Identifying Uncertainty, Learning about Productivity, and Human Capital Acquisition: A Reassessment of Labor Market Sorting and Firm Monopsony Power*

Cristina Gualdani† Elena Pastorino‡ Áureo de Paula§ Sergio Salgado¶

November 2025

Abstract

We examine the empirical content of a large class of dynamic matching models of the labor market with ex-ante heterogeneous firms and workers, symmetric uncertainty and learning about workers' productivity, and firms' monopsony power. We allow workers' human capital, acquired before and after entry into the labor market, to be general across firms to varying degrees. Such a framework nests and extends known models of worker turnover across firms, occupational choice, wage growth, wage differentials across occupations, firms, and industries, and wage dispersion across workers and over the life cycle. We establish intuitive conditions under which the model primitives are semiparametrically identified solely from data on workers' wages and jobs, despite the dynamics of these models giving rise to complex patterns of selection based on endogenously time-varying observable and unobservable characteristics of workers and firms. By relying on this identification argument, we develop a constructive estimator of the model primitives, which builds on common methods for mixture and extremal quantile regression models and displays standard properties. Through the lens of this framework, we investigate how well typical empirical wage measures of matching assortativeness and firms' wage-setting power detect the degrees of sorting and monopsony power in the labor market, respectively. We show that usual measures of sorting severely *understate* its importance because they ignore the option value of worker human capital and the information about worker productivity acquired through employment, in terms of higher future wages and improved future sorting, which is priced into current wages thus depressing them. We also demonstrate how the markdown of wages relative to output largely *overstates* firms' labor market power by ignoring that this option value, which captures future returns from acquired human capital and information, generally lowers wages. We find evidence of both of these features in U.S. data by documenting a strong degree of labor market sorting once appropriately measured and, correspondingly, a lower degree of firm monopsony power than typically documented.

Keywords: Job Matching, Inequality, Wage Dispersion, Monopsony Power, Matched Employer-Employee Data, Identification, Structural Estimation

*We are grateful to Stéphane Bonhomme, Richard Blundell, 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 title "On the Identification of Models of Uncertainty, Learning, and Human Capital Acquisition with Sorting". This work was supported by the Economic and Social Research Council (grant ES/X011186/1) and the National Science Foundation (grant no. 2017957).

†Queen Mary University of London. E-mail: c.gualdani@qmul.ac.uk.
‡Stanford University, Hoover Institution, and NBER. E-mail: epastori@stanford.edu.
§University College London. E-mail: a.paula@ucl.ac.uk.
¶Wharton School, University of Pennsylvania. E-mail: ssalgado@wharton.upenn.edu.

1 Introduction

Matching models of the labor market have been extensively used in the labor and macro economics literature to study a wide range of phenomena, including workers' occupational choice and turnover across firms, wage growth, wage differentials across occupations, firms, and industries, and wage inequality across workers and over the life cycle. At their core, these models interpret workers' career paths as the outcome of two key processes that take place as labor market experience accumulates: the *acquisition of new human capital* by workers and the gradual *learning* about workers' true productivity, which may be unknown to both workers and firms when workers first enter the labor market. Both of these processes lead workers to progressively match with the jobs and firms at which they are most productive, as workers' true and perceived productivity evolve over time.

This framework for careers and labor market sorting nests numerous models: classic ones of human capital acquisition and wage growth (Mincer, 1958, 1974; Ben-Porath, 1967; Becker, 1975); of learning and worker turnover (Jovanovic, 1979; Flinn, 1986); of static (Heckman and Honoré, 1990) and dynamic occupational choice without learning (Keane and Wolpin, 1997) and with learning (Miller, 1984); of wage dispersion across individuals and over time (Farber and Gibbons, 1996; Altonji and Pierret, 2001) and of wage differentials across occupations and industries (Gibbons et al., 2005; Gibbons and Waldman, 2006) due to human capital acquisition and learning; and many others extending these models(Jovanovic and Nyarko, 1997; Gibbons and Waldman, 1999a,b; Lange, 2007; Nagypál, 2007; Antonovics and Golan, 2012; Kahn and Lange, 2014; Pastorino, 2024). See Gibbons and Waldman (1999a) and Rubinstein and Weiss (2006) for reviews emphasizing the central role of learning about workers' productivity in determining job mobility, wage growth, and wage dispersion.

Despite the widespread use of these models to gauge the determinants of job mobility and wage inequality, their identification is difficult to establish for three well-understood reasons. First, as workers accumulate human capital and information about productivity through employment, their career paths emerge from a complex process of dynamic selection into jobs. This process is shaped by multiple worker and firm characteristics—serially correlated, endogenously time-varying, and mostly unobserved—that both capture the stocks of acquired human capital and information at each point in time and govern their subsequent evolution. Second, occupation, firm, and industry choices are by their very nature discrete, leading to the standard challenge of dynamic discrete choice models, which are known to be nonparametrically underidentified even when state variables are observed (Rust, 1994; Magnac and Thesmar, 2002). Third, since workers and firms decide on matches by intertemporally trading off the benefits and costs of alternative job opportunities, wages are typically highly nonlinear functions of these unobserved variables, which makes standard methods for fixed-effect or even interactive fixed-effect models inadequate (Bonhomme et al., 2019; Freyberger, 2018). For instance, workers may initially choose jobs at which they are less productive but that offer better opportunities to acquire new skills or learn about their ability, leading to higher future wages. In a competitive enough labor market, arbitrage then implies that equilibrium wages reflect the relative *option value* of human capital and information prospects across employment alternatives. This value is highest at intermediate levels of human capital and information, when new human capital or information may induce workers and firms to make different employment and wage decisions, yielding in general nonmonotone, and so nonlinear, wage patterns in unobservables. The inference about the primitive sources of wage inequality is further complicated by firm monopsony power, which leads to substantial deviations of wages from worker productivity (Seegmiller, 2021; Lamadon et al., 2022; Berger et al., 2026) that may vary across workers and over time, yet is often overlooked in the literature, which largely relies on single-agent or perfectly competitive models of the labor market.

In this paper, we examine the empirical content of this general class of dynamic models with ex-ante heterogeneous workers, human capital acquisition, and symmetric uncertainty and learning about workers' productivity, which we augment to incorporate imperfect labor market competition and the resulting firm monopsony power. We establish a novel result on their identification using data on only workers' jobs and wages. Our arguments rest on simple conditions that accommodate arbitrary patterns of selection on endogenously time-varying unobservables, are easy to verify, and lead to constructive estimators of the model primitives that are readily implementable using common methods for mixture models and static sample-selection models. Finally, we estimate our model on U.S. data. Two key findings emerge. First, standard measures of labor market sorting can understate its extent even in highly unequal labor markets—helping to resolve the long-standing puzzle of low measured sorting despite high wage dispersion—when they ignore that job and wage mobility are shaped by the dynamic acquisition of human capital by workers and of information about worker productivity, whose value is reflected in *future wages* but discounted in *current wages*. Our estimates suggest much stronger sorting when measured accordingly. Second, by this mechanism, wage markdowns are a poor proxy for monopsony power. Low wage-to-output ratios arise even with intense competition, because the option value of future high wages is priced into current wages, lowering them. Typical markdown measures that abstract from this force overstate firms' wage-setting power.

Formally, we study a broad class of non-stationary dynamic matching models in which a finite number of heterogeneous firms compete à la Bertrand for workers in each period over a discrete time horizon of either finite or infinite length.[1] Workers differ in three dimensions: their initial and acquired *human capital*, observed by model agents, with only the initial component observed by the econometrician; *efficiency*, a time-invariant characteristic observed by model agents but unobserved by the econometrician; and *ability*, a time-invariant characteristic initially unobserved by both model agents and the econometrician, which is gradually learned by model agents as experience accumulates. Firms differ in three dimensions observed by model agents but unobserved by the econometrician: their *output* technology—how labor produces output; *human capital* technology—how experience in the labor market and at specific firms generates new skills; and *information* technology—how output provides information about workers' ability. For example, the "stepping-stone" job of junior resident, at which a physician may not be very productive in terms of current output, may allow the physician to acquire substantial human capital or information about ability. Conversely, the "star" job of CEO, at which a manager may be highly productive, may provide little additional human capital and, due to the confounding effect of market forces, little new information about ability. As is standard in the literature, since human capital, efficiency, and ability stochastically map into a worker's output, which is observed by firms and workers, output provides a noisy signal (*performance*) about a worker's ability that firms and workers use to update their common beliefs about it.

We characterize the set of Markov perfect equilibria in this setting. We show that equilibrium wages equal the sum of a worker's expected one-period output at the firm offering the worker the second-highest expected present discounted value of wages in a period—the *second-best* firm, as in a second-price auction—and a *compensating differential*. This term amounts to either a *premium* for the foregone opportunity to acquire human capital and information at the second-best firm or a *discount* for the superior opportunity to acquire human capital and information at the first-best firm.

Econometrically, our framework can be viewed as an equilibrium dynamic generalized Roy model in which workers select into jobs based on multiple unobservables: time-varying, serially uncorrelated idiosyncratic productivity shocks; time-invariant worker efficiency; and firm and worker beliefs about worker ability, which are time-varying, serially correlated, and endogenously evolving with workers' past job choices. Together with firm technologies, worker efficiency and beliefs about ability determine both expected output and the compensating-differential component of wages. The latter, as argued, is nonlinear and nonseparable in worker efficiency, beliefs about ability, and the characteristics of firm technologies, as it captures the endogenous difference in wage returns from tomorrow onward between accepting a job today at the employing firm and at the second-best firm.

[1] Bertrand price competition provides an appealing approach to monopsony markets with differentiated labor inputs, since it allows for a nontrivial and flexible sharing rule of the surplus from a firm-worker match without the need for any of the additional parameters that, say, typical bargaining setups require, such as bargaining weights or haggling costs.

The econometric literature on Roy models provides well-known techniques to address selection on unobservables. Addressing selection on the other two unobserved dimensions central to our setting—worker efficiency and beliefs—requires a different identification strategy, though. Our approach combines classic arguments that rely on extremal quantiles of observed wages to control for selection on productivity shocks—what we refer to as the *extremal quantile step*—with a first step that identifies a mixture representation of the wage process to control for selection on worker efficiency and beliefs—what we refer to as the *mixture step*. The logic is as follows. We first represent the wage distribution, conditional on a worker's job history, current job choice, and initial human capital, as a mixture over latent worker classes indexed by a worker's efficiency and history of performance (and so beliefs). We recover mixture components and weights by exploiting a recent result by Aragam et al. (2020), which provides mild conditions for identification: the wage mixture must admit a *clusterable finite-mixture* representation. That is, it must be expressible as a finite mixture whose components are sufficiently distinct—and thus clusterable in the precise sense defined by Aragam et al. (2020)—to be identifiable. A canonical and widely used example of this class is a finite mixture of continuous mixtures of Normals, with means and variances that sufficiently differ across components, as shown by Bruni and Koch (1985) and Aragam et al. (2020). As continuous mixtures of Normals approximate any distribution arbitrarily well (Nguyen and McLachlan, 2019), we view the restriction to a clusterable finite-mixture representation as minimal. In particular, this class is well suited to modeling the selection-contaminated mixture distributions that arise in complex settings like ours in which workers dynamically sort into jobs based on multiple unobservables.

We next use the recovered mixture weights to identify the initial distribution and the law of motion of the state, in particular the joint distribution of worker efficiency and initial beliefs about ability, and how beliefs evolve over time. As experience accumulates, mixture weights also encode the distribution of employment histories across all possible observed and unobserved characteristics of workers and firms, pinning down conditional choice probabilities. Likewise, mixture components identify the wage distributions associated with each combination of worker and firm characteristics. Lastly, the extremal quantile step resolves the remaining selection on productivity shocks, yielding the deterministic component of potential wages and, in turn, expected output, the compensating differential in wages, the human capital function, and the distribution of productivity shocks.

