On the Identification of Models of Uncertainty, Learning, and Human Capital Acquisition with Sorting*

Aureo de Paula[†] Cristina Gualdani[‡] Elena Pastorino[§] Sergio Salgado[¶]

January 2025

(click here for updated version)

Abstract

We examine the empirical content of dynamic matching models of the labor market with ex ante heterogeneous firms and workers in the presence of symmetric uncertainty and learning about worker ability and workers' human capital acquisition. We allow ability and acquired human capital to be general across firms to varying degrees. We establish conditions under which the primitives of these models are semiparametrically identified based only on data on workers' jobs and wages. Through the lens of this class of models, we investigate the ability of existing empirical measures of the assortativeness of the matching between firms and workers to detect the actual degree of sorting in the labor market. We propose a new measure of matching assortativeness that accounts for the evolving uncertainty about workers' ability and workers' accumulating human capital.

Keywords: Uncertainty, Learning, Human Capital Acquisition, Matching, Estimation, Sorting, Wage Variance, Inequality

*We are grateful to Jerome Adda, Richard Blundell, Stephane Bonhomme, Xavier D'Haultfoeuille, and Han Hong for many useful comments. This is a reworked version of the paper that first circulated in October 2022 with the same title. This work was supported by the Economic and Social Research Council (grant ES/X011186/1).

[†]University College London. E-mail: a.paula@ucl.ac.uk.

[‡]Queen Mary University of London. E-mail: c.gualdani@qmul.ac.uk.

[§]Stanford University, Hoover Institution, and NBER. E-mail: epastori@stanford.edu.

[¶]Wharton School, University of Pennsylvania. E-mail: ssalgado@wharton.upenn.edu.

1 Introduction

Matching models of the labor market with heterogeneous firms and workers have been extensively used in both the labor and macro literature to study a multiplicity of phenomena ranging from occupational choice and worker turnover to wage growth, wage inequality, and sorting. Traditionally, these models account for workers' occupation decisions and career paths as resulting from a combination of two key processes—workers' human capital accumulation over time and gradual learning about workers' true productivity—governing workers' labor market experience. This general class encompasses models of wage growth (Becker, 1975; Mincer, 1958, 1974; Ben-Porath, 1967) and inequality (Heckman and Honoré, 1990), models of job choice and mobility without (à la (Keane and Wolpin, 1997) and with (Farber and Gibbons, 19996; Gibbons and Waldman, 1999a,b) learning about ability, and many more that nest or extend elements of each such class of models, including Jovanovic (1979), Miller (1984), Flinn (1986), Jovanovic and Nyarko (1997), Gibbons and Waldman (2006), Gibbons, Katz, Lemieux, and Parent (2005), Lange (2007), Nagypál (2007), Antonovics and Golan (2012), Kahn and Lange (2014), and Pastorino (2024). See Gibbons and Waldman (1999a) and Rubinstein and Weiss (2006) for reviews that emphasize the centrality of uncertainty and learning about workers' ability for explaining the dispersion of wages across workers and over time.

Despite their widespread use to isolate and interpret the determinants of wage inequality across workers and for a worker over time, the empirical content of these models is known to be hard to establish due to three well-understood challenges: the complex dynamic selection process based on past unobserved characteristics, outcomes, and choices of both firms and workers, which determines workers' jobs and wages; the discrete nature of workers' choices; and the endogenous nonlinearity of the continuous outcome of interest—a worker's wages—in the relevant primitives, due to the intertemporally optimizing behavior of both workers and firms. In this paper, we establish the semi-parametric identification of this general class of models with uncertainty and learning about workers' ability, allowing for imperfect competition among firms unlike much of the models we build on, based *only* on data on workers' jobs and wages. We then use these conditions to reconcile the empirical puzzle of why measured sorting is typically very low, even in labor markets characterized by high degrees of wage inequality, which is usually interpreted as an outcome of sorting. We finally derive constructive estimators of the primitives of these models that are easy to implement based on simple routines normally used to estimate static (Roy) models.

More formally, we consider a broad family of non-stationary dynamic matching models of

Bertrand competition between firms and workers under uncertainty and learning about worker ability. Firms compete for workers by offering wages for employment in each period, and they are heterogeneous in their technologies of output, human capital, and information production. Workers differ in their initial and acquired human capital (observable by both workers and firms), in their efficiency (observable by both workers and firms), and in their ability, which is unknown when workers enter the labor market and is gradually learned by both parties as workers accumulate experience. We characterize the equilibrium of these models and show that equilibrium wages consist of the sum of the expected output produced by the *second-best* competing firm—akin to a second-price auction mechanism—and the dynamic value of foregone opportunities for human capital accumulation and information acquisition at the other firms, functioning as a *compensating differential*.

From an econometric viewpoint, these models reduce to dynamic generalized equilibrium Roy models with selection on *unobservables*. Specifically, workers' job choices depend on all the unobservables that enter wages: worker efficiency type (time-invariant), beliefs of workers and firms about worker ability (time-varying, serially correlated, and endogenously evolving with states and choices), and productivity shocks (time-varying and serially uncorrelated). Showing identification of these models is challenging because some of the unobservables causing selection—in particular, worker efficiency type and beliefs—enter the wage equation in a potentially nonmonotonic, nonseparable, and nonmultiplicative manner, rendering the identification arguments developed for panel data models with interactive fixed effects a nonviable approach. This is because the compensating differential appearing in equilibrium wages depends on these unobservables and is the difference between two value functions—hence, endogenous dynamic payoffs with an unknown form. Furthermore, no exclusion restrictions arise from the theory of this class of models that can help us easily solve the selection issue, unlike in most Roy settings.

We show that it is possible to overcome these challenges and develop identification arguments for this rich class of models by merging a mixture-based approach with the identification strategies used for static Roy models without exclusion restrictions. Essentially, the mixture-based approach resolves selection arising from the most problematic unobservables—those that enter the wage equation in an unknown and potentially nonlinear way (as previously mentioned, the worker efficiency type and beliefs). Once this challenging source of selection is addressed, we then apply identification results from the static Roy model literature to handle selection induced by the productivity shocks, which is more standard since these shocks are serially uncorrelated and enter wages in an additively separable manner.

More precisely, we first represent the observed wage distribution as a mixture over some of the unobservables in our model. We demonstrate identification of this mixture under mild conditions—namely, that the wage distribution can be expressed as a finite mixture of mixtures of possibly uncountable families of Gaussian distributions, a specification that can virtually approximate any distribution (Bruni and Koch, 1985; Gosh and Ramamoorthi, 2003; Norets, 2010; Norets and Pelenis, 2014; Aragam, Dan, Xing, and Ravikumar, 2020). By concatenating the mixture weights over time and thus exploiting the longitudinal dimension of the data, we show that it is possible to recover the law of motion for some unobserved state variables, in particular, the worker efficiency type and beliefs. With this primitive at hand, we then identify the wage distribution conditional on job choices and state variables, including the recovered worker efficiency type and beliefs, with the exception of the productivity shocks. This distribution is yet insufficient to identify the deterministic component of wages at each firm (encompassing the technology of output production at the second-best firm and the compensating differential) because it remains contaminated by selection on the productivity shocks. However, as mentioned, this type of selection is standard in static Roy models—and having addressed all other sources of selection via the mixture step—we can now solve it by applying identification arguments from the static Roy literature. In particular, we adapt to our second-price auction setting the framework of D'Haultfoeuille and Maurel (2013), which demonstrates how to identify the deterministic wage component in a static Roy model without the need for exclusion restrictions by simply examining the extreme tail of the wage distribution. This tail can be shown not to suffer from selection on the productivity shocks, provided that such shocks are only moderately dependent across firms. Once the deterministic wage components are identified, we show how to recover the technology of output production for each firm, the compensating differential, the distribution of productivity shocks, and the conditional choice probabilities. All of these primitives are key to addressing our question about the impact of sorting on earnings inequality, which we focus on in the empirical application.

In the empirical part of the paper, we use this class of models and the econometric approach developed to measure how sorting between workers and firms affects earnings inequality in the U.S. The most widely used framework for addressing this question is that of Abowd, Kramarz, and Margolis (1999)—hereafter, AKM—which decomposes wages into worker and firm fixed effects, observable covariates, and random shocks. From these estimates (henceforth, AKM estimates), the wage variance is partitioned into contributions from worker effects, firm effects, their covariance, and the residual. The impact of sorting on earnings inequality is then gauged by the fraction of total wage variance attributable to the covariance between worker and firm effects. Empirical applications of this framework often indicate a negligible role for sorting, reflecting weak correlations between worker and firm effects. See, for instance, Song, Price, Guvenen, Bloom, and Von Wachter (2019) and Card, Heining, and Kline (2013).[1]

Building on the theoretical insights from the class of models we study, we argue that the AKM estimates of the correlation between firm and worker effects may be understated because two key forces are omitted. First, the *compensating differential* can dampen the direct impact of worker and firm characteristics on wages, as it compensates workers for the foregone future returns in human capital and information that they would have gained by accepting offers from competing firms. Second, *endogenous matching frictions*—namely, the worker's acquisition of human capital and the gradual resolution of uncertainty about ability—may prevent high-type workers from immediately joining the most productive firms. For example, workers might temporarily choose less-productive firms that offer valuable training or learning opportunities, challenging the assumption that they always sort into the most immediately productive match.

To empirically validate these theoretical conjectures, we provide both simulation-based and empirical evidence. In a Monte Carlo simulation, we generate an economy that mimics the key features of our class of models. We calibrate the simulation parameters to match key earnings moments from the Panel Study of Income Dynamics (PSID), a representative survey of U.S. households dating back to 1968, and target AKM-type moments from Song et al. (2019) derived from Social Security Administration (SSA) data. In a setting analogous to standard omitted variable bias, our findings suggest that when the compensating differential is negative in the true data-generating process—implying that workers match with firms offering human capital and information gains with higher future returns than competitors—the AKM estimates underestimate the impact of sorting on earnings inequality because the omitted compensating differential attenuates the complementarities between firm and worker characteristics in output technology. Conversely, when the compensating differential is positive—implying that workers match with firms offering human capital and information gains with lower future returns than competitors—the AKM estimates overestimate the impact of sorting because the omitted compensating differential amplifies these complementarities.

Next, we estimate our wage equation using U.S. employer-employee match data from the Longitudinal Employer-Household Dynamics (LEHD) dataset, providing quarterly earnings for all work-

[1] An important contribution by Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Bradley (2023) shows that once the AKM estimates are adjusted for biases stemming from workers' limited job mobility, the correlation between worker and firm effects generally increases.

ers across 21 states—including the largest ones—from the mid-1990s to 2022. We combine Gaussian mixture estimation with extremal quantile regression (D'Haultfoeuille, Maurel, and Zhang, 2018; D'Haultfoeuille, Maurel, Qiu, and Zhang, 2020), following our constructive identification approach. Our empirical results corroborate the findings from our simulations. In particular, the AKM estimates of the impact of sorting on earnings inequality are lower than our own estimates, thereby resolving the sorting puzzle.

To further support this key finding, we conduct an exercise designed to capture sorting in our rich class of models more comprehensively. The AKM framework measures sorting solely with respect to the worker's time-invariant efficiency type. In contrast, our setting allows workers to sort on their beliefs about ability and on accumulated human capital (endogenous matching frictions). To capture these additional dimensions, we perform a random reallocation exercise, comparing the observed earnings distribution to a counterfactual scenario in which workers and firms are matched at random, thereby eliminating any endogenous links. If sorting has a substantial impact, then disrupting these links should markedly reduce both earnings dispersion and the concentration of high earnings, as workers would no longer cluster in firms offering the highest productivity or the most valuable human capital and informational benefits. Our evidence supports all of these mechanisms. We conclude by discussing how to implement the estimators of the primitives of interest based on routines readily available and commonly used to estimate static Roy models with selection.[2]

1.1 Literature Review

This paper connects to the large literature in structural labour economics on the estimation of human capital and learning models, including Heckman (1976), Cunha and Heckman (2008), Buchinsky, Fougére, Kramarz, and Tchernis (2010), Bagger, Fontaine, Postel-Vinay, and Robin (2014), and Lamadon, Meghir, and Robin (2024); see Rubinstein and Weiss (2006) for a review.[3] Within this literature, our work is the first to provide formal identification arguments for dynamic matching models of Bertrand competition, where firms are heterogeneous in their output technologies, human capital, and information production, and workers differ in both their initial and acquired human

[2]See, for instance, the Stata routine for extremal quantile regression models, as discussed below.

[3]See Papageorgiou (2014, 2018) for search models of labor market sorting based on worker comparative advantage with learning about workers' abilities and Hagedorn, Law, and Manovskii (2017) for the identification of search models of labor market sorting based on absolute advantage. As we discuss below, our identification results could allow for search frictions and wage bargaining. For most of our discussion, we focus on the novel aspect of our work, which consists on establishing the empirical content of general matching models with uncertainty and learning about worker ability and imperfect competition among firms in the labor market. For the importance of persistent uncertainty and learning, see also the review of Keane, Ching, and Erdem (2017).

capital—observable by both workers and firms—as well as in their ability—which is unknown at labor market entry and gradually learned over time by both parties. As mentioned in the introduction with key references, this class of models nests many existing frameworks used in both the labor and macroeconomic literature to study the determinants of occupational choice, worker turnover, firm-worker sorting, wage growth, and wage inequality.

This paper relates to the literature on the identification of the Roy model, including Chamberlain (1986), Heckman (1990), Heckman and Honoré (1990), Ahn and Powell (1993), Das, Newey, and Vella (2003), Newey (2009), and D'Haultfoeuille and Maurel (2013). See French and Taber (2011) for a review. Our identification approach borrows arguments from D'Haultfoeuille and Maurel (2013), which, among those papers, provides a strategy that does not rely on excluded restrictions or continuous covariates—features typically absent in the class of models we study.

This paper connects to the literature on dynamic discrete choice models with unobserved heterogeneity, including Magnac and Thesmar (2002), Kasahara and Shimotsu (2009), Hu and Shum (2012), An, Hu, and Shum (2013), Shiu and Hu (2013), Hu, Shum, Tan, and Xiao (2015), Berry and Compiani (2023), Higgins and Jochmans (2023), and Higgins and Jochmans (2024). These papers either rely on the assumption that unobserved heterogeneity is time-invariant or allow for time-varying and serially correlated unobserved heterogeneity, but under high-level restrictions on endogenous variables (e.g., monotonicity restrictions), restrictions on the distributional support of unobserved variables relative to observed variables, or the availability of instruments. None of these assumptions apply to our framework.

This paper relates to the extensive literature on estimating the earnings distribution by building on and extending the framework of AKM. This literature includes works such as Card et al. (2013), Card, Cardoso, Heining, and Kline (2018), Bonhomme, Lamadon, and Manresa (2019), and Song et al. (2019), as well as studies that highlight the importance of correcting AKM estimates for addressing bias due to low mobility, including Abowd, Kramarz, Lengermann, and Pérez-Duarte (2004), Andrews, Gill, Schank, and Upward (2008, 2012), Kline, Saggio, and Sølvsten (2020), and Bonhomme et al. (2023).

Lastly, this paper connects to the literature on the identification of panel data models with interactive fixed effects and learning, including Freyberger (2018) and Bunting, Diegert, and Maurel (2024). The wage equation in our class of models differs in that the unobservables enter in a potentially nonlinear, nonmonotone, and nonmultiplicative manner, making the use of interactive fixed effects not viable. Moreover, unlike those papers, we allow for dynamic selection on unobservables, namely, the firms' and workers' beliefs (endogenously evolving over time and serially correlated) and the productivity shocks (time-varying and serially uncorrelated).

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3 previews our identification approach. Section 4 presents the formal identification arguments. Section 5 discusses the Monte Carlo exercises and empirical applications. Appendix A examines extensions to our framework, and Appendix B provides further details on the identification arguments. The proofs are contained in Appendix C, and Appendices D and E offer additional details on the Monte Carlo exercises.

2 Setup

We consider a canonical and well-known class of dynamic matching models of Bertrand competition under uncertainty and learning about worker ability, where firms are heterogeneous in their technologies of output, human capital, and information production; workers differ in their initial and acquired human capital, observed by both workers and firms, as well as their ability, which is unknown at the time a worker enters the labor market and is gradually learned over time by workers and firms as a worker accumulates labor market experience, based on a worker's realized output. This class of models nests many existing frameworks used in both the labor and macroeconomic literature to study the determinants of occupational choice, worker turnover, firm-worker sorting, wage growth, and wage inequality. See Section 1 for key references.

Firms. Firms produce a homogeneous good sold in a perfectly competitive market at a price normalized to 1. Each firm $d \in \mathcal{D}$, where $2 \leq |\mathcal{D}| < \infty$, operates under a constant-returns-to-scale technology in workers' labor as the only input, as described below. Firms compete for workers by offering them wages each period for their employment during that period. The model and econometric results can be extended to a setting in which firms consist of multiple jobs, where firms' employment offers to workers include both a wage and a job. As we move on, we will highlight aspects of the multi-job case that merit special attention.

Workers. Upon entering the labor market, each worker n is endowed with some time-invariant characteristics denoted as $H_{n,1}$, with support $\mathcal{H}$, which are observed by worker n, firms, and the econometrician. These may include attributes such as gender, race, and education, that capture worker n's initial human capital. Worker n has also other skills that are unobserved by the econometrician and can be distinguished into two components: e_n with support $\mathcal{E}$, which denotes worker n's efficiency, observed by worker n and firms; and θ_n with support Θ, which denotes worker n's ability, initially unknown to worker n and firms but gradually and symmetrically learned by all based on worker n's realized output through a process described in detail later. Both e_n and θ_n are viewed as general traits that can influence worker n's performance when employed at any firm d.[4]

In the model, e_n may be discrete or continuous, and can be scalar or multidimensional, with support $\mathcal{E}$. In the econometric part, we assume that $\mathcal{E}$ is finite and defer the continuous-case extension to Appendix A. We also assume that θ_n takes values in $\Theta := \{\bar{\theta}, \underline{\theta}\}$, referred to as high ($\bar{\theta}$) and low ($\underline{\theta}$) ability. This binary assumption simplifies the exposition of the learning process and can be extended to a multidimensional continuous variable at the cost of more complex notation. We maintain the same assumption in the econometric section, with further details on continuous extensions provided in Appendix A.

Human Capital. Hereafter, we use the letter t to indicate a time period, which does not represent calendar time but marks the experience of a worker in the labor market. Hence, $t = 1$ denotes the first period of worker n in the labor market.[5] Worker n accumulates human capital over time depending on the initial characteristics $(H_{n,1}, e_n, \theta_n)$ and the employment history $D_n^t := (D_{n,1}, \dots, D_{n,t})$, where $D_{n,t}$ is a random variable representing the firm employing worker n in period t, with support $\mathcal{D}$. Formally, if employed by firm d in period t, worker n with efficiency $e_n = e$ has a human capital $H_{n,t}(d, e)$ at the *end* of the period, given by

$$H_{n,t}(d, e) = a_{n,t}(d, e) + \ell_{d,e}(H_{n,1}, \kappa_{n,t}) + \epsilon_{n,t}(d, e). \tag{1}$$

In (1), $H_{n,t}(d, e)$ is determined by two components: the labor input, $\ell_{d,e}(H_{n,1}, \kappa_{n,t}) + \epsilon_{n,t}(d, e)$, and the total factor productivity, $a_{n,t}(d, e)$. As for the labor input, $\kappa_{n,t} := \kappa(H_{n,1}, D_n^{t-1})$ is a function of worker n's initial human capital $H_{n,1}$ and employment history D_n^{t-1}, known by the econometrician, which captures, for example, worker n's experience in the market and tenure at each firm. $\ell_{d,e}(\cdot)$ is a (d, e)-specific function of $H_{n,1}$ and $\kappa_{n,t}$, unknown by the econometrician and on which we place *no*

[4]This generality is essential to generate realistic job mobility patterns. If e_n and θ_n were firm-specific and independent, workers would change jobs predominantly upon poor performance, unlike in typical data where highly performing workers are observed to switch jobs both within and across firms.

[5]In practice, workers may enter a dataset a few years after their initial entry into the labor market. In our framework, this only impacts the identification of the initial distribution of beliefs about ability (prior beliefs). Specifically, our methodology recovers the prior belief about worker n's ability θ_n at the time of worker n's first appearance in the data rather than at the time of worker n's initial entry into the labor market.

restrictions. $\epsilon_{n,t}(d,e)$ is a (d,e)-specific productivity shock or amenity, unobserved by the econometrician. We detail our assumptions on the distribution of $\epsilon_{n,t}(d,e)$ for equilibrium characterization below and for identification purposes in Section 4.

Worker n's human capital $H_{n,t}(d,e)$ is affected by θ_n through the term $a_{n,t}(d,e)$. Specifically, $a_{n,t}(d,e)$ is a (d,e)-specific random variable with distribution that depends on $H_{n,1}$ and θ_n and can vary across (d,e) as well. Importantly, $a_{n,t}(d,e)$ is a *random* function of θ_n; hence, it provides only a *noisy* measure of θ_n. This aspect is crucial for ensuring a nontrivial learning process, as described later. In most employer-employee match datasets, $a_{n,t}(d,e)$ is unobserved; thus, this will be the canonical case considered in the econometric analysis. $a_{n,t}(d,e)$ is assumed to take values in the set $\mathcal{A} := \{\bar{a}, \underline{a}\}$. As for θ_n, this assumption simplifies the model description without loss and can be generalized to a multidimensional continuous variable at the expense of more complex notation. We maintain the same assumption in the econometric section, with further details on continuous extensions provided in Appendix A.[6]

Output Technology. With labor supply normalized to one, (1) represents the (potential) output $y_{n,t}(d,e)$ produced by worker n with efficiency $e_n = e$ at the *end* of t when employed by firm d,

$$y_{n,t}(d,e) = a_{n,t}(d,e) + \ell_{d,e}(H_{n,1}, \kappa_{n,t}) + \epsilon_{n,t}(d,e). \tag{2}$$

Since the index d, which affects the function $\ell_{d,e}(\cdot)$ as well as the realization and distribution of the random components $a_{n,t}(d,e)$ and $\epsilon_{n,t}(d,e)$, firms here are differentiated by their output (and human capital) technology. We discuss next how they differ in their technology of information generation.

Information Technology (Learning Process). Firms and worker n make wage offer and acceptance decisions at the *beginning* of each period t, in order to maximize the expected present discounted value of profits and wages, respectively. Before making these decisions, firms and worker n know the functions $\{\ell_{d,e}(H_{n,1}, \kappa_{n,t})\}_{d\in\mathcal{D}}$ and observe the realization of the productivity shocks $\{\epsilon_{n,t}(d,e)\}_{d\in\mathcal{D}}$ at *each* potential job d at which worker n can be employed in t. However, the random components $\{a_{n,t}(d,e)\}_{d\in\mathcal{D}}$ are only realized at the *end* of the period after production takes place.

[6] The additive separability between $a_{n,t}(d,e)$ and $\ell_{d,e}(\cdot)$ in Equation (1) is not necessary. For the purpose of the equilibrium characterisation, what is needed is that $H_{n,t}(d,e)$ is bijective in $a_{n,t}(d,e)$. This ensures that, when observing the realisation of the output $y_{n,t}(d,e)$ defined in Equation (2), worker n and firms can back out the *unique* realisation of $a_{n,t}(d,e)$ that generated it, which is key for the learning process. The additive separability of $H_{n,t}(d,e)$ with respect to $\epsilon_{n,t}(d,e)$ is common in the literature on dynamic discrete choice models, and it is exploited both for the equilibrium characterization and for our identification proof.

As a result, firms and worker n make decisions based on their expectations about $\{a_{n,t}(d,e)\}_{d\in\mathcal{D}}$, which in turn depend on their beliefs about θ_n.

