1)$ and $c \ge 0$, we have $\beta\, s(\boldsymbol{x})\, c \le \beta c$, so $(\mathcal{T}(W + c))(\boldsymbol{x}) \le (\mathcal{T}W)(\boldsymbol{x}) + \beta c$. Thus $\mathcal{T}$ satisfies Blackwell's discounting condition with modulus β.

By Blackwell's theorem, $\mathcal{T}$ is a contraction with modulus at most $\beta < 1$. The Banach fixed-point theorem then guarantees existence and uniqueness of $W^\star = \mathcal{T}W^\star$, and value-function iteration $W^{(n+1)} = \mathcal{T}W^{(n)}$ converges to $W^\star$ from any initial $W^{(0)} \in C_b(\mathcal{X})$. □

B Appendix: Estimation Results

B.1 Externally Set and Calibrated Parameters

The discount factor $\beta = 0.9$ is a standard value in the dynamic discrete choice literature. Exit probability parameters $\psi = 0.072$ (curvature) and $\bar{a} = 27.22$ (location) are calibrated by fitting a left-truncated logistic hazard model to the empirical survival data (see Figure 35 in the

Table 2: Externally Set Parameters

Parameter	Description	Empirical Moment	Value
Preferences & Demographics			
β	Discount factor		0.900
ψ	Exit curvature	Empirical exit rate	0.072
$\bar{a}$	Exit location	Empirical exit rate	27.22
Education-Gender Distribution			
Low education (L)	Total Share	Empirical shares	36.4%
Male (L_M)	Male Share	Empirical shares	24.4%
Female (L_F)	Female Share	Empirical shares	12.0%
Medium education (M)	Total Share	Empirical shares	33.3%
Male (M_M)	Male Share	Empirical shares	18.5%
Female (M_F)	Female Share	Empirical shares	14.7%
High education (H)	Total Share	Empirical shares	30.3%
Male (H_M)	Male Share	Empirical shares	13.0%
Female (H_F)	Female Share	Empirical shares	17.4%
Human Capital Accumulation			
$\boldsymbol{\omega}_j$	Task importance weights	Mihaylov and Tijdens (2019)	
CES Final-Good Aggregator			
σ	Elasticity of substitution	**?**	1.57
α_j	CES aggregator weights	Stationary equilibrium prices	Calibrated[a]
P	Aggregate price level	Estimated price level	Calibrated[a]

Notes: Most parameters are set (rather than calibrated) to match aggregate statistics or for computational tractability. [a]The CES aggregator weights α_j and the aggregate price level P are calibrated. Given wage shares s_j and prices p_j estimated from the EM algorithm, the weights are computed as $\alpha_j \propto s_j \cdot p_j^{\sigma-1}$ and normalized so that $\sum_{j \in \mathcal{O}} \alpha_j = 1$. The aggregate price level P is set to match the empirical price index level computed from the estimated prices, ensuring that the stationary equilibrium prices recovered from the model match those estimated from the data.

Data Appendix), ensuring simulated career lengths match observed panel lengths. Education-gender shares are set to the estimation sample: low-educated workers comprise 36.4% (24.4% male, 12.0% female), medium-educated 33.3% (18.5% male, 14.7% female), and high-educated 30.3% (13.0% male, 17.4% female). The task importance weights $\boldsymbol{\omega}_j$ are from Mihaylov and Tijdens (2019); ω_{jk} governs the time a worker in occupation j spends on task k, which in turn determines how task-specific experience accumulates. The CES aggregator weights α_j are calibrated from observed wage shares and estimated occupation prices, and the price index $P(p)$ is normalized to match the empirical price level.

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

Table 3 reports the full estimation results. Panel A shows the Cobb-Douglas skill elasticities λ_{jk} and experience elasticities β_{jk} for each occupation, along with idiosyncratic wage variances ς_j^2 and occupation prices p_j. For identification, Occ_7 is normalized to be purely cognitive and Occ_3 purely manual. Three groups emerge naturally: cognitive-intensive occupations (Occ_1, Occ_2, Occ_5, Occ_7) reward cognitive skills and experience more heavily; manual-intensive occupations (Occ_3, Occ_6) reward manual skills; and balanced occupations (Occ_0, Occ_4, Occ_8, Occ_9) draw on both dimensions. Experience elasticities are aligned

Table 3: Stage 1: Expectation-Maximization Estimates of Wage Equation Parameters

Panel A: Loadings, Shock Variances, Occupational Prices, and Returns to Experience

Occupation	λ_{j1} (C)	λ_{j2} (M)	ς_j^2	α_j	$\beta_{j,C}$ (C)	$\beta_{j,M}$ (M)
Occ_0	0.726	0.584	0.0257	1.278	0.186	0.019
Occ_1	0.446	0.8	0.02296	1.241	0.167	0.083
Occ_2	0.495	0.875	0.02197	1.337	0.133	0.058
Occ_3	0	1	0.02184	1.238	0.131	0.107
Occ_4	0.773	0.238	0.02676	1.342	0.135	0.029
Occ_5	0.227	1.005	0.02046	1.31	0.164	0.001
Occ_6	0	1.085	0.01599	1.272	0.127	0.076
Occ_7	1	0	0.02498	1.252	0.22	0.033
Occ_8	0.885	0.207	0.02168	1.381	0.121	0.017
Occ_9	0.427	0.796	0.02085	1.306	0.145	0.007

Panel B: Task Importance Weights by Occupation

Occupation	$\omega_{j,C}$ (Cognitive)	$\omega_{j,M}$ (Manual)	Type
Occ_0	0.936	0.064	C
Occ_1	0.201	0.799	M
Occ_2	0.566	0.434	B
Occ_3	0.015	0.985	M
Occ_4	0.322	0.678	B
Occ_5	0.712	0.288	C
Occ_6	0.117	0.883	M
Occ_7	0.998	0.002	C
Occ_8	0.846	0.154	C
Occ_9	0.432	0.568	B

Panel C: Skills Distribution by Education-Gender Groups

Education	Female	Male
Low (L)	$\begin{pmatrix}-0.131\\-0.147\end{pmatrix}, \begin{pmatrix}0.243 & 0.005\\0.005 & 0.151\end{pmatrix}$ $\rho = 0.137$	$\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}0.332 & -0.005\\-0.005 & 0.227\end{pmatrix}$ $\rho = -0.063$
Medium (M)	$\begin{pmatrix}-0.027\\-0.033\end{pmatrix}, \begin{pmatrix}0.293 & 0.02\\0.02 & 0.209\end{pmatrix}$ $\rho = 0.332$	$\begin{pmatrix}0.124\\0.075\end{pmatrix}, \begin{pmatrix}0.364 & 0.001\\0.001 & 0.278\end{pmatrix}$ $\rho = 0.007$
High (H)	$\begin{pmatrix}0.364\\0.21\end{pmatrix}, \begin{pmatrix}0.302 & 0.027\\0.027 & 0.313\end{pmatrix}$ $\rho = 0.287$	$\begin{pmatrix}0.497\\0.284\end{pmatrix}, \begin{pmatrix}0.301 & 0.019\\0.019 & 0.373\end{pmatrix}$ $\rho = 0.165$

Panel A reports Cobb-Douglas coefficients λ_{jk} (skill elasticities) and β_{jk} (experience elasticities), the idiosyncratic wage–shock variance ς_j^2, and occupational prices α_j by occupation. By normalization *Occ_7* is purely Cognitive ($\lambda_{7,1} = 1$, $\lambda_{7,2} = 0$) and *Occ_3* purely Manual ($\lambda_{3,1} = 0$, $\lambda_{3,2} = 1$). Panel B reports task importance weights ω_{jk} from the Mihaylov and Tijdens (2019) classification, which govern task aggregation and time allocation within occupations. Type indicates occupation classification based on task weights: C = Cognitive-intensive ($\omega_{j,C} \geq 2/3$), M = Manual-intensive ($\omega_{j,M} \geq 2/3$), B = Balanced. Panel C reports EM estimates of prior means (first matrix for each education group) and covariance matrices (second matrix) for Cognitive (θ_C) and Manual (θ_M) skills by education-gender group, with implied correlation ρ shown below. Low-educated male group is normalized to zero mean skills.

with skill elasticities—occupations that value a skill dimension also reward experience in it—and cognitive task experience yields generally steeper returns, consistent with Deming (2025). Panel B reports the task importance weights ω_{jk} from Mihaylov and Tijdens (2019), which govern time allocation within occupations and determine differential exposure to technological shocks. Panel C shows the estimated skill distributions by education-gender group, with low-educated males normalized to zero. Mean skills rise with education and the cognitive-manual gap widens progressively. Cross-skill correlations increase with education and are higher for females, reflecting more balanced skills among high-educated women.

B.3 Stage 2 Estimation Results: Dynamic Choice Parameters

Table 4: Stage 2: NFXP Estimates of Dynamic Choice Parameters

Switching costs: $\kappa(e) = \kappa_0 + \kappa_1 \cdot e$		*Occupation-specific amenities* η_j	
κ_0	0.979	η_0 (normalized)	0.000
κ_1	0.017	η_1	0.042
ϱ (calibrated)	0.250	η_2	-0.205
		η_3	0.060
		η_4	0.067
		η_5	0.054
		η_6	0.026
		η_7	-0.017
		η_8	-0.156
		η_9	0.119

B.4 The Learning Value of Occupations

Occupations differ not only in wage levels and returns to skills and experience, but also in how informative wages are about a worker's latent skills. In the model, after choosing occupation j the worker observes the wage-implied signal

$$s_{ijt} = w_{ijt} - p_j - \boldsymbol{\beta}_j' \log \boldsymbol{\tau}_{it} = \boldsymbol{\lambda}_j' \boldsymbol{\xi}_i + \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}(0, \varsigma_j^2),$$

where $\boldsymbol{\xi}_i$ collects latent log skills and ς_j^2 is the occupation-specific noise variance. A worker in education-gender group g enters with prior $\boldsymbol{\xi} \mid g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. I summarize the informativeness of occupation j for entrants in group g using the expected information gain from one period of wages, defined as the mutual information between skills and the signal[23]:

$$IG_{jg}^{\text{entry}} = I(\boldsymbol{\xi}; s \mid g) = \tfrac{1}{2} \log\left(1 + \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j / \varsigma_j^2\right). \tag{38}$$

The expression has a signal-to-noise interpretation: information gain is higher when occupa-

[23]Under conjugate normal updating, the posterior covariance $\boldsymbol{\Sigma}_{t+1}(j) = \boldsymbol{\Sigma}_g - \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j (\boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j + \varsigma_j^2)^{-1} \boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g$ is deterministic. The expected KL divergence from posterior to prior reduces to $\frac{1}{2} \log(\det \boldsymbol{\Sigma}_g / \det \boldsymbol{\Sigma}_{t+1})$, and the matrix determinant lemma gives (38).

tion j loads strongly on dimensions about which entrants are ex ante uncertain, and when wage noise is small. Because priors differ across groups through $\boldsymbol{\Sigma}_g$, the same occupation can deliver different learning value across groups. Cross-skill correlation enters through the quadratic form $\boldsymbol{\lambda}_j' \boldsymbol{\Sigma}_g \boldsymbol{\lambda}_j$: an occupation loading primarily on one dimension can still update beliefs about the other to the extent that skills are correlated in the prior.

Table 5: Learning Value of Occupations at Entry by Education-Gender Group

	H_Feminino	H_Masculino	L_Feminino	L_Masculino	M_Feminino	M_Masculino
Occ_0	0.810	0.835	0.492	0.665	0.694	0.781
Occ_1	0.839	0.916	0.417	0.588	0.640	0.732
Occ_2	0.935	1.014	0.488	0.672	0.728	0.824
Occ_3	0.851	1.000	0.359	0.606	0.551	0.756
Occ_4	0.642	0.636	0.444	0.626	0.594	0.709
Occ_5	0.948	1.071	0.436	0.656	0.672	0.821
Occ_6	**1.052**	**1.211**	0.494	0.783	0.721	**0.950**
Occ_7	0.768	0.765	**0.606**	**0.844**	0.744	0.921
Occ_8	0.799	0.791	0.591	0.805	**0.755**	0.893
Occ_9	0.866	0.946	0.433	0.608	0.660	0.755

Notes: Entries report IG_{jg}^{entry} in nats. Bold values denote the most informative occupation within each education-gender group.

Table 5 reports IG_{jg}^{entry} implied by the Stage 1 estimates. The occupations that maximize information acquisition are typically specialized rather than balanced. For high-educated workers and medium-educated males, the most informative occupation is Occ_6 (manual-intensive: $\lambda_{6,A} = 0$, $\lambda_{6,M} = 1.085$); for low-educated groups, it is Occ_7 (purely cognitive); for medium-educated females, it is Occ_8 (cognitive-intensive). These rankings reflect the interaction between occupational task content, group-specific prior uncertainty, and the prior correlation structure.

C Appendix: Data and Additional Figures

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

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

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

Because the estimation sample focuses on young entrants, it has a different age and experience composition than the large sample. Table 6 compares the two samples side-by-side. Crucially, the empirical patterns of occupational mobility are qualitatively consistent across the two samples. The descriptive empirical analysis and the structural model comparison use related but not identical wage objects. In the empirical section, I use residualized wages in the large sample to isolate the component of within-occupation wage variation driven by unobserved fit or comparative advantage. By contrast, the structural model is estimated on a more restricted sample and a reduced occupational state space, so it is evaluated against moments based on raw within-cluster wage ranks in the estimation sample, which constitute the natural model-consistent empirical counterpart.