This two-step identification strategy also leads to a natural estimator that combines standard finite mixture-model methods, such as maximum likelihood, with extremal quantile regression methods (Chernozhukov, 2005; D'Haultfoeuille et al., 2018). We exploit this econometric approach to measure labor market sorting, firms' monopsony power, and their implications for U.S. wage inequality. The most popular framework for measuring sorting is that of Abowd et al. (1999)—hereafter, AKM—which decomposes wages into additively separable unobserved worker and firm fixed effects, observed covariates, and residual shocks. The contribution of worker-firm sorting to wage inequality is then summarized by the fraction of wage variance attributable to the covariance between worker and firm effects. Empirical findings based on this framework typically point to a limited role for sorting, as this covariance component accounts for only a small share of overall wage dispersion; see Card et al. (2013) and Song et al. (2019).[2] We argue that neither using nor interpreting AKM estimates is straightforward in our class of models for two reasons. First, when firms have wage-setting power, wages reflect a worker's best outside option—and so (expected match) output at the second-best firm—rather than output at the employing firm. As a result, the AKM covariance statistic can be disconnected from the output complementarities that determine sorting, thus providing an inaccurate measure of it. Second, and more importantly, wages include a compensating differential that captures the dynamic value of human capital and learning opportunities across competing firms. As this term is nonlinear and nonseparable in worker and firm attributes, omitting it induces a misspecification that biases estimated worker and firm effects and, in turn, the estimated AKM covariance statistic.

A similar concern arises when measuring monopsony power. Firms' wage-setting power is often inferred from the wage markdown, defined as the ratio of wages to workers' output, with low wage-to-output ratios interpreted as evidence of weak firm competition (Seegmiller, 2021; Yeh et al., 2022; Lamadon et al., 2022). In our setting, however, the wage markdown is not directly informative about monopsony power because it conflates firms' wage-setting power with the endogenous option value, in terms of future wage returns, of the human capital and the information about ability acquired through employment, which is embedded in the equilibrium wage process precisely due to the competition among firms for workers. Namely, wages include a compensating differential capturing the future value of workers' current investment decisions, which moves current wages below or above current output and changes over the life cycle as the opportunities for investments in human capital and information, and their returns, change. Hence, wage-to-output ratios can substantially vary across workers and over time for reasons unrelated to any labor market distortions.

To shed light on this mechanism, especially on the role of the compensating differential in wages, we provide both simulation-based and empirical evidence. We first simulate an economy illustrating key features of our class of models. We choose values for the model parameters to reproduce

[2]Bonhomme et al. (2023) show that correcting AKM estimates for limited-mobility bias tends to raise the covariance component attributable to sorting, but it still accounts for a small portion of measured wage inequality.

the distribution of wages from the Panel Study of Income Dynamics (PSID). Much like in a setting with standard omitted-variable bias, our findings suggest that when the compensating differential is mostly negative under the true data-generating process—so that workers match with firms offering *more* valuable human capital and learning prospects than their competitors—the AKM covariance statistic understates the impact of sorting on wage inequality. The reason is that omitting the compensating differential removes a wage component that varies nonlinearly with worker and firm effects and depresses wages, thereby attenuating the underlying worker-firm complementarity that can be inferred from wages. This mechanism offers a structural interpretation for the recurrent and puzzling empirical finding that AKM-based sorting measures are low even in labor markets characterized by substantial wage dispersion, which canonical matching models would interpret as the outcome of a strongly assortative matching process. Through the lens of our model, this disconnect arises because the AKM decomposition omits a key component of wages—the compensating differential—thereby biasing the inferred worker-firm covariance. By contrast, a direct output-based sorting measure, such as an index of the supermodularity of output in worker and firm attributes, consistently reveals strong complementarities between firms and workers, implying large gains from sorting that increase with workers' labor market experience and when learning frictions are reduced. Relative to a scenario in which workers and firms randomly match, we find that sorting accounts for 45% of the variance of wages, more than *twice* the contribution usually estimated by the AKM covariance statistic.

Turning to monopsony power, we document similar results. Holding the simulated economy fixed, we re-estimate the wage equation on the same data but imposing gradually lower or higher values of the compensating differential. The idea is to estimate a progressively more "misspecified" wage equation in which the compensating differential is reduced or increased, mimicking what an econometrician would estimate using a model that abstracts from, or amplifies, the human capital and informational option value of employment at different firms. By forcing current wages below or above current output, this misspecified model induces large shifts in both the level and the life-cycle profile of the inferred wage markdown relative to its true level and profile, leading the econometrician to incorrectly conclude that the degree of firm competition widely varies across the estimated scenarios, even though the underlying labor market environment is unchanged. Taken together, these findings suggest that a better measure of monopsony power is an *intertemporal* wage markdown index that incorporates the full time path of the dynamic returns to employment.

Next, we estimate our model using U.S. matched employer-employee data from the Longitudinal Employer-Household Dynamics (LEHD) dataset, which reports quarterly earnings for workers in 21

U.S. states from the mid-1990s through 2022. Our empirical results—currently awaiting approval for disclosure from the U.S. Census Bureau—are qualitatively and quantitatively consistent with the simulation evidence discussed. Overall, they underscore the importance of accounting for workers' life-cycle investments in productivity and for the gradual discovery of their best matches across firms and jobs when assessing the degree of labor market sorting and the extent of firm monopsony power.

Literature Review. Our work is related to a large literature on the estimation of human capital and learning models, including Heckman (1976), Cunha and Heckman (2008), and Buchinsky et al. (2010); see Gibbons and Waldman (1999a), Rubinstein and Weiss (2006), and Keane et al. (2017) for reviews. Our paper is the first to provide formal identification arguments for dynamic matching models in which workers differ in observed and unobserved (to model agents and the econometrician) persistent characteristics and firms are heterogeneous in their output, human capital, and information technologies and have monopsony power. For the identification of frictional competitive matching models without the unobserved state dependence that human capital acquisition and learning give rise to, see Eeckhout and Kircher (2011), Hagedorn et al. (2017), and Lamadon et al. (2026).

A large literature has investigated the empirical content of the static Roy model, including Chamberlain (1986), Heckman (1990), Heckman and Honoré (1990), Ahn and Powell (1993), Das et al. (2003), Newey (2009), and D'Haultfoeuille and Maurel (2013). Our identification approach generalizes existing arguments for extremal quantile regression models (Chernozhukov, 2005; Sasaki and Wang, 2024, 2025) to account for selection on unobservables in dynamic generalized equilibrium Roy models without exclusion restrictions. D'Haultfoeuille and Maurel (2013) propose an identification procedure for classic static Roy models in the absence of exclusion restrictions for thin-tailed potential outcome distributions. By contrast, our approach accommodates wage and log-wage distributions with fat tails, such as the Pareto, lognormal, and Cauchy, which are relevant for plausibly modeling wage distributions that are highly concentrated at the top, such as the U.S. one.

Much work has explored the identification of dynamic discrete choice models with correlated unobserved states, including Kasahara and Shimotsu (2009), Hu and Shum (2012), An et al. (2013), Shiu and Hu (2013), Hu et al. (2015), Berry and Compiani (2023), Higgins and Jochmans (2023), and Higgins and Jochmans (2024). This work either assumes time-invariant unobserved heterogeneity or allows for time-varying serially correlated heterogeneity but only under high-level restrictions on endogenous variables such as monotonicity, specific distributional supports for unobserved variables relative to observed variables, or the availability of instruments. None of these conditions applies to our setting. Thus, we proceed by exploiting the information provided by wages—a continuous outcome typically not used in this literature—which allows us to identify the law of motion of unobserved state variables as well as conditional choice probabilities.

Our paper is also related to an extensive literature on sorting that builds on and extends the AKM framework, such as Card et al. (2013), Card et al. (2018), Bonhomme et al. (2019), and Song et al. (2019), and that emphasizes the importance of correcting AKM estimates to address bias due to low mobility, including Abowd et al. (2004), Andrews et al. (2008, 2012), Kline et al. (2020), and Bonhomme et al. (2023). In our framework, in which human capital accumulation and learning about ability play an important role, we show that the AKM estimates fail to capture sorting since additive worker and firm effects in a traditional (log) wage regression confound nonlinearities, imperfect firm competition, and the option value of future human capital accumulation and learning. A related mismatch between measured worker and firm effects and types in the data-generating process arises in Eeckhout and Kircher (2011), Hagedorn et al. (2017) and, recently, Borovicka and Shimer (2024) and Lamadon et al. (2026) because of search frictions. Similar to our paper, Eeckhout and Kircher (2011) and Hagedorn et al. (2017) argue that wages are nonmonotonic in firm types as they "reflect the opportunity cost of mismatch" (Eeckhout and Kircher, 2011). Borovicka and Shimer (2024) introduce "match-specific shocks" that generate selection, leading to identification challenges similar to those we address. Finally, Lamadon et al. (2026) incorporate compensating differentials for non-pecuniary aspects of jobs. In our setting, current wages incorporate the option value of human capital and information acquired on the job and reflected in future wages. Since compensating differentials arise from the accumulation of human capital and information, which is absent in the work cited, we can leverage the panel dimension of the data for identification purposes in a novel way.

Lastly, our paper connects to the literature on the identification of panel-data models with interactive fixed effects without learning (Freyberger, 2018) and with learning (Bunting et al., 2024) about worker characteristics. The wage equation typical of our class of models differs from the outcome equation in those models in that unobservables endogenously affect wages in a nonlinear, nonmonotone, and nonmultiplicative manner, which renders the use of interactive fixed-effect methods infeasible. Unlike those papers, we also allow for dynamic selection on multiple unobservables.

The rest of the paper is organized as follows. Section 2 introduces the model, Section 3 provides an overview of our identification approach, Section 4 describes the estimation procedure, Section 5 discusses illustrative Monte Carlo exercises and our empirical application, and Section 6 concludes. Appendix A contains additional tables and figures, Appendix B presents the formal identification results, Appendix C further discusses AKM-type measures of sorting, Appendix D proposes extensions of our baseline identification arguments, and Appendix E collects omitted proofs. In the Online Supplementary Material, Appendix F provides additional model details, Appendix G contains further identification extensions, and Appendix H adapts our identification arguments to search models.

2 Setup

We consider a broad class of canonical non-stationary dynamic matching models of the labor market in which a finite number of heterogeneous firms compete à la Bertrand for a large pool of workers in each period over a discrete-time horizon of either finite or infinite length. As discussed, this class of models nests many existing frameworks commonly used to study the determinants of occupational choice, worker turnover, firm-worker sorting, wage growth, and wage inequality. Settings of strategic information acquisition in markets for experience goods, in which consumers and firms learn about consumers' tastes for products over time and firms practice sophisticated forms of dynamic price discrimination, which are increasingly common in online platforms, are also an important nested case (Bergemann and Välimaki, 1996, 2006; Ching, 2010; Kehoe et al., 2024).

Workers are heterogeneous along three dimensions: their initial and acquired human capital (both observed by firms and workers, with only the initial component observed by the econometrician), efficiency (a time-invariant characteristic observed by firms and workers but unobserved by the econometrician), and ability (a time-invariant characteristic that is initially unobserved by firms, workers, and the econometrician, but is gradually learned by firms and workers as experience accumulates). Firms are also heterogeneous along three dimensions observed by both firms and workers but unobserved by the econometrician: their output technology—how labor produces output; human capital accumulation technology—how experience in the labor market and at different firms allows workers to acquire more human capital; and information technology—how output at different firms provides information about a worker's unobserved ability. Once matched, a firm-worker pair produces output, and human capital and information accumulate. Since a worker's human capital, efficiency, and ability stochastically map into output, which is publicly observed by firms and workers, output provides a noisy signal ("performance") about ability that firms and workers use to update their common beliefs about it. Each period, given workers' human capital, the common beliefs about workers' ability, and realized productivity shocks, firms offer employment to workers, workers choose among firms' offers, production occurs, new human capital and information are acquired, and the cycle repeats.

Firms. A finite number of heterogeneous firms indexed by $d \in \mathcal{D}$ produce a homogeneous output good sold in a perfectly competitive market at a price normalized to one. Each firm d operates a production technology that is constant-returns-to-scale in workers' labor, which is the only input to production, and separable across workers.[3] Firms compete for workers by offering wages each period for their employment during the period. Our model and econometric results immediately extend to settings in which firms consist of multiple jobs and, accordingly, their offers specify both a wage and a job assignment for the period; we discuss this multi-job-firm case in Appendix F.2 of the Online Supplementary Material, which is the version of the model that we take to the data. As we proceed, we highlight the features of this case that warrant special attention.