Firms and worker n learn about θ_n through a Bayesian updating process based on the common observations of $a_{n,t}(d,e)$ at the end of each period t at the employing firm d. In this precise sense, $a_{n,t}(d,e)$ represents the public noisy signal about worker n's ability θ_n that firms and worker n extract from worker n's realized output $y_{n,t}(d,e)$.[7] In the class of models we consider, learning is symmetric across firms and workers, so that *all* firms and worker n share a common belief about θ_n in each period t. Specifically, at the beginning of period $t = 1$, firms and worker n with efficiency $e_n = e$ and initial human capital $H_{n,1} = h$ have a common prior belief about θ_n taking value $\bar{\theta}$,

$$p_1(h,e) \coloneqq \Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e).$$

This prior need not coincide with the true conditional distribution of θ_n and may incorporate any learning about θ_n that has taken place before entry into the labor market, for instance, during schooling. At the end of period $t \geq 1$, firms and worker n observe the realization of $y_{n,t}(d,e)$ and thus extract the signal $a_{n,t}(d,e)$ about the worker's ability. At the beginning of period $t+1$, firms and worker n update their belief about θ_n based on $a_{n,t}(d,e)$ using Bayes' rule. Given the previous period belief, which we denote by $P_{n,t}$, and assuming that the signals $\{a_{n,t}(d,e)\}_{t,d\in\mathcal{D}}$ are conditionally independent over time, the updated belief can take two values, namely,

$$P_{n,t+1} = p_{\bar{a}}(h,d,e;P_{n,t}) = \frac{\alpha_{h,d,e}P_{n,t}}{\alpha_{h,d,e}P_{n,t} + \beta_{h,d,e}(1-P_{n,t})}, \tag{3}$$

after observing $a_{n,t}(d,e) = \bar{a}$ and

$$P_{n,t+1} = p_{\underline{a}}(h,d,e;P_{n,t}) = \frac{(1-\alpha_{h,d,e})P_{n,t}}{(1-\alpha_{h,d,e})P_{n,t} + (1-\beta_{h,d,e})(1-P_{n,t})}, \tag{4}$$

after observing $a_{n,t}(d,e) = \underline{a}$, where $\alpha_{h,d,e} \coloneqq \Pr(a_{n,t}(d,e) = \bar{a} \mid H_{n,1} = h, D_{n,t} = d, e_n = e, \theta_n = \bar{\theta})$ and $\beta_{h,d,e} \coloneqq \Pr(a_{n,t}(d,e) = \bar{a} \mid H_{n,1} = h, D_{n,t} = d, e_n = e, \theta_n = \underline{\theta})$.

Importantly, given that the parameters $\alpha_{h,d,e}$ and $\beta_{h,d,e}$ are allowed to vary across firms d, jobs can differ in their informativeness about θ_n. In particular, firms are differentiated not only in their output (and human capital) technology but also by their information technology. Consequently, the belief updating process—and thus the speed of learning about θ_n—depends on the entire history of jobs

[7]Recall that $a_{n,t}(d,e)$ is a random function of θ_n. If $a_{n,t}(d,e)$ was a deterministic function of θ_n, then the value of θ_n could be learned in one period after observing $a_{n,t}(d,e)$, and thus learning would become trivial.

experienced by worker n. The parameters $\alpha_{h,d,e}$ and $\beta_{h,d,e}$, which capture the ex-ante heterogeneous information technology of each firm, will be among the key primitives we aim to identify.

Based on this belief process, at the the beginning of period $t+1$, firms and worker n calculate worker n's *expected* output when employed by firm d as

$$\mathbb{E}\Big(y_{n,t+1}(d,e)\Big|H_{n,1}=h,\kappa_{n,t+1}=\kappa,P_{n,t+1}=p,e_n=e,\epsilon_{n,t+1}\Big)=y(d,s)+\epsilon_{n,t+1}(d,e),$$

where the variables in firms' and worker n's information set are collected into the vectors $s_{n,t+1} := (H_{n,1}, \kappa_{n,t+1}, P_{n,t+1}, e_n)$, with generic realization $s := (h, \kappa, p, e)$, and $\epsilon_{n,t+1} := (\epsilon_{n,t+1}(d,e) : d \in \mathcal{D}, e \in \mathcal{E})$. $y(d,s) := \mathbb{E}(a_{n,t}(d,e) \mid s_{n,t+1} = s) + \ell_{d,e}(h,\kappa)$ is one of the primitives we aim to identify, as it reflects the ex-ante heterogeneous output (and human capital) technology of each firm.

Equilibrium. Given the absence of complementarities in production among workers, to characterize the model's equilibrium, we can examine the competition of all firms for one worker at a time without any loss of generality. This setup is standard in equilibrium models of uncertainty, learning, and experimentation like ours. See, for instance, Bergemann and Välimaki (1996).

We adopt a refinement of the notion of Markov perfect equilibrium, which we term Robust Markov perfect equilibrium (RMPE). An RMPE consists of wage strategies for firms and an acceptance strategy for worker n, alongside a belief function such that: i) the worker maximizes the expected present discounted value of wages; ii) each firm maximizes the expected present discounted value of its profits; iii) beliefs are consistently updated according to Bayes' rule; and iv) non-employing firms are indifferent between not employing and employing the worker. Conditions i) through iii) define a standard MPE, under which multiple MPEs may exist. Condition iv) selects one of such equilibria and hence acts as a refinement condition. We provide further details on condition (iv) below. Under conditions i)-iv), an RMPE exists, is unique, and is efficient (Bergemann and Välimaki, 1996). In the extension of the model to multi-job firms, the equilibrium is typically inefficient (Pastorino, 2024). More formally, the state that firms face at the time they make their wage offers to worker n consists of $(s_{n,t}, \epsilon_{n,t})$, and the state that worker n faces at the time they make their acceptance decisions includes $(s_{n,t}, \epsilon_{n,t})$ and the collection of all firms' wage offers that worker n receives. We denote by $w_{d,n,t}$ the wage offer strategy of each generic firm d and by $\{w_{d,n,t}\}_{d\in\mathcal{D}}$ the collection of all wage offer strategies. We denote by $l_{d,n,t}$ the acceptance strategy of worker n for firm d's offer—an indicator function, taking value one if d is the employing firm and zero otherwise—and by $\{l_{d,n,t}\}_{d\in\mathcal{D}}$ the collection of all acceptance strategies.

Given the firms' wage strategies, worker n's acceptance strategy when of type $e_n = e$ satisfies

$$\tilde{W}(s_{n,t}(e), \epsilon_{n,t}(e), \{w_{d,n,t}(e)\}_{d\in\mathcal{D}}) = \max_{\{l_{d,n,t}(e)\}_{d\in\mathcal{D}}} \sum_{d\in\mathcal{D}} l_{d,n,t}(e) \times \Big[w_{d,n,t}(e)$$
$$+ \delta[1-\eta(\kappa_{n,t}, d)] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(\tilde{W}(s_{n,t+1}(e), \epsilon_{n,t+1}(e), \{w_{d,n,t+1}(e)\}_{d\in\mathcal{D}}) \mid s_{n,t}(e), d\Big) dG_e\Big]. \tag{5}$$

In (5) $s_{n,t}(e)$ and $\epsilon_{n,t}(e)$ are the vectors $s_{n,t}$ and $\epsilon_{n,t}$ with $e_n = e$, $w_{d,n,t}(e) := (w_d(s_{n,t}(e), \epsilon_{n,t}(e)),$ $l_{d,n,t}(e) := l_d(s_{n,t}(e), \epsilon_{n,t}(e), \{w_{d,n,t}(e)\}_{d\in\mathcal{D}})$, G_e is the cumulative distribution function of $\epsilon_{n,t}(e)$, δ is the discount factor, and $\eta(\kappa_{n,t}, d)$ is the probability that worker n leaves the labor market at the end of period t, given the accumulated human capital investments $\kappa_{n,t}$ and the last employing firm d.[8] Note that, in (5), we assume that for each $e \in \mathcal{E}$, $\epsilon_{n,t}(e)$ is independent of $s_{n,t}(e)$, and that the vectors $\{\epsilon_{n,t}(e)\}_t$ are i.i.d. across periods t, as is standard in dynamic models. We maintain this assumption throughout. In our framework, time persistence in the state is generated through $\kappa_{n,t}$ and $P_{n,t}$. Given worker n's acceptance strategy and competitors' wage strategies, firm d's wage strategy satisfies

$$\Pi_d(s_{n,t}(e), \epsilon_{n,t}(e)) = \max_{w_{d,n,t}(e)} \Big(l_{d,n,t}(e) \times \Big[y(d, s_{n,t}(e)) + \epsilon_{n,t}(d,e) - w_{d,n,t}(e)$$
$$+ \delta[1-\eta(\kappa_{n,t}, d)] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(\Pi_d(s_{n,t+1}(e), \epsilon_{n,t+1}(e)) \mid s_{n,t}(e), d\Big) dG_e\Big]$$
$$+ \sum_{d'\in\mathcal{D}\setminus\{d\}} l_{d',n,t}(e) \times \Big\{\delta[1-\eta(\kappa_{n,t}, d')] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(\Pi_d(s_{n,t+1}(e), \epsilon_{n,t+1}(e)) \mid s_{n,t}(e), d'\Big) dG_e\Big\}\Big). \tag{6}$$

Without condition (iv) for equilibrium, this class of models gives rise to a multiplicity of MPE. These equilibria are qualitatively similar in that they are characterized by the *same* allocations regarding which firm employs worker n in each state, resulting in the same on-path outcomes. However, these equilibria differ in the wages offered by non-employing firms; indeed, non-employing firms can offer any wage up to the point where they are indifferent between not employing and employing the worker. Condition (iv) resolves this *trivial* multiplicity by requiring that non-employing firms offer wages that make them indifferent between not employing and employing the worker. In particular, condition (iv) selects an equilibrium in a manner that is standard in the literature on trembling-hand perfect equilibrium (Selten, 1975)—we characterize the equilibrium wage equation in the next section. Specifically, if, say, firm d' employs worker n at state $(s_{n,t}(e), \epsilon_{n,t}(e))$, condition (iv) requires

[8] Although we have ignored the possibility that a worker is unemployed, in the extension of the model to multi-job firms, it would be straightforward to allow for an additional job that corresponds to the alternative of home production (non employment). We have refrained from doing so purely for simplicity, as our focus is on the dynamics of matching and wages generated by human capital and learning as mechanisms for persistent wage inequality among workers.

for any other firm d that

$$\delta[1-\eta(\kappa_{n,t},d')]\int_{\epsilon_{n,t+1}(e)}\mathbb{E}\Pi_d(\cdot|s_{n,t}(e),d')dG_e$$
$$=\max_{w_{d,n,t}(e)}\Big\{y(d,s_{n,t}(e))+\epsilon_{n,t}(d,e)-w_{d,n,t}(e)+\delta[1-\eta(\kappa_{n,t},d)]\int_{\epsilon_{n,t+1}(e)}\mathbb{E}\Pi_d(\cdot|s_{n,t}(e),d)dG_e\Big\}. \quad (7)$$

Namely, firm d must offer worker n a wage that makes firm d indifferent between not employing the worker—in which case its payoff is the left side of (7)—and employing the worker—in which case its payoff is the right side of (7).Importantly, under condition (iv), an employed worker's wage is uniquely determined—specifically, it equals the wage offered by the second-best firm plus a compensating differential, as shown in Proposition 1 below—thereby helping us sidestep the econometric challenges associated with so-called model incompleteness (Tamer, 2003).

2.1 Equilibrium Wage

We discuss here the equilibrium wage equation, starting with a comparison of our framework with standard models of imperfect competition in the labor (or output) market.

Intuition. An intuition for how wages are determined in our model can be gained by considering a static version of it with just two firms. Recall that in a static model of Bertrand competition among heterogeneous firms, the high-productivity/low-cost firm sells to a consumer at a price equal to the cost of the low-productivity/high-cost firm, making the consumer indifferent between the two sellers. Analogously, in the static version of our model, worker n's wage in period t equals the worker's expected output had the worker being hired by the competitor of the employing firm. Thus, the worker is indifferent between employment at the employing firm and employment at its competitor. In the perfectly competitive case of identical firms, the worker is paid their expected output, as the expected output at the non-employing firm is the same as at the employing firm.

In the dynamic version of our model with two firms, the same intuition holds—the worker is indifferent between the employing firm and its competitor in terms of the expected present discounted value of wages. However, additional factors must be taken into account in this dynamic setting. Specifically, recall that firms differ in both their output (and human capital) and information technologies. Moreover, the human capital and information accumulated during employment at one firm lead to future returns for a worker. Therefore, a firm where a worker can accumulate substantial human capital or information can afford to pay a lower wage while still employing the worker. Conversely, a firm that offers limited opportunities for accumulating human capital or information must offer a higher wage to attract the worker. With more than two firms, a similar argument applies; however, the difference lies in the two firms competing for a worker being those that offer the two highest expected present discounted values of wages.

Formally, we demonstrate that worker n's wage in period t equals the expected output the worker would produce if hired by the firm ranked as "second-best" in terms of expected present discounted values of wages—akin to a *second-price auction*—plus a *compensating differential* term, which is either a premium or a discount compensating the worker for the missed future returns in terms of human capital and information acquisition that would have been gained if accepting a job at the second-best firm.

Wage Equation. Consider the equilibrium ranking of firms based on the expected present discounted value of the wage they offer to worker n in period t. Focus on the two firms that provide the highest expected present discounted values of wage in this ranking. Of these, designate the "first-best" firm as the employing firm and the "second-best" as the non-employing firm. Hereafter, we typically denote them as d and d', respectively. Moreover, let $V_{d'}(s_{n,t}(e), \epsilon_{n,t}(e))$ represent the expected present discounted value of the match surplus generated by worker n and firm d' at state $(s_{n,t}(e), \epsilon_{n,t}(e))$, defined as the sum of the worker's wage value and firm d''s profit value.

Proposition 1 (Equilibrium Wage)**.** *The equilibrium wage of worker n with efficiency $e_n = e$ in period t, when d is the employing firm and d' is the second-best firm, is*

$$w_{d,n,t} \coloneqq w_{n,t}(d, d', e) = y(d', s_{n,t}(e)) + \Psi(d, d', s_{n,t}(e)) + \epsilon_{n,t}(d', e), \tag{8}$$

with

$$\begin{aligned}\Psi(d, d', s_{n,t}(e)) \coloneqq &\delta[1 - \eta(\kappa_{n,t}, d')] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}V_{d'}(s_{n,t+1}(e), \epsilon_{n,t+1}(e) | s_{n,t}(e), d') dG_e \\ &- \delta[1 - \eta(\kappa_{n,t}, d)] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}V_{d'}(s_{n,t+1}(e), \epsilon_{n,t+1}(e) | s_{n,t}(e), d) dG_e.\end{aligned}$$

According to Proposition 1, a worker's wage is the sum of three terms: $y(d', s_{n,t}(e)) + \epsilon_{n,t}(d', e)$, which is the expected per-period output at d' (after the vector of productivity shocks $\epsilon_{n,t}$ is realised), and $\Psi(d, d', s_{n,t}(e))$, which is a compensating differential. In particular, $\Psi(d, d', s_{n,t}(e))$ is the difference between two value functions: the first being the (counterfactual) future expected discounted match surplus value generated by worker n and firm d' had d' being chosen by n in period t, and the second being the future expected discounted match surplus value generated by worker n and firm d' when worker n chooses firm d in period t. Lastly, note that in the expression $w_{n,t}(d, d', e)$, the subscript (n, t) encapsulates not only the worker and time indices but also any dependence of the wage on the state $(s_{n,t}(e), \epsilon_{n,t}(e))$ that is worker-specific.

2.2 Econometric Model

The model just described is a dynamic equilibrium generalised Roy model. Equation (8) is the potential equilibrium wage equation, and hence the observed wage of worker n at time t is

$$\begin{aligned} w_{n,t} &= \sum_{(d,d')\in\mathcal{D}^2} \sum_{e\in\mathcal{E}} \mathbb{1}\{D_{n,t} = d, D'_{n,t} = d, e_n = e\} w_{n,t}(d, d', e) \\ &= \sum_{(d,d')\in\mathcal{D}^2} \sum_{e\in\mathcal{E}} \mathbb{1}\{D_{n,t} = d, D'_{n,t} = d, e_n = e\} \big[y(d', s_{n,t}(e)) + \Psi(d, d', s_{n,t}(e)) + \epsilon_{n,t}(d', e) \big], \end{aligned} \tag{9}$$

where $D_{n,t}$ is a random variable representing the employing (first-best) firm for worker n in period t, with a generic realization $d \in \mathcal{D}$, and $D'_{n,t}$ denotes the second-best firm, with a generic realization $d' \in \mathcal{D}$. Assumption 1 describes the observation scheme maintained throughout.

Assumption 1.(Data) The joint distribution of $(w_{n,t}, H_{n,1}, D_{n,t})$ is known for each period $t = 1, \ldots, T$, with $T < \infty$. $\diamond$

Assumption 1 requires the econometrician to have access to a panel of data on wages, initial attributes, and employment choices. We keep T finite and presume elsewhere that the number of workers goes to infinity. We make *minimal* data requirements to accommodate the limited information typically available in standard employer-employee match datasets. In particular, we do not rely on the availability of variables that can facilitate the identification of the learning process, such as proxies for beliefs or direct information on performance signals. In the empirical application presented in Section 5, we show how the availability of additional data can ease some identification steps. To simplify the notation, we assume that the panel is balanced; however, all econometric arguments remain valid even with an unbalanced panel.

The variables $D_{n,t}$ and $D'_{n,t}$ are allowed to depend on all the variables entering the equilibrium wage equation, some of which are not observed by the econometrician, namely e_n, $P_{n,t}$, and $\epsilon_{n,t}$. This dependency arises from the optimising behaviour of workers and firms, leading to dynamic selection on *unobservables*, which include time-varying $(P_{n,t}, \epsilon_{n,t})$ and serially correlated $(P_{n,t})$ components.

With the data in Assumption 1, we show in the following sections how to identify several primitives, which enable us to study the fundamental question of measuring the impact of sorting on earnings inequality. In particular, we identify the "deterministic" wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$—defined as the sum of the expected output (net of productivity shocks) and the compensating differential—and the distribution of the vector of productivity shocks $\epsilon_{n,t}$. Moreover, we identify the output (and human capital) technology $y(\cdot)$ and, in turn, disentangle the compensating differential $\Psi(\cdot)$ from $\varphi(\cdot)$. We also show how to identify the information technology (learning process)—fully described by the prior $p_1(\cdot)$ and the distribution of signal $a_{n,t}$ conditional on $(H_{n,1}, D_{n,t}, e_n, \theta_n)$. (To simplify the notation and without loss of generality, from now on, we write $a_{n,t}(D_{n,t}, e_n)$ simply as $a_{n,t}$.) Lastly, we prove identification of other important primitives of dynamic models, including the law of motion for the state variables $s_{n,t}$ and the distribution of job choices $D_{n,t}$ conditional on $s_{n,t}$ (conditional choice probabilities, or CCPs). Throughout, we assume the discount factor δ is known, as is standard in dynamic models.

3 Overview of Identification

A key primitive of interest for us is the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$, which we use to then separately identify the output (and human capital) technology $y(\cdot)$ and the compensating differential $\Psi(\cdot)$, and, in turn, address our question about the impact of sorting on earnings inequality. However, as previewed in Section 2.2, the identification of $\varphi(\cdot)$ is complicated by selection on the unobservables e_n, $P_{n,t}$, and $\epsilon_{n,t}$. To break down the problem and develop intuition, suppose for a moment that $D'_{n,t}$ and the state variables $s_{n,t}$ are observed in the data. Even in this simplified scenario, it is straightforward to see that we face the "standard" Roy challenge: the identification of $\varphi(\cdot)$ is confounded by selection on the shock $\epsilon_{n,t}$. Indeed,

$$\mathbb{E}(w_{n,t} \mid D_{n,t} = d, D'_{n,t} = d', s_{n,t}(e)) = \varphi(d, d', s_{n,t}(e)) + \mathbb{E}\big(\epsilon_n(d', e) \mid D_{n,t} = d, D'_{n,t} = d', s_{n,t}(e)\big), \tag{10}$$

where $\mathbb{E}(\epsilon_n(d', e) \mid D_{n,t} = d, D'_{n,t} = d', s_{n,t}(e))$ may be different from the unconditional mean $\mathbb{E}(\epsilon_n(d', e))$ as $D_{n,t}$ and $D'_{n,t}$ depend on $s_{n,t}(e)$ and $\epsilon_n(d', e)$. Consequently, it is impossible to identify $\varphi(d, d', s_{n,t}(e))$ from $\mathbb{E}(w_{n,t} \mid D_{n,t} = d, D'_{n,t} = d', s_{n,t}(e))$ alone without further assumptions.

In addition to this standard Roy challenge, we face the further complication that $D'_{n,t}$ is not observed in the data, and $s'_{n,t}$ is only partially observed—specifically, the econometrician does not know e_n and $P_{n,t}$. Therefore, the left-hand-side of Equation (10) is also unknown.

In the remainder of Section 3, we provide an informal overview of how we address these challenges. Section 4 presents the same arguments in a more technically formal manner.

3.1 Second-Best Firm

As explained in the preceding paragraph, an identification challenge is that the second-best firm $D'_{n,t}$ is not observed in the data. For now, we introduce an assumption that substantially mitigates the difficulties arising from the non-observability of $D'_{n,t}$. In Appendix A, we discuss how this assumption can be relaxed.

Assumption 2.(Choice set) In each period t, the set of firms making offers to worker n depends deterministically on the vector of state variables $s_{n,t}$ and is independent of the vector of shocks $\epsilon_{n,t}$. This choice set has a cardinality of two and is denoted by $\mathcal{D}(s_{n,t}) \subseteq \mathcal{D}$, with $|\mathcal{D}(s_{n,t})| = 2$. $\diamond$

Without Assumption 2, worker n would receive wage offers from *all* firms in $\mathcal{D}$ in each period t. The equilibrium ranking of these firms, in terms of the expected present discounted value of the wages offered, could freely depend on both $s_{n,t}$ and $\epsilon_{n,t}$. Under Assumption 2, worker n receives wage offers from only *two* firms in each period t, whose identity depends on $s_{n,t}$—hence, can vary across $(H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$—and is independent of $\epsilon_{n,t}$. Nevertheless, the equilibrium ranking of these two firms, in terms of the expected present discounted value of the wages offered, *still* depends on both $s_{n,t}$ and $\epsilon_{n,t}$. Consequently, selection on $\epsilon_{n,t}$ remains an issue.

Assumption 2 makes our model more empirically plausible by aligning it with the practical reality that workers typically receive offers from a limited number of firms. Search models also face a similar challenge in that equilibrium wages depend on the productivity of a competitor whose identity is generally unknown. The search literature addresses this issue by imposing an assumption similar to Assumption 2, in which the two firms making offers in each period are always the "incumbent" and the "competitor"—note that in search models, unlike in ours, matching frictions are exogenous, which substantially simplifies the dynamic selection of interest. In this sense, Assumption 2 aligns with established practices already found in the literature.

A key econometric implication of Assumption 2 is that the distribution of $D'_{n,t}$ conditional on $s_{n,t}$ and $D_{n,t}$ is *degenerate* at one point, meaning $D'_{n,t}$ can have *only* one realized value, which greatly facilitates dealing with this latent term. Most of our results do not require the econometrician to know the support of $D'_{n,t}$ conditional on $D_{n,t}$ and $s_{n,t}$—that is, the *identity* of the second-best firm given a first-best firm and state. Specifically, without knowing such identify, we are able to identify the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$, the distribution of the vector of productivity shocks $\epsilon_{n,t}$, the output (and human capital) technology $y(\cdot)$, the information technology, the CCPs, and the law of motion of the state. When the identity of the second-best firm is known—for example, through the use of workers' observed transition patterns, as discussed in Appendix A—we can also identify the compensating differential $\Psi(\cdot)$. Under Assumption 2, Equation (9) simplifies to

$$w_{n,t} = \sum_{e \in \mathcal{E}} \sum_{(d,d') \in \mathcal{D}(s_{n,t}(e))} \mathbb{1}\{D_{n,t} = d, D'_{n,t} = d, e_n = e\}\big[y(d', s_{n,t}(e)) + \Psi(d, d', s_{n,t}(e)) + \epsilon_{n,t}(d', e)\big]. \tag{11}$$

In the next section, we first review existing methods to identifying the Roy model, which serves as the foundation for the models of job mobility we nest. We then turn to discuss how we extend these strategies to establish the empirical content of our framework.