Table 6: Comparison of the Large Sample and the Estimation Sample

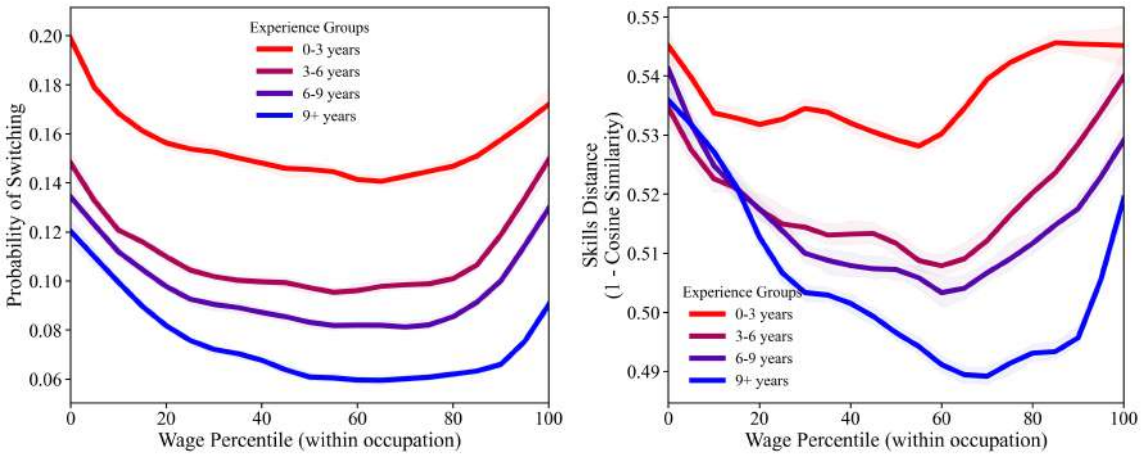
Large Sample		**Estimation Sample**	
Statistic	**Value**	**Statistic**	**Value**
Total observations	29,428,668	Total observations	2,316,035
Number of unique workers	4,425,614	Number of unique workers	283,765
Number of unique firms	618,009	Number of unique firms	131,626
Number of unique occupations	103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

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

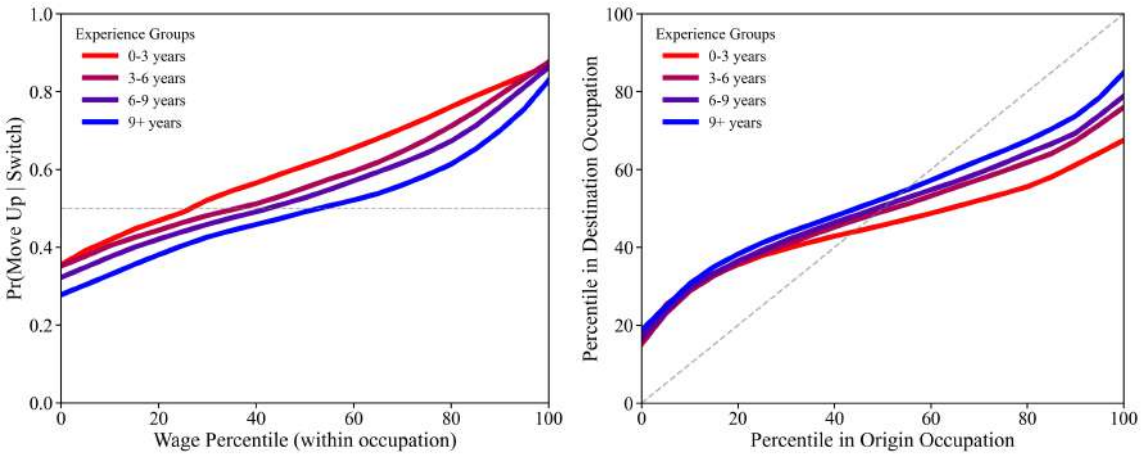
Task and Skill Classifications. The first classification, developed by Mihaylov and Tijdens (2019), classifies occupations into non-routine abstract, non-routine interactive, non-routine manual, routine cognitive, and routine manual tasks. The occupational code used in this paper is based on the ISCO88, embedded in the Portuguese code CNP1994. Starting in 2010, CPP2010 was introduced in *Quadros de Pessoal*. This code is based on ISCO08, which Mihaylov and Tijdens (2019) use. To ensure consistency across the classification change, I develop a mapping algorithm that combines the official crosswalk between CNP1994 and CPP2010 with empirical mappings based on workers who remained at the same firm during the 2009–2010 transition. This hybrid approach minimizes misclassification of occupational switches around the code change. I abstract from the routine dimension and focus on two broad categories: abstract and

manual skills. Abstract/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.

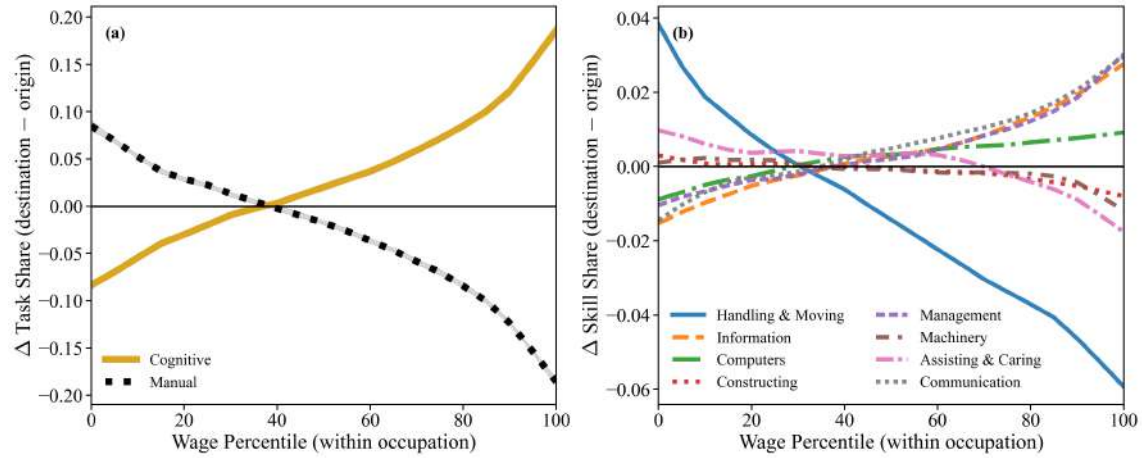
Robustness: Raw Hourly Wages. The empirical analysis in Section **??** uses residualized wages as a proxy for unobserved worker ability or match quality. Figure 22 replicates the main empirical facts from that section using raw real hourly wages to construct the within-occupation percentiles. The patterns—the pseudo U-shape in switching, the directionality in both wages and skills, and the task share changes—remain qualitatively robust when using raw wages instead of residualized wages.



(a) Probability of Switching & Skills Distance



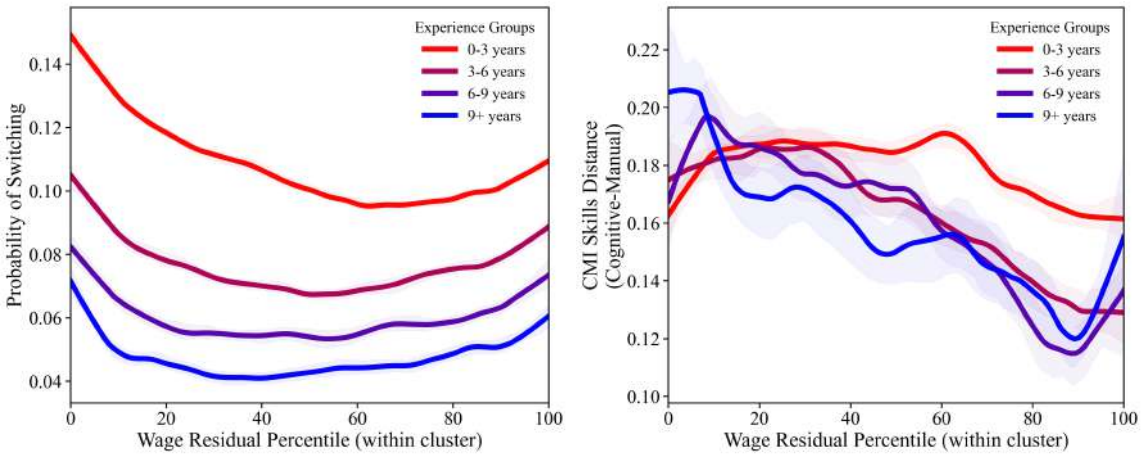
(b) Directionality by Experience



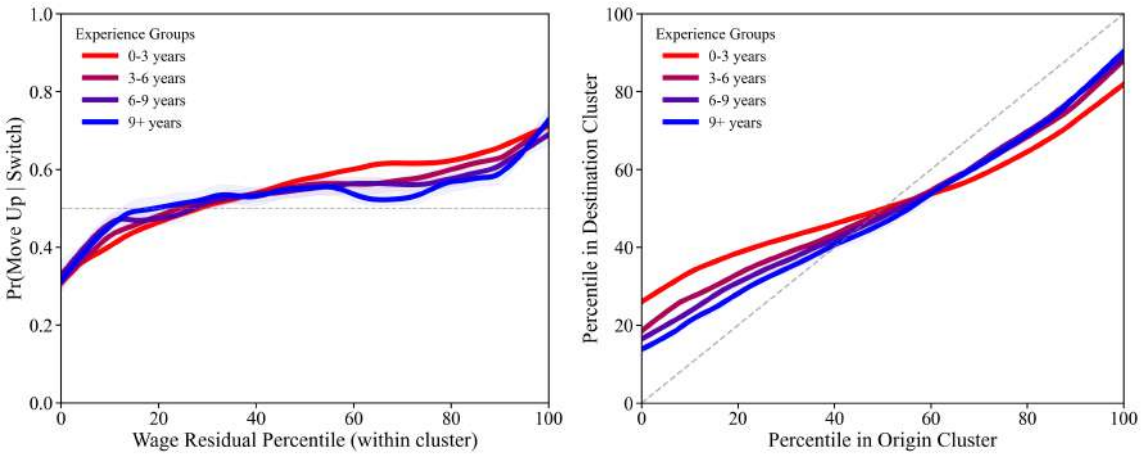
(c) Task Content Directionality (Abstract/Manual and ESCO)

Figure 22: Robustness: Raw Wages. The figure replicates the main empirical facts using raw rather than residualized wages. Panel (a) shows the probability of switching occupations and skills distance by experience. Panel (b) shows directionality and rank-to-rank by experience group. Panel (c) shows the change in abstract/manual and ESCO skill shares conditional on switching.

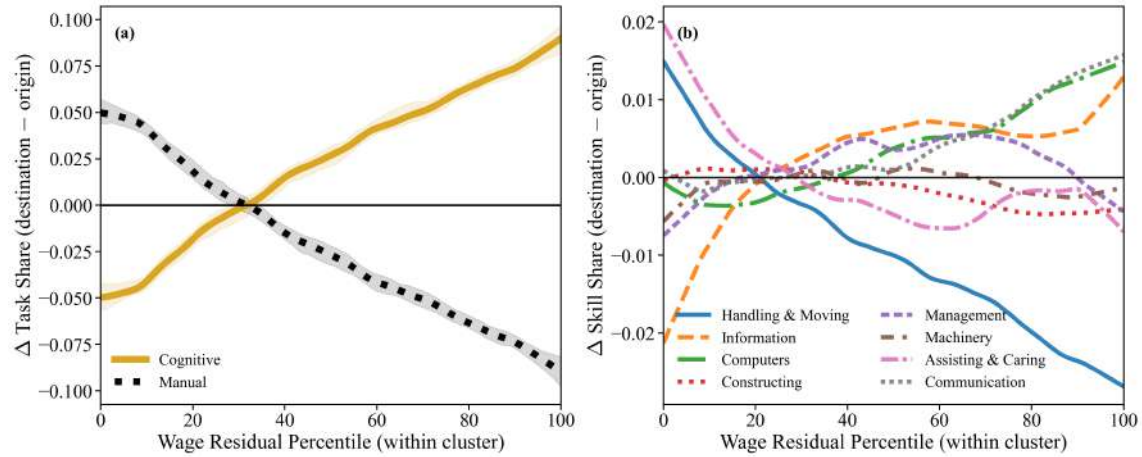
Robustness: Estimation Sample (Clusters). Furthermore, Figure 23 reproduces the main empirical analysis using the smaller estimation sample, where occupations are aggregated into ten task-based clusters. As discussed in Section **??**, this aggregation mechanically lowers mobility levels, and the smaller sample size introduces more noise. Nevertheless, the figure shows that the estimation sample still broadly matches the original empirical patterns documented in the large sample, including the U-shape in switching and the dual directionality in wages and task content. Note that for this robustness check, task distance (Panel a, right plot) is computed using the Cognitive-Manual-Interpersonal (CMI) task space rather than the full ESCO skill space.



(a) Probability of Switching & Task Distance



(b) Directionality by Experience



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

Figure 23: Robustness: Estimation Sample (Clusters). The figure replicates the main empirical facts using the aggregated estimation sample and CMI task space. Panel (a) shows switching probabilities and CMI task distance. Panel (b) shows directionality and rank-to-rank by experience group. Panel (c) shows task content directionality.

Robustness: Skill Granularity and Distance Measure. Figure 24 examines the robustness of the skills distance patterns to the level of skill aggregation. The figure reports mean (LOWESS-smoothed) and median task distance by residualized within-occupation wage percentile and experience group, using the cosine distance measure of Gathmann and Schönberg (2010). The four rows correspond to the 8-dimension ESCO classification (Matrix 1.3), the 74-dimension ESCO classification (Matrix 2.3), the full 296-dimension ESCO classification (Matrix 3.3), and the two-dimensional Cognitive–Manual space. The declining pattern of task distance with match quality is robust across all granularity levels, although the range of variation narrows at higher dimensions due to the concentration of cosine distance in high-dimensional spaces.

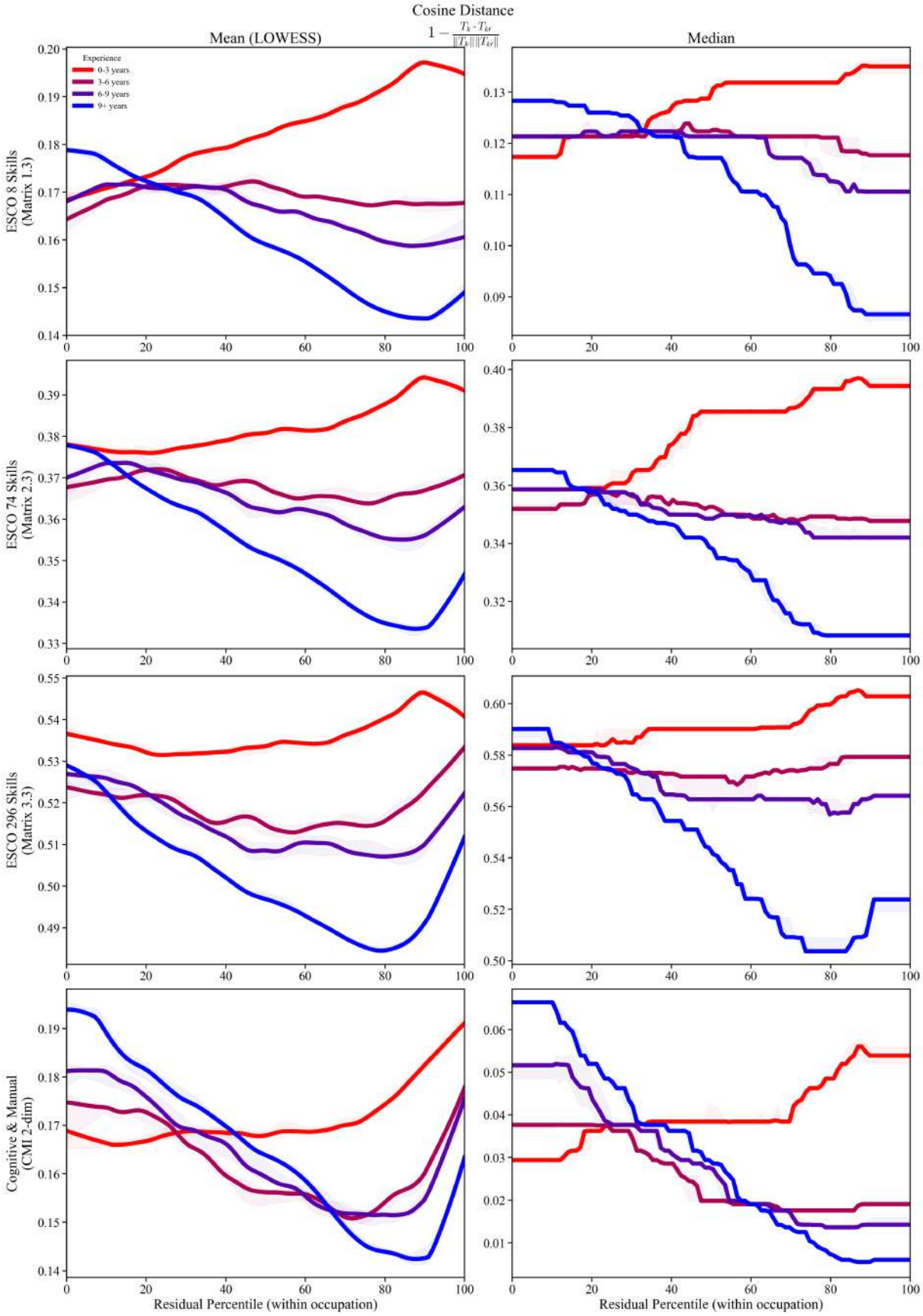


Figure 24: Robustness: Skill Granularity. Task distance (cosine distance, $1 - \text{Sim}(j, j')$, as in Gathmann and Schönberg 2010) by residualized within-occupation wage percentile and experience group. Left column: mean (LOWESS). Right column: median. Rows from top to bottom: ESCO 8 dimensions, ESCO 74 dimensions, ESCO 296 dimensions, Cognitive–Manual (2 dimensions).