Workers. We denote by t a time period, which does not represent calendar time but rather a worker's time or *experience* in the labor market, with $t=1$ denoting the first year of a worker in the market.[4] A worker is indexed by $n \in \mathbb{N}$ and endowed with time-invariant and, as described later, time-varying characteristics. As for the former, $H_{n,1}$ collects characteristics, such as gender, race, and education, that are observed by workers, firms, and the econometrician and capture a worker's human capital before entry into the labor market. Worker n is also characterized by other time-invariant skills that are unobserved by the econometrician and consist of two components: e_n, with support $\mathcal{E}$, and θ_n, with support Θ. They represent productive characteristics that can influence worker n's human capital process as well as output when employed.[5] They differ because e_n is observed by both workers and firms, whereas θ_n is initially unknown to them, but is gradually and symmetrically learned by workers and firms based on the observation of worker n's realized output at the end of each period, as described later. We think of e_n as describing an observed general level of productivity at entry into the labor market—for instance, a worker's baseline proficiency, say, knowledge of specific software, that can be gleaned from the worker's curriculum—and of θ_n as a latent productivity attribute—such as creativity, leadership or problem-solving skills, or social aptitude with clients and colleagues—that is initially uncertain and only gradually revealed as a worker is employed. Hereafter, we refer to e_n as worker's n *efficiency* and θ_n as the worker's *ability*. In the model, we impose no restrictions

[3] The assumption that the output market is perfectly competitive is without loss of generality, as we focus on firms' and workers' decisions in the labor market. That production is constant-returns-to-scale in the only input of workers' labor is likewise inconsequential here, because we do not separately model other non-labor input decisions. Maintaining that output is separable across workers, namely, that there are no complementarities in production across workers, is standard in the class of models we study and, more generally, in dynamic matching models. It rules out team production or shared-capacity and coordination constraints, which would make workers' marginal products interdependent.

[4] Some workers may first appear in the dataset after their initial entry into the labor market, which would affect the identification of the initial beliefs about their ability—the initial prior. Specifically, if workers are observed only after their labor market entry, our methodology simply recovers the prior belief about worker n's ability (denoted by θ_n below) as of worker n's first appearance in the data rather than as of the worker's labor market entry. But once the learning process is recovered, the unobserved initial prior could be backed out by Bayes' rule.

[5] The characteristics e_n and θ_n can affect worker n's performance at any firm. Some degree of generality in e_n and θ_n is essential to generate realistic job mobility patterns. In particular, if θ_n were purely firm-specific and so independent across firms, then workers would change jobs predominantly upon poor performance. In the data, by contrast, highly-performing workers are those frequently observed switching jobs both within and across firms.

on the support $\mathcal{E}$ of e_n, whereas, for simplicity of exposition, we let θ_n take values in $\Theta := \{\bar{\theta}, \underline{\theta}\}$ interpreted as high ($\bar{\theta}$) and low ($\underline{\theta}$) ability.[6] Although we ignore the possibility that a worker is not employed, it is easy to interpret one job as the alternative of non-employment.

Human Capital Process. Worker n accumulates human capital in the labor market through a process that depends on the worker's initial characteristics $(H_{n,1}, e_n, \theta_n)$ and employment history. Formally, when employed by firm $d \in \mathcal{D}$ in period t, worker n has human capital $H_{n,t}(d)$ given by

$$H_{n,t}(d) = h_d(H_{n,1}, I_n^{t-1}, e_n, \eta_{n,t}(d), \epsilon_{n,t}(d)), \tag{1}$$

where $h_d(\cdot)$ is a d-specific function (in logs) known to workers and firms but unknown to the econometrician.[7] As for the arguments of $h_d(\cdot), H_{n,1}$ and e_n were defined earlier. I_n^{t-1} is any function of worker n's employment history up to and including period $t-1$, observed by workers, firms, and the econometrician at the *beginning* of t—for example, tenure at all the firms by which worker n has been employed.[8] Note that when $h_d(\cdot)$ does not depend on I_n^{t-1}, no human capital accumulation takes place. When instead $h_d(\cdot)$ depends on I_n^{t-1}, new human capital is acquired and (1) also expresses the law of motion of human capital in cumulative form. We impose no restrictions on how (1) depends on I_n^{t-1}, allowing for a very general specification. For example, experience acquired at different firms or stages of working life can have a different impact on the stock of human capital, consistent with the notion of a flexible malleability of human capital across firms and over the life cycle.

The argument $\epsilon_{n,t}(d)$ of $h_d(\cdot)$ is an idiosyncratic, exogenous, and d-specific productivity (or amenity) shock, akin to a total factor productivity shock outside of a worker's control, whose distribution can vary across d and t. The shocks $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$ across all firms are realized at the *beginning* of period t and observed by workers and firms but unobserved by the econometrician. The argument $\eta_{n,t}(d)$ is also a d-specific random variable unobserved by the econometrician, whose distribution can depend on $(H_{n,1}, e_n, \theta_n, d)$.[9] Unlike the other arguments of $h_d(\cdot)$, $\eta_{n,t}(d)$ is observed by workers and firms at the *end* of period t after production occurs at the employing firm d—we interpret it as

[6] We consider a discrete specification for θ_n as our leading case also to allow the variance of posterior beliefs about ability, a key determinant of the variance of wages, to vary stochastically over time unlike in the standard normal prior-signal conjugate case, in which the variance of posterior beliefs deterministically declines with the number of signals observed. We can allow e_n to vary over time. For instance, it is straightforward to specify e_n as bivariate, with one dimension following a k-th-order Markov process; see Low et al. (2010) for a similar formulation.

[7] The function $h_d(\cdot)$ may also be time-varying, in which case we would express it as $h_{d,t}(\cdot)$; we omit this dependence just to simplify notation. Note that the domain of $h_d(\cdot)$ is allowed to vary over time through I_n^{t-1}, $\eta_{n,t}(d)$, and $\epsilon_{n,t}(d)$.

[8] Following Cunha et al. (2010), we can let the function I_n^{t-1} of worker n's employment history be unobserved by the econometrician and estimate it through a factor model, if proxies for human capital investments are available. Since they are not consistently so in the LEHD data we use, we treat I_n^{t-1} as observed by the econometrician.

[9] That the distribution of $\eta_{n,t}(d)$ does not depend on I_n^{t-1} is just to show how the model can be identified from very short panel data on wages and jobs. With panels increasing with I_n^{t-1}, allowing for such a dependence is straightforward.

a stochastic total factor productivity term that is on average affected by a worker's ability. That is, upon observing realized output $Y_{n,t}(d)$, defined below in (2), at the end of t, workers and firms extract from it the component $\eta_{n,t}(d)$. Given that $\eta_{n,t}(d)$ depends on θ_n, the realized value of $\eta_{n,t}(d)$ is used by workers and firms to update their beliefs about θ_n, as described later in this section. Hereafter, we refer to $\eta_{n,t}(d)$ as a *performance signal* or simply *signal*. Importantly, $\eta_{n,t}(d)$ depends on θ_n only stochastically via its distribution. Since different realizations of $\eta_{n,t}(d)$ are possible for a given θ_n, $\eta_{n,t}(d)$ provides a *noisy* signal about θ_n. If instead $\eta_{n,t}(d)$ were a deterministic function of θ_n, then learning about θ_n could be trivially complete after just one period of observation of output. For this reason, we do not allow the function $h_d(\cdot)$ to include θ_n among its arguments.

As an example, think of $\epsilon_{n,t}(d)$ as a beginning-of-period match-specific shock, such as a surge in orders for the product or service that worker n is responsible for, an unscheduled assignment, or a transitory health condition, known to workers and firms before the worker starts working in t. By contrast, $\eta_{n,t}(d)$ is an end-of-period match-specific shock revealed by realized output, say, how successfully the worker has carried out the responsibilities of the job or handled clients' needs.

We assume that $\eta_{n,t}(d)$ takes values in $\mathcal{N} := \{\bar{\eta}, \underline{\eta}\}$, interpreted as a high ($\bar{\eta}$) and a low ($\underline{\eta}$) signal. As with θ_n, this binary specification is adopted just for expositional simplicity and can be easily relaxed in the model to allow $\eta_{n,t}(d)$ to be scalar or multidimensional, either discrete or continuous. We maintain that the function $h_d(\cdot)$ is invertible in $\eta_{n,t}(d)$, as required in any well-behaved signal-extraction problem. This property ensures that after output $Y_{n,t}(d)$, defined below in (2), is observed at the end of t at the employing firm d, workers and firms can *uniquely* recover from it the realization of its component $\eta_{n,t}(d)$, through which θ_n stochastically affects $Y_{n,t}(d)$. In this case, the signal $\eta_{n,t}(d)$ about ability θ_n, which firms and workers use to learn about θ_n, is well defined.

Output Technology. Normalizing a worker's labor supply to one, (1) also represents the (log) output $Y_{n,t}(d)$ of worker n when employed by firm $d \in \mathcal{D}$ in period t,

$$Y_{n,t}(d) = h_d(H_{n,1}, I_n^{t-1}, e_n, \eta_{n,t}(d), \epsilon_{n,t}(d)), \tag{2}$$

which is realized and observed by workers and firms at the *end* of period t. Note that as the function $h_d(\cdot)$ is d-specific and the distributions (and realizations) of its random components $\eta_{n,t}(d)$ and $\epsilon_{n,t}(d)$ depend on d, firms are ex-ante differentiated by their output and human capital technologies. We discuss next how firms differ in their technologies of information generation.

Information Technology (Learning Process). As is standard in the models we nest, we focus on the case of *symmetric* learning: all firms and workers share a common belief about a worker's unknown ability θ_n in each period t. Formally, at the *beginning* of $t = 1$, firms and workers have a common prior belief that worker n is of high ability ($\theta_n = \bar{\theta}$), denoted by $P_{n,1} := \Pr(\theta_n = \bar{\theta} | H_{n,1}, e_n)$. Such a prior need not coincide with the true (conditional) distribution of θ_n and may incorporate any learning about θ_n that has occurred before entry into the labor market, for instance, based on a worker's educational attainment. At the *end* of $t \geq 1$, firms and workers observe the realization of $Y_{n,t}(d)$ at the employing firm $d \in \mathcal{D}$, extract from it the (unique) realization of the signal $\eta_{n,t}(d)$, and use it to update their common beliefs about θ_n by Bayes' rule. Since signals are independent over time conditional on a worker's ability θ_n, initial human capital $H_{n,1}$, efficiency e_n, and history of past jobs, the updated belief that $\theta_n = \bar{\theta}$ at the *beginning* of period $t + 1$ can be obtained recursively as

$$P_{n,t+1} = \begin{cases} \frac{\alpha(H_{n,1},d,e_n)P_{n,t}}{\alpha(H_{n,1},d,e_n)P_{n,t}+\beta(H_{n,1},d,e_n)(1-P_{n,t})} & \text{if } \eta_{n,t}(d) = \bar{\eta}, \\ \frac{[1-\alpha(H_{n,1},d,e_n)]P_{n,t}}{[1-\alpha(H_{n,1},d,e_n)]P_{n,t}+[1-\beta(H_{n,1},d,e_n)](1-P_{n,t})} & \text{if } \eta_{n,t}(d) = \underline{\eta}, \end{cases} \tag{3}$$

where $P_{n,t}$ is the belief that $\theta_n = \bar{\theta}$ at the beginning of period t; $\alpha(H_{n,1}, d, e_n)$ is the probability of a high signal $\bar{\eta}$ in period t at the employing firm d conditional on high ability ($\theta_n = \bar{\theta}$), $H_{n,1}$, and e_n; and $\beta(H_{n,1}, d, e_n)$ is the analogous probability conditional on low ability ($\theta_n = \underline{\theta}$), $H_{n,1}$, and e_n.

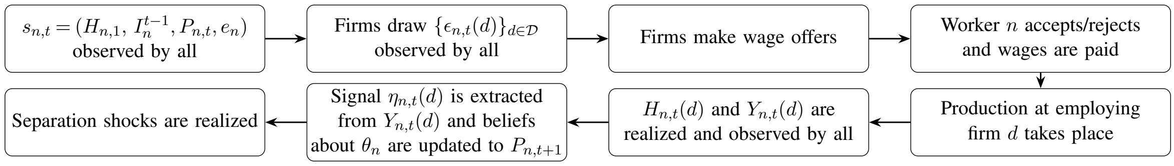
Importantly, since the probabilities $\alpha(H_{n,1}, d, e_n)$ and $\beta(H_{n,1}, d, e_n)$ may vary across firms d, performance at different firms can differ in its informativeness about θ_n.[10] Hence, firms are ex-ante differentiated not only by their output and human capital technologies but also by their information technologies. For example, observing the same high signal $\bar{\eta}$ (respectively, low signal $\underline{\eta}$) at a problem-solving job—say, troubleshooting unexpected issues—may raise (respectively, lower) more the posterior probability that a worker is of high ability than observing $\bar{\eta}$ (respectively, $\underline{\eta}$) at a highly standardized job—say, processing routine transactions. Then, output is more informative about ability in a Blackwell sense (Blackwell, 1951) in the former case than in the latter case.[11]

Expected Output. At the beginning of each period t, firms and workers base their wage and employment decisions on their knowledge of a worker's initial human capital $H_{n,1}$, investments in human capital through experience at past employing firms, as summarized by I_n^{t-1}, efficiency e_n, and productivity shocks $\{\epsilon_{n,t}(d)\}_{d \in \mathcal{D}}$. Instead, the worker's performance signal $\eta_{n,t}(d)$ about θ_n is observed by workers and firms only at the *end* of t, once it is extracted from realized output $Y_{n,t}(d)$ at the employing firm d after production takes place. Thus, at the beginning of t, firms and workers' decisions

[10] See footnote 9 for the possibility that $\alpha(\cdot)$ and $\beta(\cdot)$ also depend on I_n^{t-1}.