3.2 A Review of the Identification of the Roy Model

As discussed by French and Taber (2011), the identification of the Roy model can be achieved by employing nonparametric or semiparametric strategies. For a simple understanding of these strategies and the challenges associated with applying them to our setting, consider a *simplified static* version of the wage Equation (11), namely,

$$w_n = \sum_{d \in \{0,1\}} \mathbb{1}\{D_n = d\} w_n(d) = \sum_{d \in \{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n) + \epsilon_n(d)\big], \tag{12}$$

where we have removed the dependence on $D'_{n,t}$—hence, all the wage components previously indexed by the d' are now indexed by d only—and e_n, as well as the subscript t. $\mathcal{D} := \{0, 1\}$ represents two job alternatives and note that the compensating differential $\Psi(\cdot)$ does not arise in the static version of our model. For the purpose of this section, the vector of state variables $s_{n,t}(e)$ is replaced by the covariate(s) X_n, which are assumed to be *observed* by the econometrician. The vector of shocks $\epsilon_n := (\epsilon_n(0), \epsilon_n(1))$ is independent of X_n, as consistent with our model. As is well known, identifying the deterministic wage components $y(1, X_n)$ and $y(0, X_n)$ in Equation (12) is challenging due to selection on ϵ_n. To see why, observe that

$$\mathbb{E}(w_n \mid D_n = d, X_n) = \mathbb{E}\big(y(d, X_n) + \epsilon_n(d) \mid D_n = d, X_n\big) = y(d, X_n) + \mathbb{E}(\epsilon_n(d) \mid D_n = d, X_n),$$

where the conditional expectation $\lambda(d, X_n) := \mathbb{E}(\epsilon_n(d) \mid D_n = d, X_n)$ may differ from its unconditional counterpart $\mathbb{E}(\epsilon_n(d))$, because D_n depends on both X_n and ϵ_n. Consequently, it is impossible to recover $y(d, X_n)$ from $\mathbb{E}(w_n \mid D_n = d, X_n)$ alone without imposing further assumptions.

With Exclusion Restrictions. One way to address selection is to rewrite Equation (12) as

$$w_n = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n) + \lambda(d, X_n) + u_n(d)\big], \tag{13}$$

where $u_n(d) := \epsilon_n(d) - \lambda(d, X_n)$ and hence, by construction, $\mathbb{E}(u_n(d) \mid D_n = d, X_n) = \mathbb{E}(\epsilon_n(d))$, which is tipically normalized to zero. Then, if X_n can be split into two components, $X_n^\dagger$ and X_n^*, such that $y(d, X_n)$ depends only on $X_n^\dagger$ and $\lambda(d, X_n)$ depends only on X_n^*—often referred to as *exclusion restrictions*—it becomes possible to identify $y(d, X_n^\dagger)$, provided that certain additional assumptions on $y(\cdot)$ and $\lambda(\cdot)$ hold (Ahn and Powell, 1993; Newey, 2009; Das et al., 2003)

Another way to address selection consists of relying on worker-job-specific covariates with a sufficiently rich support that influence the wage in one job only (Heckman and Honoré, 1990). In particular, suppose we can express Equation (12) as

$$w_n = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n(d)) + \epsilon_n(d)\big], \tag{14}$$

where $X_n(d)$ is now a worker-job-specific covariate (scalar, for simplicity) that *exclusively* affects the potential wage in job d, representing another type of exclusion restriction. Consider two realizations x_1 and $\tilde{x}_1$ of $X_n(1)$ and suppose that we can correspondingly find two values x_0 and $\tilde{x}_0$ of $X_n(0)$ such that $\Pr(D_n = 1 \mid X_n = (x_0, x_1)) = \Pr(D_n = 1 \mid X_n = (\tilde{x}_0, \tilde{x}_1))$. In a setting where worker n chooses the job with the highest wage, $\mathbb{E}(\epsilon_n(1) \mid X_n = (x_0, x_1), D_n = 1) = \mathbb{E}(\epsilon_n(1) \mid X_n = (\tilde{x}_0, \tilde{x}_1), D_n = 1)$. Thus,

$$\mathbb{E}(w_n \mid X_n = (x_0, x_1), D_n = 1) - \mathbb{E}(w_n \mid X_n = (\tilde{x}_0, \tilde{x}_1), D_n = 1) = y(1, x_1) - y(1, \tilde{x}_1),$$

and so the difference $y(1, x_1) - y(1, \tilde{x}_1)$ is identified. As long as $X_n(0)$ sufficiently varies, the whole function $y(1, X_n(1))$ can be identified up to location. We can also proceed further and completely identify $y(1, X_n(1))$ as follows. Suppose $y(0, X_n(0))$ is linear and increasing in $X_n(0)$, and $X_n(0)$ has unbounded support. Then, for any realization x_1 of $X_n(1)$,

$$\lim_{x_0\to-\infty} \Pr(D_n = 1 \mid X_n = (x_0, x_1)) = 1, \tag{15}$$

and by the law of total probability,

$$\lim_{x_0 \to -\infty} \mathbb{E}(\epsilon_n(1) \mid D_n = 1, X_n = (x_0, x_1)) = \lim_{x_0 \to -\infty} \mathbb{E}(\epsilon_n(1) \mid X_n = (x_0, x_1)) = \mathbb{E}(\epsilon_n(1)).$$

Therefore, under the normalisation $\mathbb{E}(\epsilon_n(1)) = 0$,

$$\lim_{x_0 \to -\infty} \mathbb{E}(w_n \mid D_n = 1, X_n = (x_0, x_1)) = \lim_{x_0 \to -\infty} \mathbb{E}(w_n(1) \mid X_n = (x_0, x_1)) = y(1, x_1),$$

and $y(1, x_1)$ is identified from knowledge of $\lim_{x_0 \to -\infty} \mathbb{E}(w_n \mid D_n = 1, X_n = (x_0, x_1))$—hence, the phrase *identification at infinity* (Chamberlain, 1986; Heckman, 1990). In summary, Condition (15) eliminates the impact of selection from the first moment of the wage distribution of a group of individuals with extreme values of $X_n(0)$. For this group, the expected potential wage in job 1 conditional on choosing job 1, which is *observed* from the data, is equal to the unconditional expected potential wage in job 1, which is generally *unobserved* from the data.

Challenges in Our Setting. The identification strategies discussed above, which rely on exclusion restrictions, cannot be adapted to the class of models we consider, as these models do not admit such excluded restrictions. Specifically, even in the ideal scenario where all state variables are observed and can thus be treated as standard covariates, in our class of models these variables influence *both* wages and the selection-correction term $\lambda(\cdot)$. Indeed, any variable that affects wages also shapes the probability that workers choose a particular job, thereby influencing $\lambda(\cdot)$. Conversely, the probability that a worker opts for a given job determines the wages firms are willing to offer. As a result, state variables cannot be partitioned into distinct components that separately affect wages and $\lambda(\cdot)$.

Furthermore, the class of models we study lacks worker-job-specific state variables affecting the wage in one job only, which are essential for implementing the at-infinity identification strategy of Chamberlain (1986) and Heckman (1990). One might wonder about three potential candidates for such worker-job-specific variables: beliefs about a worker's ability, worker's tenure at the job, and other worker-job-specific wage components, such as a worker's distance from a job location, which must be observed in the data or identifiable. However, none of these applies to our setting. Indeed, as highlighted in Section 2, we allow ability to be *general* across jobs, rather than restricting it to be specific to a particular job. Thus, the belief about a worker's ability is represented by a *single* probability distribution, $P_{n,t}$, over the worker's possible levels of ability affecting wages at *all* jobs—rather than a collection of job-specific probability distributions over the worker's possible levels of ability, influencing each corresponding wage—and is shaped by the worker's *entire* job history. Similarly, the human capital accumulation process, which affects a worker's output, may depend on the experience gained in *all* jobs. As a result, variables such as job tenure impact wages in *all* jobs. Lastly, in the current model version, a worker's value of non-employment (non-market time) does not impact equilibrium wages. Consequently, variables such as distance from the job location are not included in the equilibrium wage equation. These could be incorporated through wage bargaining. However, they are typically difficult to observe in standard employer-employee match datasets; for instance, they are absent in the LEHD dataset.[9][10]

Without Exclusion Restrictions. D'Haultfoeuille and Maurel (2013) (hereafter DM) proposed an alternative method to identify "at infinity" the deterministic wage component in the Roy model that does not require worker-job-specific regressors affecting the wage in one job only. Instead, they impose sufficient conditions on the distribution of unobserved heterogeneity that remove or, more generally, attenuate the impact of selection on the right *extreme tail* of the distribution of wages in a job. This approach also accommodates heteroskedasticity, allowing for the identification of the conditional variance of potential wages. For an intuitive understanding of how this approach works and how it can be extended to our framework, consider the simplified wage Equation (12) enriched with a conditional heteroskedasticity term, $\sigma(d, X_n) > 0$:[11]

$$w_n = \sum_{d \in \{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n) + \sigma(d, X_n)\epsilon_n(d)\big]. \tag{16}$$

We focus, as above, on the identification of the deterministic wage component in job 1, $y(1, X_n)$. We assume that $\epsilon_n(1)$ has unbounded support—however, the arguments can also be applied to $\epsilon_n(1)$ with bounded support—and that $\epsilon_n(1)$ and $\epsilon_n(0)$ are "moderately" dependent—a restriction formalized by DM in their Corollary 4.1, which we extend to our setting in Section 4.3. Then, it can be shown that

$$\lim_{w \to +\infty} \Pr(D_n = 1 \mid X_n = x, w_n(1) = w) = q_1 > 0 \quad \text{for every } x, \tag{17}$$

[9]In Appendix A, we discuss how the identification argument developed in this section also applies to models with search and matching frictions in which wages are bargained and a worker's value of non-employment (non-market time) affects paid wages, unlike in our framework.

[10]We emphasize that job-specific covariates fixed at the worker level (e.g., firm size or revenues) are often available in datasets, yet typically do not provide enough variation for identification in the Roy model.

[11]In the wage Equation (12), we do not have a term $\sigma(d, X_n)$ multiplying $\epsilon_{n,t}(d)$. Nevertheless, we illustrate the approach of DM in its full generality, including $\sigma(d, X_n)$. This is because the potential for extending DM to address selection issues in dynamic equilibrium generalized Roy models goes beyond the class of models we consider, encompassing search models that are inherently characterized by heteroskedasticity and are particularly focused on identifying it. We provide a more detailed explanation of this point in Appendix A.

with q_1 *invariant* across realizations x of X_n and unknown by the econometrician. Differently from (15), in (17), we condition on an extreme value of the potential wage—hence, an extreme value of the shock $\epsilon_n(1)$, as X_n is fixed at x—rather than an extreme value of the covariate X_n. Moreover, the conditional probability of choosing job 1 in (17) is not required to be equal to one but rather to be just strictly positive for reasons that will become clear below. Now, for a given realization x of X_n, observe that Condition (17) implies

$$\lim_{w \to +\infty} \frac{\Pr(D_n = 1, w_n(1) \geq w \mid X_n = x)}{q_1 \Pr(w_n(1) \geq w \mid X_n = x)} = 1. \tag{18}$$

By (18), $\Pr(D_n = 1, w_n(1) \geq w \mid X_n = x)$—which is observed in the data—and $\Pr(w_n(1) \geq w \mid X_n = x)$—which is not observed in the data due to selection—are *asymptotically equivalent* (up to q_1) as w gets large. Therefore, (17) essentially reduces—when $0 < q_1 < 1$—or altogether removes—when $q_1 = 1$—the impact of selection on the right extreme tail of the wage distribution in job 1. (Note that both the numerator and denominator of (18) converge to zero as $w \to +\infty$. (18) states that they converge to zero at the *same* rate.) Hereafter, we use the standard symbol $\sim$ to denote asymptotic equivalence, so that (18) is simply $\Pr(D_n = 1, w_n(1) \geq w \mid X_n = x) \sim q_1 \Pr(w_n(1) \geq w \mid X_n = x)$. Letting S be the survival function of $\epsilon_n(1)$, (18) can be written as

$$\Pr\left(D_n = 1, w_n(1) \geq w \mid X_n = x\right) \sim q_1 S\left(\frac{w - y(1,x)}{\sigma(1,x)}\right). \tag{19}$$

Assuming that for some other realization $\bar{x} \neq x$ of X_n,

$$y(1,\bar{x}) = 0 \quad \text{and} \quad \sigma(1,\bar{x}) = 1, \tag{20}$$

as location and scale normalisations, Condition (17) similarly implies that

$$\Pr\left(D_n = 1, w_n(1) \geq \frac{w - y(1,x)}{\sigma(1,x)} \mid X_n = \bar{x}\right) \sim q_1 S\left(\frac{w - y(1,x)}{\sigma(1,x)}\right). \tag{21}$$

By combining (19) and (21), we obtain that

$$\Pr(D_n = 1, w_n(1) \geq w \mid X_n = x) \sim \Pr\left(D_n = 1, w_n(1) \geq \frac{w - y(1,x)}{\sigma(1,x)} \mid X_n = \bar{x}\right), \tag{22}$$

which is the key identifying equation for $y(1,x)$ and $\sigma(1,x)$.To elaborate, as mentioned, $\Pr(D_n =$

$1, w_n(1) \geq w | X_n = x)$ is observed for every w. Thus, $y(1,x)$ and $\sigma(1,x)$ are identified if

$$\Pr(D_n = 1, w_n(1) \geq w \mid X_n = x) \sim \Pr(D_n = 1, w_n(1) \geq sw + u \mid X_n = \bar{x}), \tag{23}$$

only holds at $s = \frac{1}{\sigma(1,x)}$ and $u = -\frac{y(1,x)}{\sigma(1,x)}$. To show identification, take $s > 0$ and u such that (23) holds. Let $\tilde{w} := \sigma(1,x)w + y(1,x)$. By (19),

$$\Pr\left(D_n = 1, w_n(1) \geq \tilde{w} \mid X_n = x\right) \sim q_1 S\left(\frac{\tilde{w} - y(1,x)}{\sigma(1,x)}\right)$$

and

$$\Pr(D_n = 1, w_n(1) \geq s\tilde{w} + u \mid X_n = \bar{x}) \sim q_1 S(s\tilde{w} + u).$$

Therefore, using (23) and the definition of $\tilde{w}$,

$$S(w) \sim S(t(w + \nu)), \tag{24}$$

where $t := s\sigma(1,x)$ and $\nu := \frac{1}{\sigma(1,x)}(y(1,x) + u/s)$. By Lemma 2.1 of DM, it can be shown that if $\mathbb{E}(|\epsilon_n(1)|) < +\infty$, then (24) holds only if the *multiplicative* term t is equal to one, that is, $s = 1/\sigma(1,x)$, which implies that $\sigma(1,x)$ is identified; see Appendix C for an informal proof of Lemma 2.1 of DM. Next, we rewrite (24) using $t = 1$ and obtain

$$S(w) \sim S(w + \nu), \tag{25}$$

where $\nu := \frac{1}{\sigma(1,x)}(y(1,x) + u\sigma(1,x)) = \frac{y(1,x)}{\sigma(1,x)} + u$. To use again Lemma 2.1 of DM to identify $y(1,x)$, we write (25) more explicitly as

$$\Pr(\epsilon_n(1) \geq w) \sim \Pr\left(\epsilon_n(1) \geq w + \frac{y(1,x)}{\sigma(1,x)} + u\right). \tag{26}$$

Multiplying by some $\beta > 0$, and taking the exponential of, each side of the inequalities defining the events whose probabilities (26) compares, we obtain

$$\tilde{S}\left(\tilde{w} \exp\left(\beta\left(\frac{y(1,x)}{\sigma(1,x)} + u\right)\right)\right) \sim \tilde{S}(\tilde{w}), \tag{27}$$

where $\tilde{w} := \exp(\beta w)$ and $\tilde{S}$ is the survival function of $\exp(\beta\epsilon_n(1))$. Using again Lemma 2.1 of DM, if $\mathbb{E}(\exp(\beta\epsilon_n(1))) < +\infty$ for some $\beta > 0$, then (27) holds only if the *multiplicative* term $\exp\left(\beta\left(\frac{y(1,x)}{\sigma(1,x)} + u\right)\right)$ is equal to one, that is, $u = -y(1,x)/\sigma(1,x)$, which implies that $y(1,x)$ is identified. The assumption $\mathbb{E}(\exp(\beta\epsilon_n(1))) < +\infty$ for some $\beta > 0$ accommodates distributions with tails heavier than those of the Normal distribution, such as the Laplace and logistic distributions, which is important for realistically modelling income distributions. Furthermore, this assumption implies $\mathbb{E}(|\epsilon_n(1)|) < \infty$, a condition used to identify $\sigma(1, x)$, as mentioned earlier.

Challenges in Our Setting. The approach proposed by DM addresses a key challenge in our framework—namely, the absence of exclusion restrictions—an aspect common to the models we consider. However, it relies on two key assumptions that merit further analysis. The first assumption states that the variables affecting the deterministic wage of each job are covariates observed by the researcher, so that the probability $\Pr(D_n = 1, w_n(1) \geq w \mid X_n = x)$—key element of DM's identifying Equation (22)—is directly known from the data. This is not the case in our setting, as some components of the vector of state variables $s_{n,t} := (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$—specifically, $P_{n,t}$ and e_n—are unobserved. Therefore, we must identify the distribution of $(D_{n,t}, w_{n,t})$ conditional on $s_{n,t}$ in a preliminary step.

The second assumption states that the model must be structured so that, if the wage of job d becomes arbitrarily large while the covariates X_n remain fixed, the probability of choosing job d stays strictly positive, as evidenced by Condition (17). This condition does not typically hold in settings whose equilibrium pricing rule resembles a second-price auction, such as ours. To be precise, in our framework, the equilibrium wage of job d is the expected output at the second-best firm d', $y(d', s_{n,t}(e)) + \epsilon_{n,t}(d', e)$, plus a compensating differential, $\Psi(d, d', s_{n,t}(e))$. Consequently, letting the wage of job d go to $+\infty$, while the state variables $s_{n,t}(e)$ remain fixed, effectively pushes the productivity shock of the *second*-best firm, $\epsilon_{n,t}(d', e)$—and, in turn, the expected output produced by such firm and the wage it could offer—to $+\infty$, thereby potentially altering the equilibrium ranking of firms and making the former first-best firm d no longer the top choice for worker n.

Therefore, we next outline how we address the two issues mentioned above and extend DM's approach to apply it to our framework.[12][13]

[12]Some papers not discussed in our (incomplete) literature overview of the Roy model show that the deterministic component of wages can be identified without exclusion restrictions and at-infinity arguments, provided the we observe at least as many continuous worker attributes as there are alternatives (see, for instance, (Lee and Lewbel, 2013; Kim and Lee, 2025).) However, standard employer-employee match datasets, such as the LEHD dataset, do not contain continuous worker attributes, making this identification strategy infeasible in our setting.

[13]Our overview has focused on the static Roy model. Dynamic extensions of these arguments often rely on additional simplifying assumptions, such as directional and irreversible choices and the presence of absorbing states. For instance, in the schooling context described by Taber (2000), students acquire one degree at a time, once a degree is earned, it cannot be revoked, and withdrawing from a degree program effectively precludes reentry. None of these restrictions apply to our framework, nor are required for our identification arguments.

3.3 Our Identification Approach

Here, we sketch our identification argument by class of primitives of interest. We present the formal argument in the next section. The reader will see that we establish identification in our rich class of models by simply building on existing identification arguments developed for static models and using only information on job choices and wages. Our approach does not require imposing restrictions on endogenous variables—such as monotonicity conditions—or on the dynamics of states, choices, and outcomes—such as sufficient job mobility as in the AKM. Instead, it relies on restrictions that allow for arbitrary patterns of selection based on endogenously time-varying unobservables, are easy to verify, impose minimal data requirements, and ultimately lead to a constructive estimator.

Information Technology, Deterministic Wage Component, and Distribution of Productivity Shocks. As discussed in the previous section, a first challenge in implementing DM's approach for identifying the deterministic wage $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$ is that some components of the vector of state variables $s_{n,t} := (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ are unobserved. More precisely, according to DM's identifying Equation (22), their strategy hinges on knowledge of the distribution of (D_n, w_n) conditional on X_n. Using the notation of our original dynamic wage Equation (11), we would need to know the distribution of $(D_{n,t}, w_{n,t})$ conditional on $s_{n,t}$, which the sampling process does not directly reveal because $P_{n,t}$ and e_n are unobserved. Consequently, before we can use DM's approach, we must identify this distribution. We accomplish this in two steps. In the first step, we express the distribution of $w_{n,t}$ conditional on $(H_{n,1}, D_n^t)$—which is known from the data under Assumption 1—as a mixture over worker n's efficiency e_n and signal history $a_n^{t-1} := (a_{n,1}, \ldots, a_{n,t-1})$:

$$\begin{aligned}\Pr(w_{n,t} \mid H_{n,1}, D_n^t) = \sum_{e \in \mathcal{E}} \sum_{(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}} & \Pr\left(w_{n,t} \mid H_{n,1}, D_n^t, e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1})\right) \\ & \times \Pr\left(e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1}) \mid H_{n,1}, D_n^t\right).\end{aligned} \tag{28}$$

Using exiting results on the identification of mixture models, we identify the mixture weights and components in (28) (Proposition 2). In the second step, we use the fact that the vector of state variables $s_{n,t}$ is essentially a deterministic function of the variables $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$. Consequently, by suitably combining the wage mixture weights and components, we can identify the distribution of $(D_{n,t}, w_{n,t})$ given $s_{n,t}$, which is the crucial starting ingredient in DM's approach (Proposition 3).

To be more precise on the latter point, by concatenating the mixture weights across periods, we identify the prior belief function

$$p_1(H_{n,1}, e_n), \tag{29}$$

and the conditional signal distribution

$$\Pr(a_{n,t-1} \mid H_{n,1}, D_{n,t-1}, e_n, \theta_n). \tag{30}$$

By combining (29) and (30), we obtain the belief $P_{n,t}$ using Bayes' rule, as in (3) and (4). Thus, we identify the model's learning process (information technology). In turn, if $|\mathcal{E}|$ is also known, we identify the support $\mathcal{S}_t$ of $s_{n,t}$ and, in particular, the *map* from realizations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realizations of $s_{n,t}$.[14] Moreover, from the mixture weights, we identify the distribution

$$\Pr(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1}). \tag{31}$$

Lastly, from the mixture components and weights, we identify the distribution

$$\Pr(D_{n,t}, w_{n,t} \mid H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1}). \tag{32}$$

Therefore, by combining (31) and (32) with knowledge of the map from $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to $s_{n,t}$, we are able to identify the distribution of $(D_{n,t}, w_{n,t})$ conditional $s_{n,t}$.