Skills Distance: Summary Statistics. Table 7 reports summary statistics for the skills distance measure defined in equation (1). Panel A shows the distribution of pairwise distances across all 115 three-digit ISCO-88 occupations mapped to the ESCO skill classification (6,555 unique pairs). The mean pairwise distance is 0.71 (std. 0.19), indicating that the average pair of occupations has substantially different skill requirements. Panel B restricts to the 2.13 million actual occupation switches observed in the data. The mean distance for switchers is 0.52 (std. 0.22, median 0.58), well below the unconditional mean—confirming that workers tend to switch to occupations with similar skill profiles. The 10th percentile of switcher distances is 0.21, while the 90th percentile is 0.80.

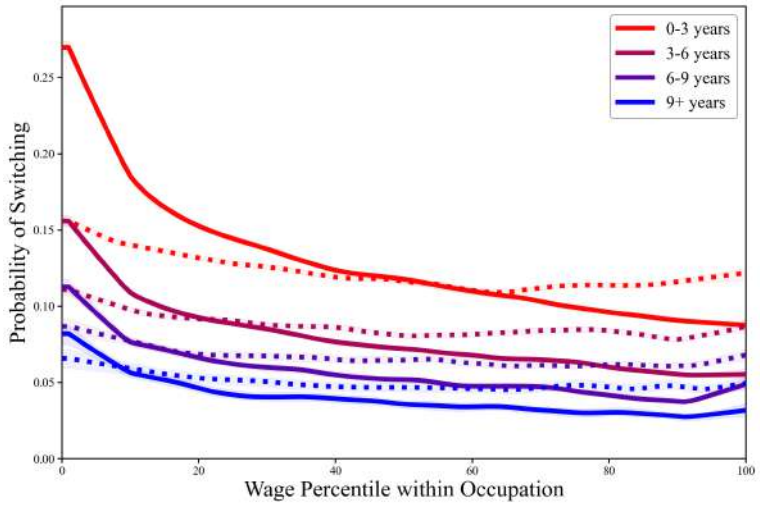
The most distant occupation pair is glass and ceramics kiln operators and shoe cleaning and street services workers ($d = 0.99$). The most similar pair (excluding identical occupations) is stall and market salespersons and street food vendors ($d = 0.04$). Women switch slightly more distant occupations on average (mean $d = 0.53$) than men (mean $d = 0.51$). Panel C reports the three most common occupational switches by education group. The most popular move for low-educated workers (education levels 1–2) is between mining and quarrying labourers and building frame workers ($d = 0.21$; 23,000 transitions), while the most common move for medium-educated workers (level 3) is between accounting clerks and bookkeepers ($d = 0.36$; 11,000 transitions). For high-educated workers (levels 4–5), the most frequent switches are between accounting clerks and bookkeepers ($d = 0.36$) and between physical science technicians and architects and engineers ($d = 0.43$). Panels D and E decompose the most common switches by gender and by gender $\times$ education. Men's most frequent move is between labourers and building frame workers—a short-distance switch within manual occupations—while women's most frequent move is between accounting clerks and bookkeepers. These gender differences are largely driven by occupational sorting: low-educated men concentrate in construction and manufacturing, while low-educated women concentrate in domestic and care services.

Table 7: Skills Distance Summary Statistics. Cosine distance $d(j, j') = 1 - \text{Sim}(j, j')$ computed using the 296-dimension ESCO skill classification. ISCO-88 three-digit occupation codes.

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

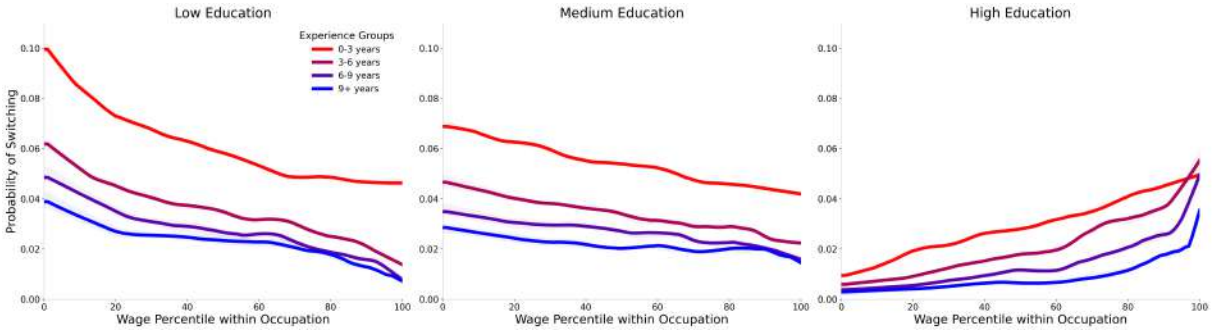
Panel C: Most common occupation switch by group

Group	Occupation switch	d	N
Low educ. (1–2)	Labourers → 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	Labourers → Building frame workers	0.21	24,311
Women	Accounting clerks → Bookkeepers	0.36	16,774
Men, Low	Labourers → 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



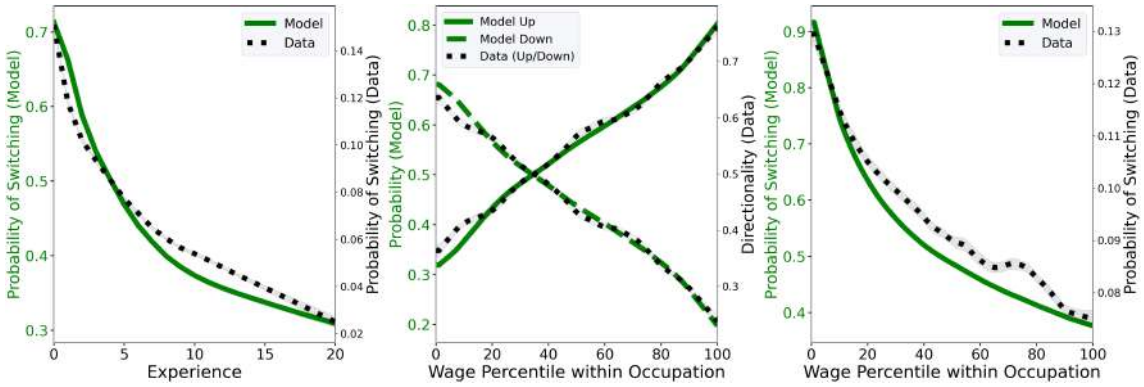
(a) Pseudo U-shape by experience

Figure 25: Probability of switching occupation (vertical axis) by wage percentile within occupation (horizontal axis), for four experience ranges (0–3, 3–6, 6–9, 9+ years). Solid lines = model predictions; dotted lines = empirical patterns. Colors range from red (youngest) to blue (oldest).

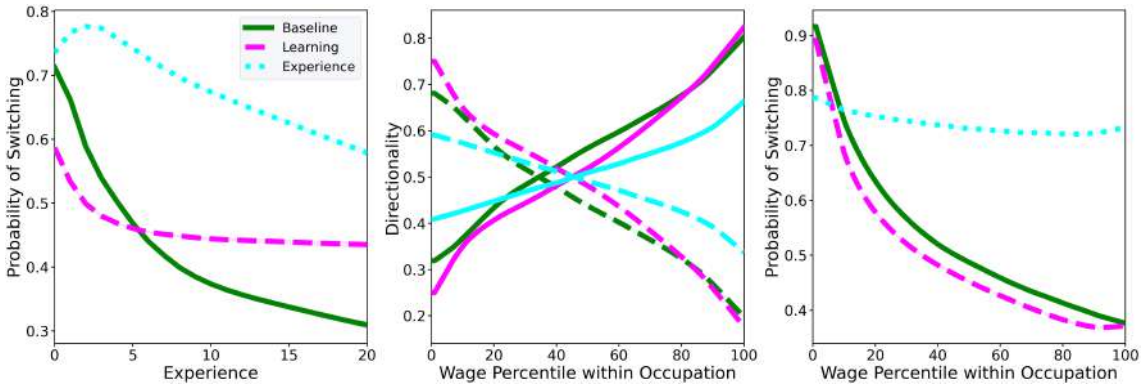


(a) Switching probability by wage percentile, experience, and education

Figure 26: Joint probability of switching occupation and education group (vertical axis) by wage percentile within occupation (horizontal axis). Decomposition: $\Pr(\text{switch} = 1 \mid \text{rank}_{i(j,t)}) = \sum_{g \in \{L,M,H\}} \Pr(\text{switch} = 1,\ \text{educ} = g \mid \text{rank}_{i(j,t)})$. Colors indicate education groups (Low, Medium, High). Mobility at the bottom is driven by low-educated workers; high-educated workers exhibit higher switching at the top.

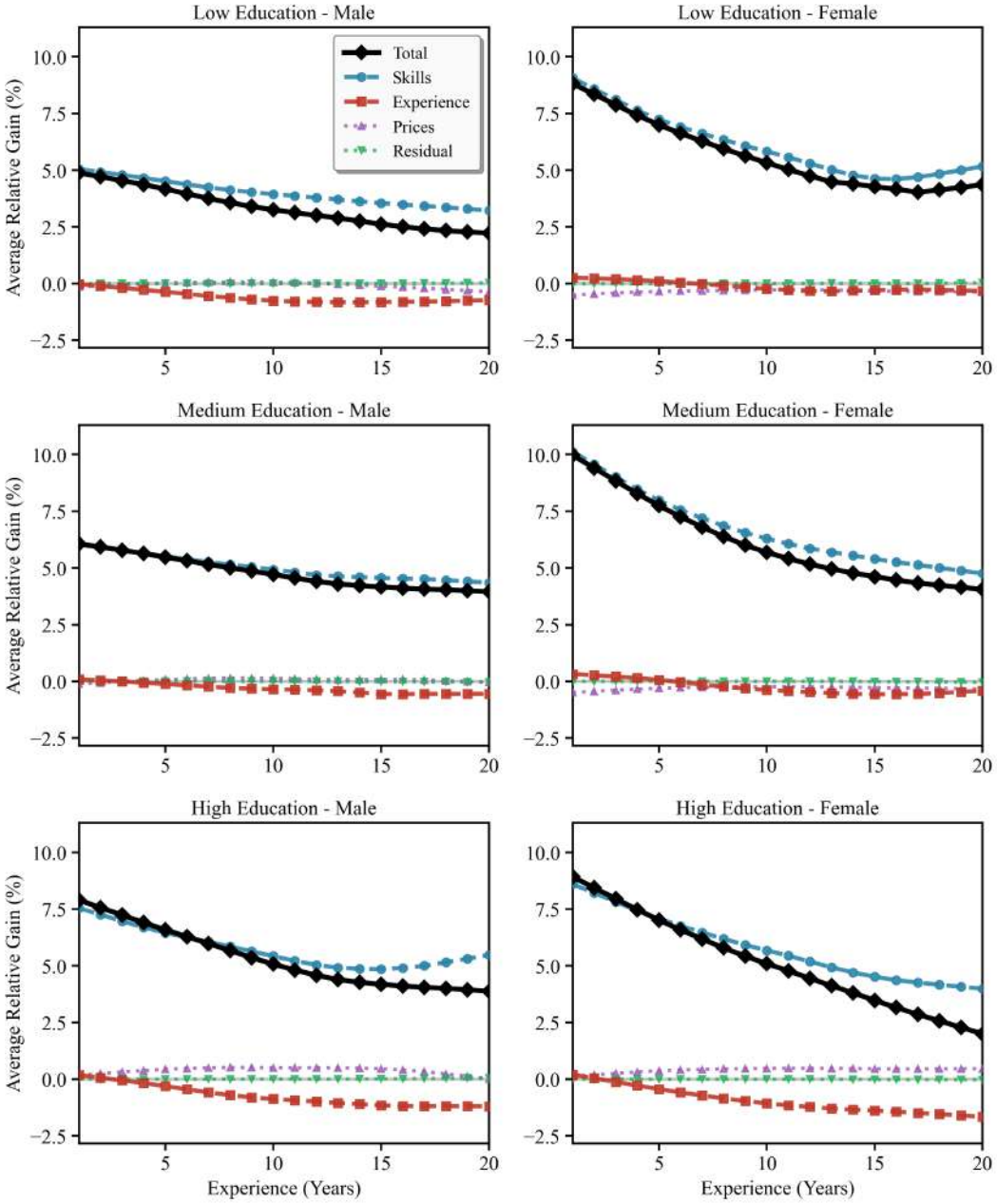


(a) Key Occupational Mobility Patterns: Model vs Data



(b) Occupational Mobility Channels Decomposition in the Simple Model

Figure 27: Both subfigures contain three panels: (i) hazard rate of switching by experience (horizontal axis: years; vertical axis: probability), (ii) directionality of moves by wage percentile (horizontal axis: percentile; vertical axis: probability of up/down move), and (iii) overall switching probability by wage percentile (horizontal axis: percentile; vertical axis: probability). *Subfigure 27(a)*: Model predictions vs empirical moments in a simplified model without switching costs, amenities, or idiosyncratic shocks; left y-axis = model, right y-axis = data. *Subfigure 27(b)*: Counterfactual simulations under Baseline (green), Learning (magenta), and Experience (cyan).



(a) Decomposition of wage gains by education and gender

Figure 28: Each panel shows a specific education-gender group (horizontal axis: years of experience; vertical axis: percent wage gain). Four channels: Skills (sorting), Experience (human capital accumulation), Prices (GE effects), and Residual. Low-educated workers, particularly women, experience the largest costs from information frictions. The skill sorting channel dominates across all groups.