[11] We refer here to the stochastic order introduced by Blackwell (1951) whereby action (experiment) k is *more informative* than action (experiment) k' if the posterior beliefs about the relevant state of the world, here ability, reached if k is chosen are a mean-preserving spread of the posterior beliefs reached if k' is chosen.

Figure 1: Timing of Stage Game in a Period



are made given their expectations about $\{Y_{n,t}(d)\}_{d\in\mathcal{D}}$, which are informed by their beliefs about θ_n, since $Y_{n,t}(d)$ depends on θ_n through $\eta_{n,t}(d)$.[12] More formally, let $\epsilon_{n,t} := (\epsilon_{n,t}(d) : d \in \mathcal{D})$ denote the vector of shocks realized and observed by firms and workers, and let $s_{n,t} := (H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$ collect the other variables observed by firms and workers, at the beginning of t. Then,

$$\mathbb{E}(Y_{n,t}(d)|s_{n,t}, \epsilon_{n,t}) = \mathbb{E}_{\eta_{n,t}(d)}[h_d(H_{n,1}, I_n^{t-1}, e_n, \eta_{n,t}(d), \epsilon_{n,t}(d))|s_{n,t}, \epsilon_{n,t}] := y(d, s_{n,t}, \epsilon_{n,t}(d)) \quad (4)$$

is worker n's *expected* output at each firm $d \in \mathcal{D}$, where the expectation is over the signal $\eta_{n,t}(d)$ given $(s_{n,t}, \epsilon_{n,t})$ and so over ***high***, $\overline{Y}_{n,t}(d) = h_d(H_{n,1}, I_n^{t-1}, e_n, \bar{\eta}, \epsilon_{n,t}(d))$, and *low*, $\underline{Y}_{n,t}(d) = h_d(H_{n,1},$-$I_n^{t-1}, e_n, \underline{\eta}, \epsilon_{n,t}(d))$, output, with $y(d, s_{n,t}, \epsilon_{n,t}(d)) = \underline{Y}_{n,t}(d) + [\overline{Y}_{n,t}(d) - \underline{Y}_{n,t}(d)]P_{n,t}$.

Separation. At the end of each period t, workers leave the labor market considered with probability $\varsigma(H_{n,1}, I_n^{t-1}, d)$, which depends on their initial human capital $H_{n,1}$, human capital investments while in the labor market as summarized by I_n^{t-1}, and last employing firm $d \in \mathcal{D}$.

Timing. Figure 1 illustrates the timing of the stage game in each period t. At the start of period t, productivity shocks are realized. Next, firms simultaneously submit wage offers to workers for the period. Then, each worker decides which offer to accept and wages are paid. Finally, production occurs; human capital and output are realized; beliefs about ability are updated based on the signal extracted from realized output; and separation shocks are realized. Note that since firms commit to the period wage offers they make, the timing of wage payments within the period is immaterial.

Equilibrium. Given the absence of complementarities in production among workers, in characterizing equilibrium, we focus without loss on the competition among all firms for one worker at a time. We adopt as equilibrium concept a refinement of the notion of Markov perfect equilibrium, which we term Robust Markov perfect equilibrium (RMPE).[13] 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

[12] Neither realized output in (2) nor expected output in (4) explicitly depend on worker n's ability θ_n, but the frequency of high and low output depends on θ_n through the signal probabilities $\alpha(H_{n,1}, d, e_n)$ and $\beta(H_{n,1}, d, e_n)$ defined above.

[13] With one-job firms, this concept is equivalent to the notion of *cautious* MPE introduced by Bergemann and Välimaki (1996) for a game between two one-product firms. We extend it to the case of multi-job firms in the general version of our model (Appendix F.2 of the Online Supplementary Material). See also Pastorino (2024) and Kehoe et al. (2024).

discounted value of profits; (*iii*) beliefs are 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. As multiple MPEs may exist, condition (*iv*) selects one equilibrium with the desirable robustness properties that we discuss below. An RMPE exists, is unique, but does not need to be efficient in general (Bergemann and Välimaki, 1996).[14]

More formally, the state that firms face when making their wage offers to worker n, after productivity shocks $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$ are realized at all firms $d \in \mathcal{D}$, consists of $(s_{n,t}, \epsilon_{n,t})$, with $s_{n,t} = (H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$, whereas the state that worker n faces when deciding on firms' wage offers consists of $(s_{n,t}, \epsilon_{n,t})$ and the collection of offers received.[15] We denote by $w_{n,t,d} := w_d(s_{n,t}, \epsilon_{n,t})$ the wage offer strategy of a generic firm d. We denote by $l_{n,t,d} := l_d(s_{n,t}, \epsilon_{n,t}, \{w_{n,t,d}\}_{d\in\mathcal{D}})$ 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 at a given state so that workers accept at most one offer each period. Given firms' strategies, worker n's strategy satisfies

$$\begin{aligned}\tilde{W}(s_{n,t}, \epsilon_{n,t}, \{w_{n,t,d}\}_{d\in\mathcal{D}}) = \max_{\{l_d\}_{d\in\mathcal{D}}} \sum\nolimits_{d\in\mathcal{D}} l_d \times \Big\{ w_{n,t,d} \\ + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d)] \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[\tilde{W}(s_{n,t+1}, \epsilon_{n,t+1}, \{w_{n,t+1,d}\}_{d\in\mathcal{D}}) | s_{n,t}, d\Big] dF_{\epsilon_{n,t+1}} \Big\}.\end{aligned} \tag{5}$$

In (5), l_d is equal to one if worker n accepts firm d's offer and zero otherwise, δ is the discount factor, and $\varsigma(H_{n,1}, I_n^{t-1}, d)$ is the probability that worker n exogenously leaves the labor market at the end of period t, given the worker's initial human capital $H_{n,1}$, past human capital investments I_n^{t-1}, and last employing firm $d \in \mathcal{D}$. The conditional expectation $\mathbb{E}(\cdot)$ is taken over the value of the posterior beliefs in $t+1$, $(P_{n,t+1}, 1 - P_{n,t+1})$, and $F_{\epsilon_{n,t}}$ is the cumulative distribution function (CDF) of the vector of productivity shocks $\epsilon_{n,t}$. We assume that $\epsilon_{n,t}$ is independent of (a) $(s_{n,t-1}, \epsilon_{n,t-1}, \{w_{n,t-1,d}\}_{d\in\mathcal{D}}, \{l_{n,t-1,d}\}_{d\in\mathcal{D}})$ conditional on $s_{n,t}$; and (b) $s_{n,t}$. Condition (a) requires $\epsilon_{n,t}$ to be a true innovation in the sense that it is independent of all lagged variables given $s_{n,t}$, and condition (b) requires $\epsilon_{n,t}$ to be completely exogenous with respect to the period state $s_{n,t}$. These two conditions, which we maintain throughout, are standard in dynamic discrete choice models. Nevertheless, persistence in the state arises over time through I_n^{t-1}, observed by the econometrician, and $(P_{n,t}, e_n)$, unobserved by the econometrician, with $P_{n,t}$ serially correlated and endogenously evolving with worker n's job history, and e_n invariant over time. This rich structure of persistent

[14] Whereas efficiency is easy to establish in the two-firm one-job case, it does not necessarily extend to the case of more than two firms or firms with multiple jobs. See Appendix F of the Online Supplementary Material.

[15] In the finite-horizon case, a worker's labor market experience is part of the state. Since experience is already included in I_n^{t-1}, we consider here interchangeably the infinite-horizon case or the finite-horizon case with $t < T$.

unobserved heterogeneity captured by $(P_{n,t}, e_n)$ raises unique challenges from an identification perspective. We return to this point in Section 3 when discussing the empirical content of our model.

Given worker n's and its competitors' strategies, the strategy of a generic firm $d \in \mathcal{D}$ satisfies

$$\begin{aligned}\Pi_d(s_{n,t}, \epsilon_{n,t}) = \max_w \Big(l_{n,t,d} \times \Big\{ y(d, s_{n,t}, \epsilon_{n,t}(d)) - w \\ + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d)] \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[\Pi_d(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, d\Big] dF_{\epsilon_{n,t+1}} \Big\} \\ + \sum\nolimits_{d' \in \mathcal{D} \setminus \{d\}} l_{n,t,d'} \Big\{ \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d')] \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[\Pi_d(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, d'\Big] dF_{\epsilon_{n,t+1}} \Big\} \Big).\end{aligned} \tag{6}$$

Without condition (*iv*) in the definition of equilibrium, multiple MPE arise that are qualitatively similar in that they are characterized by the same allocations of workers to firms and paid wages, thus 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 level at which they are indifferent between not employing and employing a worker, without affecting on-path allocations or payoffs. Condition (*iv*) resolves this trivial multiplicity by requiring that non-employing firms offer wages that make them exactly indifferent between not employing and employing worker n.[16] Formally, condition (*iv*) requires that if firm $d \in \mathcal{D}$ employs worker n at state $(s_{n,t}, \epsilon_{n,t})$, then

$$\begin{aligned}\delta[1-\varsigma(H_{n,1}, I_n^{t-1}, d)] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_{d'}(\cdot) | s_{n,t}, d] dF_{\epsilon_{n,t+1}} \\ = \max_w \Big\{ y(d', s_{n,t}, \epsilon_{n,t}(d')) - w + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d')] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_{d'}(\cdot) | s_{n,t}, d'] dF_{\epsilon_{n,t+1}} \Big\},\end{aligned} \tag{7}$$

for any other firm $d' \in \mathcal{D}$. 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 unique, as shown in Proposition 1 below.

2.1 Equilibrium Wages

Here, we derive the equilibrium wage equation, starting by providing an informal intuition for it.

Intuition. Recall that in a static model of Bertrand price competition for a consumer between two firms with heterogeneous output technologies, the high-productivity (low-cost) firm sells to the consumer at a price equal to the cost of the low-productivity (high-cost) firm, making the consumer indifferent between the two sellers (Tirole, 1988). Analogously, in the static version of our model

[16] In particular, condition (*iv*) selects an equilibrium in a manner analogous to trembling-hand perfection (Selten, 1975). On the issue of robustness of MPEs and the refinement introduced by condition (*iv*), see also the discussion of Definition 4 in Bergemann and Välimaki (1996) for the two-firm one-job case.

with known worker ability and two firms with heterogeneous output technologies, a worker's wage at the employing (more productive) firm equals the worker's (expected) output at the competing (less productive) firm, making the worker indifferent between employment at the two firms. When firms have identical technologies—and productivity shocks—workers are paid their output, since the worker produces the same output at the employing and non-employing firms.

In the dynamic version of our model with unknown worker ability and two firms that differ in their output, human capital, and information technologies, the same indifference condition holds: in equilibrium, a worker must be indifferent between the employing firm and its competitor. However, a worker's preferences over the two firms now take into account the potentially different future wage returns from the human capital and the information that can be acquired through employment at each firm. But then a firm at which a worker can acquire substantial new human capital or information about ability can afford to pay a *lower* wage. Conversely, a firm offering limited opportunities for human capital or information acquisition must pay a *higher* wage. With more than two firms, a similar argument applies. The two firms effectively competing for a worker in a period are those offering the two highest expected present discounted values of wages. Like in a second-price auction, a worker's wage in period t then equals the expected output the worker would produce if hired by the firm ranked by the worker as "second best"—namely, offering the second-highest expected present discounted value of wages in the period—plus a *compensating differential*. The latter is a *premium* for the missed future wage returns from the human capital and the information that would have been acquired at the second-best firm or a *discount* for the greater future wage returns from the human capital and the information acquired at the first-best (employing) firm.