Having obtained the distribution of $(D_{n,t}, w_{n,t})$ given $s_{n,t}$ from the steps above, the next challenge in using DM's approach to recover the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ is ensuring that the following limiting probability

$$\lim_{w \to +\infty} \Pr(D_{n,t} = d \mid s_{n,t}(e) = s,\ w_{n,t}(d, C(d, s), e) = w), \tag{33}$$

which is an adaptation of the probability in Condition (17) to our setting, remains strictly positive, where $C(d, s)$ (short for "complement") denotes the firm in worker n's choice set $\mathcal{D}(s)$ that is not the first-best firm d, under Assumption 2. However, in our environment—unlike a standard Roy model—the relevant shock affecting the observed wage $w_{n,t}$ is the shock of the second-best firm. Consequently, increasing $w_{n,t}$ to $+\infty$—or, equivalently to its finite upper bound if assuming a bounded support—while holding the state variables $s_{n,t}(e)$ fixed at s effectively pushes the second-

[14] Note that $\mathcal{S}_t$ can potentially vary across time periods because of $\kappa_{n,t}$, which captures worker n's human capital investments up to the beginning of period t and, hence, can include variables such as worker n's tenure at each firm, as previously mentioned. For example, if $t = 3$ and $\kappa_{n,4}$ includes worker n's tenure at each firm, then that tenure variable cannot take values strictly greater than three with positive probability. However, at $t = 4$, it is possible for this variable to take on values strictly greater than three, namely up to four.

best firm's productivity shock—and thus its expected output and wage offer—to $+\infty$. This alters the equilibrium firm ranking and causes the former first-best firm to no longer be the top choice for worker n, thereby driving the probability in (33) to zero. To see this more clearly, recall that the model's equilibrium is efficient, meaning that a worker's choice of firm is the solution to a planning problem, namely, the problem of a planner choosing a job for each worker in each period. Therefore, job choices maximize the expected present discounted value of output. In turn, recalling that Assumption 2 specifies the cardinality of worker n's choice set $\mathcal{D}(s_{n,t}(e))$ as two, we have

$$\begin{aligned}
&\Pr(D_{n,t} = d \mid s_{n,t}(e) = s,\, w_{n,t}(d, C(d,s), e) = w) \\
&= \Pr\big(Y(d,s) + \epsilon_{n,t}(d,e) \geq Y(C(d,s),s) + \epsilon_{n,t}(C(d,s),e) \\
&\qquad\qquad\qquad \mid s_{n,t}(e) = s,\, \varphi(d, C(d,s), s) + \epsilon_{n,t}(C(d,s),e) = w\big) \\
&= \Pr\big(Y(d,s) + \epsilon_{n,t}(d,e) \geq Y(C(d,s),s) + w - \varphi(d, C(d,s), s) \\
&\qquad\qquad\qquad \mid \epsilon_{n,t}(C(d,s),e) = w - \varphi(d, C(d,s), s)\big),
\end{aligned} \tag{34}$$

where $Y(d,s) + \epsilon_{n,t}(d,e)$ is the expected present discounted value of output for firm d in state s after productivity shocks have realized. The first equality uses the equilibrium expression for $w_{n,t}(d, C(d,s), e)$ in the conditioning event, and the second replaces $\epsilon_{n,t}(C(d,s),e)$ in the inequality with its conditional value by the exogeneity of the productivity shock. Further assume, for the sake of argument, that the productivity shocks are independent. Then, (34) simplifies to

$$\begin{aligned}
&\Pr(D_{n,t} = d \mid s_{n,t}(e) = s,\, w_{n,t}(d, C(d,s), e) = w) \\
&= \Pr\big(Y(d,s) + \epsilon_{n,t}(d,e) \geq Y(C(d,s),s) + w - \varphi(d, C(d,s), s)\big).
\end{aligned} \tag{35}$$

We can see now that sending w to $+\infty$ causes the probability in (35) to become zero, illustrating how the original DM's construction fails in our second-price auction-like scenario.

To address this issue, we send w to $-\infty$ in (33)—or, equivalently, to its finite lower bound if assuming a bounded support—rather than to $+\infty$. When w takes extremely negative values, Equation (35) approaches one, thus resolving our problem and allowing us to apply DM's arguments by focusing on the *left* extreme tail of the wage distribution, instead of the right extreme tail as in their original construction (see Proposition 4).

Lastly, having identified the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$ through the steps above, we use it to extract the distribution of the productivity shocks from the wage mixture in (28). (Lemma 6). We highlight that such a distribution would not be identified based on DM's arguments alone. Indeed, the distribution of unobserved heterogeneity remains unidentified in their framework—in particular, its first moment, thus preventing the identification of mean wages. We can identify it in our framework by nesting DM within our mixture argument.

Law of Motion of the State and CCPs. The wage mixture weights in (28) not only help us identify the deterministic wage components but are also key to identifying other objects of interest, such as the law of motion of the state variables, $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$, and the CCPs, $\Pr(D_{n,t} \mid s_{n,t})$ (Proposition 3). Intuitively, recall that the mixture weights essentially determine the distribution of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ in each period, which in turn governs the state variables $s_{n,t}$ and subsequently the occupation choices. Therefore, by appropriately combining these weights across periods, it becomes natural to recover the law of motion of $s_{n,t}$ and the CCPs. Notably, in contrast to the typical method of deriving the CCPs from agents' discrete choices in dynamic models, here the CCPs are identified from the continuous part of the model, that is, the wage distribution.

Output Technology and Compensating Differential. In addition to identifying the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$, we also distinguish within $\varphi(\cdot)$ the output (and human capital) technology, captured by $y(\cdot)$, from the compensating differential $\Psi(\cdot)$. In particular, leveraging the equilibrium efficiency discussed in Section 2, we represent the market-wide equilibrium allocation problem as a single-agent (the social planner) decision-theoretic problem. Therefore, based on knowledge of the CCPs and the distribution of the productivity shocks, we identify $y(\cdot)$ using the standard arguments of dynamic decision problems as in Magnac and Thesmar (2002) (Proposition 7). With $y(\cdot)$ identified, then $\Psi(\cdot)$ can be backed out residually from $\varphi(\cdot)$ (Proposition 8).

4 Formal Identification Argument

We now formally illustrate the identification approach previewed in Section 3.

4.1 Wage Mixture

Using the law of total probability, we represent the distribution of $w_{n,t}$ conditional on $(H_{n,1}, D_n^t)$ as a finite mixture over worker n's efficiency e_n and signal history a_n^{t-1}:

$$\begin{aligned}\Pr(w_{n,t} \mid H_{n,1}, D_n^t) = \sum_{e \in \mathcal{E}} \sum_{(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}} & \Pr\left(w_{n,t} \mid H_{n,1}, D_n^t, e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1})\right) \\ & \times \Pr\left(e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1}) \mid H_{n,1}, D_n^t\right).\end{aligned} \tag{36}$$

In Equation (36), for each $(e,a) \in \mathcal{E} \times \mathcal{A}^{t-1}$, $\Pr(w_{n,t} \mid H_{n,1}, D_n^t, e_n = e, a_n^{t-1} = (a_1,\ldots,a_{t-1}))$ is called a mixture component and $\Pr(e_n = e, a_n^{t-1} = (a_1,\ldots,a_{t-1}) \mid H_{n,1}, D_n^t)$ is called a mixture weight. In this section, we show the identification of these components and weights.

Assumption 3.(Finite Mixture of Gaussian Mixtures) (i) Given $h \in \mathcal{H}$, $(d_1,\ldots,d_t) \in \mathcal{D}^t$, $e \in \mathcal{E}$, and $(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}$, let C denote the identity of worker n's second-best firm at time t, uniquely determined under Assumption 2.[15] For each $t \geq 1$, $h \in \mathcal{H}$, $(d_1,\ldots,d_t) \in \mathcal{D}^t$, $e \in \mathcal{E}$, and $(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}$, the productivity shock $\epsilon_{n,t}(C,e)$ conditional on $H_{n,1} = h$, $D_n^t = (d_1,\ldots,d_t)$, $e_n = e$, and $a_n^{t-1} = (a_1,\ldots,a_{t-1})$ is distributed as mixture of a, possibly uncountable, family of Gaussian distributions with means and variances taking values within a known compact set $\mathcal{G}_t \subset \mathbb{R} \times \mathbb{R}_+$:

$$\begin{aligned} f(\epsilon_{n,t}(C,e) \mid H_{n,1} = h, D_n^t = (d_1,\ldots,d_t), &e_n = e, a_n^{t-1} = (a_1,\ldots,a_{t-1})) \\ &= \int_{(\mu,\sigma^2)\in\mathcal{G}_t} \mathcal{N}(\mu,\sigma^2) d\pi(\mu,\sigma^2; h, d_1,\ldots,d_t, e, a_1,\ldots,a_{t-1}), \end{aligned}$$

where $\pi(\cdot; h, d_1,\ldots,d_t, e, a_1,\ldots,a_{t-1})$ is the mixture weight function, which can vary across values $(h, d_1,\ldots,d_t, e, a_1,\ldots,a_{t-1})$. (ii) $\mathcal{E}$ and $\mathcal{A}$ are finite sets with known cardinalities. $\diamond$

Proposition 2 (Wage Mixture)**.** *Under Assumptions 1 to 3, the wage mixture weights* $\Pr(w_{n,t} | H_{n,1} = h, D_n^t = (d_1,\ldots,d_t), e_n = e, a_n^{t-1} = (a_1,\ldots,a_{t-1}))$ *and components* $\Pr(e_n = e, a_n^{t-1} = (a_1,\ldots,a_{t-1}) | H_{n,1} = h, D_n^t = (d_1,\ldots,d_t))$ *are identified for each* $t \geq 1$, $h \in \mathcal{H}$, $(d_1,\ldots,d_t) \in \mathcal{D}^t$, $e \in \mathcal{E}$, *and* $(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}$ *such that* $\Pr(H_{n,1} = h, D_n^t = (d_1,\ldots,d_t)) > 0$.

We demonstrate Proposition 2 using the identification result from Bruni and Koch (1985). Specifically, under Assumption 1, the left-hand side of Equation (36) is known from the data. Furthermore, under Assumptions 2 and 3, Equation (36) is a *finite mixture of mixtures of possibly uncountable families of Gaussian distributions*, where the means and variances take values within a compact set.[16] In this scenario, Bruni and Koch (1985) establishes that the weights and components are identified (Section 4.c of Bruni and Koch (1985)). In the remainder of the section, we examine Assumption 3 in more depth.

Assumption 3 (i) imposes that the conditional distribution of the productivity shock of the second-best firm at each time t is a mixture of a, possibly uncountable, family of Gaussian distributions. To

[15]Recall that $s_{n,t}$ is essentially a deterministic function of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ in our framework. Thus, if $(H_{n,1}, D_n^t, e_n, a_n^{t-1})$ is fixed to a specific realization, then the realization of $s_{n,t}$ is also fixed and, consequently, the identity of worker n's second-best firm at time t is uniquely determined by Assumption 2.

[16]We emphasize that we do *not* assume that Equation (36) is a finite Gaussian mixture.

understand the rationale behind this assumption, recall that the literature offers two main approaches for the nonparametric identification of mixture models. The first approach uses exclusion restrictions, that is, variables that enter either the mixture weights or the mixture components, but not both (Henry, Kitamura, and Salanié, 2014; Compiani and Kitamura, 2016; Jochmans, Henry, and Salanié, 2017). The second approach considers the distribution of the entire vector of wages $(w_{n,1}, \ldots, w_{n,T})$, rather than focusing on the wage distribution at each time t as in (36), and relies on assumptions that simplify the temporal dependence of wage observations, such as conditional independence and Markovianity (Hall and Zhou, 2003; Allman, Matias, and Rhodes, 2009; Kasahara and Shimotsu, 2009; Bonhomme, Jochmans, and Robin, 2016a,b). Neither approach is suitable for our framework. Specifically, in the class of models considered, exclusion restrictions do not arise: any variable that affects the conditional distribution of efficiency and signals also affects the conditional distribution of wages, and vice versa. Furthermore, the wage observations are neither conditionally independent over time nor subject to Markovian constraints; more broadly, due to the underlying human capital and learning processes, nothing remains invariant enough over time for the wage time series to be useful for identification.

Given these considerations, we resort to imposing parametric restrictions on the wage distribution in each period to identify (36). At the same time, we aim for these parametric restrictions to be as weak as possible, considering that the mixture components in (36) are "contaminated" by selection on $\epsilon_{n,t}$. In fact, note that each mixture component of (36) is essentially the distribution of the productivity shock of the second-best firm, conditional on $(H_{n,1}, D_n^t, e_n, a_n^{t-1})$ and shifted by the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$. For the reasons explained in Section 3.2, due to selection, we should not expect this conditional distribution to be equal to the unconditional one, nor to have a "standard" parametric form, such as Normal or Gumbel. Instead, this conditional distribution is endogenously determined by how workers and firms make decisions within the model. Therefore, we must rely on assumptions that allow for very flexible distributions of the mixture components. In particular, we model each mixture component as a mixture of a, possibly uncountable, family of Gaussian distributions, which is known to effectively approximate *any* distribution if the support for its means and variances ($\mathcal{G}_t$ in our notation) is large enough. This is shown by Gosh and Ramamoorthi (2003), Norets (2010), and Norets and Pelenis (2014).

Despite the extreme flexibility of the class of mixtures embraced by Assumption 3, Proposition 2 can also be extended to nonparametric mixture families. In particular, Aragam et al. (2020) propose a criterion known as the "clusterability" condition, which is sufficient for identification. Intuitively, this condition essentially requires that the mixture components are "sufficiently distinct," as quantified by an appropriate distance measure—a notion that can already be found in Teicher (1961, 1963)'s earlier discussion of mixture identifiability. Not only is this condition met in the setting described by Bruni and Koch (1985), but it applies to a wide range of other mixtures. Assumption 3(ii) requires that the supports $\mathcal{E}$ and $\mathcal{A}$ of e_n and $a_{n,t}$ be finite with known cardinalities. In Appendix A, we explain that this assumption can be readily relaxed to know only an upper bound on $\mathcal{E}$ and $\mathcal{A}$. We have refrained from doing so in the main text for simplicity of exposition. In the same appendix, we also discuss the extent to which e_n and $a_{n,t}$ can be allowed to vary continuously and be multidimensional. Note that we do *not* impose that the mixture weights of (36) are strictly positive for each $e \in \mathcal{E}$ and $(a_1, \ldots, a_{t-1}) \in \mathcal{A}^{t-1}$. In fact, these weights could be zero due to Assumption 2. We elaborate on this further in Appendix B.

Finite mixture models are always identified up to the labeling of mixture components because the likelihood is invariant under permutation of these components. In our setting, pinning down the labeling is important as we must combine the mixture weights across different periods t to proceed with the next steps of the identification arguments. In Appendix B, we discuss how the labeling indeterminacy is resolved.

4.2 Information Technology, Law of Motion of the State, and CCPs

In this section, we show how to identify the prior belief function $p_1(H_{n,1}, e_n)$ and the conditional signal distribution $\Pr(a_{n,t} \mid H_{n,1}, D_{n,t}, e_n, \theta_n)$—which allow us to compute $P_{n,t}$ and thus fully describe the information technology—by concatenating the wage mixture weights across periods. Based on these results, we show how to identify the unconditional distribution of the vector of state variables $s_{n,t}$ and the distribution of $(D_{n,t}, w_{n,t})$ given $s_{n,t}$. Finally, using the wage mixture weights again, we show how to identify the law of motion of the vector of state variables, $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$, and the CCPs, $\Pr(D_{n,t} \mid s_{n,t})$.

Assumption 4.(Support of Ability and Signals) $\Theta := \{\bar{\theta}, \underline{\theta}\}$ and $\mathcal{A} := \{\bar{a}, \underline{a}\}$. $\diamond$

Assumption 5.(Conditional Signal Distribution) (i) For each period $1 \leq t \leq T - k$ and integer $k > 0$,

$$\Pr(a_{n,t}, \ldots, a_{n,t+k} \mid H_{n,1}, D_{n,t}, \ldots, D_{n,t+k}, e_n, \theta_n) = \prod_{j=t}^{t+k} \Pr(a_{n,j} \mid H_{n,1}, D_{n,j}, e_n, \theta_n).$$

(ii) The distribution of $a_{n,t}$ conditional on $(H_{n,1}, D_{n,t}, e_n, \theta_n)$ is time-invariant. In other words, for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$, the following probabilities remain the same across periods t:

$$\alpha_{h,d,e} = \Pr(a_{n,t} = \bar{a} \mid H_{n,1} = h, D_{n,t} = d, e_n = e, \theta_n = \bar{\theta}),$$

and

$$\beta_{h,d,e} = \Pr(a_{n,t} = \bar{a} \mid H_{n,1} = h, D_{n,t} = d, e_n = e, \theta_n = \underline{\theta}).$$

(iii) For each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$, $\alpha_{h,d,e} > \beta_{h,d,e}$. ◇

Assumption 6.(Full State Support) For each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$, there exist three consecutive periods $\tau_{h,d,e}, \tau_{h,d,e}+1, \tau_{h,d,e}+2$ in the data such that:

(i) $\Pr(H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d) > 0$.

(ii) $\Pr(e_n = e \mid H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d) > 0$.

(iii) $\Pr(a_{n,\tau_{h,d,e}} = a_{\tau_{h,d,e}}, a_{n,\tau_{h,d,e}+1} = a_{\tau_{h,d,e}+1}, a_{n,\tau_{h,d,e}+2} = a_{\tau_{h,d,e}+2} \mid H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d, e_n = e) > 0$ for every $(a_{\tau_{h,d,e}}, a_{\tau_{h,d,e}+1}, a_{\tau_{h,d,e}+2}) \in \mathcal{A}^3$.

These periods $\tau_{h,d,e}, \tau_{h,d,e}+1, \tau_{h,d,e}+2$ include the starting periods $1,2,3$ for some $d \in \mathcal{D}$. ◇

Assumption 7.(Length of Panel) Let $T^* := \max_{(h,d,e)\in\mathcal{H}\times\mathcal{D}\times\mathcal{E}} \ \tau_{h,d,e}$. Then, $T \geq T^* + 3$. ◇

Proposition 3. *(Information Technology, Law of Motion of the State, and CCPs) Let Assumptions 1 to 7 hold. Then: (I) the prior belief function* $p_1(h,e)$ *and the conditional signal distribution components* $\{\alpha_{h,d,e}, \beta_{h,d,e}\}$ *are identified for each* $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$*. (II) The unconditional distribution of the state* $s_{n,t} := (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ *and its support* $\mathcal{S}_t$ *are identified for each* $t \geq 1$*. (III)* $\Pr(D_{n,t} = d, w_{n,t} \leq w \mid s_{n,t} = s)$ *is identified for each* $d \in \mathcal{D}$*,* $w \in \mathbb{R}$*,* $s \in \mathcal{S}_t^0$*, and* $t \geq 1$*, where* $\mathcal{S}_t^0 \subseteq \mathcal{S}_t$ *denote the subset of* $\mathcal{S}_t$ *collecting the realizations of* $s_{n,t}$ *with strictly positive probability. (IV) The law of motion of the vector of state variables,* $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$*, is identified for each* $t \geq 2$*. (V) The conditional choice distribution,* $\Pr(D_{n,t} \mid s_{n,t})$*, is identified for each* $t \geq 1$*.*

Assumption 4 restricts the supports of Θ and $\mathcal{A}$ to have cardinality two, as assumed in the model description in Section 2. In Appendix A, we discuss how this assumption can be relaxed to allow for continuous/multidimensional θ_n and $a_{n,t}$. Assumption 5 (i) imposes that the signals are conditionally independent, while Assumption 5 (ii) requires the conditional signal distribution to be time-invariant. Assumption 5 (iii) orders $\alpha_{h,d,e}$ and $\beta_{h,d,e}$, and is a natural condition because high-ability types are more likely to produce high signals. Under Assumption 6, there is a strictly positive mass of workers remaining at each job d for three periods in a sufficient number of states. This assumption can be verified from the data and wage mixture weights, as detailed in Appendix C. Assumption 7 requires the panel to contain enough periods to verify the conditions imposed by Assumption 6.

We now give the reader an intuition on how we use these assumptions to show the identification of the information technology in part (I) of Proposition 3 and refer to Appendix C for the proof of the other parts. Let $(h, d, e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ such that Assumption 6 is satisfied for $(\tau_{h,d,e}, \tau_{h,d,e}+1, \tau_{h,d,e}+2) = (1, 2, 3)$. We proceed in three steps. First, under Assumptions 6 (i)–(ii) and 7, we establish the identification of the conditional distribution of signal history at $t = 3$ for workers who remain at job d throughout all three periods. Specifically, $\Pr(a_n^3 = (a_1, a_2, a_3) \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e)$ is identified for all $(a_1, a_2, a_3) \in \mathcal{A}^3$ by combining the wage mixture weights in periods 3 and 4—hence, the requirement $T \geq 4$ of Assumption 7. These wage mixture weights are identified by Proposition 2 under Assumptions 1 to 3 and 6 (i). Assumption 6 (ii) is used to avoid dealing with zero probability events, which would hinder the identification of $\Pr(a_n^3 = (a_1, a_2, a_3) \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e)$ when combining the wage mixture weights. Similarly, we establish the identification of the conditional distribution of the signal at $t = 1$ for workers who are at job d in the first period. Specifically, $\Pr(a_{1,n} = a_1 \mid H_{n,1} = h, D_{n,1} = d, e_n = e)$ is identified for all $a_1 \in \mathcal{A}$ by combining the wage mixture weights in periods 1 and 2.

Second, using Assumptions 4 and 5 (i)-(ii), we represent $\Pr(a_n^3 \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e)$—identified in the first step above—as a binomial mixture over θ_n via Bayes' rule. The two components of this binomial mixture are given by $\alpha_{h,d,e}$ and $\beta_{h,d,e}$. Similarly, we represent $\Pr(a_{1,n} \mid H_{n,1} = h, D_{n,1} = d, e_n = e)$–identified in the first step above—as a Bernoulli mixture over θ_n via Bayes' rule. The two components of this Bernoulli mixture are given by $\alpha_{h,d,e}$ and $\beta_{h,d,e}$; the two weights are given by $p_1(h, d, e)$ and $1 - p_1(h, d, e)$.

Third, under Assumption 6 (iii), we show that $\alpha_{h,d,e}$ and $\beta_{h,d,e}$ are identified from knowledge of $\Pr(a_n^3 \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e)$ up to labelling based on Blischke (1964, 1978). Under Assumption 5 (iii), this labelling indeterminacy is solved. In turn, $p_1(h, d, e)$ is identified from knowledge of $\alpha_{h,d,e}$, $\beta_{h,d,e}$, and $\Pr(a_{1,n} \mid H_{n,1} = h, D_{n,1} = d, e_n = e)$.

Observe that Assumption 6 cannot hold for every firm $d \in \mathcal{D}$ during the same starting periods $(\tau_{h,d,e}, \tau_{h,d,e} + 1, \tau_{h,d,e} + 2) = (1, 2, 3)$, due to Assumption 2. This is because, under Assumption 6, firm d must appear in the choice set of worker n for "sufficiently enough" state realizations across periods 1,2,3 to satisfy conditions (i) to (iii) of Assumption 6, thus limiting the presence of other firms d' in such choice sets and, in turn, the identifiability of $\alpha_{h,d',e}$ and $\beta_{h,d',e}$. To overcome this issue, we must impose that conditions (i) to (iii) of Assumption 6 hold for each $d \in \mathcal{D}$ at *some*, possibly (h, d, e)-specific, triplets of consecutive periods $\tau_{h,d,e}, \tau_{h,d,e} + 1, \tau_{h,d,e} + 2$.

4.3 Deterministic Wage Component and Distribution of Productivity Shocks

In this section, we show how to identify the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$ and the distribution of the vector of shocks $\epsilon_{n,t}$ by combining Proposition 3 with DM's approach. Introducing some notation is useful to formalise the result. Given $(d, e) \in \mathcal{D} \times \mathcal{E}$ and $t \geq 1$, define $\mathcal{S}_t^0(e) \subseteq \mathcal{S}_t^0$ as the subset of state realizations in $\mathcal{S}_t^0$ with e_n set to e. Further, let $\mathcal{S}_t^0(d, e) \subseteq \mathcal{S}_t^0(e)$ be the subset of realizations s of $s_{n,t}(e)$ such that $\Pr(D_{n,t} = d \mid s_{n,t}(e) = s) > 0$. $\mathcal{S}_t^0(e)$ and $\mathcal{S}_t^0(d, e)$ are known by Proposition 3. Lastly, let $C(d, s_{n,t}(e)) := \mathcal{D}(s_{n,t}(e)) \setminus \{d\}$ denote the firm in worker n's choice set $\mathcal{D}(s_{n,t}(e))$ that is not the first-best firm d. Such a firm is uniquely determined under Assumption 2 (but its identity is unknown to the econometrician).