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

Table 8: Distribution of Wage Gains from Occupational Reallocation

Education	Gender	Exp	Share	Mean	P5	P25	Median	P75	P95
Low	F	1	35.4%	22.3	-2.0	7.3	16.4	30.0	65.8
Low	F	10	41.3%	8.9	-10.7	-2.3	6.0	16.6	39.9
Low	M	1	26.7%	17.7	-2.6	5.3	13.9	25.5	50.5
Low	M	10	34.6%	8.0	-11.5	-3.5	5.0	16.0	37.5
Medium	F	1	37.6%	21.6	-1.9	6.9	15.3	29.4	63.7
Medium	F	10	41.7%	8.7	-10.5	-2.0	6.4	16.0	37.0
Medium	M	1	29.6%	17.6	-1.2	6.1	13.4	24.2	49.4
Medium	M	10	38.2%	8.7	-9.5	-1.8	6.4	15.7	36.4
High	F	1	35.2%	15.2	-3.4	4.7	11.3	21.5	46.8
High	F	10	38.1%	6.3	-11.3	-2.0	5.5	13.1	27.5
High	M	1	33.3%	15.3	-2.5	5.1	11.4	21.6	45.7
High	M	10	38.4%	6.7	-10.7	-1.8	5.2	13.3	29.3

Notes: The table reports the distribution of wage gains for workers who reallocate to different occupations in the counterfactual scenario with perfect information relative to the baseline economy with learning. The sample is partitioned by education level ('Low', 'Medium', 'High'), gender (M=Male, F=Female), and years of labor market experience. *Share* reports the percentage of workers in each cell who choose a different occupation in the counterfactual equilibrium. Wage gains are calculated as the percentage difference in wages between the counterfactual and baseline scenarios for reallocated workers only.

Table 9: AI Exposure Shock: Parameter Changes by Occupation Cluster

	AI Exposure				Parameter Changes ($\delta = 1.00$)			
Cluster	α_C	α_M	ω_C	ω_M	$\Delta\omega_C$	$\Delta\omega_M$	$\Delta\lambda_C$	$\Delta\beta_C$
0	0.478	0.0	0.117	0.883	-44.7%	$+5.9\%$	-44.7%	-44.7%
1	0.444	0.0	0.936	0.064	-4.9%	$+71.0\%$	-4.9%	-4.9%
2	0.409	0.0	0.322	0.678	-31.9%	$+15.2\%$	-31.9%	-31.9%
3	0.388	0.0	0.712	0.288	-15.4%	$+38.2\%$	-15.4%	-15.4%
4	0.356	0.0	0.015	0.985	-35.3%	$+0.5\%$	-35.3%	-35.3%
5	0.448	0.0	0.846	0.154	-11.1%	$+61.0\%$	-11.1%	-11.1%
6	0.434	0.0	0.998	0.002	-0.2%	$+77.2\%$	-0.2%	-0.2%
7	0.485	0.0	0.566	0.434	-29.0%	$+37.8\%$	-29.0%	-29.0%
8	0.392	0.0	0.201	0.799	-34.0%	$+8.5\%$	-34.0%	-34.0%
9	0.421	0.0	0.432	0.568	-29.2%	$+22.2\%$	-29.2%	-29.2%
Mean	0.426	0.0	0.515	0.485	-23.6%	—	-23.6%	-23.6%

Notes: The table reports AI exposure and resulting parameter changes by occupation cluster following the Eloundou et al. (2024) framework. α_C and α_M denote the weighted average AI exposure for cognitive and manual skills respectively, computed from 296 ESCO Level 3 skills classified into exposure categories *E0–E3*. All manual skills have *E0* (no exposure) since LLMs cannot perform physical tasks, hence $\alpha_M = 0$ for all clusters. ω_C and ω_M are baseline skill importance weights calibrated from Mihaylov and Tijdens (2019). $\Delta\omega_k$ denotes the percentage change in skill weight after applying the shock $\omega_k^{\text{new}} = \omega_k(1 - \delta \cdot \alpha_k)$ and renormalizing to sum to one. The cognitive skill weight decreases while the manual skill weight increases mechanically due to renormalization. $\Delta\lambda_C$ and $\Delta\beta_C$ equal $\Delta\omega_C$ since skill loadings and experience returns scale proportionally with skill weights under the Cobb-Douglas aggregation. Clusters with high baseline cognitive share (e.g., Cluster 6 with $\omega_C = 0.998$) show large percentage increases in manual share due to the small baseline, though the absolute change remains modest.

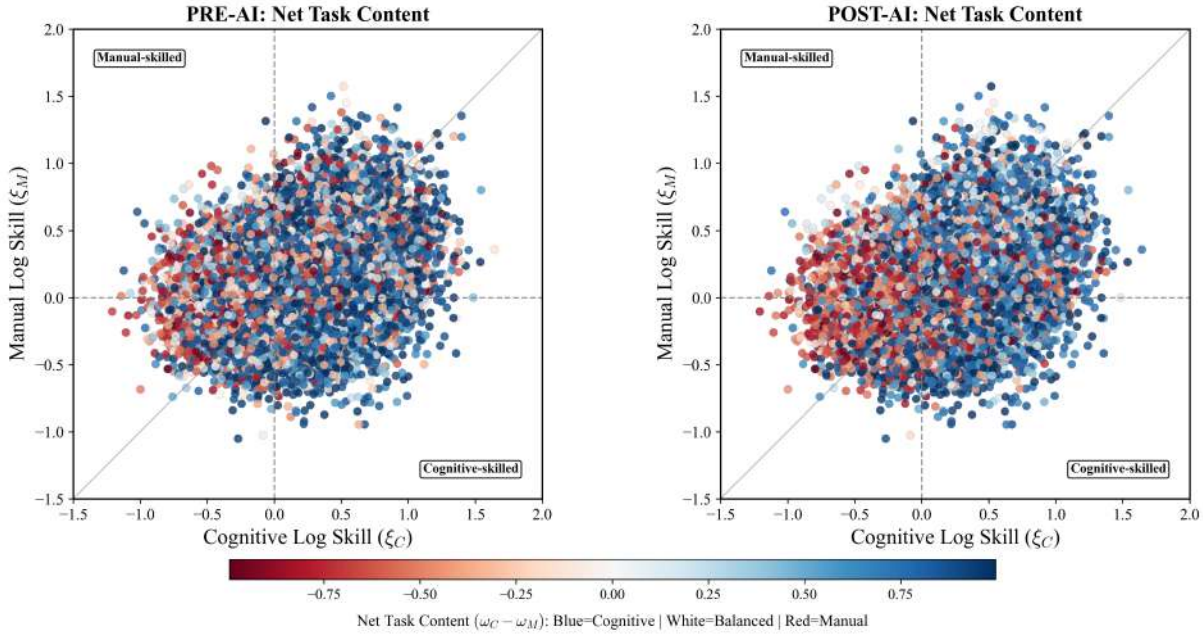


Figure 29: Net task content $(\omega_C - \omega_M)$ before and after AI adoption. Each point represents a simulated worker positioned according to their true log cognitive skill ξ_C (horizontal axis) and true log manual skill ξ_M (vertical axis). Color indicates the average net task importance of occupations held over the lifecycle: blue indicates cognitive-specialized occupations ($\omega_C > \omega_M$), white indicates balanced occupations ($\omega_C \approx \omega_M$), and red indicates manual-specialized occupations ($\omega_C < \omega_M$). *Left panel*: Pre-AI baseline. *Right panel*: Post-AI equilibrium. The comparison reveals that low-wage workers (those with manual comparative advantage, above the diagonal) sort into more manual-intensive occupations after AI, while high-wage cognitive specialists (below the diagonal) maintain their sorting into cognitive-intensive occupations despite reduced returns.

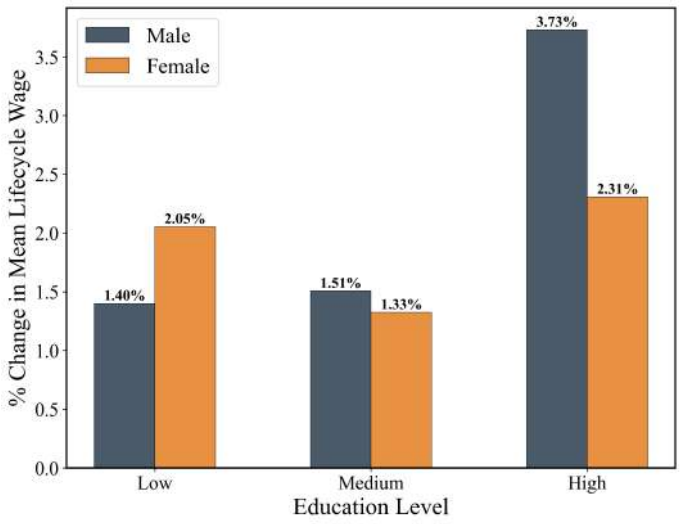


Figure 30: Direct effects of AI on wages by demographic group. Bars show the percent change in mean lifecycle wage by education-gender category (horizontal axis), illustrating the heterogeneous distributional consequences of AI-driven task transformation across demographic groups.

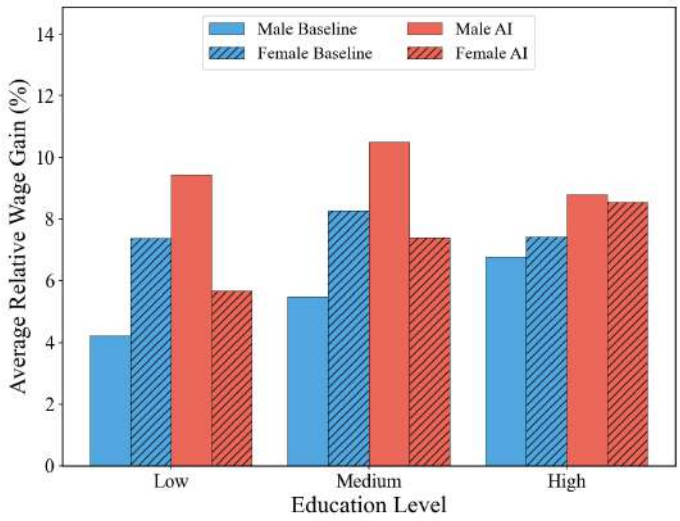


Figure 31: Value of information by education-gender group under AI. Bars compare the average lifecycle wage gain from removing information frictions (vertical axis: percent) in the baseline economy versus the post-AI economy, disaggregated by education level and gender. The AI shock increases the value of information mainly for men, reflecting the disproportionate increase in returns to manual skills in which men tend to have higher skill levels.

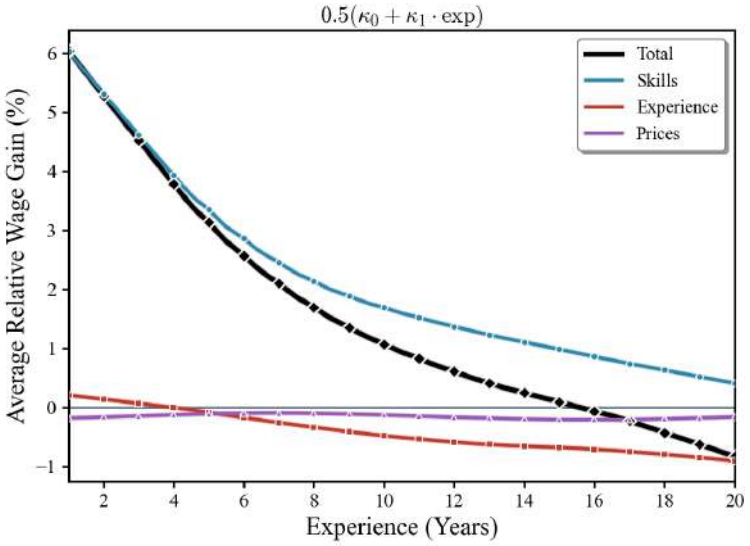


Figure 32: Decomposition of the lifecycle wage cost of imperfect information under 50%-reduced switching costs ($0.5(\kappa_0 + \kappa_1 \cdot \exp)$). Horizontal axis: years of potential experience; vertical axis: average relative wage gain (%). The total wage gap (black) is decomposed into Skills (blue), Experience (red), and Prices (purple) channels. The Skills channel remains positive throughout—workers with low switching costs keep improving their occupational sorting as they learn. The Experience channel turns negative and deepens over time: workers under imperfect information accumulated experience in a different occupation early in their career, and because output depends on experience in both task dimensions, this cross-task experience turns out to be productive later on. After approximately fourteen years, the experience advantage offsets the remaining sorting disadvantage and the total wage gap turns negative.

Figure 33: LLM Classification Output: Example Skill Classifications by AI Exposure

E0 – No Exposure


```
{
 "esco_skill": "moving and lifting",
 "esco_code": "S5.5.1",
 "dimension": "Manual",
 "exposure": "E0",
 "rationale": "Physical task"
}
{
 "esco_skill": "negotiating",
 "esco_code": "S4.7.2",
 "dimension": "Cognitive",
 "exposure": "E0",
 "rationale": "Human interaction required"
}
```


E1 – Direct LLM


```
{
 "esco_skill": "translating and
     interpreting",
 "esco_code": "S4.8.3",
 "dimension": "Cognitive",
 "exposure": "E1",
 "rationale": "Core LLM capability"
}
```


E2 – LLM + Tools


```
{
 "esco_skill": "programming computer
     systems",
 "esco_code": "S3.4.1",
 "dimension": "Cognitive",
 "exposure": "E2",
 "rationale": "Needs IDE integration"
}
```


E3 – Vision AI


```
{
 "esco_skill": "monitoring quality of
     products",
 "esco_code": "S1.4.1",
 "dimension": "Cognitive",
 "exposure": "E3",
 "rationale": "Visual inspection required"
}
```


Summary

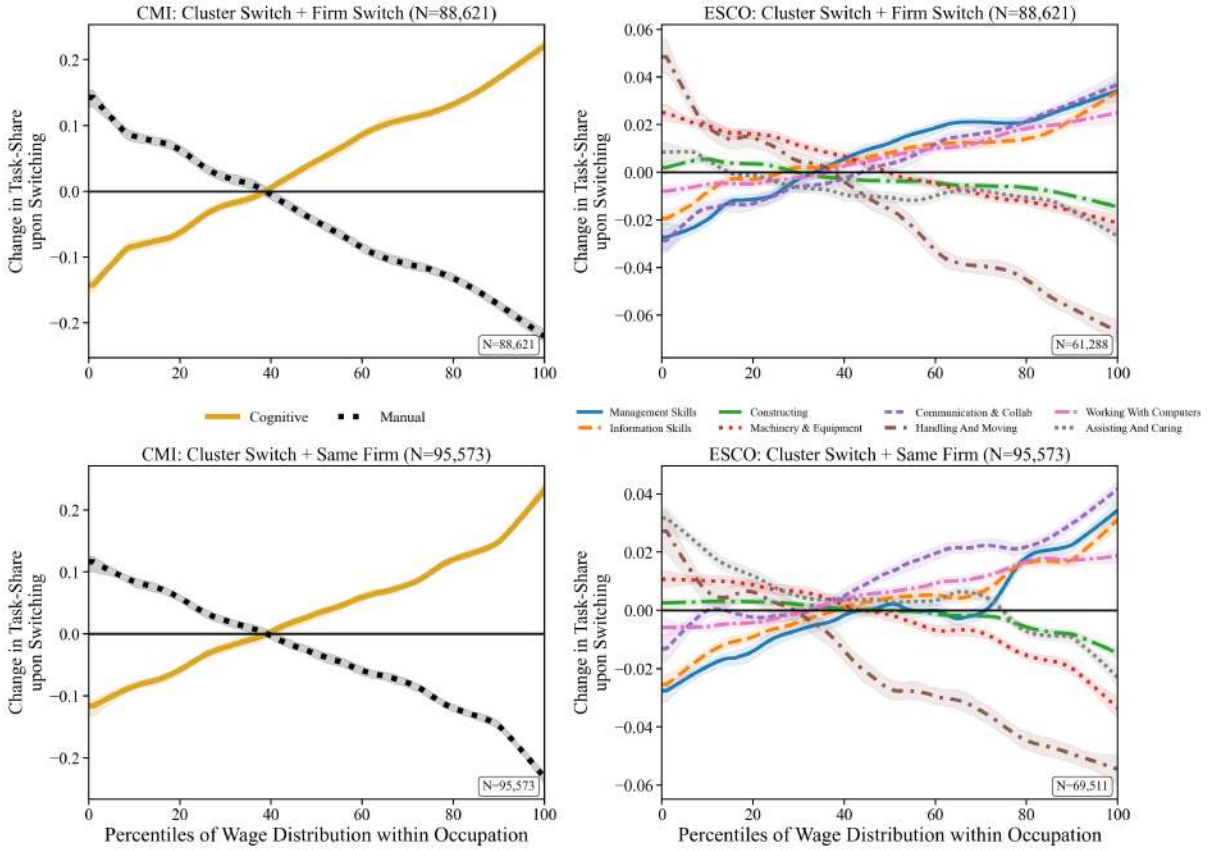

```
{
 "total_skills": 296,
 "manual": 138, "cognitive": 158,
 "E0": 175, "E1": 41, "E2": 66, "E3": 14
}
```