Formal Argument. Given the state $(s_{n,t}, \epsilon_{n,t})$ in period t, with $\epsilon_{n,t} := (\epsilon_{n,t}(d) : d \in \mathcal{D})$, consider the equilibrium ranking of firms based on the expected present discounted values of wages implied by their wage offers to worker n in t. Denote the firm whose offer yields the highest such value by $D_{n,t} := D(s_{n,t}, \epsilon_{n,t})$, namely, the *first-best firm* or employing firm, and the firm whose offer yields the second-highest such value by $D'_{n,t} := D'(s_{n,t}, \epsilon_{n,t})$, namely, the second-best firm or best non-employing firm, both expressed as functions of $(s_{n,t}, \epsilon_{n,t})$. Let $V_d(s_{n,t}, \epsilon_{n,t})$ be the expected present discounted value of match surplus or *match surplus value* generated by worker n and firm d at state $(s_{n,t}, \epsilon_{n,t})$, defined as the sum of the worker's expected present discounted value of wages and firm d's expected present discounted value of profits. We now formally characterize equilibrium wages.

Proposition 1 (Equilibrium Wage). *At state $(s_{n,t}, \epsilon_{n,t})$ in period t, worker n's equilibrium wage is*

$$w_{n,t} := w(s_{n,t}, \epsilon_{n,t}) = y(D'_{n,t}, s_{n,t}, \epsilon_{n,t}(D'_{n,t})) + \Psi(D_{n,t}, D'_{n,t}, s_{n,t}), \quad \textit{with} \tag{8}$$

$$
\begin{aligned}
\Psi(D_{n,t}, D'_{n,t}, s_{n,t}) :=& \delta[1-\varsigma(H_{n,1}, I_n^{t-1}, D'_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, D'_{n,t}] dF_{\epsilon_{n,t+1}} \\
& - \delta[1-\varsigma(H_{n,1}, I_n^{t-1}, D_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, D_{n,t}] dF_{\epsilon_{n,t+1}}.
\end{aligned}
$$

According to Proposition 1, a worker's wage is the sum of two terms: $y(D'_{n,t}, s_{n,t}, \epsilon_{n,t}(D'_{n,t}))$, which is the one-period expected output at the second-best firm $D'_{n,t}$ after the vector of productivity shocks $\epsilon_{n,t}$ is realized, and $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$, a *compensating differential*, which is the difference between two continuation match surplus values. The first such value is the future match surplus value that would be generated by worker n and firm $D'_{n,t}$ from period $t+1$ on, had the worker been employed by firm $D'_{n,t}$ in t—off the equilibrium path. The second such value is the future match surplus value generated by worker n and firm $D'_{n,t}$ from period $t+1$ on, after the worker is employed by firm $D_{n,t}$ in t—on the equilibrium path. The term $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$ compensates the worker for any difference in the prospects of human capital or information acquisition, or in the risk of exogenous separation, between firms $D_{n,t}$ and $D'_{n,t}$. In fact, this term would be zero if the processes of human capital and information acquisition as well as the risk of exogenous separation were identical across firms. In this case, employment at any firm would entail the same *future* wage returns, and workers would simply be paid their expected output at the second-best firm. In general, however, wages can be lower or higher than expected output due to the compensating differential.

Compensating Differential. We characterize two properties of the compensating differential that are key to our econometric strategy: its nonmonotonicity and its sign with respect to $s_{n,t}$. The nonmonotonicity is most transparent when this differential is mostly governed by differences across firms in the information that output provides about ability, that is, when the human capital process and exogenous separation rates are sufficiently similar across firms. Intuitively, the value of signals about ability is largest when uncertainty is highest, that is, for a prior $P_{n,t}$ far from zero and one, and so is the option value of the information acquired at the first-best firm or lost at the second-best firm.[17]

Proposition 2 (Nonmonotonicity of Compensating Differential). *When exogenous separation rates $\varsigma(H_{n,1}, I_n^{t-1}, d)$ for each $H_{n,1}$ and I_n^{t-1} and the human capital process are sufficiently similar across firms, the compensating differential in wages converges to zero as $P_{n,t}$ converges to zero or one.*

Similarly, a clear characterization of the sign of the compensating differential can be obtained when exogenous separation rates and the human capital process are sufficiently similar across firms. In this situation, differences in the option value of employment across firms are solely due to differences in the information that their output provides about ability. In particular, whenever the job at the

[17] An analogous result holds when firms differ only in their human capital technologies.

employing firm $D_{n,t}$ is *more* informative about ability than the job at the competing firm $D'_{n,t}$ in the Blackwell sense, the compensating differential is *negative*, whereas whenever it is *less* informative, the compensating differential is *positive*. Intuitively, firm $D_{n,t}$ pays a worker less than under static competition when employment at it leads to greater learning about ability—in this case, the worker experiences an informational *gain* with employment at $D_{n,t}$ that is discounted in the paid wage. Instead, firm $D_{n,t}$ pays a premium over the competitive wage when employment at it forces the worker to forgo some learning about ability—in this case, the worker experiences an informational *loss* with employment at $D_{n,t}$ and hence is compensated for it. We prove this result in the simplest possible, and empirically relevant, case of two firms in the labor market consisting each of two or more jobs. The proof straightforwardly extends to the general case.

Proposition 3 (Sign of Compensating Differential)**.** *Suppose that* $|\mathcal{D}| = 2$*, each firm consists of more than one job, and* $P_{n,t} \in (0,1)$*. When exogenous separation rates* $\varsigma(H_{n,1}, I_n^{t-1}, d)$ *for each* $H_{n,1}$ *and* I_n^{t-1} *and the human capital process are sufficiently similar across firms, the compensating differential in wages is negative (resp., positive) whenever performance at firm* $D_{n,t}$ *is more (resp., less) informative about ability in the Blackwell sense than performance at firm* $D'_{n,t}$*.*

2.2 Implications for Measuring Labor Market Sorting and Monopsony Power

In this section, we discuss the issues related to the measurement of worker sorting across firms and the markdown of wages relative to workers' output that arise in our framework.

Labor Market Sorting. Fundamentally, the patterns of labor market sorting are governed by the complementarity between firms and workers in production, as determined by how the (expected) output $y(\cdot)$ of a firm-worker match depends on the latent characteristics e_n of a worker and of a worker's employing firm $D_{n,t}$, after controlling for the worker and firm attributes observed in the data.[18] The stronger the complementarity in production between workers' and firms' residual characteristics, the greater the potential for assortative matching between firms and workers, and, if the market is sufficiently competitive, the greater the degree of labor market sorting.

Traditional statistical measures of sorting as in AKM represent the dependence of output, and so wages, on e_n and $D_{n,t}$ through a linear and additively separable specification,

$$w_{n,t} = e_n + \psi(D_{n,t}) + g(X_{n,t};\beta) + \varepsilon_{n,t}, \tag{9}$$

[18] Ability θ_n is another worker attribute that is unobserved by the econometrician. Unlike e_n, however, θ_n is initially *unknown* to firms and workers and is only gradually learned over time. As a result, firms and workers do not directly sort on θ_n but rather on their beliefs $(P_{n,t}, 1 - P_{n,t})$ about θ_n. Conceptually, one could thus include $P_{n,t}$, rather than θ_n, alongside e_n in the calculation of a sorting measure. For simplicity of exposition, however, we do not do so here. In Section 5, we discuss the role of uncertainty about θ_n for sorting patterns.

where e_n is a worker fixed effect; $\psi(D_{n,t})$ is a fixed effect for the employing firm $D_{n,t}$ in t; $g(\cdot;\beta)$ is a parametric function of observable worker characteristics $X_{n,t}$, namely, $H_{n,1}$ and I_n^{t-1}, with possibly additional covariates that may measure beliefs $(P_{n,t}, 1-P_{n,t})$ about worker ability $\theta_n \in \{\overline{\theta}, \underline{\theta}\}$;[19] and $\varepsilon_{n,t}$ is an exogenous scalar shock.The parameters $(\{e_n\}_n, \{\psi(d)\}_d, \beta)$ can be estimated by OLS under appropriate normalizations and corrections for the potential bias due to low worker mobility across firms. Sorting is then commonly summarized by the share of the variance of wages accounted for by the covariance between worker and firm effects, $\mathrm{S}_{\mathrm{AKM}} := 2\mathrm{Cov}(e_n, \psi(D_{n,t}))/\mathrm{Var}(w_{n,t})$, and calculated using the OLS estimates of worker and firm effects, $\widehat{\mathrm{S}}_{\mathrm{AKM}} := 2\mathrm{Cov}(\widehat{e}_n, \widehat{\psi}(D_{n,t}))/\mathrm{Var}(w_{n,t})$. We next argue that for our class of models, $\mathrm{S}_{\mathrm{AKM}}$ is conceptually an inaccurate descriptor of the degree of labor market sorting and that $\widehat{\mathrm{S}}_{\mathrm{AKM}}$ is a biased estimator of it.

S_{AKM} *Inaccurate Descriptor of Sorting.* This measurement issue is two-fold. First, as is known, $\mathrm{S}_{\mathrm{AKM}}$ is an appropriate measure of sorting when the linear, additively separable approximation to wages in (9), in terms of worker and firm fixed effects, provides a good description of the underlying output technology—such a description is exact when the output technology is log-linear in workers' and firms' latent attributes. Since we do not impose this functional form in (2) and (4), in general (9) may offer a poor approximation to the wage process in our setting, and $\mathrm{S}_{\mathrm{AKM}}$ may thus be an inaccurate descriptor of sorting. Second, in models of imperfect competition such as ours in which firms have wage-setting power, a worker's wage reflects the output that a worker would produce at their best outside option—the second-best firm—rather than the worker's output at the employing firm, as shown in Proposition 1. As a result, wages may not well capture how unobserved worker and employing-firm attributes contribute to match output. In such a setting, inferring labor market sorting indirectly from wage data, as S_{AKM} does, can thus be misleading.[20]

Remark 1 (S_{AKM} Inaccurate Descriptor of Sorting)**.** *When the output technology is not log-linear in worker and firm latent attributes or in imperfectly competitive labor markets in which wages reflect a worker's productivity at the (best) competitor of the worker's employer, S_{AKM} may fail to recover the underlying complementarity in production between worker and firm latent attributes.*

$\widehat{S}_{\mathrm{AKM}}$ *Biased Estimator of* S_{AKM}. Even when $\mathrm{S}_{\mathrm{AKM}}$ is an accurate descriptor of the degree of sorting, we now argue that $\widehat{S}_{\mathrm{AKM}}$ is nonetheless a biased estimator of S_{AKM} when the data are generated by a wage equation as in (8). It is well known that $\widehat{S}_{\mathrm{AKM}}$ can be biased for S_{AKM} when

[19] Merely for simplicity of exposition, we assume—for the purposes of this section only—that the researcher can observe, or accurately proxy for, $P_{n,t}$. Under this assumption, $P_{n,t}$ can be interpreted as a standard covariate. Our identification arguments developed in Section 3 are more general, as they allow $P_{n,t}$ to be unobserved.

[20] As an alternative to $\mathrm{S}_{\mathrm{AKM}}$, in the empirical analysis of Section 5, we use a classic sorting measure based on the supermodularity of output in worker and firm unobserved characteristics.

worker mobility is limited (Andrews et al., 2008) and when workers' job choices $D_{n,t}$ depend on unobservables $\varepsilon_{n,t}$ (Bonhomme et al., 2023). Here, we emphasize an additional source of bias: the specification in (9) omits the compensating differential $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$ that enters the wage in (8). This compensating differential is an unknown function of $D_{n,t}$, $D'_{n,t}$, and $s_{n,t} := (H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$ that depends on e_n and $D_{n,t}$ in a nonlinear and nonseparable way.[21] Accounting for this compensating differential would require allowing the function $g(\cdot;\beta)$ to depend flexibly on e_n and $D_{n,t}$ in a manner that cannot be well accommodated by the log-linear AKM representation of worker and firm fixed effects in (9) nor by an interactive fixed-effect extension. The resulting misspecification—using (9) in place of (8)—biases the estimator of worker and firm effects and, in turn, $\widehat{S}_{\mathrm{AKM}}$.