Assumption 8.(Location Normalisation) For each $(d, e) \in \mathcal{D} \times \mathcal{E}$ and $t \geq 1$, $\varphi(d, C(d, \bar{s}_t), \bar{s}_t) = 0$ for some $\bar{s}_t \in \mathcal{S}_t^0(d, e)$ known by the econometrician. ◇

Assumption 9.(Identical Distribution) For each $e \in \mathcal{E}$, the productivity shocks $\{\epsilon_{n,t}(d, e)\}_{d \in \mathcal{D}}$ are identically distributed across d. ◇

Assumption 10.(Exogeneity) For each $e \in \mathcal{E}$, $\epsilon_{n,t}(e)$ is independent of $s_{n,t}(e)$. ◇

Assumption 11.(Tail) For each $(d, e) \in \mathcal{D} \times \mathcal{E}$, $\inf(\text{Supp}(\epsilon_{n,t}(d, e))) = -\infty$ and $\mathbb{E}(\exp(-\beta\epsilon_{n,t}(d, e)) < +\infty$ for some $\beta > 0$. ◇

Assumption 12.(Independence in the Limit) For each $(d, e) \in \mathcal{D} \times \mathcal{E}$ and $t \geq 1$,

$$\lim_{w \to -\infty} \Pr(D_{n,t} = d \mid s_{n,t}(e) = s, w_{n,t}(d, C(d, s), e) = w) = q_{d,e} > 0,$$

for each $s \in \mathcal{S}_t^0(d, e)$ and for some unknown $0 < q_{d,e} \leq 1$, independent of s. ◇

Proposition 4. *(Deterministic Wage) Under Assumptions 1 to 12, the deterministic wage component* $\varphi(d, C(d, s), s)$ *is identified for each* $(d, e, s) \in \mathcal{D} \times \mathcal{E} \times \mathcal{S}_t^0(d, e)$ *and* $t \geq 1$.

Assumptions 8 to 12 adapt DM's restrictions to our context, as previewed in Section 3.2—note that we do not have a scale term multiplying the productivity shock in the wage equation in our case, and thus, we are only interested in the identification of the location term $\varphi(\cdot)$. By combining these assumptions with knowledge of the conditional distribution of $(D_{n,t}, w_{n,t})$ given $s_{n,t}$, established by part (III) of Proposition 3, Proposition 4 shows how to identify the deterministic wage component $\varphi(\cdot)$. In the remainder of the section, we further discuss Assumptions 8 to 12.

Assumption 8 introduces a location normalization at $\bar{s}_t$, akin to Condition (20) discussed in Section 3.2. The subscript t in $\bar{s}_t$ highlights that the normalization point can differ across time periods. Assumption 8 can be replaced by imposing $\varphi(d, C(d, \bar{s}_t), \bar{s}_t) = c_t$ with c_t known. Moreover, it can be replaced by a weaker condition when the function $\varphi(\cdot)$ is parameterized with time-invariant parameters, as discussed in Section 5, by leveraging the longitudinal dimension of the data.

Assumption 9 is not used by DM and is needed in our framework due to the dependence of the expression for $w_{n,t}$ on the second-best firm. Indeed, recall that the relevant shock affecting $w_{n,t}$ is the shock of the second-best firm, whose identity, as per Assumption 2, can change with each realization of $s_{n,t}(e)$. Namely, if d is the first-best firm and s represents a realization of $s_{n,t}(e)$, the shock entering $w_{n,t}$ is $\epsilon_{n,t}(C(d,s), e)$. Consequently, when applying DM's survival function arguments—beginning with Equation (19)—we must use the survival function of $\epsilon_{n,t}(C(d,s), e)$. If the distribution of such shock is allowed to vary across firms, then the survival function would also vary depending on s, which would compromise the validity of the approach.

Assumption 10 is a standard exogeneity condition. It has already been imposed to characterize equilibrium in Section 2, but we restate it here for completeness, as this is the first place where we use it explicitly for identification purposes. Similarly to Assumption 9, Assumption 10 ensures that the survival function used in DM's arguments does not vary across realizations of $s_{n,t}(e)$.

Assumption 11 (i) requires $\epsilon_{n,t}(d, e)$ to have large support, particularly to accommodate very negative values. However, all the arguments can be easily generalized to the case where $\epsilon_{n,t}(d, e)$ has a finite and known lower bound. Assumption 11(ii) imposes restrictions on the left tail of the distribution of $\epsilon_{n,t}(d, e)$ by requiring that the probability of observing very low values decays more slowly than any exponential rate. As previewed in Section 3.2, DM use this condition to demonstrate identification of the deterministic wage component by examining the behavior of the wage distribution's extreme tails.

Assumption 12 is the key identifying assumption of DM and is the analogue of Condition (17) in Section 3.2. Observe that the quantity $q_{d,e}$ is not indexed by t because the vector of shocks $\epsilon_{n,t}(e)$—which essentially determines the probability limit—is assumed to be identically distributed across time periods t with joint distribution G_e; see the equilibrium characterization in Section 2. For the same reason, Assumptions 9 to 11 are not imposed for each $t \geq 1$ as this would be redundant. Lemma

5 shows that Assumption 12 is satisfied if the productivity shocks are "moderately" dependent across jobs, similarly to what is established by Corollary 4.1 of DM for a standard Roy model.

Lemma 5. *(Moderate Dependence) For each* $(d, e, s) \in \mathcal{D} \times \mathcal{E} \times \mathcal{S}_t^0(d, e)$*, let*

$$\lim_{u \to -\infty} \Pr(\epsilon_{n,t}(d, e) \geq a + u \mid \epsilon_{n,t}(C(d, s), e) = u) = q_{d,e} > 0 \quad \forall a \in \mathbb{R}. \tag{37}$$

Then, Assumption 12 holds. Moreover, if the shocks $\{\epsilon_{n,t}(d, e)\}_{d \in \mathcal{D}}$ *are independently distributed across* d*, then* $q_{d,e}$ *in (37) is equal to one, and hence, Assumption 12 holds.*

We highlight that, in Assumptions 11 and 12, we focus on the left extreme tails of the shock and wage distributions, unlike in the original DM's construction which considers the right extreme tails. This approach ensures that the quantity $q_{d,e}$ remains strictly positive. Indeed, as detailed at the end of Section 3.2, when w approaches $+\infty$, $q_{d,e}$ falls to zero due to the equilibrium pricing mechanism in our model, which resembles a second-price auction.

Lastly, Lemma 6 shows that the distribution of the productivity shocks is identified by combining knowledge of $\varphi(\cdot)$ with the wage mixture Equation (36) and an independence assumption.

Assumption 13.(Independent Distribution) For each $e \in \mathcal{E}$, the productivity shocks $\{\epsilon_{n,t}(d, e)\}_{d \in \mathcal{D}}$ are independently distributed across d. ◇

Lemma 6. *(Distribution of Productivity Shocks) Under Assumptions 1 to 12, the (marginal) distribution of* $\epsilon_{n,t}(d, e)$ *is identified for each* $(d, e) \in \mathcal{D} \times \mathcal{E}$*. If Assumption 13 also holds, then the (joint) distribution* G_e *of the vector* $\epsilon_{n,t}(e) := (\epsilon_{n,t}(d, e) : d \in \mathcal{D})$ *is identified for each* $e \in \mathcal{E}$*.*

4.4 Output Technology and Compensating Differential

Once the deterministic component of wages $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$ is identified, what is left to identify is the output (and human capital) technology, $y(\cdot)$, and the compensating differential, $\Psi(\cdot)$. Here we argue how we can recover $y(\cdot)$ building on standard arguments for the identification of dynamic discrete choice models. With $y(\cdot)$ identified, then $\Psi(\cdot)$ can be backed out residually from $\varphi(\cdot)$.

Intuitively, based on our characterization of equilibrium and its efficiency, it is immediate to show that a worker's choice of firm is the solution to a planning problem, namely, the problem of a planner choosing a job for each worker in each period. Formally, defining by $S(s_{n,t}(e), \epsilon_{n,t}(e))$ the expected present discounted value of the output produced by worker n of type $e_n = e$ or, equivalently, the expected present discounted value of the social surplus generated by worker n at state $(s_{n,t}(e), \epsilon_{n,t}(e))$, this value is given by

$$S(s_{n,t}(e), \epsilon_{n,t}(e)) = \max_{d \in \mathcal{D}} \Big[y(d, s_{n,t}(e)) + \epsilon_{n,t}(d, e) \\ + \delta[1 - \eta(\kappa_{n,t}, d)] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(S(s_{n,t+1}(e), \epsilon_{n,t+1}(e) \mid s_{n,t}(e), d \Big) dG_e \Big].$$

Hence, the market-wide equilibrium allocation problem—namely, the equilibrium matching of workers to firms—reduces to a single-agent decision-theoretic problem. Note that the exogenous separation rate $\eta(\kappa_{n,t}, d)$ is straightforwardly non-parametrically identified by the fraction of employed workers at firm d with given $\kappa_{n,t}$ who leave the market at the end of each period. Furthermore, by Proposition 3 (V) and Lemma 6, the CCPs $\Pr(D_{n,t} \mid s_{n,t}(e))$ and the distribution G_e of the vector of shocks $\epsilon_{n,t}(e)$ are identified. Therefore, $y(s_{n,t}(e))$ can be identified following Magnac and Thesmar (2002).

Proposition 7 (Output and Human Capital Technology)**.** *The exogenous separation rate $\eta(\kappa_{n,t}, d)$ is identified by the fraction of workers with accumulated human capital $\kappa_{n,t}$ leaving the market at the end of period t after employment at firm d. Then, under Assumptions 1 to 13, the output (and human capital) technology $y(d, s)$ is identified for each $(d, e, s) \in \mathcal{D} \times \mathcal{E} \times \mathcal{S}_t^0(d, e)$ and $t \geq 1$ up to its value at one firm at each state.*

Proposition (7) identifies the output (and human capital) technology under the usual set of normalizations imposed in dynamic models. To clarify the mechanics of such normalizations, recall that $\mathcal{S}_t^0(d, e)$ is the set of realizations s of $s_{n,t}(e)$ such that $\Pr(D_{n,t} = d \mid s_{n,t}(e) = s) > 0$. Therefore, if $s \in \mathcal{S}_t^0(d, e)$ and $s \notin \mathcal{S}_t^0(\tilde{d}, e)$ for any other firm $\tilde{d} \neq d$, then the normalization asks to know $y(d, s)$ and Proposition 7 does not tell us anything about $y(\tilde{d}, s)$ for each $\tilde{d} \neq d$. Conversely, if $s \in \mathcal{S}_t^0(d, e)$ and $s \in \mathcal{S}_t^0(\tilde{d}, e)$ for some firm $\tilde{d} \neq d$, then the normalization asks to know $y(d, s)$, and Proposition 7 allows us to identify $y(\tilde{d}, s)$.

Recent work has demonstrated that depending on the counterfactual of interest, the requirement to know specific points of the function $y(\cdot)$ in Proposition 7 may not be a normalization but rather a substantive restriction with significant implications for the empirical conclusions we draw (Kalouptsidi, Scott, and Souza-Rodrigues, 2021; Kalouptsidi, Kitamura, Lima, and Souza-Rodrigues, 2024). The same papers show that removing such mischaracterized "normalizations" leads to the set identification of $y(\cdot)$ and counterfactuals. These arguments can be directly applied to our setting should the reader choose to dispense with the normalizations imposed by Proposition 7. We do not encounter this issue in our empirical application of Section 5 because we estimate a parametric version of our model, which simplifies many identification steps.

When firms consist of multiple jobs, the equilibrium can be inefficient, as noted in Section 2. Even so, the identification of $y(\cdot)$ proceeds analogously to Proposition 7. Specifically, the main difference in the multi-job case is that the market-wide equilibrium allocation problem does not solve the planning problem but instead solves the pseudo-planning problem of maximizing the match surplus for each firm $d \in \mathcal{D}$. In this scenario, the one-period surplus when firm d does not employ the worker equals the deterministic component of the wage paid by the employing firm. Since the latter is identified by Proposition 4, standard dynamic discrete choice arguments applied to each pseudo-planning problem can once again be used to establish the identification of $y(\cdot)$.

With $y(\cdot)$ known, then $\Psi(\cdot)$ can be identified by subtracting $y(\cdot)$ from $\varphi(\cdot)$. To perform this subtraction operation, the researcher must know the identity of the second-best firm—for example, through the use of workers' observed transition patterns, as discussed in Appendix A. In fact, recall that the deterministic wage component when firm d is the employing firm in state $s_{n,t}(e)$, $\varphi(d, C(d, s_{n,t}(e)), s_{n,t}(e))$, is the expected output at the second-best firm (net of productivity shocks), $y(C(d, s_{n,t}(e)), s_{n,t}(e))$, plus the compensating differential, $\Psi(d, C(d, s_{n,t}(e)), s_{n,t}(e))$. Therefore, to compute $\Psi(d, C(d, s_{n,t}(e)), s_{n,t}(e))$ from knowledge of $\varphi(\cdot)$ and $y(\cdot)$, we must know the identity of the second-best firm $C(d, s_{n,t}(e))$ so that we subtract the "correct" $y(C(d, s_{n,t}(e)), s_{n,t}(e))$ from $\varphi(d, C(d, s_{n,t}(e)), s_{n,t}(e))$. This is summarized in the next proposition.

Proposition 8 (Compensating Differential)**.** *Under Assumptions 1 to 13, the compensating differential* $\Psi(d, C(d, s), s)$ *is identified for each* $(d, e, s) \in \mathcal{D} \times \mathcal{E} \times \mathcal{S}_t^0(d, e)$ *and* $t \geq 1$ *provided that the identity of the second-best firm,* $C(d, s) := \mathcal{D}(s) \setminus \{d\}$*, is known.*

4.5 Discussion: Longitudinal vs. Cross-Sectional Dimension

In this section, we provide a high-level overview of our identification strategy, focusing on how it leverages both the longitudinal and cross-sectional dimensions of the data. First, we use the longitudinal dimension to identify the information technology, the law of motion of the state, the CCPs, and the distribution of $(D_{n,t}, w_{n,t})$ conditional on $s_{n,t}$. Indeed, these primitives are identified by concatenating the wage mixture weights across periods. Next, based on knowledge of the distribution of $(D_{n,t}, w_{n,t})$ given $s_{n,t}$, we identify the deterministic wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$ by applying DM's approach in each period, thereby drawing on the cross-sectional dimension. Because $\varphi(\cdot)$ is left nonparametrically specified and depends on $s_{n,t}$—whose support $\mathcal{S}_t$ may vary across periods due to $\kappa_{n,t}$—$\varphi(\cdot)$ is effectively a time-varying function and must therefore be identified in each period, making the longitudinal dimension less helpful. In the empirical application of Section 5, we parameterise $\varphi(\cdot)$ with time-invariant parameters and show how to leverage the longitudinal dimension to avoid the location normalisation of Assumption 8 when identifying the deterministic wage component. Finally, we once again leverage the dynamic dimension of the model to identify the output (and human capital) technology $y(\cdot)$, and in turn, the compensating differential $\Psi(\cdot)$.

A strength of our identification approach is its limited reliance on workers' mobility across jobs over time. However, a certain degree of heterogeneous variation in job choices—akin to job mobility—facilitates the identification of the output (and human capital) technology $y(\cdot)$ and the compensating differential $\Psi(\cdot)$. Regarding $y(\cdot)$, recall that the standard normalizations imposed in Proposition 7, which require knowing the value taken by the function $y(\cdot)$ at one firm for each state, lead to nontrivial identification if, at the same state, workers can choose to be employed at least two different firms with strictly positive probability, as also explained in Section 4.4. As for $\Psi(\cdot)$, remember that for a given state realization $s \in \mathcal{S}_t^0(e)$, the compensating differential of firm d with respect to firm d', $\Psi(d, d', s)$, is obtained by subtracting the expected output $y(d', s)$—identified from observing d' as the first-best firm at state s—from the deterministic wage component $\varphi(d, d', s)$—identified from observing d' as the second-best firm at state s. Therefore, to identify $\Psi(d, d', s)$, d' must possibly be both a first-best and second-best choice for worker n in state s with a strictly positive probability.

Although job mobility plays a limited role, on the other hand, we emphasize that we rely on a worker's job retention for at least three periods to identify the information technology, which in turn is key to pinning down all the other primitives, including the law of motion of the state, the CCPs, the deterministic wage component, the distribution of productivity shocks, the output (and human capital) technology, and the compensating differential.

5 Empirics: The Impact of Sorting on Earnings Inequality

In this section, we use our class of models and econometric approach to empirically measure how sorting between workers and firms affects earnings inequality in the U.S. The most commonly used framework for this question is that of AKM, which decomposes wages into worker and firm fixed effects, observable covariates, and random shocks. From these estimates, the wage variance is partitioned into contributions from worker effects, firm effects, the covariance between them, and the residual. The impact of sorting on earnings inequality is then gauged by the fraction of total wage variance attributable to the covariance between worker and firm effects. Empirical applications of this framework often point to a negligible role for sorting, reflecting weak correlations between worker and firm effects.

Building on the theoretical insights from the class of models studied, we argue that the AKM estimates of the correlation between firm and worker effects may be understated because two key forces are omitted. First, the *compensating differential* can dampen the direct impact of worker and firm characteristics on wages, because it compensates a worker for the missed future returns in human capital and information that would have been gained by accepting competing firms' offers. Second, *endogenous matching frictions*, namely the worker's acquisition of human capital and the gradual resolution of uncertainty about ability, may prevent high-type workers from joining the most productive firms immediately. For instance, workers might temporarily choose less-productive firms that offer valuable training or learning opportunities, challenging the assumption that they always sort into the most immediately productive match.

To empirically validate these conjectures coming from the theory of our class of models, we provide both simulation-based evidence and empirical evidence.

Monte Carlo Simulation. We simulate an economy based on a data-generating process (DGP) that captures the main features of our class of models, while introducing a few simplifications to facilitate direct comparison with the AKM framework. Specifically, we remove the wage equation's dependence on the second-best firm and assume away selection on $\epsilon_{n,t}$, since neither is present in AKM. Under these simplifications, workers' wages follow Equation (9), parameterized as:

$$\begin{aligned} w_{n,t} \; = & \sum_{d\in\mathcal{D}}\sum_{e\in\mathcal{E}} \mathbb{1}\{D_{n,t}=d,\, e_n=e\} \times \Big[e+\beta_0(d)+\beta_1(d,e)\,H_{n,1}+\beta_2(d,e)\,\kappa_{n,t}+\beta_3(d,e)\,P_{n,t} \\ & + \Psi\big(H_{n,1},\, \kappa_{n,t},\, P_{n,t};\, \psi(d,e)\big)+\epsilon_{n,t}(d,e)\Big], \end{aligned} \tag{38}$$

where the output technology $y(\cdot)$ consists of an AKM-style sum of worker and firm effects, $e+\beta_0(d)$, plus first-order terms in $H_{n,1}, \kappa_{n,t}, P_{n,t}$ governed by the parameters $\beta_1(d,e)$, $\beta_2(d,e)$, and $\beta_3(d,e)$.The compensating differential $\Psi(\cdot)$ is approximated by a truncated Taylor expansion, which includes higher-order and interaction (cross) terms in $H_{n,1}, \kappa_{n,t}$, and $P_{n,t}$, governed by the parameters $\psi(d,e)$. $H_{n,1}$ consists of gender and education, while $\kappa_{n,t}$ incorporates age. We calibrate the wage parameters and other simulation features to match key earnings moments from PSID, a representative survey of U.S. households dating back to 1968. These moments encompass wage growth, life-cycle patterns (both first and higher-order), inequality, and concentration. We also include as targets the AKM-type moments from Song et al. (2019), which derive from SSA. This calibration ensures that our simulated economy reflects both the broader U.S. earnings distribution and the key features highlighted by the AKM framework. Additional details are provided in Appendix D.

As mentioned, the literature centered on the AKM framework measures the impact of sorting on earnings inequality based on firm and worker (linear) complementarities in the output technology $y(\cdot)$. Correspondingly, this literature focuses on the fraction of the total wage variance attributable to the covariance between worker and firm effects or, by the notation of the wage equation in (38),

$$\rho := \frac{\text{Cov}(e_n, \beta_0(D_{n,t}))}{\text{Var}(w_{n,t})}.$$

Assuming that the econometrician has access to a short panel of data on $w_{n,t}$, $H_{n,1}$, $\kappa_{n,t}$, and $P_{n,t}$ from the simulated economy—for simplicity, we assume that beliefs about workers' ability are observed—the AKM estimate of ρ, denoted by $\hat{\rho}_{\text{AKM}}$, is obtained by estimating the wage equation

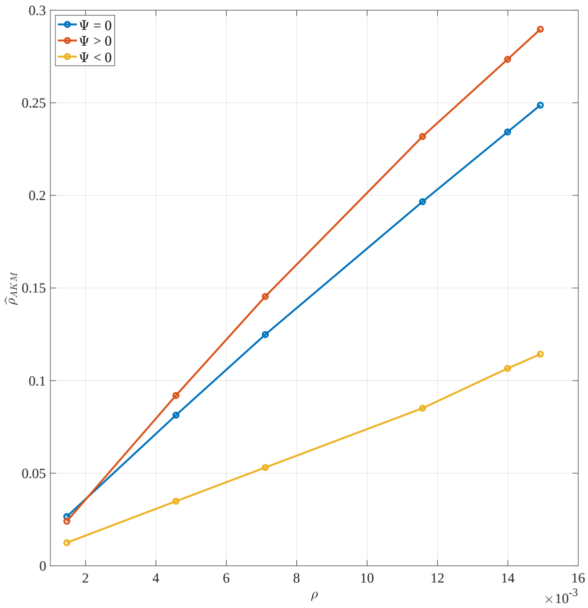
$$w_{n,t} = \sum_{d \in \mathcal{D}} \sum_{e \in \mathcal{E}} \mathbb{1}\{D_{n,t} = d, e_n = e\} \Big[e + \beta_0(d) + \beta_1 H_{n,1} + \beta_2 \kappa_{n,t} + \beta_3 P_{n,t} + \epsilon_{n,t}(d,e) \Big], \quad (39)$$

where the compensating differential $\Psi(\cdot)$ is omitted and the parameters β_1, β_2, and β_3 are assumed to be invariant across (d, e). Our findings suggest that when $\Psi(\cdot)$ is negative—implying that workers match with firms offering human capital and information gains with future returns higher than competitors—$\hat{\rho}_{\text{AKM}}$ underestimates ρ because the omitted $\Psi(\cdot)$ attenuates the firm and worker complementarities in output technology $y(\cdot)$. Conversely, when $\Psi(\cdot)$ is positive—implying that workers match with firms offering human capital and information gains with future returns lower than competitors—$\hat{\rho}_{\text{AKM}}$ overestimates ρ because the omitted $\Psi(\cdot)$ enhances the firm and worker complementarities in output technology $y(\cdot)$.

Figure 1 illustrates these patterns. On the horizontal axis, we plot the increasing values of ρ used to generate our data, while on the vertical axis we report the corresponding AKM estimates. The blue line corresponds to the case $\Psi(\cdot) = 0$ in the true DGP—though it does not perfectly coincide with the 45-degree line because, even though $\Psi(\cdot) = 0$ in the true DGP, the parameters $\beta_1(d, e)$, $\beta_2(d, e)$, and $\beta_3(d, e)$ still vary across (d, e), whereas AKM incorrectly treats them as invariant across (d, e). The red line shows results whenworkers $\Psi(\cdot) > 0$ (leading to an upward bias), and the yellow line shows results when $\Psi(\cdot) < 0$ (leading to a downward bias).[17]

[17]Throughout our analysis, we adjust standard AKM estimates for low-mobility bias, following the methodology of

Figure 1: Comparison of True vs. AKM Estimates of ρ



Empirical Application. We estimate the wage Equation (38) using U.S. employer-employee match data, namely LEHD data. This rich dataset provides quarterly earnings for all workers across 21 states—including California, Florida, and Pennsylvania—from the mid-1990s to 2022. We directly observe each worker's current firm, wage, gender, education, and age. Performance measures—in the model's notation, signal $a_{n,t}$—are not directly observed and we built a procedure to infer them from workers' variable pay. The idea is that the quantiles of the variable pay distribution identify performance measures to the extent that variable pay is monotonic in performance. Based on these extracted performance measures, we are able to estimate $P_{n,t}$ for each worker n and period t and so treat $P_{n,t}$ as a "covariate" in the subsequent wage estimation step. Note that this construction of $P_{n,t}$ is not mandatory in our more general identification framework, where we show how to identify the distribution of $P_{n,t}$ from the wage mixture. Additional details on the procedure used to construct $P_{n,t}$ are provided in Appendix E. As in the simulations, we omit the wage equation's dependence on the second-best firm to maintain closer alignment with the AKM framework. Unlike in the simulations, however, we now allow for selection on $\epsilon_{n,t}$, consistent with our original class of models.