Notes: Excerpts from the JSON classification of 296 ESCO Level 3 skills by Claude Opus 4.5 following Eloundou et al. (2024). Each entry contains: `esco_skill` (skill name), `esco_code` (ESCO identifier), `dimension` (Manual or Cognitive), `exposure` (E0–E3), and `rationale`. Categories: *E0* = no exposure; *E1* = direct LLM; *E2* = LLM + tools/data; *E3* = vision AI. All manual skills receive *E0*. Full file in replication materials.

years at different firms, but the hiring date for the second job is the same as the hiring date for the first job we drop the entire history for a worker. We also drop individuals who have missing education and we drop workers with education decreasing over time. The sample is further restricted to full-time workers (30h per week or 130h per month) that earns at least the minimum wage. Following Fonseca et al. (2018), we drop workers in occupations and industries associated with the primary sector, given that the survey is optional for those. We also drop workers in public administration and defense[24] and drop a worker history if he/she has been in sectors with less than 100 workers reported or in occupations with less than 100 workers reported in a given year. As we are also interested in occupational mobility within firms, we drop workers in firms with less than 3 employees. Regarding industry codes, the CAE Rev. 3 (Classificação Portuguesa de Actividades Económicas, Revisão 3) replaced the CAE Rev 2.1 (Classificação Portuguesa de Actividades Económicas, Revisão 2.1) used since 1995[25]. While for occupations, the National Classification of Occupations 1994 (CNP/94) was used from 1994 until 2009 and was then replaced by the Portuguese Classification of Occupations 2010 (CPP/2010) that is embedded in the ISCO08 international classifications. I develop an

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

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



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

Figure 34: Change in task shares (vertical axis) by within-occupation wage percentile (horizontal axis) for occupation switchers, comparing those who stay at the same firm versus those who switch firms. Based on the abstract/manual task classification and ESCO skill measures.

algorithm to minimize the number of occupational switch missclassifcation in the years around the occupational code change.

C.1 Exit Probability Calibration

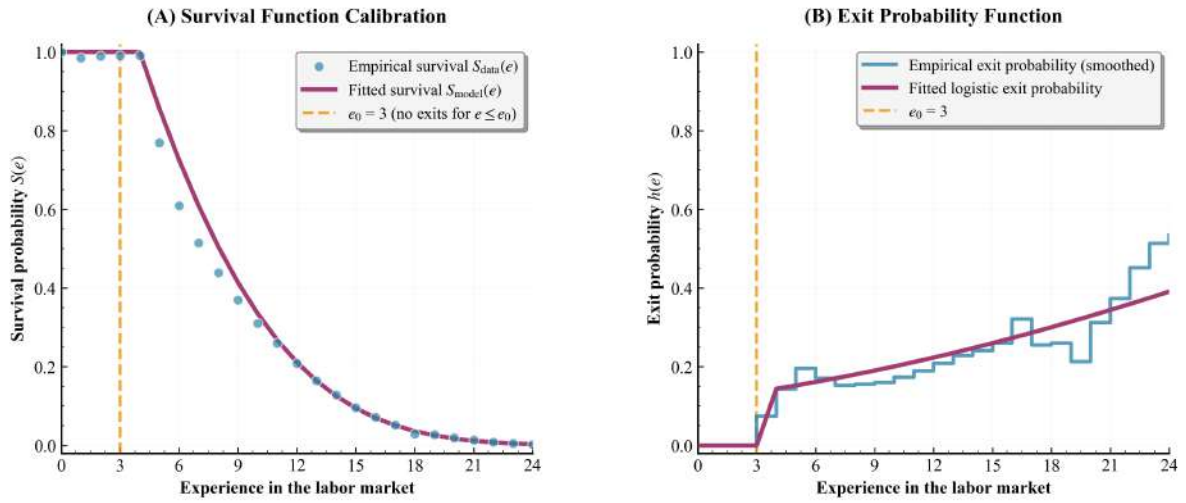
The exit probability parameters (ψ and $\bar{a}$) in the external parameters table are set using a left-truncated logistic hazard model. Figure 35 shows the calibration results, comparing empirical survival and exit probability patterns with the fitted model.

C.2 Empirical evidence on wage changes upon occupational switches

C.2.1 Event–study with stayers as control group

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

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



(a) Exit probability calibration

Figure 35: *Panel (A)*: Empirical vs fitted survival probabilities (horizontal axis: experience; vertical axis: survival probability). *Panel (B)*: Empirical vs fitted exit probabilities (horizontal axis: experience; vertical axis: exit probability). Vertical dashed line = truncation point $e_0 = 3$. Fitted model: left-truncated logistic hazard with $\psi = 0.072$ and $\bar{a} = 27.22$.

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

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

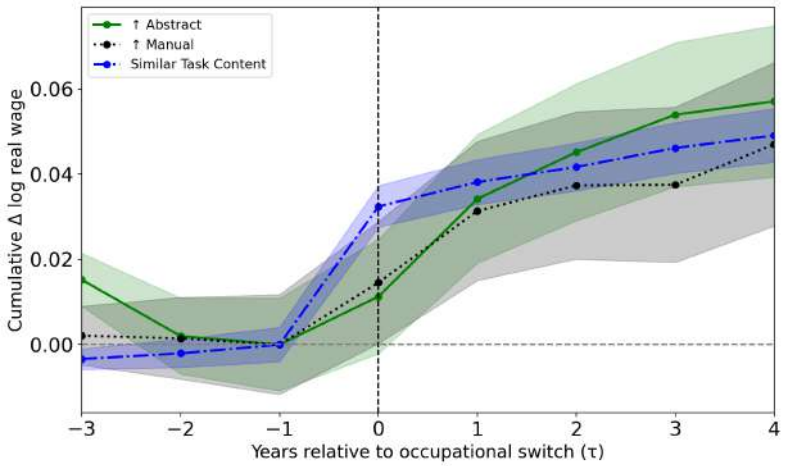
$$\underbrace{\left(\Delta \log w_{it+\tau}\right)^{\circ}}_{\text{demeaned by year } t} = \alpha_\tau + \sum_{g} \beta_{\tau g} D_{it}^{g} + \theta_\tau' X_{it}^{\circ} + \varepsilon_{it\tau}, \tag{39}$$

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

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

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

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



(a) Cumulative wage changes around occupational switches

Figure 36: Cumulative Δ log real wage (vertical axis) by event time relative to switch (horizontal axis: years from -3 to $+4$), normalized at $\tau = -1$. Lines: ↑ Abstract (green, solid), ↑ Manual (black, dotted), Similar Task Content (blue, dash-dot). Shaded areas = 95% confidence bands clustered by worker.

C.2.2 Wage changes by initial position within occupation

Workers at different positions in their origin occupation's wage distribution experience markedly different wage trajectories upon switching. Figure 37 shows cumulative wage changes for switchers across the initial wage percentile distribution.

For each switcher, we calculate their wage percentile within their origin occupation (in the year before switching) and track cumulative log wage changes over the following four years. The plot reveals a strong negative relationship: workers initially at the bottom of their occupation's wage distribution experience cumulative gains exceeding 30%, while those at the top initially lose wages but recover to modest gains of approximately 10% by year four. This pattern persists across all time horizons and suggests that occupational switching is particularly beneficial for workers who were poorly matched to their origin occupation.

D Appendix: Model

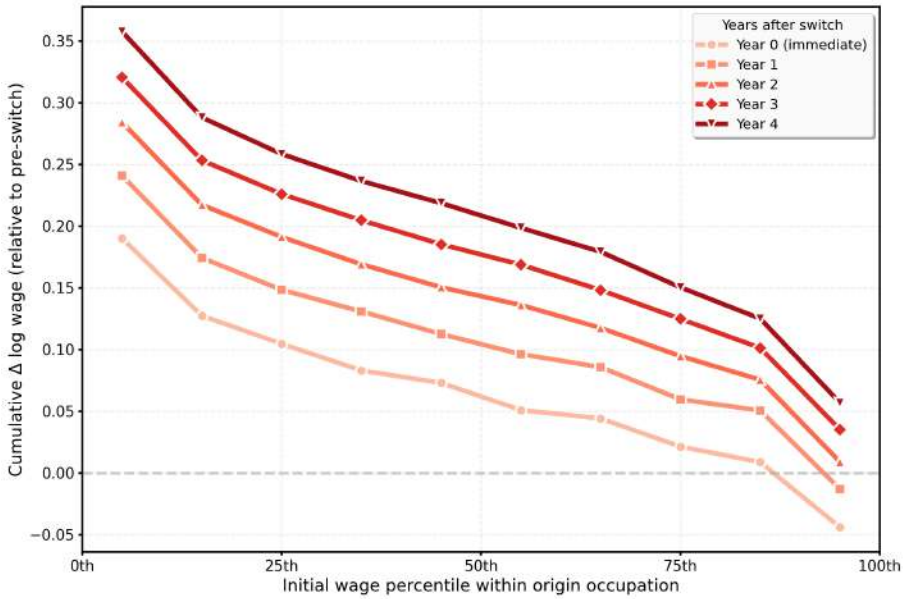
This appendix provides additional derivations and describes the computational algorithms used to solve the model.

D.1 Firm's Time Allocation Problem

This section derives the optimal allocation of worker time across tasks within an occupation. Following **?**, I assume that the firm controls the worker's time allocation to maximize current output, and that this allocation is non-contractible.

Consider a worker with skills $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ and task-specific experience $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$ employed in occupation j. Task-level output is:

$$y_{jk} = \ell_{jk} \cdot \theta_k^{\lambda_{jk}} \cdot \tau_k^{\beta_{jk}}, \tag{40}$$



(a) Cumulative wage changes by initial wage percentile

Figure 37: Cumulative log wage change (vertical axis) by initial wage percentile within origin occupation (horizontal axis), for Years 0–4 after switching (separate lines). Workers at lower percentiles experience larger wage gains from switching.

where $\ell_{jk} \in [0,1]$ is the fraction of unit time allocated to task k, subject to $\sum_{k=1}^{K} \ell_{jk} = 1$, and λ_{jk} and β_{jk} are the primitive elasticities of skill and experience in task k within occupation j. Tasks aggregate with Cobb-Douglas weights $\omega_{jk} > 0$ satisfying $\sum_{k=1}^{K} \omega_{jk} = 1$:

$$Y_j = \prod_{k=1}^{K} y_{jk}^{\omega_{jk}}. \tag{41}$$

The firm's problem is to choose $\{\ell_{jk}\}_{k=1}^{K}$ to maximize Y_j subject to $\sum_k \ell_{jk} = 1$. Taking logs:

$$\log Y_j = \sum_{k=1}^{K} \omega_{jk} \left[\log \ell_{jk} + \lambda_{jk} \log \theta_k + \beta_{jk} \log \tau_k \right]. \tag{42}$$

Forming the Lagrangian $\mathcal{L} = \log Y_j - \mu \left(\sum_k \ell_{jk} - 1 \right)$ and taking first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \ell_{jk}} = \frac{\omega_{jk}}{\ell_{jk}} - \mu = 0 \quad \Rightarrow \quad \ell_{jk} = \frac{\omega_{jk}}{\mu}. \tag{43}$$

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

$$\ell_{jk}^{*} = \omega_{jk}. \tag{44}$$

Key property: The optimal time allocation is *independent of worker skills and experience*. It depends only on the task structure ω_{jk} of the occupation. This result follows from the Cobb-Douglas specification and greatly simplifies the analysis.

D.2 Stationary Equilibrium Algorithm

The stationary equilibrium is characterized by constant occupation prices $\boldsymbol{p} = (p_j)_{j\in\mathcal{O}}$ such that markets clear in every period. We work with log prices $\tilde{\boldsymbol{p}} = (\log p_j)_{j\in\mathcal{O}}$. Given primitives $(\boldsymbol{\lambda}, \boldsymbol{b})$ and CES demand parameters $(\sigma, \boldsymbol{\phi})$, the algorithm proceeds as follows:

Simulation-based computation via the ergodic theorem. The model induces a Markov chain over worker states $x = (\boldsymbol{h}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, e, g)$. Because workers exit stochastically and are replaced by entrants drawn from the initial distribution $\pi_g^0(\cdot)$, the chain repeatedly regenerates. This demographic turnover makes the process aperiodic and ensures that initial conditions are forgotten over time. Under these conditions, the chain admits a unique invariant distribution π^* and is ergodic. Therefore, for any integrable function f,

$$\frac{1}{N \cdot T_{\text{sim}}} \sum_{i=1}^{N} \sum_{t=t_{\text{burn}}+1}^{T} f(x_{it}) \xrightarrow{\text{a.s.}} \mathbb{E}_{\pi^*}[f(x)],$$

so the empirical mean over workers and post-burn-in periods consistently estimates the corresponding stationary expectation. Occupation outputs Y_j are computed by applying this result to workers' wage contributions conditional on choosing occupation j. The demographic turnover induced by the exit hazard implies that the influence of initial conditions vanishes well before the burn-in period ends.