Remark 2 ($\widehat{S}_{\mathrm{AKM}}$ Biased Estimator of S_{AKM})**.** *When the compensating differential is nonzero, AKM-type representations of the wage process yield bias in the estimators of worker and firm effects and, in turn, in $\widehat{S}_{\mathrm{AKM}}$ relative to S_{AKM}. Appendix C derives analytical expressions for this bias.*

Wage Markdown and Monopsony Power. The characteristics of the wage process in our framework have also implications for how firms' monopsony power affects wages and thus for how it should be measured. A growing literature emphasizes the importance of firms' wage-setting power in U.S. labor markets for the level and dispersion of wages across workers, and documents a high degree of monopsony power as implied by a low *wage markdown*, namely, a low ratio of a worker's wage to a worker's output (Seegmiller, 2021; Yeh et al., 2022; Lamadon et al., 2022; Berger et al., 2026). A common reading of this evidence is that U.S. labor markets are highly noncompetitive. The issue is that imperfect firm competition not only depresses wages relative to workers' productivity but also potentially distorts firms' employment choices and workers' labor supply decisions.

Through the lens of our framework, however, such a conclusion may not be warranted for two reasons. First, workers can dynamically acquire both human capital and information about their ability. These processes of human capital and information accumulation lead to a natural force that depresses current wages when workers are employed by firms at which substantial human capital and information can be acquired. Put differently, the option value of higher *future wages*—stemming from the investment in human capital and information that employment affords—is priced into *current wages* via the compensating differential. In this case, a low wage markdown, especially early in the life cycle, can simply reflect high future wage returns to current investments in human capital

[21] From (8), it is apparent that the compensating differential is an endogenous function of e_n and $D_{n,t}$ through the second-best firm's match surplus value. This value is given by the sum of the values of the intertemporal maximization problems that define worker and (second-best) firm best responses illustrated above. These problems and their values depend nonlinearly not only on the state but also on all the primitives governing worker and firm productivity. Thus, both match surplus value and the compensating differential are generally highly nonlinear in worker and firm attributes.

and information rather than labor market distortions. Indeed, the greater this option value, the more negative the compensating differential, and the smaller the markdown of wages relative to output.

Second, even in a static version of our framework ($\delta = 0$) in which the compensating differential is zero, workers are *not* paid their output at their employing firms. Instead, their wages reflect the output they would have produced at their second-most preferred firms, namely, their most-preferred firms among those whose offers they rejected, which isalways lower than output at the first-best firm in the static case. Thus, a low wage markdown can simply capture a worker's lower productivity at the second-best firm, namely, a low outside option, instead of any wage-setting frictions. Importantly, unlike in usual models of monopsony, the wage markdown in our framework can be greater than one. That is, whenever the compensating differential in wages is positive, wages are *higher* than workers' (expected) output. Intuitively, when workers sacrifice better human capital or informational opportunities by working at their employers, they are compensated for this loss. Although firms have market power, wages are then *marked up* relative to output.

Remark 3 (Wage Markdown and Monopsony Power)**.** *When the compensating differential is negative (resp., positive), wages are marked down (resp., up) relative to a worker's expected output (inclusive of productivity shocks) at the employing firm. When the discount factor δ is small, the compensating differential is correspondingly small and wages are marked down relative to a worker's expected output (inclusive of productivity shocks).*

3 Overview of Identification

We provide here an overview of identification. We present the formal argument in Appendix B.

3.1 Econometric Model

Our model can be cast as an equilibrium dynamic generalized Roy model. As is standard in this empirical literature, we impose that wages are additively separable in the idiosyncratic productivity shock $\epsilon_{n,t}(d)$ by assuming that the human capital function $h_d(\cdot)$ in (1) and so the output function $Y_{n,t}(d)$ in (2) are (in logs) additively separable in $\epsilon_{n,t}(d)$—one instance is when human capital and output technologies are CES, with $\epsilon_{n,t}(d)$ a stochastic total factor productivity term.

Assumption 1 (Additive Separability)**.** *For each $d \in \mathcal{D}$, the output function $Y_{n,t}(d)$ defined in* (2) *satisfies* $Y_{n,t}(d) = h_d(H_{n,1}, I_n^{t-1}, e_n, \eta_{n,t}(d)) + \epsilon_{n,t}(d)$.

Under Assumption 1, worker n's observed (log) wage in period t is

$$w_{n,t} = \sum\nolimits_{(d,d')\in\mathcal{D}^2} \mathbb{1}\{D_{n,t} = d,\ D'_{n,t} = d'\}\big[y(d', s_{n,t}) + \Psi(d, d', s_{n,t}) + \epsilon_{n,t}(d')\big], \tag{10}$$

where $w_{n,t}(d, d') \coloneqq y(d', s_{n,t}) + \Psi(d, d', s_{n,t}) + \epsilon_{n,t}(d')$ denotes worker n's *potential wage* in period t when the first- and second-best firms are d and d', respectively.

Assumption 2 (Data)**.** *The joint distribution of* $\{w_{n,t}, H_{n,1}, D_{n,t}\}_{t=1}^{T}$ *is known.*

Assumption 2 describes the observation scheme maintained throughout. We assume access to a panel dataset of workers' wages $w_{n,t}$, initial characteristics $H_{n,1}$, and employment choices $D_{n,t}$. We keep T finite and allow the number of workers to be arbitrarily large (short panel). As mentioned in Section 2, the index t denotes labor market experience rather than calendar time, with $t = 1$ corresponding to a worker's first year in the market—see footnote 4 for the case in which a worker enters the dataset several years after entering the market. Likewise, $t = T$ does not represent the last year of a worker's experience but rather the finite observation window available in a given dataset. For most of our arguments, no minimum panel length is required. The only exception concerns the identification of the learning process, for which we require $T \geq 4$; see Appendix B.2 for details. For ease of notation, we assume that the panel is balanced. However, all arguments remain valid even with an unbalanced panel. Throughout, we maintain that e_n, $P_{n,t}$, and $\epsilon_{n,t}$ are unobserved by the econometrician. We do not require the econometrician either to observe performance signals $\{\eta_{n,t}(d)\}_{d\in\mathcal{D}}$ or output $\{Y_{n,t}(d)\}_{d\in\mathcal{D}}$ or to have access to credible proxies for these objects. In Section 4, we discuss how the availability of such information can nonetheless simplify estimation.

Objects of Interest. We semiparametrically identify *(i)* the law of motion of the state $s_{n,t}$, $\Pr(s_{n,t}| D_{n,t-1}, s_{n,t-1})$, including the learning process about a worker's ability θ_n characterized by the prior belief $(P_{n,1}, 1 - P_{n,1})$ about θ_n being high ($\bar{\theta}$) or low ($\underline{\theta}$) and by the distribution of signals conditional on θ_n—that is, the probabilities $\alpha(H_{n,1}, D_{n,t}, e_n)$ and $\beta(H_{n,1}, D_{n,t}, e_n)$ defined in Section 2; *(ii)* the conditional choice probabilities (CCPs), $\Pr(D_{n,t}|s_{n,t})$; *(iii)* the deterministic component of potential wages, $\varphi(D_{n,t}, D'_{n,t}, s_{n,t}) \coloneqq y(D'_{n,t}, s_{n,t}) + \Psi(D_{n,t}, D'_{n,t}, s_{n,t})$, defined as the sum of expected output $y(D'_{n,t}, s_{n,t})$ (net of productivity shocks) and the compensating differential $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$; and *(iv)* the expected output $y(D'_{n,t}, s_{n,t})$ and the human capital functions $\{h_d(H_{n,1}, I_n^{t-1}, e_n, \eta_{n,t}(d))\}_{d\in\mathcal{D}}$. Finally, we parametrically identify the distribution of the productivity shocks $\epsilon_{n,t} \coloneqq (\epsilon_{n,t}(d) : d \in \mathcal{D})$. Throughout, as is standard, we treat the discount factor δ as known. Recovering all these primitives is key to addressing the empirical questions about the degree of labor market sorting and the extent of firm monopsony power that we tackle in Section 5.

3.2 A Two-Step Approach

The identification of the model's primitives is complex because of workers' dynamic selection into jobs through the choice of $D_{n,t}$ based on multiple unobservables: *(i)* the idiosyncratic productivity shocks $\epsilon_{n,t}$,which are time varying and serially uncorrelated; *(ii)* worker efficiency e_n, which is time invariant;and *(iii)* workers' and firms' beliefs $(P_{n,t}, 1-P_{n,t})$ about worker ability θ_n, which are time varying, serially correlated, and endogenously evolving with workers' past job choices. Our approach builds on arguments for Roy models that leverage extremal quantiles of observed wages to recover the deterministic component of potential wages—the *extremal quantile step*—with a first step that is based on the identification of a mixture model representing the wage process—the *mixture step*.

The logic is as follows. We first represent the conditional distribution of wages $w_{n,t}$ given workers' job history $D_n^{t-1} := (D_{n,1}, \ldots, D_{n,t-1})$, current job choice $D_{n,t}$, and initial human capital $H_{n,1}$ as a mixture over latent workers' classes indexed by their efficiency e_n and history of performance signals $\eta_n^{t-1} := (\eta_{n,1}, \ldots, \eta_{n,t-1})$, where $\eta_{n,\tau} := \eta_{n,\tau}(D_{n,\tau})$ for each $\tau \in \{1, \ldots, t-1\}$. Under mild conditions, we show how to recover the components and weights of this mixture. We then exploit that the state $s_{n,t} := (H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$ evolves as a deterministic function of the same objects $(H_{n,1}, D_n^{t-1}, e_n, \eta_n^{t-1})$ indexing mixture weights to identify from these weights the law of motion of the state, $\Pr(s_{n,t}|D_{n,t-1}, s_{n,t-1})$, including the learning process.[22] Over time, these weights also encode the distribution of workers' employment histories for all possible combinations of worker and firm observed and unobserved characteristics and so pin down the CCPs, $\Pr(D_{n,t}|s_{n,t})$. Likewise, mixture components characterize the distributions of wages for all possible combinations of worker and firm observed and unobserved characteristics, which identify $\Pr(w_{n,t}|D_{n,t}, s_{n,t})$. Importantly, this mixture step resolves the issue of selection on the persistent unobservables e_n and $P_{n,t}$ part of $s_{n,t}$, since it recovers how they govern job choices and wages.

Once job choice and wage distributions are identified through the mixture step, the extremal quantile step resolves the remaining selection on the idiosyncratic shock $\epsilon_{n,t}$ and thereby recovers the deterministic component of potential wages $\varphi(\cdot)$ and, in turn, expected output $y(\cdot)$, the compensating differential $\Psi(\cdot)$, the human capital functions $\{h_d(\cdot)\}_{d\in\mathcal{D}}$, and the distribution of $\epsilon_{n,t}$.

In what follows, Sections 3.3 and 3.4 provide intuition for these two mixture and extremal quantile steps, respectively, based on a *simplified* version of the general wage equation in (10),

$$w_{n,t} = \sum\nolimits_{d\in\mathcal{D}} \mathbb{1}\{D_{n,t} = d\}\big[\varphi(d, s_{n,t}) + \epsilon_{n,t}(d)\big], \tag{11}$$

[22]Since I_n^{t-1} is a function of D_n^{t-1}, $P_{n,1}$ is a function of $(H_{n,1}, e_n)$, and $\{P_{n,t}\}_{t=2}^T$ is a function of η_n^{t-1}, $P_{n,1}$, $\alpha(H_{n,1}, D_{n,t-1}, e_n)$, and $\beta(H_{n,1}, D_{n,t-1}, e_n)$, conditioning on $(H_{n,1}, D_n^{t-1}, e_n, \eta_n^{t-1})$ amounts to conditioning on $s_{n,t}$.

which abstracts from the dependence of wages on the second-best firm $D'_{n,t}$. Hereafter, we denote by $w_{n,t}(d) := \varphi(d, s_{n,t}) + \epsilon_{n,t}(d)$ worker n's potential wage in period t when employed by firm d. Section 3.5 explains how we handle $D'_{n,t}$ in the general wage equation in (10).

3.3 The Mixture Step

The mixture step of our identification approach is a novel argument, which allows us to account for the complex selection patterns on multiple worker and firm unobservable characteristics that arise in our class of models. In this section, we focus on the main intuition behind this step.