To estimate the wage Equation (38), we assume that $\epsilon_{n,t}(d, e)$ is normally distributed conditional on $e_n = e$. Consequently, if there was no selection on $\epsilon_{n,t}$, the distribution of $w_{n,t}$ conditional on $(D_{n,t}, H_{n,1}, \kappa_{n,t}, P_{n,t})$ would be a finite (because $\mathcal{E}$ is finite) mixture of Normal distributions. In that case, we could simply use the Stata command `fmm` to estimate the wage parameters, since it

Bonhomme et al. (2023).

combines the estimation of Normal mixtures with OLS. On the other hand, if there were no worker efficiency types e_n, we could rely on the Stata package `eqregsel` to run extremal quantile regressions that account for selection on $\epsilon_{n,t}$ (D'Haultfoeuille et al., 2018, 2020). To accommodate both aspects simultaneously, we have created a Stata package integrating `fmm` with `eqregsel`.[18]

Our empirical results (currently awaiting approval from the Census Bureau before disclosure) corroborate the findings from our simulations. In particular, we see that the AKM estimates of ρ based on the wage Equation (39) fall below our own estimates of this parameter. This is because the estimated $\Psi(\cdot)$ is, on average, negative, suggesting that workers tend to match primarily with firms offering human capital and informational gains associated with high future returns. This finding indicates that we have resolved the puzzle of low sorting.

Next, we support this key finding with an exercise designed to capture global sorting in our rich class of models. By construction, ρ measures sorting exclusively with respect to the worker time-invariant efficiency type e_n. However, in our setting, workers may also sort on their beliefs about ability θ_n and on accumulated human capital (endogenous matching frictions). To capture these additional sorting dimensions, we perform a random reallocation exercise, comparing the observed earnings distribution to a counterfactual scenario in which workers and firms are matched at random, eliminating any endogenous links. If sorting indeed has a substantial impact, then disrupting these links should markedly reduce both earnings dispersion and the concentration of high earnings, since workers would no longer cluster in the firms offering the greatest productivity or the most valuable human-capital and informational benefits. Our preliminary evidence supports all of these mechanisms.

6 Conclusion

In this paper, we examine the empirical content of a generalized class of matching models of the labor market that incorporate learning about ability and human capital acquisition. These models have been widely applied to study worker turnover, occupational choice, wage differentials across occupations and industries, and worker career paths. We provide a novel argument establishing that these models are identified under intuitive conditions. Our approach relies solely on information from

[18] With the parameterization of Equation (38), we can eliminate the need for the location normalization imposed by Assumption 8 by exploiting the longitudinal dimension of the data. Without a longitudinal dimension, the `eqregsel` package would require that the intercept $e + \beta_0(d)$ be normalized to zero to satisfy Assumption 8, thereby precluding its estimation. By leveraging repeated observations over time, however, we can identify $e + \beta_0(d)$ and further disentangle e from $\beta_0(d)$ through the standard AKM normalization based on connectivity.

job choices and wages, without requiring additional variables that could facilitate identification of the learning process (e.g., proxies for beliefs or direct measures of performance signals). Moreover, we do not impose restrictions on endogenous variables or on the dynamics of states, choices, and outcomes. Instead, our identification arguments rest on conditions that allow for arbitrary patterns of selection based on endogenously time-varying unobservables, are easy to verify, impose minimal data requirements, and yield a constructive estimator, as implemented in our empirical application. Using this framework, we revisit an outstanding puzzle regarding the role of worker sorting in wage inequality. Specifically, we demonstrate that ignoring the dynamics of the matching process driven by learning about ability—and the resulting compensating differentials when firms differ in the learning and human capital acquisition opportunities they offer—can lead to a systematic underestimation of the importance of worker sorting for wage inequality.

References

ABOWD, J. M., F. KRAMARZ, P. LENGERMANN, AND S. PÉREZ-DUARTE (2004): "High Wage Workers and High Wage Firms," *Working Paper*.

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

AHN, H. AND J. L. POWELL (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

ALLMAN, E. S., C. MATIAS, AND J. A. RHODES (2009): "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *The Annals of Statistics*, 37, 3099–3132.

AN, Y., Y. HU, AND M. SHUM (2013): "Identifiability and Inference of Hidden Markov Models," *Working Paper*.

ANDREWS, M. J., L. GILL, T. SCHANK, AND R. UPWARD (2008): "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society A*, 171, 673–697.

——— (2012): "High Wage Workers Match with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias," *Economics Letters*, 117, 824–827.

ANTONOVICS, K. AND L. GOLAN (2012): "Experimentation and Job Choice," *Journal of Labor Economics*, 30, 333–366.

ARAGAM, B., C. DAN, E. P. XING, AND P. RAVIKUMAR (2020): "Identifiability of Nonparametric Mixture Models and Bayes Optimal Clustering," *The Annals of Statistics*, 48, 2277–2302.

BAGGER, J., F. FONTAINE, F. POSTEL-VINAY, AND J.-M. ROBIN (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104, 1551–1596.

BECKER, G. S. (1975): "Human Capital : A Theoretical and Empirical Analysis, with Special Reference to Education, 2nd Edition," *New York: National Bureau of Economic Research*.

BEN-PORATH, Y. (1967): “The Production of Human Capital and the Life Cycle of Earnings,” *Journal of Political Economy*, 75, 352–365.

BERGEMANN, D. AND J. VÄLIMAKI (1996): “Learning and Strategic Pricing,” *Econometrica*, 64, 1125–1149.

BERRY, S. T. AND G. COMPIANI (2023): “An Instrumental Variable Approach to Dynamic Models,” *The Review of Economic Studies*, 90, 1724–1758.

BLISCHKE, W. R. (1964): “Estimating the Parameters of Mixtures of Binomial Distributions,” *Journal of the American Statistical Association*, 59, 510–528.

——— (1978): “Mixtures of Distributions,” *International Encyclopedia of Statistics*, 1, 174–180.

BONHOMME, S., K. HOLZHEU, T. LAMADON, E. MANRESA, M. MOGSTAD, AND S. BRADLEY (2023): “How Much Should We Trust Estimates of Firm Effects and Worker Sorting?” *Journal of Labor Economics*, 41, 291–322.

BONHOMME, S., K. JOCHMANS, AND J.-M. ROBIN (2016a): “Estimating Multivariate Latent-Structure Models,” *The Annals of Statistics*, 44, 540–563.

——— (2016b): “Non-Parametric Estimation of Finite Mixtures from Repeated Measurements,” *Journal of the Royal Statistical Society*, 78, 211–229.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2019): “A Distributional Framework for Matched Employer Employee Data,” *Econometrica*, 87, 699–739.

BRUNI, C. AND G. KOCH (1985): “Identifiability of Continuous Mixtures of Unknown Gaussian Distributions,” *The Annals of Probability*, 13, 1341–1357.

BUCHINSKY, M., D. FOUGÉRE, F. KRAMARZ, AND R. TCHERNIS (2010): “Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States,” *The Review of Economic Studies*, 77, 972–1001.

BUNTING, J., P. DIEGERT, AND A. MAUREL (2024): “Heterogeneity, Uncertainty and Learning: Semiparametric Identification and Estimation,” *NBER Working Paper 32164*.

CARD, D., A. R. CARDOSO, J. HEINING, AND P. KLINE (2018): “Firms and Labor Market Inequality: Evidence and Some Theory,” *Journal of Labor Economics*, 36, S13 – S70.

CARD, D., J. HEINING, AND P. KLINE (2013): “Workplace heterogeneity and the rise of West German wage inequality,” *The Quarterly journal of economics*, 128, 967–1015.

CHAMBERLAIN, G. (1986): “Asymptotic Efficiency in Semi-Parametric Models with Censoring,” *Journal of Econometrics*, 32, 189–218.

COMPIANI, G. AND Y. KITAMURA (2016): “Using Mixtures in Econometric Models: A Brief Review and Some New Results,” *Econometrics Journal*, 19, C95–C127.

CUNHA, F. AND J. J. HECKMAN (2008): “Formulating, Identifying, and Estimating the Technology of cognitive and noncognitive skill formation,” *Journal of Human Resources*, 43, 738–782.

DAS, M., W. K. NEWEY, AND F. VELLA (2003): “Nonparametric Estimation of Sample Selection Models,” *The Review of Economic Studies*, 70, 33–58.

D’HAULTFOEUILLE, X. AND A. MAUREL (2013): “Another Look at the Identification au Infinity of Sample Selection Models,” *Econometric Theory*, 29, 213–224.

D'HAULTFOEUILLE, X., A. MAUREL, X. QIU, AND Y. ZHANG (2020): "Estimating selection models without an instrument with Stata," *The Stata Journal*, 20, 297–308.

D'HAULTFOEUILLE, X., A. MAUREL, AND Y. ZHANG (2018): "Extremal Quantile Regressions for Selection Models and the Black–White Wage Gap," *Journal of Econometrics*, 203, 129–142.

FARBER, H. S. AND R. GIBBONS (19996): "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 111, 1007–1047.

FLINN, C. J. (1986): "Wages and Job Mobility of Young Workers," *Journal of Political Economy*, 94, S88–S110.

FRENCH, E. AND C. TABER (2011): "Identification of Models of the Labor Market," In *Handbook of Labor Economics*, 4, Part A, O. Ashenfelter and D. Card, Eds., North–Holland, 537–617.

FREYBERGER, J. (2018): "Non-parametric Panel Data Models with Interactive Fixed Effects," *Review of Economic Studies*, 85, 1824–1851.

GIBBONS, R., L. KATZ, T. LEMIEUX, AND D. PARENT (2005): "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 23, 681–723.

GIBBONS, R. AND M. WALDMAN (1999a): "Careers in Organizations: Theory and Evidence," In *Handbook of Labor Economics*, Volume IIIB, O. Ashenfelter and D. Card, Eds., Amsterdam, North–Holland, 2373–2437.

——— (1999b): "A Theory of Wage and Promotion Dynamics inside Firms," *Quarterly Journal of Economics*, 114, 1321–1358.

——— (2006): "Enriching a Theory of Wage and Promotion Dynamics inside Firms," *Journal of Labor Economics*, 24, 59–107.

GOSH, J. AND V. RAMAMOORTHI (2003): *Bayesian Nonparametrics*, Springer.

HAGEDORN, M., T. H. LAW, AND I. MANOVSKII (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85, 29–65.

HALL, P. AND X.-H. ZHOU (2003): "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *The Annals of Statistics*, 31, 201–224.

HECKMAN, J. J. (1976): "A Life-Cycle Model of Earnings, Learning, and Consumption," *Journal of Political Economy*, 84, S9–S44.

——— (1990): "Varieties of Selection Bias," *The American Economic Review, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association*, 80, 313–318.

HECKMAN, J. J. AND B. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.

HENRY, M., Y. KITAMURA, AND B. SALANIÉ (2014): "Partial Identification of Finite Mixtures in Econometric Models," *Quantitative Economics*, 5, 123–144.

HIGGINS, A. AND K. JOCHMANS (2023): "Identification of mixtures of dynamic discrete choices," *Journal of Econometrics*, 237, 105462.

——— (2024): "Learning Markov Processes with Latent Variables," *Forthcoming at Econometric Theory*.

HU, Y. AND M. SHUM (2012): “Nonparametric Identification of Dynamic Models with Unobserved State Variables,” *Journal of Econometrics*, 171, 32–44.

HU, Y., M. SHUM, M. TAN, AND R. XIAO (2015): “A Simple Estimator for Dynamic Models with Serially Correlated Unobservables,” *CAEPR Working Paper 2015-017*.

JOCHMANS, K., M. HENRY, AND B. SALANIÉ (2017): “Inference on Two-Component Mixtures under Tail Restrictions,” *Econometric Theory*, 33, 610–635.

JOVANOVIC, B. (1979): “Job Matching and the Theory of Turnover,” *Journal of Political Economy*, 87, 972–990.

JOVANOVIC, B. AND Y. NYARKO (1997): “Stepping-Stone Mobility,” *Carnegie-Rochester Confer- ence Series on Public Policy*, 46, 289–325.

KAHN, L. B. AND F. LANGE (2014): “Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures,” *Review of Economic Studies*, 81, 1575–1613.

KALOUPTSIDI, M., Y. KITAMURA, L. LIMA, AND E. SOUZA-RODRIGUES (2024): “Counterfactual Analysis for Structural Dynamic Discrete Choice Models,” *Forthcoming at The Review of Economic Studies*.

KALOUPTSIDI, M., P. T. SCOTT, AND E. SOUZA-RODRIGUES (2021): “Identification of Counterfactuals in Dynamic Discrete Choice Models,” *Quantitative Economics*, 12, 351–402.

KASAHARA, H. AND K. SHIMOTSU (2009): “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices,” *Econometrica*, 77, 135–175.

KEANE, M. P., A. CHING, AND T. ERDEM (2017): “Empirical models of learning dynamics: A survey of recent developments,” In *Handbook of Marketing Decision Models*, Berend Wierenga and Ralf van der Lans, Eds., New York City, Springer, 223–257.

KEANE, M. P. AND K. I. WOLPIN (1997): “The Career Decisions of Young Men,” *Journal of Political Economy*, 105, 473–522.

KIM, D. AND Y. J. LEE (2025): “Point-identifying Semiparametric Sample Selection Models with No Excluded Variable,” *Working Paper CWP07/25*.

KLINE, P., R. SAGGIO, AND M. SØLVSTEN (2020): “Leave-out Estimation of Variance Components,” *Econometrica*, 88, 1859–1898.

LAMADON, T., C. MEGHIR, AND J.-M. ROBIN (2024): “Labor Market Matching, Wages, and Amenities,” *Working Paper*.

LANGE, F. (2007): “The Speed of Employer Learning,” *Journal of Labor Economics*, 25, 1–35.

LEE, S. AND A. LEWBEL (2013): “Nonparametric Identification of Accelerated Failure Time Competing Risks Models,” *Econometric Theory*, 29, 905–919.

MAGNAC, T. AND D. THESMAR (2002): “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 70, 801–816.

MILLER, R. A. (1984): “Job Matching and Occupational Choice,” *Journal of Political Economy*, 92, 1086–1120.

MINCER, J. A. (1958): “Investment in Human Capital and Personal Income Distribution,” *Journal of Political Economy*, 66, 281–302.

——— (1974): "Schooling, Experience, and Earnings," *NBER Books*.

NAGYPÁL, E. (2007): "Learning by Doing vs. Learning about Match Quality: Can We Tell Them Apart?" *Review of Economic Studies*, 74, 537–566.

NEWEY, W. K. (2009): "Two-Step Series Estimation of Sample Selection Models," *Econometrics Journal*, 12, S217–S229.

NORETS, A. (2010): "Approximation of Conditional Densities by Smooth Mixtures of Regressions," *Annals of Statistics*, 38, 1733–1766.

NORETS, A. AND J. PELENIS (2014): "Posterior Consistency in Conditional Density Estimation by Covariate Dependent Mixtures," *Econometric Theory*, 30, 606–646.

PAPAGEORGIOU, T. (2014): "Learning Your Comparative Advantages," *Review of Economic Studies*, 81, 1263–1295.

——— (2018): "Large Firms and Within Firm Occupational Reallocation," *Journal of Economic The- ory*, 174, 184–223.

PASTORINO, E. (2024): "Careers in Firms: The Role of Learning about Ability and Human Capital Acquisition," *Journal of Political Economy*, 132, 1994–2073.

RUBINSTEIN, Y. AND Y. WEISS (2006): "Post Schooling Wage Growth: Investment, Search and Learning," In *Handbook of the Economics of Education*, Volume I, Eric Hanushek and Finis Welch, Eds., Amsterdam, North–Holland, 1–67.

SELTEN, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *Int Journal of Game Theory*, 4, 25–55.

SHIU, J.-L. AND Y. HU (2013): "Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates," *Journal of Econometrics*, 175, 116–131.

SONG, J., D. J. PRICE, F. GUVENEN, N. BLOOM, AND T. VON WACHTER (2019): "Firming up inequality," *The Quarterly journal of economics*, 134, 1–50.

TABER, C. (2000): "Semiparametric Identification and Heterogeneity in Discrete Choice Dynamic Programming Models," *Journal of Econometrics*, 96, 201–229.

TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

TEICHER, H. (1961): "Identifiability of Mixtures," *Ann. Math. Statist.*, 32, 244–248.

——— (1963): "Identifiability of Finite Mixtures," *Ann. Math. Statist.*, 34, 1265–1269.

A Extensions of Identification Argument

We discuss here extensions of our identification framework.

Support $\mathcal{E}$ of Efficiency e_n. To identify the wage mixture (36), Assumption 3(ii) requires that $\mathcal{E}$ is finite with a known cardinality. This result can be readily extended to the case where the econometrician knows only an upper bound E^* on the cardinality of $\mathcal{E}$, since the identification result of Bruni and Koch (1985) accommodates the possibility of some mixture weights being zero. See their Equation (4.9) in Section 4.c where the mixture weights α_i are only required to be ≥ 0.

If e_n is continuous (and potentially multidimensional), then the wage mixture (36) becomes a *continuous* mixture of continuous Gaussian mixtures (instead of a *finite* mixture of continuous Gaussian mixtures), making identification more challenging. This impasse can be easily resolved by assuming away selection on $\epsilon_{n,t}$. In that case, Assumption 3 can be replaced by requiring that the *unconditional* distribution of $\epsilon_{n,t}$ is Normal. Since $D_{n,t}$ is now independent of $\epsilon_{n,t}$, the distribution of $\epsilon_{n,t}$ conditional on $D_{n,t}$ equals its unconditional distribution and is therefore also Normal. Under this simplification, the wage mixture (36) is a continuous mixture of Normals, whose identification is established by Bruni and Koch (1985)'s Theorem 1. The same theorem also allows e to be multi-dimensional.

Support $\mathcal{A}$ of Signal $a_{n,t}$. As for $\mathcal{E}$, Proposition 2 can also be adapted to cases where the econometrician only knows an upper bound A^* on the cardinality of $\mathcal{A}$, as well as to scenarios where $a_{n,t}$ is continuous (and potentially multidimensional) and there is no selection on $\epsilon_{n,t}$.

To identify the learning process in Proposition 3, Assumption 4 (i) imposes that $\mathcal{A}$ has a cardinality of two. As explained in Section 4.2, this restriction enables us to represent the signal distribution as a *binomial* mixture over the unobserved ability θ_n, which is identified based on Blischke (1964, 1978). This assumption can be extended to include other cardinalities and potentially continuous/multidimensional $a_{n,t}$, provided that the signal distribution remains an identifiable mixture. For instance, if $a_{n,t}$ is distributed as a continuous and multivariate Gaussian mixture conditional on θ_n, then the signal distribution would then be a finite mixture of continuous and multivariate Gaussian mixtures (finite because Θ is finite), which remains identifiable according to Bruni and Koch (1985), as discussed in their Section 4.9.

Support Θ of Ability $\theta_{n,t}$. To identify the learning process in Proposition 3, Assumption 4(i) requires that Θ has cardinality two. This restriction allows us to model the signal distribution as a binomial mixture over the unobserved ability θ_n with *two* components. The binomial aspect arises because $\mathcal{A}$ has cardinality two, and the two components of this binomial mixture correspond to the cardinality of Θ. This mixture is identifiable, as shown by Blischke (1964) and Blischke (1978), provided that the number of periods where workers are observed at each given job d is at least $2r - 1 = 3$, where $r = |\Theta| = 2$ represents the number of mixture components (see Appendix C for more details). Keeping $\mathcal{A}$ of cardinality two, Assumption 4(i) can be extended to any finite Θ, requiring an increase in the number of observation periods to meet the new lower bound $2r - 1$. Going beyond the finite case, if both θ_n and a_n are continuous and multidimensional, and $a_{n,t}$ follows a multivariate Normal distribution conditional on θ_n, then the signal distribution is a continuous mixture of multivariate Normals, identified by Bruni and Koch (1985)'s Theorem 1.

Dispensing with Assumption 2. Our identification results rely on Assumption 2, under which, in each period t, worker n receives wage offers from only two firms. The identities of these firms are deterministically dependent on $s_{n,t}$ and independent of $\epsilon_{n,t}$. It is possible to entirely dispense with this assumption and nonparametrically identify the workers' acceptance strategy, which in turn, pinpoints the identity of the second-best firm. In practice, we operationalize this identification through the use of workers' observed transition patterns.

From Imperfect Competition to Search Frictions. DM's methods extend beyond our class of imperfectly competitive models to encompass standard search frameworks as well. For example, consider a standard wage equation of search models inspired by the output technology of Bagger et al. (2014). For a given worker n and firm d pair, and following our notation where possible, the wage is:

$$w_{n,t}(d) = \omega \gamma_{d,t}^{\alpha} H_{n,1}^{\beta} \varepsilon_{n,t}(d) + (1 - \omega)(1 - \delta) U(H_{n,1}),$$

where $H_{n,1}$ is human capital (assumed fixed over time for simplicity), ω is the worker's bargaining weight, $\gamma_{d,t}$ is the firm productivity, α and β are parameters of interest, and

$$U(H_{n,1}) = z + \delta \mathbb{E}(f(S(\cdot); \omega, \alpha, \beta, g_{\epsilon}))$$

is the value of unemployment, with $S(\cdot)$ denoting the match surplus function, g_{ϵ} the distribution of $\varepsilon_{n,t}(d)$, and $f(\cdot)$ a known function of $S(\cdot)$ which depends on $(\omega, \alpha, \beta, g_{\epsilon})$.

Suppose $\gamma_{d,t}$, ω, and g_ϵ are known. Then, the DM approach can be used to identify the scale term $\omega\gamma_{d,t}^{\alpha_d}H_{n,1}^{\beta_d}$ and the location term $(1-\omega)(1-\delta)U(H_{n,1})$. Knowing the scale term at two values of $H_{n,1}$ allows us to identify α_d and β_d. With the location term identified, we can recover $U(H_{n,1})$. Once α_d and β_d are known, we can compute $\delta\mathbb{E}(f(S(\cdot);\omega,\alpha_d,\beta_d,f_\epsilon))$ and finally identify $z = U(H_{n,1}) - \delta\mathbb{E}(f(S(\cdot);\omega,\alpha_d,\beta_d,g_\epsilon))$. We summarize this argument in the following proposition.

Proposition 9. *Suppose that ω and δ are known. Then, then the parameters α and β are identified up to a level normalization.*

Proof. Assume without loss of generality that $H_{n,1}$ can take the value one. Consider period t and firm d. Assume $\gamma_{d,t}^{\alpha} = c$ with c known and different from 1. Then, the scale term $\omega\gamma_{d,t}^{\alpha}H_{n,1}^{\beta}$ is known at $H_{n,1} = 1$. Also assume that the location term $(1-\omega)(1-\delta)U(H_{n,1})$ is known at $H_{n,1} = 1$. Consequently, the location and scale normalizations of DM are satisfied and, under the rest of DM's assumptions, we can identify α, β, $\gamma_{d,t}$, and $U(H_{n,1})$.