- **Step 0:** Set an initial guess for log prices $\tilde{\boldsymbol{p}}^{(0)}$, typically uniform across occupations. Compute the target price index P^* from empirical prices.
- **Step 1:** Given current log prices $\tilde{\boldsymbol{p}}^{(n)}$, solve the worker's stationary dynamic programming problem to obtain value function coefficients $\Phi^{(n)}$.
- **Step 2:** Using the optimal policy derived from $\Phi^{(n)}$, simulate N workers for T periods. Discard the first t_{burn} periods; the demographic turnover from the exit hazard ensures convergence to the ergodic distribution well before the burn-in ends.
- **Step 3:** For each occupation j, compute aggregate output Y_j as the cross-sectional and time average of workers' wage contributions conditional on $d = j$, over the post-burn-in sample.
- **Step 4:** Under CES demand with elasticity σ and weights $\boldsymbol{\phi}$, compute implied prices:

$$\hat{p}_j = \phi_j^{1/\sigma} Y_j^{-1/\sigma} P^*.$$

- **Step 5:** Apply a damped update:

$$\tilde{\boldsymbol{p}}^{(n+1)} = (1-\lambda)\tilde{\boldsymbol{p}}^{(n)} + \lambda \log \hat{\boldsymbol{p}},$$

  where $\lambda \in (0,1)$ is the damping parameter.
- **Step 6: Check convergence.** If $\|\tilde{\boldsymbol{p}}^{(n+1)} - \tilde{\boldsymbol{p}}^{(n)}\| < \epsilon$, stop. Otherwise, return to Step 1.

D.3 Transition Path Algorithm

The transition path algorithm computes the perfect foresight equilibrium between pre-shock and post-shock stationary equilibria. Let T denote the transition horizon. The economy starts at $t = 0$ in the pre-shock stationary equilibrium with primitives $(\boldsymbol{\lambda}, \boldsymbol{b})$ and log prices $\tilde{\boldsymbol{p}}_{\text{init}}$. At $t = 0$, an unanticipated permanent shock changes the primitives to $(\boldsymbol{\lambda}', \boldsymbol{b}')$. The economy transitions toward the post-shock stationary equilibrium with log prices $\tilde{\boldsymbol{p}}_{\text{final}}$. Workers have perfect foresight about the price path.

- **Step 0:** Initialization.
  - Solve for pre-shock and post-shock stationary equilibria to obtain $(\tilde{\boldsymbol{p}}_{\text{init}}, \Phi_{\text{init}})$ and $(\tilde{\boldsymbol{p}}_{\text{final}}, \Phi_{\text{final}})$.
  - Initialize the price path by linear interpolation:

$$\tilde{\boldsymbol{p}}_t^{(0)} = \left(1 - \frac{t+1}{T+1}\right) \tilde{\boldsymbol{p}}_{\text{init}} + \frac{t+1}{T+1} \tilde{\boldsymbol{p}}_{\text{final}}, \quad t = 0, \ldots, T-1.$$

  - Compute initial worker states by simulating the economy at pre-shock stationary prices.
- **Step 1:** Given current log prices $\{\tilde{\boldsymbol{p}}_t^{(n)}\}_{t=0}^{T-1}$:
  - Set terminal condition $\Phi_T = \Phi_{\text{final}}$.
  - For $t = T-1, \ldots, 0$, compute value function coefficients by projecting the Bellman operator:

$$\Phi_t = \Pi\Big[\max_{j \in \mathcal{O}} \Big\{ r_j(\boldsymbol{x}; \tilde{\boldsymbol{p}}_t^{(n)}) + \beta(1 - \delta(e)) \mathbb{E}[V_{t+1}(\boldsymbol{x}') \mid \boldsymbol{x}, j] \Big\}\Big].$$

- **Step 2:** Starting from pre-shock initial states, simulate N workers forward for T periods using the time-varying policy functions from $\{\Phi_t\}$ and the post-shock primitives $(\boldsymbol{\lambda}', \boldsymbol{b}')$.
- **Step 3:** Compute implied prices. For each period t, compute occupation outputs Y_{jt} and implied prices:

$$\hat{p}_{jt} = \phi_j^{1/\sigma} Y_{jt}^{-1/\sigma} P^*.$$

  Apply moving average smoothing to reduce simulation noise.
- **Step 4:** Apply a damped update with Anderson acceleration:

$$\tilde{\boldsymbol{p}}_t^{(n+1)} = (1 - \lambda_n) \tilde{\boldsymbol{p}}_t^{(n)} + \lambda_n \log \hat{\boldsymbol{p}}_t.$$

  Use adaptive damping: reduce λ_n when oscillation is detected, increase when progress is steady.
- **Step 5:** Compute the mean absolute deviation:

$$d^{(n)} = \frac{1}{T \cdot |\mathcal{O}|} \sum_{t,j} \left| \tilde{p}_{jt}^{(n+1)} - \tilde{p}_{jt}^{(n)} \right|.$$

Also check that tail prices converge to $\tilde{\boldsymbol{p}}_{\text{final}}$. If both criteria are satisfied, stop. Otherwise, return to Step 1.

E Appendix: Estimation and Numerical Methods

E.1 EM Algorithm: Computational Details

We start with a guess for

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

where q denotes the EM iteration. For each occupation $j \in \{1, \dots, J\}$, $p_j \in \mathbb{R}$ is an occupation fixed effect interpreted as an occupation-specific price, $\boldsymbol{\lambda}_j \in \mathbb{R}^K$ are factor loadings on K latent skill components, $\boldsymbol{\beta}_j \in \mathbb{R}^K$ are linear returns to log cumulative task tenure, and ς_j^2 is the occupation-specific variance of idiosyncratic wage shocks. Workers belong to groups $g(i) \in \{1, \dots, G\}$ (education$\times$gender in the application). We preprocess task tenures as $\tau_{it,k} \leftarrow 1 + \tau_{it,k}$ so that $\log(\tau_{it,k})$ is well-defined. In the structural model skills enter wages multiplicatively, and with log wages it is convenient to work with log skills. We therefore define the latent state as $\boldsymbol{\xi}_i := \log \boldsymbol{\theta}_i \in \mathbb{R}^K$. The group prior is

$$\boldsymbol{\xi}_i \mid g(i) = g \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g). \tag{45}$$

The (log) wage equation is

$$w_{ijt} = \alpha_{j(t)} + \boldsymbol{\xi}_i' \boldsymbol{\lambda}_{j(t)} + \boldsymbol{\beta}_{j(t)}' \log(\boldsymbol{\tau}_{it}) + \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \sim \mathcal{N}\big(0, \varsigma_{j(t)}^2\big), \tag{46}$$

where $\alpha_j(t) = \log p_j(t)$ and $\boldsymbol{\tau}_{it} \in \mathbb{R}^K$ are observed cumulative task tenures (already shifted by $+1$ prior to logging).

E.1.1 E-step

Conditional on parameters $\boldsymbol{\Gamma}^q$, the posterior of $\boldsymbol{\xi}_i$ is Gaussian. Let the worker-specific likelihood implied by (46) be

$$p(w_i \mid \boldsymbol{\xi}_i, \boldsymbol{\Gamma}^q) = \prod_{t=t_0(i)}^{T(i)} \frac{1}{\sqrt{2\pi\, \varsigma_{j(t)}^{2,q}}} \exp\left(-\frac{1}{2\varsigma_{j(t)}^{2,q}} \Big[w_{ijt} - \alpha_{j(t)}^q - \boldsymbol{\xi}_i' \boldsymbol{\lambda}_{j(t)}^q - (\boldsymbol{\beta}_{j(t)}^q)' \log(\boldsymbol{\tau}_{it}) \Big]^2 \right),$$

and the prior density by (45). Bayes' rule gives

$$p(\boldsymbol{\xi}_i \mid w_i, \boldsymbol{\Gamma}^q) \propto p(w_i \mid \boldsymbol{\xi}_i, \boldsymbol{\Gamma}^q)\, p(\boldsymbol{\xi}_i \mid g(i), \boldsymbol{\Gamma}^q),$$

and completing the square in $\boldsymbol{\xi}_i$ yields the closed-form posterior.

Define the residual net of the occupation log price and task-tenure component,

$$M_{ijt}^q := w_{ijt} - \alpha_{j(t)}^q - (\boldsymbol{\beta}_{j(t)}^q)' \log(\boldsymbol{\tau}_{it}). \tag{47}$$

Then the posterior covariance and mean are

$$\Xi_i^q = \left((\Sigma_{g(i)}^q)^{-1} + \sum_{t=t_0(i)}^{T(i)} \frac{1}{\varsigma_{j(t)}^{2,q}} \lambda_{j(t)}^q \lambda_{j(t)}^{q\prime} \right)^{-1}, \tag{48}$$

$$\zeta_i^q = \Xi_i^q \left((\Sigma_{g(i)}^q)^{-1} \mu_{g(i)}^q + \sum_{t=t_0(i)}^{T(i)} \frac{1}{\varsigma_{j(t)}^{2,q}} \lambda_{j(t)}^q \, M_{ijt}^q \right), \tag{49}$$

$$\xi_i \mid w_i, \Gamma^q \sim \mathcal{N}(\zeta_i^q, \Xi_i^q). \tag{50}$$

In the code, we store the posterior mean ζ_i^q and posterior covariance Ξ_i^q for each worker.

The expected wage log-likelihood (the Q-function wage component) can be written as

$$E_q[L^w] = -\sum_{it} \sum_j d_{ijt} \left[\tfrac{1}{2} \log(2\pi \varsigma_j^2) + \frac{1}{2\varsigma_j^2} E_q\Big[\big(w_{ijt} - \alpha_j - \xi_i' \lambda_j - \beta_j' \log(\tau_{it}) \big)^2 \Big] \right], \tag{51}$$

where d_{ijt} is the occupation indicator. Using (47) with $\Gamma = \Gamma^q$ and the posterior moments (50),

$$E_q\Big[\big(w_{ijt} - \alpha_j - \xi_i' \lambda_j - \beta_j' \log(\tau_{it}) \big)^2 \Big] = (M_{ijt}^q)^2 - 2 M_{ijt}^q \, \lambda_j' \zeta_i^q + (\zeta_i^q)' \lambda_j \lambda_j' \zeta_i^q + \operatorname{tr}(\lambda_j \lambda_j' \Xi_i^q). \tag{52}$$

E.1.2 M-step

The M-step chooses Γ^{q+1} to maximize

$$Q(\Gamma; \Gamma^q) = E_q\big[\log p(\{\xi_i\}_i, \{w_{ijt}\}_{it} \mid \Gamma) \big] = E_q[L^{\text{prior}}] + E_q[L^w],$$

where $E_q[\cdot]$ is taken under (50). The updates separate into (i) prior parameters by group and (ii) wage-equation parameters by occupation.

Prior parameters. Maximizing $E_q[L^{\text{prior}}]$ over (μ_g, Σ_g) yields

$$\mu_g^{q+1} = \frac{1}{N_g} \sum_{i \in g} \zeta_i^q, \tag{53}$$

$$\Sigma_g^{q+1} = \frac{1}{N_g} \sum_{i \in g} \Big[\Xi_i^q + (\zeta_i^q - \mu_g^{q+1})(\zeta_i^q - \mu_g^{q+1})' \Big]. \tag{54}$$

Wage-equation parameters. For each occupation j, define the stacked regressor and parameter vectors

$$\mathbf{z}_{it} = \begin{bmatrix} 1 \\ \xi_i \\ \log(\tau_{it}) \end{bmatrix} \in \mathbb{R}^{1+2K}, \qquad \phi_j = \begin{bmatrix} \alpha_j \\ \lambda_j \\ \beta_j \end{bmatrix} \in \mathbb{R}^{1+2K}.$$

The first-order conditions of $E_q[L^w]$ imply the normal equations

$$\left(\sum_{it} d_{ijt} \, E_q[\mathbf{z}_{it} \mathbf{z}_{it}'] \right) \phi_j = \sum_{it} d_{ijt} \, w_{ijt} \, E_q[\mathbf{z}_{it}], \tag{55}$$

with posterior expectations (using $\boldsymbol{\xi}_i \sim \mathcal{N}(\boldsymbol{\zeta}_i^q, \boldsymbol{\Xi}_i^q)$)

$$E_q[\mathbf{z}_{it}] = \begin{bmatrix} 1 \\ \boldsymbol{\zeta}_i^q \\ \log(\boldsymbol{\tau}_{it}) \end{bmatrix}, \qquad E_q[\mathbf{z}_{it}\mathbf{z}_{it}'] = \begin{bmatrix} 1 & (\boldsymbol{\zeta}_i^q)' & \log(\boldsymbol{\tau}_{it})' \\ \boldsymbol{\zeta}_i^q & \boldsymbol{\zeta}_i^q(\boldsymbol{\zeta}_i^q)' + \boldsymbol{\Xi}_i^q & \boldsymbol{\zeta}_i^q \log(\boldsymbol{\tau}_{it})' \\ \log(\boldsymbol{\tau}_{it}) & \log(\boldsymbol{\tau}_{it})(\boldsymbol{\zeta}_i^q)' & \log(\boldsymbol{\tau}_{it})\log(\boldsymbol{\tau}_{it})' \end{bmatrix}.$$

In computation, we stack all observations belonging to occupation j into a design matrix

$$\mathbf{W}_j = \begin{bmatrix} \mathbf{1} & \boldsymbol{\zeta}^q & \log(\boldsymbol{\tau}) \end{bmatrix} \in \mathbb{R}^{N_j \times (1+2K)},$$

where $\boldsymbol{\zeta}^q$ denotes the posterior mean of the corresponding worker repeated across that worker's observations. The matrix of normal equations can be written as $\mathbf{W}_j'\mathbf{W}_j$ plus an additive correction that inserts $\sum \boldsymbol{\Xi}_i^q$ into the $(\boldsymbol{\lambda}_j, \boldsymbol{\lambda}_j)$ block:

$$\mathbf{S}_j = \mathbf{W}_j'\mathbf{W}_j + \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \displaystyle\sum_{t=1}^{N_j} \boldsymbol{\Xi}_{i(t)}^q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{b}_j = \mathbf{W}_j'\mathbf{y}_j,$$

and we solve

$$(\mathbf{S}_j + \rho\,\mathbf{I})\,\boldsymbol{\phi}_j^{q+1} = \mathbf{b}_j, \tag{56}$$

with a small ridge parameter $\rho > 0$ for numerical stability. This yields $(\alpha_j^{q+1}, \boldsymbol{\lambda}_j^{q+1}, \boldsymbol{\beta}_j^{q+1})$ jointly for each j. If some $\boldsymbol{\lambda}_j$ are held fixed for normalization, (56) is partitioned and solved only for the remaining free parameters conditional on the fixed $\boldsymbol{\lambda}_j$.

Residual variances. Given $(\alpha_j^{q+1}, \boldsymbol{\lambda}_j^{q+1}, \boldsymbol{\beta}_j^{q+1})$, define

$$M_{ijt}^{q+1} = w_{ijt} - \alpha_j^{q+1} - (\boldsymbol{\beta}_j^{q+1})'\log(\boldsymbol{\tau}_{it}), \qquad \text{for observations with } d_{ijt} = 1.$$

The expected squared residual is

$$\widetilde{R}_{ijt}^{2,q+1} = (M_{ijt}^{q+1})^2 - 2M_{ijt}^{q+1}(\boldsymbol{\lambda}_j^{q+1})'\boldsymbol{\zeta}_i^q + (\boldsymbol{\zeta}_i^q)'\boldsymbol{\lambda}_j^{q+1}(\boldsymbol{\lambda}_j^{q+1})'\boldsymbol{\zeta}_i^q + \mathrm{tr}(\boldsymbol{\lambda}_j^{q+1}(\boldsymbol{\lambda}_j^{q+1})'\boldsymbol{\Xi}_i^q),$$

and the variance update is the within-occupation average

$$\varsigma_j^{2,q+1} = \frac{\displaystyle\sum_{it} d_{ijt}\,\widetilde{R}_{ijt}^{2,q+1}}{\displaystyle\sum_{it} d_{ijt}}. \tag{57}$$

The algorithm iterates the E-step (48)–(50) and the M-step (53)–(57) until convergence. In the numerical implementation, convergence is assessed using either a parameter-distance criterion based on the mean squared change in the stacked parameter vector or an absolute tolerance on the change in the Q-function.