Wage Mixture. For any firm $d \in \mathcal{D}$, by the law of total probability,

$$\Pr(w_{n,t} \leq w | H_{n,1}, D_n^{t-1}, D_{n,t} = d) = \sum_{(e,\eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}} \Pr(e_n = e, \eta_n^{t-1} = \eta^{t-1} | H_{n,1}, D_n^{t-1}, D_{n,t} = d) \times \Pr(w_{n,t} \leq w | H_{n,1}, D_n^{t-1}, D_{n,t} = d, e_n = e, \eta_n^{t-1} = \eta^{t-1}), \tag{12}$$

where $\mathcal{E}$ and $\mathcal{N}$ are the supports of e_n and $\eta_{n,t}$, respectively; $\eta_{n,\tau} := \eta_{n,\tau}(D_{n,\tau})$ for each $\tau \leq t-1$ and $\eta_n^{t-1} := (\eta_{n,1}, \ldots, \eta_{n,t-1})$; e and $\eta^{t-1} := (\eta_1, \ldots, \eta_{t-1})$ denote generic realizations of e_n and η_n^{t-1}, respectively; $\Pr(w_{n,t} \leq w | \cdot, e_n = e, \eta_n^{t-1} = \eta^{t-1})$, suppressing its dependence on the observables, denotes a representative mixture component; and $\Pr(e_n = e, \eta_n^{t-1} = \eta^{t-1} | \cdot)$ denotes the associated weight. We maintain that $|\mathcal{E}| < \infty$ and $|\mathcal{N}| < \infty$ so that (12) is a finite mixture. With $\mathcal{E}$ finite, e_n should be viewed as a worker "finite fixed effect" in the terminology of Bonhomme et al. (2019).[23]

Identifying this wage mixture amounts to uniquely recovering its components and weights from knowledge of the left side of (12). It is well known that mixture models such as (12) are not identified without additional restrictions. Intuitively, for a K-component mixture, the unknowns comprise K component distributions $F_1, \ldots, F_K$ and $K-1$ independent weights $(\pi_1, \ldots, \pi_K)$ satisfying $\sum_{k=1}^{K} \pi_k = 1$. By contrast, in the data, only a single distribution is observable—the left side of (12). Such a decomposition is thus undetermined, and additional assumptions are required for identification. A large literature provides sufficient conditions for the identification of mixture models. Some work has imposed a parametric structure on mixture components by restricting each of them to belong to a known, usually common, family, such as Normal or Gumbel; see Teicher (1961) for a classic reference. Other work has developed nonparametric or semiparametric identification strategies, which follow two main approaches. The first one relies on exclusion restrictions, that is, on the existence of variables that affect either mixture weights or components but not both, and thus help

[23] Finiteness of $\mathcal{E}$ is a standard specification (see Keane and Wolpin, 1997 for a classic example). It accommodates multidimensional efficiency types while restricting them to take finitely many values. Since the model has been described for a finite $\mathcal{N}$, we maintain this formulation in the econometric analysis as well. Appendix D discusses how our econometric approach can be extended to the case in which e_n and $\eta_{n,t}$ are continuous and multidimensional.

recover them separately (Henry et al., 2014; Compiani and Kitamura, 2016; Jochmans et al., 2017). The second approach exploits the joint distribution of the wage vector $(w_{n,1}, \dots, w_{n,T})$, rather than the wage distribution at each time t as in (12), and leverages assumptions about the intertemporal dependence structure of wages—such as conditional independence or Markovianity—together with the restriction that the number of latent classes is constant over time (Hall and Zhou, 2003; Allman et al., 2009; Kasahara and Shimotsu, 2009; Bonhomme et al., 2016a,b).

None of these approaches applies to our framework. Starting with the standard parametric approach, as the simplified wage equation in (11) makes clear, each mixture component $\Pr\big(w_{n,t} \leq w|\cdot, e_n, \eta_n^{t-1}\big)$ is essentially governed by the distribution of the shock $\epsilon_{n,t}(d)$ *conditional* on $(D_{n,t} = d, s_{n,t})$. Because workers select into jobs $D_{n,t}$ based on $s_{n,t}$ and the vector of shocks $\epsilon_{n,t}$, the distribution of $\epsilon_{n,t}(d)$ conditional on $(D_{n,t} = d, s_{n,t})$ is the outcome of workers and firms' intertemporal job and wage decisions. Thus, it will typically differ from the unconditional distribution of $\epsilon_{n,t}(d)$ and may not admit a standard parametric form.[24] We must then rely on identifying assumptions that accommodate highly flexible mixture component distributions.

The existing nonparametric or semiparametric approaches described are not suitable for our framework either. In the class of models we consider, exclusion restrictions do not naturally arise: any variable that affects the conditional distribution of worker-efficiency types and performance signals—and so mixture weights—also affects the conditional distribution of wages—and so mixture components—and viceversa. Also, although our model has a Markovian structure in that serial dependence in wages arises only from the state $s_{n,t}$, the number of latent classes of the period-t mixture representing equilibrium wages is not constant over time, unlike in the literature on mixture models that exploits the joint distribution of $(w_{n,1}, \dots, w_{n,T})$. In fact, the number of latent classes grows rapidly over time with the length of the signal history η_n^{t-1}.

In light of these considerations, we exploit a recent result by Aragam et al. (2020), which provides mild conditions for identification by requiring that the wage mixture in (12) admit a *clusterable finite-mixture* representation. Specifically, the wage distribution must be expressible as a finite mixture whose components $\{\Pr(w_{n,t} \leq w|\cdot, e_n = e, \eta_n^{t-1} = \eta^{t-1})\}_{(e,\eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}}$ are sufficiently distinct—and so clusterable in the precise sense defined by Aragam et al. (2020)—to be identifiable. A canonical and widely used example of this class is a finite mixture of continuous mixtures of Normals, with means and variances that differ sufficiently across components (Bruni and Koch, 1985; Aragam et al.,

[24] The unconditional distribution of $\epsilon_{n,t}(d)$ need not be closed under conditioning. By "closure under conditioning", we mean that if a random vector ϵ follows a parametric family F_θ, then for any selection event A defined in terms of ϵ, the conditional distribution $\epsilon|A$ still belongs to the same family, that is, $\epsilon|A \sim F_{\theta'}$ for some θ'. This requirement is satisfied only by a few families of distributions, such as i.i.d. type-I extreme value (Gumbel) distributions.

2020). In this case, each mixture component of (12) is itself a continuous mixture of Normals,

$$\Pr\big(w_{n,t} \leq w|\cdot, e_n = e, \eta_n^{t-1} = \eta^{t-1}\big) = \int \Phi\left(\frac{w-\mu}{\sigma}\right) d\pi_{e,\eta^{t-1}}\big(\mu, \sigma^2|\cdot\big),$$

with the mixing distributions $\{\pi_{e,\eta^{t-1}}(\mu, \sigma^2|\cdot)\}_{(e,\eta^{t-1})\in\mathcal{E}\times\mathcal{N}^{t-1}}$ satisfying a *non-overlapping support* condition in that for any two distinct indices $(e, \eta^{t-1}) \neq (e', \eta'^{t-1})$, the supports of the corresponding mixing laws over (μ, σ^2) are disjoint, that is, $\mathrm{supp}\big(\pi_{e,\eta^{t-1}}\big) \cap \mathrm{supp}\big(\pi_{e',\eta'^{t-1}}\big) = \emptyset$.[25]

Intuitively, finite mixtures of Normals are known to be identified (Teicher, 1963). Bruni and Koch (1985) show that this logic extends to finite mixtures of continuous mixtures of Normals, provided that the parameters (μ, σ^2) that govern each mixture component are from distinct (non-overlapping) regions of the parameter space across components. This condition rules out any observationally equivalent re-partitioning, or clustering, of the overall mixture, guaranteeing the unique recovery of its components and weights. Aragam et al. (2020) further extend this result to clusterable finite mixtures, which include the setting of Bruni and Koch (1985) as a special case. Since continuous mixtures of Normals are known to approximate any distribution arbitrarily well (Nguyen and McLachlan, 2019), we view the restriction to a clusterable finite-mixture representation as minimal. In particular, this class is well suited to modeling the selection-contaminated mixture distributions that arise in complex settings like ours when workers dynamically sort into jobs $D_{n,t}$ based on $\epsilon_{n,t}$.

As is well known, finite-mixture models are identified only up to a labeling of their weights and components, since their likelihood is invariant to permutations of them. Labeling the weights and components in (12) with respect to the latent variables (e_n, η_n^{t-1}) is needed *only* for the first four periods of the data so as to identify the learning process. Indeed, identifying the latter entails combining the mixture weights of the first four periods, which in turn requires a consistent labeling of these weights across these periods; see Appendix B. Once the learning process is identified, the remaining primitives—namely, the law of motion of the state, the CCPs, the deterministic wage component, the distribution of productivity shocks, expected output, and the human capital function—can be recovered without further linking mixture weights and components over time. To consistently label weights over the first four periods, we use the variances of the mixture components to order weights with respect to e_n, and the means of the mixture components to order weights with respect to η_n^{t-1}. The latter ordering follows from monotonicity conditions that are natural for the class of models we study. As higher-ability workers are more likely to produce high output ($\alpha(H_{n,1}, D_{n,t}, e_n) > \beta(H_{n,1}, D_{n,t}, e_n)$), expected output increases with the belief $P_{n,t}$ that a worker is of high ability for given $D_{n,t}$, $s_{n,t}$, and

[25] This condition requires that the means and the variances of the mixture components differ sufficiently across them, but it does *not* require that mixture components have disjoint supports. In fact, their supports may arbitrarily overlap.

$\epsilon_{n,t}(D_{n,t})$. In semiparametric contexts such as those we consider in Section 5, this monotonicity suffices for ordering purposes. In general, wages can be shown to increase with $P_{n,t}$ if the compensating differential is not too large, as we find in the data.[26]

State Law of Motion and CCPs. The weights of the mixture in (12) yield several key primitives of interest. In particular, these weights allow us to recover the distribution of performance signals conditional on a worker's ability θ_n—that is, the probabilities $\alpha(H_{n,1}, D_{n,t}, e_n)$ and $\beta(H_{n,1}, D_{n,t}, e_n)$ defined in Section 2—and the prior belief $(P_{n,1}, 1 - P_{n,1})$ about ability θ_n being high ($\bar{\theta}$) or low ($\underline{\theta}$). We then pin down the posteriors $\{P_{n,t}\}_{t=2}^{T}$ by recursively updating $P_{n,1}$ via (3), and so identify the learning process. The mixture weights are also instrumental in identifying the CCPs, $\Pr(D_{n,t}|s_{n,t})$.

To see this latter point, consider period $t = 2$ and note that by Bayes' rule,

$$\Pr(D_{n,2}|H_{n,1}, D_{n,1}, e_n, \eta_{n,1}) = \frac{\Pr(e_n, \eta_{n,1}|H_{n,1}, D_{n,1}, D_{n,2}) \Pr(D_{n,2}|H_{n,1}, D_{n,1})}{\Pr(e_n, \eta_{n,1}|H_{n,1}, D_{n,1})}, \tag{13}$$

where $\Pr(e_n, \eta_{n,1}|H_{n,1}, D_{n,1}, D_{n,2})$, the identified mixture weight in $t = 2$, is known at this stage, $\Pr(e_n, \eta_{n,1}|H_{n,1}, D_{n,1})$ is immediately identified from it, and $\Pr(D_{n,2}|H_{n,1}, D_{n,1})$ is known by Assumption 2. Thus, the left side of (13) is identified. Observe that $\Pr(H_{n,1}, D_{n,1}, e_n, \eta_{n,1})$ is also identified from the mixture weight in $t = 2$. Once $\Pr(D_{n,2}|H_{n,1}, D_{n,1}, e_n, \eta_{n,1})$, $\Pr(H_{n,1}, D_{n,1}, e_n, \eta_{n,1})$, and the mapping from realizations of $(H_{n,1}, D_{n,1}, e_n, \eta_{n,1})$ to realizations of $s_{n,2}$ are known—this mapping is identified from the law of motion of $s_{n,t}$ as discussed—we can recover $\Pr(D_{n,2}|s_{n,2})$ from them. Importantly, in contrast to the standard approach in dynamic discrete choice models—which identifies CCPs directly from discrete choices—here CCPs are recovered from the *continuous* portion of the data, namely, the wage distribution, through its mixture representation.

3.4 The Extremal Quantile Step

We first provide an overview and then a formal argument for the simplified wage equation in (11), which is new in that it applies to both thin- and fat-tailed selected outcome distributions and so accommodates wage distributions—such as that of the United States—with substantial top inequality.[27]

[26] By (4), the discussion following it, and the fact that $\alpha(H_{n,1}, D_{n,t}, e_n) > \beta(H_{n,1}, D_{n,t}, e_n)$, a higher $P_{n,t}$ leads on average to higher output. This first-order stochastic dominance relationship between beliefs and output translates into an analogous one between beliefs and wages that is well supported by the data (see the reviews by Rubinstein and Weiss, 2006 and Waldman, 2012). Thus, empirically, any non-monotonicity of the compensating differential with respect to beliefs does not reverse the overall monotonicity of wages with respect to beliefs, which we use to label mixture components. This logic, stated for the first four periods of our panel, extends to all subsequent periods.