Next, assume that the support of $\varepsilon_{n,\tau}(d)$ is independent of d—namely, it is common across firms—and bounded. Then, comparing the largest wage of workers with the same $H_{n,1}$ employed by two different firms d, d' in period t—for one of which $\gamma_{d,t}^{\alpha}$ is known—identifies $\gamma_{d',t}^{\alpha}$, and thus $\gamma_{d',t}$, given that α has already been identified.

If $\gamma_{d,t} := \gamma_d$ for each firm d (firm fixed effect), we can stop here. Otherwise, we move to period $\tau \neq t$. Assume $\gamma_{d,\tau} = r$ with r known for some firm d. Then, comparing the largest wage of workers with the same $H_{n,1}$ employed by two different firms d, d' in period τ—for one of which $\gamma_{d,\tau}^{\alpha}$ is known—identifies $\gamma_{d',\tau}^{\alpha}$, and thus $\gamma_{d',\tau}$, given that α has already been identified. □

B Details on the Identification of the Wage Mixture

We provide here omitted details about the identification of the model.

B.1 Zero Mixture Weights

As highlighted in Section 4.1, to show Proposition 2, we do not impose that the mixture weights of Equation (36) are strictly positive for each $e \in \mathcal{E}$ and $(a_1, \ldots, a_{t-1}) \in \mathcal{A}^{t-1}$. In fact, these weights could be zero due to Assumption 2. Specifically, recall that in our framework, the vector of state variables $s_{n,\tau}$ is essentially a deterministic function of $(H_{n,1}, D_n^{\tau-1}, e_n, a_n^{\tau-1})$ for each period $\tau \leq t$. Furthermore, under Assumption 2, the choice set of worker n is a deterministic function of $s_{n,\tau}$ for each $\tau \leq t$. Therefore, the choice set of worker n is a deterministic function of $(H_{n,1}, D_n^{\tau-1}, e_n, a_n^{\tau-1})$ for each $\tau \leq t$. Critically, this choice set is also assumed to have a cardinality of two. This implies that the joint probability $\Pr(H_{n,1}, D_n^t, e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1}))$ may be zero if the employment history D_n^t is not compatible with $(H_{n,1}, e, a_1, \ldots, a_{t-1})$ because some of the jobs appearing in D_n^t are not in worker n's choice sets over some periods $\tau \leq t$.[19] In turn, the mixture weight $\Pr(e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1}) \mid H_{n,1}, D_n^t)$ would be zero as well. A fundamental advantage of the result from Bruni and Koch (1985) is that it allows some of the mixture weights to be zero and identifies which of them are indeed zero.

B.2 Labelling of the Mixture Components

Using the result from Bruni and Koch (1985), we identify the mixture Model (36) up to the labeling of mixture components. To resolve the labeling indeterminacy, it is sufficient to know an injective map between (e_n, a_n^{t-1}) and the pairs of mixture components and weights, $(e, a_1, \ldots, a_{t-1}) \in \mathcal{E} \times \mathcal{A}^{t-1} \mapsto \{\Pr(w_{n,t} \mid H_{n,1}, D_n^t, e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1})), \Pr(e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1})|$-$H_{n,1}, D_n^t)\}$. In our setting, this injective map can be constructed by examining the moments of the mixture components, for example, by using their variances to order them with respect to e_n and their means to order them with respect to a_n^{t-1}.

C Omitted Proofs

Proof of Proposition 3, Part (I). The proof is organised in three steps. In Step 1, we establish the identification of the conditional distribution of signal history, $\Pr(a_n^t \mid H_{n,1}, D_n^t, e_n)$. In Step 2, we represent $\Pr(a_n^t \mid H_{n,1}, D_n^t, e_n)$ as a binomial mixture over θ_n. In Step 3, we demonstrate the identification of $\{p_1(h, e), \alpha_{h,d,e}, \beta_{h,d,e}\}$ from the weights and components of such binomial mixture for each $(h, d, e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$.

Step 1: Identification of $\Pr(a_n^t \mid H_{n,1}, D_n^t, e_n)$. Let $t = 1$. Hence, $\Pr(a_n^t \mid H_{n,1}, D_n^t, e_n)$ simplifies to $\Pr(a_{n,1} \mid H_{n,1}, D_{n,1}, e_n)$. Let $(h, d_1, d_2) \in \mathcal{H} \times \mathcal{D}^2$ denote a realization of $(H_{n,1}, D_{n,1}, D_{n,2})$. Let $(e, a) \in \mathcal{E} \times \mathcal{A}$ denote a realization of $(e_n, a_{n,1})$. Decompose the joint probability $\Pr(H_{n,1} =$

[19] To clarify with an example, consider worker n with initial conditions $e_n = e$, $H_{n,1} = h$. At time $\tau = 1$, let $s_{n,1} = (h, p_1(h, e), e) := s'$. Assuming the choice set $\mathcal{D}(s') := \{f_1, f_2\}$, suppose worker n chooses firm f_1 and performs well, resulting in a high signal $a_{1,n}(f_1, e) = \bar{a}$. Subsequently, at $\tau = 2$, the state updates to $s_{n,2} = (h, \kappa(f_1), p_2^{\bar{a}}(h, f_1, e; p_1), e) := s''$, with a new choice set $\mathcal{D}(s'') = \{f_2, f_3\}$. Then, the probability $\Pr(H_{n,1} = h, D_n^2 = (f_3, f_1), e_n = e, a_{1,n}(f_3, e) = \bar{a}) = 0$ because f_3 is not in $\mathcal{D}(s')$ and f_2 is not in $\mathcal{D}(s'')$.

$h, D_{n,1} = d_1, D_{n,2} = d_2, e_n = e, a_{n,1} = a_1)$ as follows:

$$
\begin{aligned}
\Pr(H_{n,1} = h, &D_{n,1} = d_1, D_{n,2} = d_2, e_n = e, a_{n,1} = a_1) = \\
&\Pr(D_{n,2} = d_2 \mid H_{n,1} = h_1, D_{n,1} = d_1, e_n = e, a_{n,1} = a_1) \\
&\times \Pr(a_{n,1}(D_{n,1} e_n) = a_1 \mid H_{n,1} = h, D_{n,1} = d_1, e_n = e) \\
&\times \Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d_1) \\
&\times \Pr(H_{n,1} = h_1, D_{n,1} = h_1).
\end{aligned} \tag{40}
$$

The left-hand side of (40), $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2, e_n = e, a_{n,1} = a_1)$, is identified by Proposition 2 if $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2) > 0$, which can be verified from the data under Assumption 1. Specifically, by Proposition 2, the wage mixture weight at $t = 2$ is identified:

$$
\Pr(e_n = e, a_{1,n} = a_1 \mid H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2). \tag{41}
$$

From knowledge of $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2)$ under Assumption 1, we can then identify the joint probability $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2, e_n = e, a_{n,1} = a_1)$.

Similarly, $\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d_1)$ in the right-hand side of (40) is identified by Proposition 2 if $\Pr(H_{n,1} = h, D_{n,1} = d_1) > 0$, which can be verified from the data under Assumption 1. In fact, by Proposition 2, the wage mixture weight at $t = 1$ is identified:

$$
\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d_1). \tag{42}
$$

From knowledge of $\Pr(H_{n,1} = h, D_{n,1} = d)$ under Assumption 1, we can then identify the conditional probability $\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d)$.

Lastly, $\Pr(H_{n,1} = h, D_{n,1} = d)$ in the right-hand side of (40) is known from the data under Assumption 1. Hence, based on (40), the product

$$
\begin{aligned}
&\Pr(D_{n,2} = d_2 \mid H_{n,1} = h, D_{n,1} = d_1, e_n = e, a_{n,1} = a_1) \\
&\times \Pr(a_{n,1} = a_1 \mid H_{n,1} = h, D_{n,1} = d, e_n = e),
\end{aligned} \tag{43}
$$

is identified.

Write (40) for all realizations d_2 of $D_{n,2}$ such that $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2) > 0$, and identify the product (43) for all such d_2's. By summing up (43) across these d_2's, we obtain:

$$\begin{aligned}
&\Pr(a_{n,1} = a_1 \mid H_{n,1} = h, D_{n,1} = d, e_n = e) \\
&= \sum_{d_2 \in \mathcal{D}} \Pr(D_{n,2} = d_2 \mid H_{n,1} = h, D_{n,1} = d, e_n = e, a_{n,1} = a_1) \\
&\times \Pr(a_{n,1} = a_1 \mid H_{n,1} = h, D_{n,1} = d, e_n = e),
\end{aligned} \tag{44}$$

which is, therefore, identified.

Remark on mixture labelling: Note that, to perform the final summation step (44), we must have established a labelling for the wage mixture weights at $t = 1$ with respect to $e \in \mathcal{E}$ and at $t = 2$ with respect to $(e, a_1) \in \mathcal{E} \times \mathcal{A}$.

Remark on zero probabilities: By Proposition 2, the wage mixture weight (42) is identified if $\Pr(H_{n,1} = h, D_{n,1} = d_1) > 0$. Similarly, the wage mixture weight (41) is identified if $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2) > 0$. Hence, the focus is on $(h, d_1, d_2) \in \mathcal{H} \times \mathcal{D}^2$ such that $\Pr(H_{n,1} = h, D_{n,1} = d_1, D_{n,2} = d_2) > 0$. Moreover, the wage mixture weights (41) and (42), identified by Proposition 2, may be zero due to Assumption 2. If 41 is zero and 42 is strictly positive, then the product 43 is identified and, in particular, it is equal to zero. Conversely, if both 41 and 42 are zero, then 43 is not identified. Consequently, the probability $\Pr(a_{n,1} = a_1 \mid H_{n,1} = h, D_{n,1} = d, e_n = e)$ is identified for $e \in \mathcal{E}$ such that $\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d_1) > 0$.

To conclude: The arguments above can be applied for any $1 \leq t \leq T - 1$ to establish the identification of

$$\Pr(a_n^t = (a_1, \ldots, a_t) \mid H_{n,1} = h, D_n^t = (d_1, \ldots, d_t), e_n = e),$$

for all $(a_1, \ldots, a_t) \in \mathcal{A}^t$, $h \in \mathcal{H}$, $(d_1, \ldots, d_t) \in \mathcal{D}^t$, and $e \in \mathcal{E}$ such that $\Pr(H_{n,1} = h, D_n^t = (d_1, \ldots, d_t)) > 0$ and $\Pr(e_n = e \mid H_{n,1} = h, D_n^t = (d_1, \ldots, d_t)) > 0$. The latter strictly positive probability condition can be verified by summing over $(a_1, \ldots, a_{t-1}) \in \mathcal{A}^{t-1}$ the wage mixture weights in period t which are identified by Proposition 2:

$$\begin{aligned}
&\Pr(e_n = e \mid H_{n,1} = h, D_n^t = (d_1, \ldots, d_t)) \\
&= \sum_{(a_1, \ldots, a_{t-1}) \in \mathcal{A}^{t-1}} \Pr(a_n^{t-1} = (a_1, \ldots, a_{t-1}), e_n = e \mid H_{n,1} = h, D_n^t = (d_1, \ldots, d_t)).
\end{aligned}$$

<u>*Step 2: Representation of* $\Pr(a_n^t \mid H_{n,1}, D_n^t, e_n)$ *as a binomial mixture.*</u> Using Bayes rule and As-

sumptions 4 and 5 (ii), for any $a_1 \in \mathcal{A}$, we have that:

$$\begin{aligned}\Pr(a_{n,1} = a_1 \mid H_{n,1} = h, D_{n,1} = d, e_n = e) =& \alpha_{h,d,e}^{\mathbb{1}\{a_1=\bar{a}\}}(1-\alpha_{h,d,e})^{\mathbb{1}\{a_1=\underline{a}\}} p_1(h,e) \\ &+ \beta_{h,d,e}^{\mathbb{1}\{a_1=\bar{a}\}}(1-\beta_{h,d,e})^{\mathbb{1}\{a_1=\underline{a}\}}(1-p_1(h,e)).\end{aligned} \tag{45}$$

Equation (45) is a Bernoulli mixture with two components. Moving to $t > 1$ and using Assumption 5 (i), for any $(a_1, \ldots, a_t) \in \mathcal{A}^t$, we have that:

$$\begin{aligned}\Pr(a_n^t = (a_1, \ldots, a_t) \mid H_{n,1} = h, &D_{n,1} = \cdots = D_{n,t} = d, e_n = e) \\ &= \alpha_{h,d,e}^{\sum_{\tau=1}^t \mathbb{1}\{a_\tau=\bar{a}\}}(1-\alpha_{h,d,e})^{t-\sum_{\tau=1}^t \mathbb{1}\{a_\tau=\underline{a}\}} q(h,d,e) \\ &+ \beta_{h,d,e}^{\sum_{\tau=1}^t \mathbb{1}\{a_\tau=\bar{a}\}}(1-\beta_{h,d,e})^{t-\sum_{\tau=1}^t \mathbb{1}\{a_\tau=\underline{a}\}}(1-q(h,d,e)),\end{aligned} \tag{46}$$

where $q(h,d,e) := \Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, D_{n,1} = \cdots = D_{n,t} = d, e_n = e)$. Equation (46) is a binomial mixture with two components and t trials.

<u>*Step 3: Identification of* $p_1(h,e)$, $\alpha_{h,d,e}$, *and* $\beta_{h,d,e}$.</u> Let $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$. For any $(a_1, a_2, a_3) \in \mathcal{A}^3$, write (46) at $t = 3$:

$$\begin{aligned}\Pr(a_n^3 = (a_1, a_2, a_3) \mid H_{n,1} = h, &D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e) \\ &= \alpha_{h,d,e}^{\sum_{\tau=1}^3 \mathbb{1}\{a_\tau=\bar{a}\}}(1-\alpha_{h,d,e})^{3-\sum_{\tau=1}^3 \mathbb{1}\{a_\tau=\underline{a}\}} q(h,d,e) \\ &+ \beta_{h,d,e}^{\sum_{\tau=1}^3 \mathbb{1}\{a_\tau=\bar{a}\}}(1-\beta_{h,d,e})^{3-\sum_{\tau=1}^3 \mathbb{1}\{a_\tau=\underline{a}\}}(1-q(h,d,e)),\end{aligned} \tag{47}$$

The left-hand side of (47) is identified by Step 1 from combining the wage mixture weights in periods 3 and 4, under three conditions: (1) $\Pr(H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d) > 0$, which can be verified from the data under Assumption 1; (2) $\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d) > 0$, which can be verified from the wage mixture weights in period 3, as explained in Step 1; (3) $T \geq 4$, which is guaranteed by Assumption 7.

Following Blischke (1964, 1978), the weights and components of the binomial mixture (47), $\{\alpha_{h,d,e}, \beta_{h,d,e}, q(h,d,e)\}$, are identified under two conditions: (1) the number of trials is greater than or equal to $2r - 1$, where r is the number of the binomial mixture components; (2) the weights and components of the binomial mixture (47) are strictly positive and strictly less than one. Regarding condition (1), in our case, $r = 2$. Therefore, we need the observation of histories for at least $2r - 1 = 3$ periods, hence our focus on $t = 3$ in (47). Regarding condition (2), we must have that $\Pr(a_n^3 = (a_1, a_2, a_3) \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e) > 0$ for each $(a_1, a_2, a_3) \in \mathcal{A}^3$, which can be verified from Step 1 using the wage mixture weights at $t = 3$ and $t = 4$.

Under such conditions, $\alpha_{h,d,e}$ and $\beta_{h,d,e}$ are identified from the knowledge of $\{\Pr(a_n^3 = (a_1, a_2, a_3) \mid$

$H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e)\}_{(a_1,a_2,a_3)\in\mathcal{A}^3}$, up to labeling with respect to θ_n. Under Assumption 5 (iii), $\alpha_{h,d,e}$ and $\beta_{h,d,e}$ are identified without any labeling indeterminacy with respect to θ_n. In turn, $p_1(h, e)$ is identified from the Bernoulli mixture (45).

Conditions to identify $\alpha_{h,d,e}$ and $\beta_{h,d,e}$. Let $(h, d, e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$. To sum up, in Step 3, we impose the following conditions to identify $\alpha_{h,d,e}$ and $\beta_{h,d,e}$:

(i) $\Pr(H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d) > 0$.

(ii) $\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d) > 0$.

(iii) $\Pr(a_n^3 = (a_1, a_2, a_3) \mid H_{n,1} = h, D_{n,1} = D_{n,2} = D_{n,3} = d, e_n = e) > 0$ for each $(a_1, a_2, a_3) \in \mathcal{A}^3$.

Observe that conditions (i) to (iii) *cannot* hold for *every* firm $d \in \mathcal{D}$ during the same starting periods 1,2,3, due to Assumption 2. This is because firm d must appear in the choice set of worker n for "sufficiently enough" state realizations across periods 1,2,3 to satisfy conditions (i) to (iii)—particularly condition (iii), which is the most demanding—thus limiting the presence of other firms d' in such choice sets and, in turn, the identifiability of $\alpha_{h,d',e}$ and $\beta_{h,d',e}$.

To overcome this issue, we must impose that conditions (i) to (iii) hold for each $d \in \mathcal{D}$ at *some*, possibly (h, d, e)-specific, triplets of consecutive periods $\tau_{h,d,e}, \tau_{h,d,e} + 1, \tau_{h,d,e} + 2$:

(i) $\Pr(H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d) > 0$.

(ii) $\Pr(e_n = e \mid H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d) > 0$.

(iii) $\Pr(a_{n,\tau_{h,d,e}} = a_{\tau_{h,d,e}}, a_{n,\tau_{h,d,e}+1} = a_{\tau_{h,d,e}+1}, a_{n,\tau_{h,d,e}+2} = a_{\tau_{h,d,e}+2} \mid H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d, e_n = e) > 0$ for each $(a_{\tau_{h,d,e}}, a_{\tau_{h,d,e}+1}, a_{\tau_{h,d,e}+2}) \in \mathcal{A}^3$.

As is clear from Step 2, these triplets of consecutive periods $\tau_{h,d,e}, \tau_{h,d,e}+1, \tau_{h,d,e}+2$ must include the initial periods $1, 2, 3$ for some d in order to identify the prior $p_1(h, e)$.

These conditions are summarised in Assumption 6.

Verification of conditions (i) to (iii). Condition (i) can be verified from the data under Assumption 1.

Condition (ii) can be verified from the wage mixture weights at time $\tau_{h,d,e} + 2$ identified by

Proposition 2. Namely, observe that:

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d\big)$$

$$= \sum_{(d_1,\ldots,d_{\tau_{h,d,e}-1})\in\mathcal{D}^{\tau_{h,d,e}-1}} \sum_{(a_1,\ldots,a_{\tau_{h,d,e}+1})\in\mathcal{A}^{\tau_{h,d,e}+1}} \Big[$$

$$\Pr\big(a_n^{\tau_{h,d,e}+1} = (a_1,\ldots,a_{\tau_{h,d,e}+1}),\ e_n = e \mid H_{n,1} = h,\ D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}),$$

$$D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d\big)$$

$$\times \Pr\big(D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}) \mid H_{n,1} = h,\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d\big)\Big],$$

where $\Pr(a_n^{\tau_{h,d,e}+1} = (a_1,\ldots,a_{\tau_{h,d,e}+1}), e_n = e \mid H_{n,1} = h, D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}), D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d)$ is wage mixture weight at $\tau_{h,d,e}+2$ identified by Proposition 2 and $\Pr(D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}) \mid H_{n,1} = h, D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d)$ is known from the data under Assumption 1.

Condition (iii) can be verified from Step 1 based on the knowledge of the wage mixture weights at times $\tau_{h,d,e}+2$ and $\tau_{h,d,e}+3$, as identified by Proposition 2; hence, the reason for Assumption 7. Namely, observe that:

$$\Pr\big(a_{n,\tau_{h,d,e}} = a_{\tau_{h,d,e}},\ a_{n,\tau_{h,d,e}+1} = a_{\tau_{h,d,e}+1},\ a_{n,\tau_{h,d,e}+2} = a_{\tau_{h,d,e}+2}$$

$$\mid H_{n,1} = h,\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d,\ e_n = e\big)$$

$$= \sum_{(d_1,\ldots,d_{\tau_{h,d,e}-1})\in\mathcal{D}^{\tau_{h,d,e}-1}} \sum_{(a_1,\ldots,a_{\tau_{h,d,e}+1})\in\mathcal{A}^{\tau_{h,d,e}+1}} \Pr\big(a_n^{\tau_{h,d,e}-1} = (a_1,\ldots,a_{\tau_{h,d,e}-1}),\ a_{n,\tau_{h,d,e}} = a_{\tau_{h,d,e}},$$

$$a_{n,\tau_{h,d,e}+1} = a_{\tau_{h,d,e}+1},\ a_{n,\tau_{h,d,e}+2} = a_{\tau_{h,d,e}+2} \mid H_{n,1} = h,\ D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}),$$

$$D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d,\ e_n = e\big)$$

$$\times \Pr\big(D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}) \mid D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d,\ e_n = e\big),$$

where

$$\Pr(a_n^{\tau_{h,d,e}-1} = (a_1,\ldots,a_{\tau_{h,d,e}-1}), a_{n,\tau_{h,d,e}} = a_{\tau_{h,d,e}}, a_{n,\tau_{h,d,e}+1} = a_{\tau_{h,d,e}+1}, a_{n,\tau_{h,d,e}+2} = a_{\tau_{h,d,e}+2}$$

$$\mid H_{n,1} = h, D_n^{\tau_{h,d,e}-1} = (d_1,\ldots,d_{\tau_{h,d,e}-1}), D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d, e_n = e)$$

is known by Step 1 from the wage mixture weights at $\tau_{h,d,e}+2$ and $\tau_{h,d,e}+3$ identified by Proposition

2, and

$$
\begin{aligned}
&\Pr\big(D_n^{\tau_{h,d,e}-1} = (d_1, \ldots, d_{\tau_{h,d,e}-1}) \mid H_{n,1} = h,\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d,\ e_n = e\big) \\
&= \Bigg[\Bigg(\sum_{(a_1,\ldots,a_{\tau_{h,d,e}+1}) \in \mathcal{A}^{\tau_{h,d,e}+1}} \Pr\big(a_n^{\tau_{h,d,e}+1} = (a_1, \ldots, a_{\tau_{h,d,e}+1}),\ e_n = e \\
&\qquad \mid H_{n,1} = h,\ D_n^{\tau_{h,d,e}-1} = (d_1, \ldots, d_{\tau_{h,d,e}-1}),\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d\big)\Bigg) \\
&\quad \times \Pr\big(D_n^{\tau_{h,d,e}-1} = (d_1, \ldots, d_{\tau_{h,d,e}-1}) \mid H_{n,1} = h,\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d\big)\Bigg] \\
&\quad \Big/ \Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,\tau_{h,d,e}} = D_{n,\tau_{h,d,e}+1} = D_{n,\tau_{h,d,e}+2} = d\big).
\end{aligned}
$$

is known from the wage mixture weights at $\tau_{h,d,e} + 2$ identified by Proposition 2. □

Proof of Proposition 3, Part (II). Part (II) of Proposition 3 follows from part (I). Specifically, by combining $p_1(h, e)$ with $\{\alpha_{h,d,e}, \beta_{h,d,e}\}$, we can compute the belief $P_{n,t}$ for every possible realization of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ using Bayes' rule, as outlined in Equations (3) and (4) in Section 2. Given that the cardinality of $\mathcal{E}$ is known under Assumption 3, we have identified the support $\mathcal{S}_t$ of $s_{n,t}$ and, in particular, the map from realizations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realizations of $s_{n,t}$. Moving to the unconditional distribution of $s_{n,t}$, $\Pr(e_n, a_n^{t-1} \mid H_{n,1}, D_n^t)$ is identified from the wage mixture weights at time t by Proposition 2 for conditioning events $(H_{n,1}, D_n^t)$ having strictly positive probability. By combining this with the probability $\Pr(H_{n,1}, D_n^t)$ identified from the data, we identify the joint probability $\Pr(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$. Knowing the map from realizations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realizations of $s_{n,t}$, we identify the unconditional distribution of $s_{n,t}$ from knowledge of $\Pr(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$. □