E.1.3 Likelihood Evaluation

The observed–data wage likelihood integrates out latent log skills $\boldsymbol{\xi}_i$ under the group prior. For a given parameter vector

$$\boldsymbol{\Gamma} = \Big\{(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)_{g=1}^{G},\ (\alpha_j)_{j=1}^{J},\ (\boldsymbol{\lambda}_j)_{j=1}^{J},\ (\boldsymbol{\beta}_j)_{j=1}^{J},\ (\varsigma_j^2)_{j=1}^{J}\Big\},$$

the per–worker contribution is

$$\tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) = \int_{\mathbb{R}^K} \left\{ \prod_{t=1}^{T_i} \prod_{j=1}^{J} \Big[\phi\Big(w_{ijt};\ \alpha_j + \boldsymbol{\xi}_i' \boldsymbol{\lambda}_j + \boldsymbol{\beta}_j' \log(\boldsymbol{\tau}_{it}),\ \varsigma_j^2 \Big) \Big]^{d_{ijt}} \right\} \phi\big(\boldsymbol{\xi}_i; \boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)}\big)\, d\boldsymbol{\xi}_i, \tag{58}$$

where $\phi(\cdot; \mu, \varsigma^2)$ denotes the univariate Normal density and $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the multivariate Normal density. The sample log-likelihood is $\sum_{i=1}^{N} \log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma})$.

A Monte Carlo approximation draws $\boldsymbol{\xi}_i^{(q)} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_{g(i)}, \boldsymbol{\Sigma}_{g(i)})$ for $q = 1, \ldots, Q$ and replaces the integral in (58) by a simple average:

$$\tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \approx \frac{1}{Q} \sum_{q=1}^{Q} \prod_{t=1}^{T_i} \prod_{j=1}^{J} \Big[\phi\Big(w_{ijt};\ \alpha_j + \boldsymbol{\xi}_i^{(q)\prime} \boldsymbol{\lambda}_j + \boldsymbol{\beta}_j' \log(\boldsymbol{\tau}_{it}),\ \varsigma_j^2 \Big) \Big]^{d_{ijt}}. \tag{59}$$

In computation we avoid multiplying densities. For each draw define the time–summed log contribution

$$S_i^{(q)}(\boldsymbol{\Gamma}) = \sum_{t=1}^{T_i} \sum_{j=1}^{J} d_{ijt}\ \log \phi\Big(w_{ijt};\ \alpha_j + \boldsymbol{\xi}_i^{(q)\prime} \boldsymbol{\lambda}_j + \boldsymbol{\beta}_j' \log(\boldsymbol{\tau}_{it}),\ \varsigma_j^2 \Big),$$

where the univariate Normal log-density is evaluated as

$$\log \phi(w; \mu, \varsigma^2) = -\tfrac{1}{2} \log(2\pi\varsigma^2) - \tfrac{(w-\mu)^2}{2\varsigma^2}.$$

Then the Monte Carlo log-likelihood is computed by the log-mean-exp identity,

$$\log \tilde{\mathcal{L}}_i^w(\boldsymbol{\Gamma}) \approx \log\Big(\tfrac{1}{Q} \sum_{q=1}^{Q} e^{S_i^{(q)}(\boldsymbol{\Gamma})} \Big) = \operatorname{logsumexp}\big(\{ S_i^{(q)}(\boldsymbol{\Gamma}) \}_{q=1}^{Q} \big) - \log Q, \tag{60}$$

with the stabilized evaluation

$$\operatorname{logsumexp}(a_1, \ldots, a_Q) = a_{\max} + \log \sum_{q=1}^{Q} e^{a_q - a_{\max}}, \qquad a_{\max} = \max_q a_q.$$

To stabilize comparisons across parameter values, we use common random numbers (CRN). For each worker i we fix a sequence of standard–normal draws $\boldsymbol{Z}_i^{(q)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and map them through the current group prior:

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

Hence the simulation noise is held fixed while $\boldsymbol{\xi}_i^{(q)}$ updates coherently with $(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$. When increasing Q, we append additional draws rather than resampling.

With R distinct initializations $\{\mathbf{\Gamma}^{(0,r)}\}_{r=1}^{R}$, we run the EM algorithm to convergence for each start, yielding limit points $\widehat{\mathbf{\Gamma}}^{(r)}$. For each converged solution we evaluate the Monte Carlo sample log-likelihood using (60) at a fixed draw size Q (and the same CRN across r),

$$\widehat{\ell}_Q(\mathbf{\Gamma}) := \sum_{i=1}^{N} \log \tilde{\mathcal{L}}_i^w(\mathbf{\Gamma}).$$

The selected estimator is the maximizer across starts,

$$\widehat{\mathbf{\Gamma}} = \widehat{\mathbf{\Gamma}}^{(r^\star)}, \qquad r^\star \in \arg\max_{1\le r\le R} \widehat{\ell}_Q\big(\widehat{\mathbf{\Gamma}}^{(r)}\big).$$

Using common random numbers ensures that differences in $\widehat{\ell}_Q(\widehat{\mathbf{\Gamma}}^{(r)})$ across starts primarily reflect fit rather than simulation noise.

E.2 NFXP Algorithm: Computational Details

E.2.1 Gaussian–Hermite Quadrature for Multivariate Normal Log Skills

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

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

where $\tilde{\boldsymbol{\xi}}_t \in \mathbb{R}^K$ and $\tilde{\mathbf{\Sigma}}_t \in \mathbb{R}^{K\times K}$ are the posterior mean and covariance of log skills $\boldsymbol{\xi} \equiv \log \boldsymbol{\theta}$, $\boldsymbol{\tau}_t \in \mathbb{R}_+^K$ are task tenures, and $d_{t-1} \in \mathcal{O}$ is last period's occupation (an explicit state variable). If the worker chooses occupation $m \in \mathcal{O}$ at time t, she pays a contemporaneous switching cost $\mathcal{C}(m, d_{t-1}, e_t) \ge 0$ (with $\mathcal{C}(d_{t-1}, d_{t-1}, e_t) = 0$), which only shifts the current deterministic index (wage plus amenities) and does not affect belief updating or the law of motion for tenures.

Conditional on choosing m at t, the end-of-period log wage is

$$\log W_{mt} = p_{mt} + \boldsymbol{\lambda}_m^\top \boldsymbol{\xi} + \boldsymbol{\beta}_m^\top \log \boldsymbol{\tau}_t + \varepsilon_{mt}, \qquad \varepsilon_{mt} \sim \mathcal{N}(0, \varsigma_m^2),$$

where p_{mt} is the log occupation price, $\boldsymbol{\lambda}_m \in \mathbb{R}^K$ are factor loadings, and $\boldsymbol{\beta}_m \in \mathbb{R}^K$ are the linear returns to $\log \boldsymbol{\tau}_t$. Define the demeaned signal

$$s_{mt} \;\equiv\; \log W_{mt} - p_{mt} - \boldsymbol{\beta}_m^\top \log \boldsymbol{\tau}_t = \boldsymbol{\lambda}_m^\top \boldsymbol{\xi} + \varepsilon_{mt}. \tag{61}$$

The linear–Gaussian update yields

$$\tilde{\boldsymbol{\xi}}_{t+1} = \tilde{\boldsymbol{\xi}}_t + \mathbf{G}_{mt}\Big(s_{mt} - \boldsymbol{\lambda}_m^\top \tilde{\boldsymbol{\xi}}_t\Big), \tag{62}$$

$$\tilde{\mathbf{\Sigma}}_{t+1} = \tilde{\mathbf{\Sigma}}_t - \mathbf{G}_{mt} \boldsymbol{\lambda}_m^\top \tilde{\mathbf{\Sigma}}_t, \tag{63}$$

$$\mathbf{G}_{mt} = \tilde{\mathbf{\Sigma}}_t \boldsymbol{\lambda}_m \Big(\boldsymbol{\lambda}_m^\top \tilde{\mathbf{\Sigma}}_t \boldsymbol{\lambda}_m + \varsigma_m^2\Big)^{-1}, \qquad s_{zz} \equiv \boldsymbol{\lambda}_m^\top \tilde{\mathbf{\Sigma}}_t \boldsymbol{\lambda}_m + \varsigma_m^2. \tag{64}$$

Tenures evolve deterministically as $\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{b}_m$, and the next "previous occupation" component of the state becomes $d_t = m$.

Given $(\tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t)$ and choice m, the next covariance $\tilde{\boldsymbol{\Sigma}}_{t+1}$ is deterministic:

$$\tilde{\boldsymbol{\Sigma}}_{t+1} = \left(\tilde{\boldsymbol{\Sigma}}_t^{-1} + \varsigma_m^{-2}\,\boldsymbol{\lambda}_m\boldsymbol{\lambda}_m^\top\right)^{-1} = \tilde{\boldsymbol{\Sigma}}_t - \frac{\tilde{\boldsymbol{\Sigma}}_t\boldsymbol{\lambda}_m\boldsymbol{\lambda}_m^\top\tilde{\boldsymbol{\Sigma}}_t}{\boldsymbol{\lambda}_m^\top\tilde{\boldsymbol{\Sigma}}_t\boldsymbol{\lambda}_m + \varsigma_m^2}. \tag{65}$$

By contrast, the next posterior mean $\tilde{\boldsymbol{\xi}}_{t+1}$ is random ex ante because it depends on the yet-unseen innovation $s_{mt} - \boldsymbol{\lambda}_m^\top\tilde{\boldsymbol{\xi}}_t$. Taking expectations in (62) and using $\mathbb{E}[\,s_{mt} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, m\,] = \boldsymbol{\lambda}_m^\top\tilde{\boldsymbol{\xi}}_t$ gives

$$\mathbb{E}\left[\tilde{\boldsymbol{\xi}}_{t+1} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, m\right] = \tilde{\boldsymbol{\xi}}_t.$$

Its ex-ante covariance is

$$\operatorname{Var}\left(\tilde{\boldsymbol{\xi}}_{t+1} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, m\right) = \mathbf{G}_{mt}\,s_{zz}\,\mathbf{G}_{mt}^\top = \frac{\tilde{\boldsymbol{\Sigma}}_t\boldsymbol{\lambda}_m\boldsymbol{\lambda}_m^\top\tilde{\boldsymbol{\Sigma}}_t}{\boldsymbol{\lambda}_m^\top\tilde{\boldsymbol{\Sigma}}_t\boldsymbol{\lambda}_m + \varsigma_m^2} \;\equiv\; \boldsymbol{\Sigma}_{t+1}^{(m)}. \tag{66}$$

Combining (65)–(66) yields the law-of-total-variance identity

$$\tilde{\boldsymbol{\Sigma}}_t = \underbrace{\tilde{\boldsymbol{\Sigma}}_{t+1}}_{\text{uncertainty after observing the wage (deterministic)}} + \underbrace{\boldsymbol{\Sigma}_{t+1}^{(m)}}_{\text{variance of the next posterior mean (random)}}, \qquad \boldsymbol{\Sigma}_{t+1}^{(m)} = \tilde{\boldsymbol{\Sigma}}_t - \tilde{\boldsymbol{\Sigma}}_{t+1}. \tag{67}$$

Let G be any measurable functional of the next-period state. Because the next "previous occupation" is $d_t = m$, the continuation term conditional on choosing m is

$$\mathbb{E}\left[G(\tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, d_t = m) \,\middle|\, \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, d_{t-1}, m\right] = \int_{\mathbb{R}^K} G\big(\boldsymbol{z},\, \tilde{\boldsymbol{\Sigma}}_{t+1}(m),\, \boldsymbol{\tau}_t + \boldsymbol{b}_m,\, m\big)\, \phi_K\big(\boldsymbol{z}; \tilde{\boldsymbol{\xi}}_t, \boldsymbol{\Sigma}_{t+1}^{(m)}\big)\, d\boldsymbol{z}. \tag{68}$$

Switching costs only enter the current period's deterministic index (wage + amenities − cost) and therefore lie outside the integral in (68); the integral averages solely over the randomness in the next posterior mean.

Let $\boldsymbol{\Sigma}_{t+1}^{(m)} = \boldsymbol{U}_{t+1}^{(m)}\boldsymbol{U}_{t+1}^{(m)\top}$ be the (lower-triangular) Cholesky factorization. With the change of variables

$$\boldsymbol{x} = \big(\boldsymbol{U}_{t+1}^{(m)}\big)^{-1}\big(\boldsymbol{z} - \tilde{\boldsymbol{\xi}}_t\big)/\sqrt{2} \qquad \Longleftrightarrow \qquad \boldsymbol{z} = \tilde{\boldsymbol{\xi}}_t + \sqrt{2}\,\boldsymbol{U}_{t+1}^{(m)}\boldsymbol{x},$$

the Jacobian gives $d\boldsymbol{z} = 2^{K/2}\,|\det \boldsymbol{U}_{t+1}^{(m)}|\,d\boldsymbol{x}$ and

$$\phi_K\big(\boldsymbol{z}; \tilde{\boldsymbol{\xi}}_t, \boldsymbol{\Sigma}_{t+1}^{(m)}\big)\, d\boldsymbol{z} = \pi^{-K/2}\, e^{-\boldsymbol{x}^\top\boldsymbol{x}}\, d\boldsymbol{x}.$$

Hence the Gaussian–Hermite transform of (68) is

$$\int_{\mathbb{R}^K} G\big(\boldsymbol{z}, \tilde{\boldsymbol{\Sigma}}_{t+1}(m), \boldsymbol{\tau}_t + \boldsymbol{b}_m, m\big)\, \phi_K\big(\boldsymbol{z}; \tilde{\boldsymbol{\xi}}_t, \boldsymbol{\Sigma}_{t+1}^{(m)}\big)\, d\boldsymbol{z} = \pi^{-K/2} \int_{\mathbb{R}^K} G\Big(\tilde{\boldsymbol{\xi}}_t + \sqrt{2}\,\boldsymbol{U}_{t+1}^{(m)}\boldsymbol{x},\, \tilde{\boldsymbol{\Sigma}}_{t+1}(m),\, \boldsymbol{\tau}_t + \boldsymbol{b}_m,\, m\Big)\, e^{-\boldsymbol{x}^\top\boldsymbol{x}}\, d\boldsymbol{x}. \tag{69}$$

Let $\{(\nu_r, \omega_r)\}_{r=1}^N$ be the univariate Gauss–Hermite nodes and weights satisfying $\int_{\mathbb{R}} g(x)e^{-x^2}dx \approx \sum_{r=1}^N \omega_r\, g(\nu_r)$. For K dimensions,

$$\int_{\mathbb{R}^K} g(\boldsymbol{x})\, e^{-\boldsymbol{x}^\top\boldsymbol{x}}\, d\boldsymbol{x} \;\approx\; \sum_{r_1=1}^{N} \cdots \sum_{r_K=1}^{N} \big(\omega_{r_1} \cdots \omega_{r_K}\big)\, g\big(\nu_{r_1}, \ldots, \nu_{r_K}\big).$$

Applying this quadrature to (69) yields a numerically stable approximation to the continuation

expectation in (68).

Let $s(\boldsymbol{x}_t) \equiv 1 - \delta\big(\sum_k \tau_{kt}\big) \in (0,1)$ be the survival factor. Define the choice-specific deterministic index at state $\boldsymbol{x}_t$ by

$$\nu_m(\boldsymbol{x}_t) \equiv \mathbb{E}\big[\log W_{mt} \mid \tilde{\boldsymbol{\xi}}_t, \tilde{\boldsymbol{\Sigma}}_t, \boldsymbol{\tau}_t, p_{mt}\big] + \eta_m - \mathcal{C}(m, d_{t-1}, e_t) + \beta\, s(\boldsymbol{x}_t)\, \mathbb{E}\big[W(\tilde{\boldsymbol{\xi}}_{t+1}, \tilde{\boldsymbol{\Sigma}}_{t+1}, \boldsymbol{\tau}_{t+1}, d_t = m) \mid \boldsymbol{x}_t, m\big],$$

where the continuation expectation is approximated via (69). Then the integrated value is

$$W(\boldsymbol{x}_t) = \varrho \log \sum_{m \in \mathcal{O}} \exp\Big(\nu_m(\boldsymbol{x}_t)/\varrho\Big),$$

where the only role of switching costs in this subsection is to subtract $\mathcal{C}(m, d_{t-1}, e_t)$ from the static component of $\nu_m(\boldsymbol{x}_t)$; the Gaussian–Hermite integral itself averages exclusively over the randomness in the next posterior mean.

E.2.2 Occupation–Specific Integrated Values and Fixed Point

Let $J := |\mathcal{O}|$. We work with occupation-specific integrated values

$$W_j(\boldsymbol{x}) \;\equiv\; W\big(\tilde{\boldsymbol{\mathcal{F}}}, \boldsymbol{\tau}, d_{t-1} = j\big), \qquad \boldsymbol{x} := (\tilde{\boldsymbol{\mathcal{F}}}, \boldsymbol{\tau}) \in \mathcal{X},$$

i.e., the log-sum value before idiosyncratic taste shocks when the lagged occupation equals j. Let the survival factor be $s(\boldsymbol{x}) = 1 - \delta(T) \in (0,1)$, where $T = \sum_{k=1}^{K} \tau_k$. For a prospective choice $m \in \mathcal{O}$, define the deterministic index

$$\begin{aligned} \nu_{m|j}(\boldsymbol{x}) = \mathbb{E}\big[u(w_m(\boldsymbol{\tau})) \mid \tilde{\boldsymbol{\mathcal{F}}}\big] + z_m - \mathcal{C}(m, j, e) \\ + \beta\, s(\boldsymbol{x})\, \mathbb{E}\big[\, W_m(\boldsymbol{x}') \;\big|\; \boldsymbol{x}, m\big], \end{aligned} \tag{70}$$

where $\mathcal{C}(m, j, e) = \mathbf{1}\{m \neq j\}\,\big(\gamma_0^c + \gamma_1^c e + \gamma_2^c e^2\big)$, and the continuation expectation is evaluated by Gaussian–Hermite quadrature using the mapping $\tilde{\boldsymbol{\xi}}' = \tilde{\boldsymbol{\xi}} + \sqrt{2}\, \boldsymbol{U}_{t+1}^{(m)} \boldsymbol{x}^{(\mathrm{gh})}$ and $\boldsymbol{\tau}' = \boldsymbol{\tau} + \boldsymbol{b}_m$. With i.i.d. EV–I$(0, \varrho)$ shocks, the occupation-specific integrated values satisfy, for each $j \in \mathcal{O}$,

$$W_j(\boldsymbol{x}) = \tilde{\gamma}\, \varrho + \varrho\, \log\left(\sum_{m \in \mathcal{O}} \exp\Big(\nu_{m|j}(\boldsymbol{x})/\varrho\Big) \right), \qquad \boldsymbol{x} \in \mathcal{X}, \tag{71}$$

where $\tilde{\gamma} = 0$ under the mean-zero Gumbel normalization (or $\tilde{\gamma} = \gamma \simeq 0.57721$ under the standard parameterization). Define $W(\boldsymbol{x}) = (W_1(\boldsymbol{x}), \ldots, W_J(\boldsymbol{x}))^\top$. The Bellman operator $\mathcal{T}$ maps W into $\mathcal{T}W$ by applying (71) componentwise at all $(\boldsymbol{x}, j)$.

At state $(\boldsymbol{x}, j)$, the CCPs are

$$\mathbb{P}\big(d_t = m \mid \boldsymbol{x}, j\big) = \frac{\exp\big(\nu_{m|j}(\boldsymbol{x})/\varrho\big)}{\sum_{\ell \in \mathcal{O}} \exp\big(\nu_{\ell|j}(\boldsymbol{x})/\varrho\big)}, \qquad m \in \mathcal{O}. \tag{72}$$

E.2.3 Smolyak (Sparse–Grid) Approximation Integrated Value Function

The continuous state $\boldsymbol{x} = (\tilde{\boldsymbol{\xi}}, \boldsymbol{\tau}) \in \mathbb{R}^{2K}$ can be high-dimensional. We approximate each $W_j(\cdot)$ on a sparse grid using the Smolyak method (Malin et al. 2011) with Chebyshev polynomials.

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

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

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

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

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

$$\widehat{W}_j^{d,\mu}(\boldsymbol{x}) = \sum_{d \le |\mathbf{i}| \le d+\mu} (-1)^{d+\mu-|\mathbf{i}|} \binom{d-1}{d+\mu-|\mathbf{i}|} \, W_j^{\mathbf{i}}(\boldsymbol{x}; \boldsymbol{\phi}^{(j,\mathbf{i})}). \tag{73}$$

E.2.4 Computation of the Occupation–Specific Fixed Point

We compute the fixed point of (71) jointly for all $j \in \mathcal{O}$ using value iteration with sparse-grid interpolation.

1. Choose bounds for $\boldsymbol{x} = (\tilde{\boldsymbol{\xi}}, \boldsymbol{\tau})$ and build the sparse grid $\mathcal{H}^{d,\mu}$.
2. Select GH order N and tabulate the univariate nodes/weights $\{(\nu_r, \omega_r)\}_{r=1}^{N}$.
3. Initialize all Smolyak coefficient blocks $\{\boldsymbol{\phi}^{(j,\mathbf{i})}\}$ for $j \in \mathcal{O}$, $d \le |\mathbf{i}| \le d+\mu$.

Given current approximations $\{\widehat{W}_m^{d,\mu}\}_{m \in \mathcal{O}}$, update each component W_j at each grid node $\boldsymbol{x} \in \mathcal{H}^{d,\mu}$:

1. For each prospective choice $m \in \mathcal{O}$, form the deterministic index

$$v_{m|j}(\boldsymbol{x}) = \mathbb{E}[u(w_m(\boldsymbol{\tau})) \mid \tilde{\boldsymbol{\mathcal{F}}}] + z_m - \mathcal{C}(m, j, e) + \beta \, s(\boldsymbol{x}) \, \mathbb{E}\left[\widehat{W}_m^{d,\mu}(\boldsymbol{x}') \mid \boldsymbol{x}, m \right],$$

   where the inner expectation is approximated by GH using (69) and the mapping $\tilde{\boldsymbol{\xi}}' = \tilde{\boldsymbol{\xi}} + \sqrt{2}\, \boldsymbol{U}_{t+1}^{(m)} \boldsymbol{x}^{(\mathrm{gh})}$ and $\boldsymbol{\tau}' = \boldsymbol{\tau} + \boldsymbol{b}_m$.

2. Compute

$$\widetilde{W}_j(\boldsymbol{x}) = \tilde{\gamma}\, \varrho + \varrho \, \log\left(\sum_{m \in \mathcal{O}} \exp\{ v_{m|j}(\boldsymbol{x}) / \varrho \} \right).$$

3. Update coefficients $\{\boldsymbol{\phi}^{(j,\mathbf{i})}\}$ so that the interpolant $\widehat{W}_j^{d,\mu}(\boldsymbol{x})$ (via (73)) matches $\widetilde{W}_j(\boldsymbol{x})$ at all nodes $\boldsymbol{x} \in \mathcal{H}^{d,\mu}$.

Iterate until the uniform difference across occupations falls below the tolerance:

$$\max_{j \in \mathcal{O}} \sup_{\boldsymbol{x} \in \mathcal{H}^{d,\mu}} \left| \widehat{W}_{j,\mathrm{new}}^{d,\mu}(\boldsymbol{x}) - \widehat{W}_{j,\mathrm{old}}^{d,\mu}(\boldsymbol{x}) \right| < \mathrm{tol}.$$

For small ϱ, compute $\varrho \log\left(\sum_m e^{v_m/\varrho}\right)$ stably by shifting with $v_{\max} = \max_m v_m$:

$$\begin{aligned}\varrho \log\Big(\sum_m e^{v_m/\varrho}\Big) &= \varrho \log\Big(e^{v_{\max}/\varrho} \sum_m e^{(v_m - v_{\max})/\varrho}\Big) \\ &= v_{\max} + \varrho \log\Big(\sum_m e^{(v_m - v_{\max})/\varrho}\Big).\end{aligned}$$

E.2.5 Policy Functions, CCPs, and Simulation

Given $\{\widehat{W}_j^{d,\mu}\}_{j\in\mathcal{O}}$, the CCPs at $(\boldsymbol{x}, j)$ follow from (72). To simulate a worker's path:

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

$$\log W_{d_t t} = p_{d_t t} + \boldsymbol{\lambda}_{d_t}^\top \boldsymbol{\xi} + \boldsymbol{\beta}_{d_t}^\top \log \boldsymbol{\tau}_t + \varepsilon_{d_t,t}.$$

   Equivalently, the signal in (61) is $s_{d_t t} = \log W_{d_t t} - p_{d_t t} - \boldsymbol{\beta}_{d_t}^\top \log \boldsymbol{\tau}_t$.
3. Update beliefs via (62)–(64) using $m = d_t$.
4. Update tenures: $\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}_t + \boldsymbol{b}_{d_t}$.
5. Update the discrete state: $j \leftarrow d_t$ (next period's lagged occupation).

E.2.6 Differentiable NFXP Implementation (JAX)

Estimation of the dynamic parameters $\boldsymbol{\Delta}$ requires a nested fixed point (NFXP) approach (Rust 1987): for each candidate $\boldsymbol{\Delta}$, the inner loop solves the worker's dynamic program to obtain continuation values, and the outer loop evaluates the choice log-likelihood. One likelihood evaluation is the composition

$$\boldsymbol{\Delta} \longmapsto \Phi^\star(\boldsymbol{\Delta}) \longmapsto P_{\boldsymbol{\Delta}}(\cdot \mid x) \longmapsto \ell^c(\boldsymbol{\Delta}),$$

where $\Phi^\star(\boldsymbol{\Delta})$ denotes the sparse-grid coefficients of the integrated value function at the fixed point. I implement this pipeline in JAX, exploiting reverse-mode automatic differentiation (AD) to obtain exact gradients of the outer objective without finite differences.

The taste-shock scale $\varrho = 0.25$ is held fixed throughout (see Section **??**). The optimizer works with an unconstrained vector θ, mapped to the remaining structural parameters via smooth transformations (softplus for positivity constraints on κ_0, κ_1; amenities anchored at $\eta_0 = 0$). This yields $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\theta)$ with well-defined derivatives throughout.

The central challenge is differentiating through the inner fixed point. A standard tolerance-based iteration introduces variable-length control flow that is incompatible with AD. I instead run a *fixed* number K of Bellman iterations on the sparse-grid coefficients:

$$\Phi^{(k+1)} = T(\Phi^{(k)}; \boldsymbol{\Delta}), \qquad k = 0, \ldots, K-1,$$

where T is the Bellman operator applied at the sparse-grid nodes, with continuation values computed via interpolation and Gaussian–Hermite quadrature. With K fixed, $\Phi^K(\boldsymbol{\Delta})$ is a finite composition of differentiable maps, making the entire inner loop amenable to AD. K is chosen large enough that Φ^K is within numerical tolerance of $\Phi^\star$.

Given Φ^K, the implied CCPs follow the multinomial logit form:

$$P_{\boldsymbol{\Delta}}(d = j \mid x) = \frac{\exp(v_j(x; \boldsymbol{\Delta}, \Phi^K)/\varrho)}{\sum_{m \in \mathcal{O}} \exp(v_m(x; \boldsymbol{\Delta}, \Phi^K)/\varrho)},$$

where $v_j(\cdot)$ includes current expected utility and the discounted continuation value. The choice log-likelihood is

$$\ell^c(\boldsymbol{\Delta}) = \sum_{i=1}^{N} \log P_{\boldsymbol{\Delta}}(d_i \mid x_i),$$

evaluated in fixed-size batches to keep GPU memory bounded.

The outer objective $\mathcal{L}(\theta) \equiv -\ell^c(\boldsymbol{\Delta}(\theta))$ depends on θ both directly through $\boldsymbol{\Delta}$ and indirectly through the inner solution $\Phi^K(\boldsymbol{\Delta})$. The total derivative is

$$\nabla_\theta \mathcal{L}(\theta) = -\left[\frac{\partial \ell^c}{\partial \boldsymbol{\Delta}} + \frac{\partial \ell^c}{\partial \Phi}\frac{\partial \Phi^K}{\partial \boldsymbol{\Delta}}\right]\frac{\partial \boldsymbol{\Delta}}{\partial \theta}.$$

Since $\Phi^K(\boldsymbol{\Delta}) = T(\cdot; \boldsymbol{\Delta}) \circ \cdots \circ T(\cdot; \boldsymbol{\Delta})(\Phi^{(0)})$, reverse-mode AD computes $\partial \Phi^K / \partial \boldsymbol{\Delta}$ by propagating adjoints backward through the K Bellman steps. Letting $\bar{\Phi}^{(k)}$ denote the adjoint at step k, the backward recursion is

$$\bar{\Phi}^{(k)} = \left(\frac{\partial T}{\partial \Phi}\Big|_{\Phi^{(k)}, \boldsymbol{\Delta}}\right)^\top \bar{\Phi}^{(k+1)}, \qquad k = K-1, \ldots, 0,$$

with gradient contributions accumulated at each step:

$$\nabla_{\boldsymbol{\Delta}} \ell^c \;\leftarrow\; \nabla_{\boldsymbol{\Delta}} \ell^c + \left(\frac{\partial T}{\partial \boldsymbol{\Delta}}\Big|_{\Phi^{(k)}, \boldsymbol{\Delta}}\right)^\top \bar{\Phi}^{(k+1)}.$$

The resulting gradient is an analytical derivative of the NFXP objective for the chosen K, with no finite-difference approximation. To manage memory, intermediate Bellman iterates are recomputed during the backward pass (checkpointing) rather than stored. The full NFXP evaluation—inner fixed-point iteration plus batched likelihood on $N \simeq 2.4$ million observations—runs in approximately 31 minutes on a single NVIDIA H100 GPU node.