Proof of Proposition 3, Part (III). Part (III) of Proposition 3 follows from part (II). In fact, the mixture components and weights at time t, $\Pr(w_{n,t} \mid H_{n,1}, D_n^t, e_n, a_n^{t-1})$ and $\Pr(e_n, a_n^{t-1} \mid H_{n,1}, D_n^t)$, are identified by Proposition 2. By combining these two with knowledge of $\Pr(H_{n,1}, D_n^t)$—from the data—and $\Pr(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$—shown in the previous paragraph—we identify $\Pr(D_{n,t}, w_{n,t} \mid H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ for conditioning events $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ having strictly positive probability. By combining knowledge of $\Pr(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$, $\Pr(D_{n,t}, w_{n,t} \mid H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$, the distribution of $s_{n,t}$, and the map from realizations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realizations of $s_{n,t}$, we identify $\Pr(D_{n,t} = d, w_{n,t} \leq w \mid s_{n,t} = s)$ for each $d \in \mathcal{D}$, $w \in \mathbb{R}$, and $s \in \mathcal{S}_t^0$.[20] □

Proof of Proposition 3, Part (IV). First, recall that $s_{n,t} := (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ and $\kappa_{n,t}$ is a deterministic function of D_n^{t-1}. Therefore, $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$ is simply $\Pr(P_{n,t} \mid D_{n,t-1}, H_{n,1}, P_{n,t-1}, e_n)$. Second, by Proposition 3, we can compute $P_{n,t}$ for every realization of $(D_{n,t-1}, H_{n,1}, P_{n,t-1}, e_n)$. Moreover, from Equations (3) and (4) in Section 2, $\Pr(P_{n,t} \mid D_{n,t-1}, H_{n,1}, P_{n,t-1}, e_n)$ is nothing but $\Pr(a_{n,t-1} \mid H_{n,1}, D_{n,t-1}, e_n)$. The latter probability can be computed from $\Pr(a_n^{t-1} \mid H_{n,1}, D_n^{t-1}, e_n)$ and $\Pr(H_{n,1}, D_n^{t-1}, e_n)$. $\Pr(a_{n,t-1} \mid H_{n,1}, D_{n,t-1}, e_n)$ is shown to be identified in Step 1 of the proof of Proposition 3 (I) by concatenating the wage mixture weights at times $t-1$ and t. $\Pr(H_{n,1}, D_n^{t-1}, e_n)$ is identified from the wage mixture weights at time t. Therefore, $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$ is identified. □

Proof of Proposition 3, Part (V). Let $t = 1$. We proceed in two steps. First, observe that

$$\Pr(D_{n,1} \mid H_{n,1}, e_n) = \frac{\Pr(e_n \mid H_{n,1}, D_{n,1}) \Pr(H_{n,1}, D_{n,1})}{\Pr(H_{n,1}, e_n)}. \tag{48}$$

$\Pr(e_n \mid H_{n,1}, D_{n,1})$ is identified from the wage mixture weights at $t = 1$ by Proposition 2. $\Pr(H_{n,1}, D_{n,1})$ is known from the data under Assumption 1. $\Pr(H_{n,1}, e_n)$ is computed by combining $\Pr(e_n \mid H_{n,1}, D_{n,1})$ and $\Pr(H_{n,1}, D_{n,1})$. Hence, (48) is identified. Moreover, $p_1(H_{n,1}, e_n)$ is identified by part (I) of Proposition 3—in particular, we know the map from realizations of $(H_{n,1}, e_n)$ to realizations of $p_1(H_{n,1}, e_n)$. Thus, we identify $\Pr(D_{n,1} \mid H_{n,1}, e_n, p_1(H_{n,1}, e_n))$, which is $\Pr(D_{n,1} \mid s_{n,1})$.

Consider the case $2 \leq t \leq T$. We proceed in three steps. First, we identify $\Pr(D_{n,t} \mid H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ from the wage mixture weights at t and $t-1$, as explained below. Second, $\Pr(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ is identified from the wage mixture weights at time t, as explained in Section 4.2. Third, recalling that $s_{n,t}$ is a deterministic function of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ and is identified by part (III) of Proposition 3—in particular, we know the map from realizations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realizations of $s_{n,t}$—we recover $\Pr(D_{n,t} \mid s_{n,t})$, based on the first two steps.

To clarify the identification of $\Pr(D_{n,t} \mid H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ for $2 \leq t \leq T$, consider, for

[20] Observe that the joint probability $\Pr(H_{n,1} = h, D_n^{t-1} = (d_1, \ldots, d_{t-1}), e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1}))$ may be zero due to Assumption 2 either because $\Pr(H_{n,1} = h, D_n^{t-1} = (d_1, \ldots, d_{t-1})) = 0$—which can be directly verified from the data—or because $\Pr(e_n = e, a_n^{t-1} = (a_1, \ldots, a_{t-1}) \mid H_{n,1} = h, D_n^{t-1} = (d_1, \ldots, d_{t-1})) = 0$—which can be ascertained by examining the wage mixture weights in t identified by Proposition 2. In light of this, $\mathcal{S}_t^0 \subseteq \mathcal{S}_t$ introduced in Part (III) of Proposition 3 denotes the subset of $\mathcal{S}_t$ that collects the realizations of $s_{n,t}$ with strictly positive probability.

example, $t = 2$. This conditional probability can be expressed as:

$$\begin{aligned}
&\Pr(D_{n,2} \,| H_{n,1}, D_{n,1}, e_n, a_{n,1}) \\
&\qquad = \frac{\Pr(H_{n,1}, D_n^2, e_n, a_{n,1})}{\Pr(a_{n,1} \mid H_{n,1}, D_{n,1}, e_n)\Pr(e_n \mid H_{n,1}, D_{n,1})\Pr(H_{n,1}, D_{n,1})}.
\end{aligned} \tag{49}$$

As shown in Step 1 of the proof of Part (I) of Proposition 3, $\Pr(H_{n,1}, D_n^2, e_n, a_{n,1})$, $\Pr(e_n \mid H_{n,1}, D_{n,1})$, and $\Pr(a_{n,1} \mid H_{n,1}, D_{n,1}, e_n)$ are identified from the wage mixture weights at $t = 1$ and $t = 2$. $\Pr(H_{n,1}, D_{n,1})$ is known from the data under Assumption 1. Therefore, $\Pr(D_{n,2} \mid H_{n,1}, D_{n,1}, e_n, a_{n,1})$ is also identified. Similar arguments can be applied to other periods $2 \le t \le T$. □

Proof of Proposition 4. Let $(d, e) \in \mathcal{D} \times \mathcal{E}$ and $t \ge 1$. Assumption 12 implies

$$\lim_{w \to -\infty} \frac{\Pr(D_{n,t} = d, w_{n,t}(d, C(d, s), e) \le w \mid s_{n,t}(e) = s)}{q_{d,e}\Pr(w_{n,t}(d, C(d, s), e) \le w \mid s_{n,t}(e) = s)} = 1 \quad \forall s \in \mathcal{S}_t^0(d, e). \tag{50}$$

Fix $s \in \mathcal{S}_t^0(d, e)$. By (50) and using Assumptions 8, 9, and 10, and we have that:

$$\Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, s), e) \le w \mid s_{n,t}(e) = s\right) \sim q_{d,e}F_e\left(w - \varphi(d, C(d, s), s)\right), \tag{51}$$

and

$$\Pr(D_{n,t}{=}d, w_{n,t}(d, C(d, \bar{s}_t), e) \le w{-}\varphi(d, C(d, s), s) \mid s_{n,t}(e){=}\bar{s}_t) \sim q_{d,e}F_e(w{-}\varphi(d, C(d, s), s)), \tag{52}$$

where the symbol "$\sim$" denotes asymptotic equivalence as $w \to -\infty$. By combining (51) and (52), we obtain that

$$\begin{aligned}
&\Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, s), e) \le w \mid s_{n,t}(e) = s\right) \\
&\qquad \sim \Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, \bar{s}_t), e) \le w - \varphi(d, C(d, s), s) \mid s_{n,t}(e) = \bar{s}_t\right),
\end{aligned}$$

which is the key identifying equation for $\varphi(d, C(d, s), s)$. Namely, $\varphi(d, C(d, s), s)$ is identified if

$$\begin{aligned}
&\Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, s), e) \le w \mid s_{n,t}(e) = s\right) \\
&\qquad \sim \Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, \bar{s}_t), e) \le w + u \mid s_{n,t}(e) = \bar{s}_t\right),
\end{aligned} \tag{53}$$

only holds at $u = -\varphi(d, C(d, s), s)$.

To show identification, take u such that (53) holds. Let $\tilde{w} := w + \varphi(d, C(d, s), s)$. By (51),

$$\Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, s), e) \le \tilde{w} \mid s_{n,t}(e) = s\right) \sim q_{d,e}F_e\left(\tilde{w} - \varphi(d, C(d, s), s)\right), \tag{54}$$

and

$$\Pr\left(D_{n,t} = d, w_{n,t}(d, C(d, \bar{s}_t), e) \leq \tilde{w} + u \mid s_{n,t}(e) = \bar{s}_t\right) \sim q_{d,e} F_e(\tilde{w} + u), \tag{55}$$

By (53), (54), and (55),

$$F_e\left(\tilde{w} - \varphi(d, C(d, s), s)\right) \sim F_e(\tilde{w} + u). \tag{56}$$

Replacing $\tilde{w} := w + \varphi(d, C(d, s), s)$ in (56), we obtain

$$F_e(w) \sim F_e(w + \varphi(d, C(d, s), s) + u),$$

which more explicitly can be rewritten as

$$\Pr(\epsilon_{n,t}(e, C(d, s)) \leq w) \sim \Pr(\epsilon_{n,t}(e, C(d, \bar{s}_t)) \geq w + \varphi(d, C(d, s), s) + u). \tag{57}$$

Multiplying both sides of (57) by some $\beta > 0$ and applying the exponential function yields that

$$\tilde{F}_e\left(w \exp\left(\beta\left(\varphi(d, C(d, s), s) + u\right)\right)\right) \sim \tilde{F}_e(w), \tag{58}$$

where $\tilde{F}_e$ is the CDF of $\exp(\beta \epsilon_{n,t}(e, C(d, s)))$. By Lemma 2.1 of DM, under Assumption 11, (58) holds only if the *multiplicative* term $\exp\left(\beta\left(\varphi(d, C(d, s), s) + u\right)\right)$ is equal to one, that is, if

$$u = -\varphi(d, C(d, s), s),$$

which implies that $\varphi(d, C(d, s), s)$ is identified.

The same procedure can be repeated for each $s \in \mathcal{S}_t^0(d, e)$, $(d, e) \in \mathcal{D} \times \mathcal{E}$, and $t \geq 1$. □

Proof of Lemma 5. Recall that the equilibrium of the model is efficient. In an efficient equilibrium, job choices maximize the expected present discounted value of output, and, as shown in Equation (34), for a given $(d, e, s) \in \mathcal{D} \times \mathcal{E} \times \mathcal{S}_t^0(d, e)$, we have that:

$$\begin{aligned} &\Pr(D_{n,t} = d \mid s_{n,t}(e) = s, w_{n,t}(d, C(d, s), e) = w) \\ &= \Pr\big(\epsilon_{n,t}(d, e) \geq Y(C(d, s), s) - Y(d, s) + w - \varphi(d, C(d, s), s) \\ &\qquad\qquad \mid \epsilon_{n,t}(C(d, s), e) = w - \varphi(d, C(d, s), s)\big). \end{aligned} \tag{59}$$

Using (59), Assumption 12 states that:

$$\lim_{w\to-\infty} \Pr\big(\epsilon_{n,t}(d,e) \geq Y(C(d,s),s) - Y(d,s) + w - \varphi(d,C(d,s),s) \\ \mid \epsilon_{n,t}(C(d,s),e) = w - \varphi(d,C(d,s),s)\big) = q_{d,e} \quad \forall s \in \mathcal{S}_t^0(d,e). \tag{60}$$

Replacing $u := w - \varphi(d,C(d,s),s)$ in (61), we obtain:

$$\lim_{u\to-\infty} \Pr\big(\epsilon_{n,t}(d,e) \geq Y(C(d,s),s) - Y(d,s) + u \mid \epsilon_{n,t}(C(d,s),e) = u\big) = q_{d,e} \quad \forall s \in \mathcal{S}_t^0(d,e). \tag{61}$$

A sufficient condition condition for (61) is

$$\lim_{u\to-\infty} \Pr(\epsilon_{n,t}(d,e) \geq a + u \mid \epsilon_{n,t}(C(d,s),e) = u) = q_{d,e} > 0 \quad \forall a \in \mathbb{R},$$

as desired. □

Proof of Lemma 6. Fix $e \in \mathcal{E}$. By Proposition 3 (parts (III) and (V)), knowing both the distribution of $(D_{n,t}, w_{n,t})$ conditional on $s_{n,t}(e)$ and the distribution of $D_{n,t}$ conditional on $s_{n,t}(e)$ identifies the distribution of $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t}(e))$. Let $(d,s) \in \mathcal{D} \times \mathcal{S}_t^0(d,e)$. The distribution of $w_{n,t}$ conditional on $(D_{n,t} = d,\ s_{n,t}(e) = s)$ corresponds to the distribution of $\epsilon_{n,t}(C(d,s),e) + \varphi(d,C(d,s),e)$ conditional on $(D_{n,t} = d,\ s_{n,t}(e) = s)$. Knowing $\varphi(d,C(d,s),e)$ from Proposition 4 then allows us to isolate the distribution of $\epsilon_{n,t}(C(d,s),e)$ under the same conditioning, and hence to determine its unconditional distribution.

Under Assumption 9, this distribution does not vary across $d \in \mathcal{D}$. Therefore, the (marginal) distribution of $\epsilon_{n,t}(d,e)$ is identified for every $d \in \mathcal{D}$. Furthermore, if the shocks $\{\epsilon_{n,t}(d,e)\}$ are independently distributed across d, then the (joint) distribution G_e of the vector of shocks $\epsilon_{n,t}(e) := (\epsilon_{n,t}(d,e) : d \in \mathcal{D})$ is also identified. □

Proof of Lemma 2.1 of DM. Let S be the survival function of a random variable Y having full support on $\mathbb{R}$. Suppose that:

1. $\mathbb{E}\big(\max\{0,Y\}^p\big) < +\infty$ for some $p > 0$.

2. h is a function such that $h(y) \sim y$ as $y \to +\infty$.

3. $S(y) \sim S\big(\kappa\, h(y)\big)$ as $y \to +\infty$, for some $\kappa > 0$.

The claim is that κ must equal 1. We now present an informal proof of this claim.

Assume, by contradiction, that $\kappa \neq 1$. We will show that this contradicts the finiteness of $\mathbb{E}\big(\max\{0, Y\}^p\big)$. First, note that $S(y)$ cannot vanish exactly like y^{-p} if $\kappa \neq 1$. To see this, suppose that $S(y) \sim y^{-p}$. Then, since $h(y) \sim y$, we also have

$$S\big(\kappa\, h(y)\big) \sim (\kappa\, h(y))^{-p} = \kappa^{-p}\, y^{-p}.$$

But by assumption, $S(y) \sim S\big(\kappa\, h(y)\big)$. Hence,

$$y^{-p} \sim S(y) \sim S\big(\kappa\, h(y)\big), \sim \kappa^{-p}\, y^{-p}.$$

The only way to avoid contradiction here is if $\kappa^{-p} = 1$, meaning $\kappa = 1$. Thus, if $\kappa \neq 1$, $S(y)$ cannot behave precisely like y^{-p} asymptotically.

Second, we show that if $S(y)$ decays either more slowly or more quickly than y^{-p}, then $\mathbb{E}(\max\{0,$-$Y\}^p)$ must be infinite. Specifically, if $S(y)$ decays more slowly than y^{-p}, then the integral

$$\int_0^\infty p\, y^{p-1} S(y)\, dy$$

diverges, which implies $\mathbb{E}\big(\max\{0, Y\}^p\big) = +\infty$. On the other hand, if $S(y)$ decays more quickly than y^{-p}, one might guess this ensures convergence of the same integral. However, the assumption $S(y) \sim S(\kappa\, y)$ (from Assumptions 2 and 3) implies that $S(y)$ changes very little when y is scaled by κ, preventing $S(y)$ from sufficiently dominating y^{-p} to make the integral converge. Therefore, the only remaining possibility is $\kappa = 1$. This completes the argument. $\square$

D Details on Monte Carlo Simulation

Data are generated by tracking 1,000,000 workers over 40 periods (from age 25 to 65). Upon entering the labor market, each worker is assigned an efficiency type e from K_1 possible values. To assign these values, the interval $[-2\sigma_1, 2\sigma_1]$ is divided into $K_1 - 1$ equal subintervals, with grid points $\{a_1, \ldots, a_{K_1}\}$ (with $\sigma_1 > 0$). For each worker, a number is drawn from the Normal distribution $\mathcal{N}(0, \sigma_1)$; if the number falls within the interval $[a_{j-1}, a_j]$, the worker is assigned efficiency type a_j; if the number falls below a_1, the worker is assigned efficiency type a_1; and if the number falls above a_{K_1}, the worker is assigned type a_{K_1}. Workers are also assigned gender, education level, and age. In addition, each worker is given an ability type θ_n from two possible levels: high ability $\bar{\theta}$ and low ability $\underline{\theta}$.

The simulation includes 200 firms. Each firm is assigned a productivity type from K_2 possible values, determined by drawing from a Normal distribution $\mathcal{N}(0, \sigma_2)$ in a manner similar to that for workers. The set $\mathcal{D}$ now denotes the collection of labels for the firm productivity types (as opposed to firm identities in the original model), where a generic label $d \in \mathcal{D}$ is associated with the productivity type $\beta_0(d) \in \{b_1, \ldots, b_{K_2-1}\}$. Note that Bonhomme et al. (2019) also considers economies with a finite number of worker and firm types.

To generate mobility in the economy, for each worker in each period a draw is made from a Bernoulli distribution with parameter p. If the draw equals 1, an additional number is drawn from a Normal distribution $\mathcal{N}(e \times \mu, \sigma_2)$ to determine the type of firm the worker moves to, where $\mu > 0$ and e is the worker efficiency type. The parameter μ significantly influences sorting; when μ is high, workers with high efficiency tend to match with firms having high productivity.

Finally, beliefs are generated from a uniformly distributed prior and updated via Bayes' rule by drawing high and low signals in each period, with high signals being more likely for workers with high ability $\bar{\theta}$ and low signals more likely for those with low ability $\underline{\theta}$.

The components of the wage Equation (38) are specified as

$$\begin{aligned}
\beta_1(d,e)H_{n,1} &= \beta_{1,0}\exp(e)\text{edu_high}_n + \beta_{1,1}\exp(e)\text{gender}_n,\\
\beta_2(d,e)\kappa_{n,t} &= \beta_{2,0}\exp(e) + \beta_{2,1}\exp(e)\text{age}_n + \beta_{2,2}\exp(e)\text{age}_n^2,\\
\beta_3(d,e)P_{n,t} &= \beta_3\exp(e)P_{n,t},
\end{aligned}$$

where edu_high_n and gender_n are education (college/noncollege) and gender dummies, and

$$\begin{aligned}
&\Psi(H_{n,1},\,\kappa_{n,t},\,P_{n,t};\,\psi(d,e)) = \psi_{1,1}(d)\text{age}_n^3 + \psi_{1,2}(d)\text{age}_n^4 + \psi_{2,2}(d)(P_{n,t})^2 + \psi_{2,3}(d)(P_{n,t})^3\\
&+ \psi_{2,4}(d)(P_{n,t})^4 + \psi_{3,1}(d)(P_{n,t})\,\text{age} + \psi_{3,2}(d)(P_{n,t})\,\text{age}_n^2 + \psi_{3,3}(d)(P_{n,t})\,\text{age}_n^3 + \psi_{3,4}(d)(P_{n,t})\,\text{age}_n^4\\
&+ \psi_{4,1}(d)(P_{n,t}^2)\,\text{age} + \psi_{4,2}(d)(P_{n,t}^2)\,\text{age}_n^2 + \psi_{4,3}(d)(P_{n,t}^2)\,\text{age}_n^3 + \psi_{4,4}(d)(P_{n,t}^2)\,\text{age}_n^4\\
&+ \psi_{5,1}(d)(P_{n,t})\,\text{edu_high}_n + \psi_{5,2}(d)(P_{n,t}^2)\,\text{edu_high}_n + \psi_{5,3}(d)(P_{n,t}^3)\,\text{edu_high}_n\\
&+ \psi_{5,4}(d)(P_{n,t}^4)\,\text{edu_high}_n,
\end{aligned}$$

where each $\psi_{i,j}(d) := \psi_{i,j}\exp(\beta_0(d))$. The moments from PSID used to calibrate the simulation parameters are:

- the variance, skewness, and kurtosis of log-earnings;

- the variance of log-earnings at three ages:30, 50, and 65 years old;
- the variance of log earning within cells defined by age-gender-education groups. Age groups are 5-year groups for a total of 30 cells;
- the variance of two-years log-earnings growth;
- The growth of average log earnings between 30 and 50 years old;
- The growth of average log earnings between 50 and 65 years old;
- the college premium;
- the gender gap.

We also include the share of the variance accounted for by the worker effect, firm effect, and their covariance, based on the AKM estimates from Song et al. (2019), that is, the share of log earnings variance accounted for by workers fixed effects, the share accounted for by firm fixed effects, and the share accounted for the covariance term (sorting).

E Details on Empirical Application

We describe how we construct workers' variable pay and compute $P_{n,t}$. First, given a firm k and quarter q, we select individuals working full-time at firm k if their earnings exceed the full-time minimum wage for the quarter, i.e., $12 \times 5 \times 8 \times \underline{w}$, where $\underline{w}$ is the federal minimum wage (approximately 3,500 USD). For these individuals, we retain only those who remain at firm k for at least 6 quarters. Second, we define the fraction of variable pay as follows:

- For each worker n in the selected sample, we compute a 5-quarter moving average centered on quarter q (i.e., two quarters before and two quarters after q). Denote this average by $\overline{w}_{k,n,q}$.
- For each quarter q, we calculate $r_{k,n,q} = w_{k,n,q} - \overline{w}_{k,n,q}$, where $w_{k,n,q}$ is the observed wage of worker n in quarter q at firm k.
- We identify a positive jump, denoted by $r^{+}_{k,n,q} = r_{k,n,q}$, if the following conditions are met:
  - $r_{k,n,q}/\overline{w}_{k,n,q} \geq 0.1$ (i.e., the jump is at least 10% of the moving average income),
  - $|r_{k,n,q-1}/\overline{w}_{k,n,q}| < 0.1$ and $|r_{k,n,q+1}/\overline{w}_{k,n,q}| < 0.1$ (ensuring the jump is isolated rather than part of a permanent increase).

- A negative jump $r^-_{k,n,q}$ is identified similarly.

Third, we define three objects:

- Annual wage: $w_{d,n,t} = \sum_q \sum_k w_{k,n,q}$, where d is the firm among the jobs k held by the worker that provided the highest wage in that year.
- Positive variable pay: $r^+_{d,n,t} = \sum_q \sum_k r^+_{k,n,q}$.
- Negative variable pay: $r^-_{d,n,t} = \sum_q \sum_k r^-_{k,n,q}$.

Given these three objects, we define the fraction of variable pay over total income as $\rho_{d,n,t} = \frac{r^+_{d,n,t}}{w_{d,n,t}}$, considering only the positive jumps. This yields a distribution of $\rho_{d,n,t}$ for each year t and firm d across all workers. In turn, a worker is assigned a positive performance signal (equal to 1) if they are in the top quartile of the $\rho_{d,n,t}$ distribution within a year; all other workers receive a signal of zero. Finally, $P_{n,t}$ is generated using a uniform prior and updated via Bayes' rule using the signals derived from the $\rho_{d,n,t}$ distribution.