[27] D'Haultfoeuille and Maurel (2013) identify the deterministic wage component in a static single-agent Roy model by exploiting the extremal tails of the distribution of wages. Whereas those authors focus on thin-tailed wage distributions—see their Assumption 2—we allow for wage and log-wage distributions with fat tails. Briefly, the difference between their argument and ours is as follows. D'Haultfoeuille and Maurel (2013) fix a very high wage *level* w' and consider the share of workers with wages above w', namely, the corresponding tail probability. They then infer the deterministic

Deterministic Wage Component: Intuition. Through the mixture step, we identify several primitives of interest, including the law of motion of the state $s_{n,t}$, the distribution of job choices $D_{n,t}$ conditional on the state $s_{n,t}$ (CCPs), and the distribution of wages $w_{n,t}$ *conditional* on job choices $D_{n,t}$ and the state $s_{n,t}$. These distributions alone, though, are not sufficient to identify the wage components $\{(\varphi(d, s_{n,t})\}_{d \in \mathcal{D}}$ because workers choose jobs $D_{n,t}$ based on the vector of shocks $\epsilon_{n,t}$—that is, there is selection on $\epsilon_{n,t}$. Addressing this endogeneity problem, which is the focus of the literature on the Roy model, motivates our use of extremal quantile methods in the next step.

Nonparametric or semiparametric identification arguments for static Roy models typically leverage *excluded regressors with rich support* that shift wages in only one job—job-specific excluded regressors—so that for workers experiencing extreme values of such regressors, the choice of that job becomes nearly certain and independent of their unobserved characteristics (Chamberlain, 1986; Heckman, 1990). For such workers, the expected wage *conditional* on the chosen job coincides with the *unconditional* expected (potential) wage, thus eliminating selection and allowing for the direct recovery of the deterministic wage component $\varphi(\cdot)$ from observed wages. Dynamic extensions of these arguments usually rely on both job-specific excluded regressors and additional simplifying assumptions, such as directional and irreversible choices or absorbing states. For instance, in the schooling context studied by Taber (2000), students acquire grades and degrees one at a time, an earned grade or degree cannot be revoked, and withdrawing from school effectively precludes re-enrollment.

In our framework, these dynamic restrictions do not apply and job-specific excluded regressors do not naturally arise. In particular, two seemingly natural candidates for job-specific excluded regressors fail to meet this requirement: beliefs about a worker's ability, once identified via the mixture step, and a worker's job tenure. Indeed, since ability θ_n is *general* across jobs and our setting is one of common learning, beliefs about a worker's ability are summarized by a *single* (probability) distribution, $(P_{n,t}, 1 - P_{n,t})$, over a worker's possible ability level, $\bar{\theta}$ or $\underline{\theta}$, that affects wages at *all* jobs—rather than a collection of job-specific distributions over a worker's job-specific level of ability affecting wages at each corresponding job. Hence, beliefs cannot serve as a job-specific excluded regressor. Similarly, a worker's job tenure, which is part of I_n^{t-1}, influences both human capital and information accumulation and, in turn, affects wages at *all* jobs.[28]

wage component $\varphi(\cdot)$ from how that share changes when w' changes as $\varphi(\cdot)$ varies. With fat tails, the share above such a large w' changes very little when w' changes, so $\varphi(\cdot)$ is hard to pin down, which explains their focus on thin-tailed distributions. Our approach instead starts by fixing the tail *probability*—a small enough one, for instance, the top 1%—and looking at the wage level w' corresponding it. We then infer $\varphi(\cdot)$ directly from how w' changes as $\varphi(\cdot)$ varies. As w' moves one-to-one with $\varphi(\cdot)$, $\varphi(\cdot)$ can be identified without requiring thin tails.

[28]Other work has shown that the deterministic component of wages in static Roy models can be identified without job-specific excluded regressors, provided we observe at least as many continuous worker attributes as job alternatives; see, for instance, Lee and Lewbel (2013) and Kim and Lee (2025). However, employer–employee matched datasets

To overcome these challenges, we apply a logic akin to that proposed by Chamberlain (1986) and Heckman (1990) for the static Roy model, but exploit the rich support of the wage rather than of any job-specific excluded regressors. Intuitively, at an extremal quantile of the distribution of wages at a given job for a given state and time period, the realized productivity shock at that job must have been exceptionally high, making it by far the best assignment for any worker paid such a wage. For such workers, the extremal wage quantile *conditional* on the chosen job then coincides with the *unconditional* extremal (potential) wage quantile—again, selection is effectively eliminated—so the deterministic wage component $\varphi(\cdot)$ can be easily recovered from observed wages.

Deterministic Wage Component: Argument. Consider the simplified wage equation in (11). For the rest of this section, assume that for each $t \in \{1, \dots, T\}$, the distribution of $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$ and the distribution of $D_{n,t}$ conditional on $s_{n,t}$ (CCPs) are identified, as ensured by the mixture step. The goal now is to recover from them the deterministic component of potential wages, $\varphi(d, s)$, for each firm d and realization s of $s_{n,t}$. The argument relies on two main assumptions. First, we impose a *tail regularity condition* by requiring that the extreme right tails of the potential wage distribution $w_{n,t}(d)|s_{n,t} = s$ and the observed wage distribution $w_{n,t}|(D_{n,t} = d, s_{n,t} = s)$ can be inverted into extreme quantiles—namely, the relevant wage CDFs are continuous and strictly increasing sufficiently far in the right tail. For simplicity, we assume an unbounded right support for both potential and observed wages.[29] Second, we impose a *tail limit condition* on worker selection into any firm d by assuming that there exists an unknown constant $q_t(d) \in (0, 1]$ such that for every s, the probability $\Pr(D_{n,t} = d|s_{n,t} = s, w_{n,t}(d) > w)$ converges to $q_t(d)$ as w grows arbitrarily large.

This condition requires that as the potential wage at firm d becomes arbitrarily large, the probability of workers' selection into firm d converges to a strictly positive constant $q_t(d)$ *invariant across states* s. In particular, it rules out *(i)* lack of convergence of the selection probability, and *(ii)* degenerate tail selection, whereby the probability of choosing firm d, conditional on a high potential wage, converges to zero. Appendix G.1 of the Online Supplementary Material shows that for this tail limit restriction to be satisfied, it is sufficient that the productivity shocks $\{\epsilon_{n,t}(d)\}_{d \in \mathcal{D}}$ be only moderately dependent—see Corollary 4.1 of D'Haultfoeuille and Maurel (2013) for an analogous condition. Although restricting the dependence among shocks in the static Roy model may be unde-

such as the LEHD record worker attributes like education in coarse, discrete categories, and do not contain continuous worker attributes. The belief $(P_{n,t}, 1 - P_{n,t})$, once identified via the mixture step, could be treated as an approximately continuous regressor, but, as discussed, it is a *single* probability distribution over a worker's general ability θ_n. Thus, it does not provide the required number of continuous regressors. Honoré and Hu (2020) characterize bounds on the parameters of sample-selection models in the absence of excluded regressors.

[29] If the right endpoints are finite, then the same reasoning goes through by taking limits at the endpoints. Footnote 49 in Appendix B.4 in the paper and Appendix G.2 in the Online Supplementary Material provide further details about the case with bounded support, including a proof identifying the support endpoints of potential wages and shocks.

sirable, as these shocks are typically the sole source of unobserved heterogeneity and thus the only latent feature through which potential wages can be correlated, for the class of models we study, this assumption is less consequential. Dependence across potential wages is captured by other latent state variables (e_n and $P_{n,t}$) so productivity shocks can be treated as akin to residual independent errors.

Given these two assumptions, we show that $\varphi(d, s)$ is identified in three steps. First, from Bayes' rule and the identity $\Pr(w_{n,t} > w | D_{n,t} = d, s_{n,t} = s) = \Pr(w_{n,t}(d) > w | D_{n,t} = d, s_{n,t} = s)$, it follows

$$\Pr(w_{n,t} > w | D_{n,t} = d, s_{n,t} = s) = \frac{\Pr(D_{n,t} = d | s_{n,t} = s,\, w_{n,t}(d) > w)}{\Pr(D_{n,t} = d | s_{n,t} = s)} \Pr(w_{n,t}(d) > w | s_{n,t} = s). \quad (14)$$

Letting $w \to \infty$ and using the tail limit condition, (14) implies that

$$\Pr(w_{n,t} > w | D_{n,t} = d, s_{n,t} = s) \;\sim\; c_t(d, s) \Pr(w_{n,t}(d) > w | s_{n,t} = s), \quad (15)$$

where $c_t(d, s) := \lim_{w\to\infty} \Pr(D_{n,t} = d | s_{n,t} = s, w_{n,t}(d) > w) / \Pr(D_{n,t} = d | s_{n,t} = s)$ equals $q_t(d)$/-$\Pr(D_{n,t} = d | s_{n,t} = s)$, exists, and is nonzero. By (15), the extremal right-tail probabilities of *observed* selected wages (left side of (15)) are proportional to those of *unobserved* potential wages (right side of (15)), with nonzero finite proportionality factor $c_t(d, s)$. Second, by the tail regularity condition, (15) can be inverted to obtain a simple relationship between extreme τ-quantiles of observed wages (left side of (16)) and of potential wages (right side of (16)) as τ approaches 1,

$$Q_{w_{n,t}|D_{n,t}=d, s_{n,t}=s}(\tau) = \varphi(d, s) + Q_{\epsilon_{n,t}(d)}\Big(1 - \tfrac{1-\tau}{c_t(d,s)} + o_s(1-\tau)\Big), \quad (16)$$

where the right side of (16) uses that $w_{n,t}(d) = \varphi(d, s) + \epsilon_{n,t}(d)$. Third, we apply (16) first at state s and quantile $\tau_{d,s,t}$ and then at state $\bar{s}$ and quantile $\tau_{d,\bar{s},t}$, with $\tau_{d,s,t}$ and $\tau_{d,\bar{s},t}$ satisfying $1 - (1 - \tau_{d,s,t})/c_t(d, s) = 1 - (1 - \tau_{d,\bar{s},t})/c_t(d, \bar{s})$ and the normalization $\varphi(d, \bar{s}) = 0$.[30] Subtracting the resulting two expressions term by term and noting that the difference between the remainder terms vanishes as $\tau_{d,s,t}$ and $\tau_{d,\bar{s},t}$ converge to 1, we obtain $Q_{w_{n,t}|D_{n,t}=d, s_{n,t}=s}(\tau_{d,s,t}) - Q_{w_{n,t}|D_{n,t}=d, s_{n,t}=\bar{s}}(\tau_{d,\bar{s},t}) = \varphi(d, s)$. Since this quantile difference can be calculated from the data, $\varphi(d, s)$ is identified.[31]

As shown in Appendix G.3 of the Online Supplementary Material, this identification approach extends to wage specifications in which the shock $\epsilon_{n,t}(d)$ in (11) is multiplied by a scale function

[30] As in standard Roy models, wages are identified only up to an additive constant, here $\varphi(d, \bar{s})$. Alternatively, the error term can be normalized to have zero unconditional mean or median (French and Taber, 2011).

[31] For this difference to serve as an identifying condition, the indices $\tau_{d,s,t}$ and $\tau_{d,\bar{s},t}$ must be known. We can freely pick any $\tau_{d,\bar{s},t}$ and then set $\tau_{d,s,t}$ equal to $1 - c_t(d, s)(1 - \tau_{d,\bar{s},t})/c_t(d, \bar{s})$. One can next use the tail limit condition and (15), under which $c_t(d, s) = q_t(d)/\Pr(D_{n,t} = d | s_{n,t} = s)$, to calculate the ratio $c_t(d, s)/c_t(d, \bar{s})$ from the data as

$$\frac{c_t(d, s)}{c_t(d, \bar{s})} = \frac{q_t(d)/\Pr(D_{n,t} = d | s_{n,t} = s)}{q_t(d)/\Pr(D_{n,t} = d | s_{n,t} = \bar{s})} = \frac{\Pr(D_{n,t} = d | s_{n,t} = \bar{s})}{\Pr(D_{n,t} = d | s_{n,t} = s)}.$$

$\sigma(d, s_{n,t})$ as, say, in wage equations typical of search models, in which wages are inherently heteroskedastic because the parameters of the output technology enter as shifters of the error-term distribution (Bagger et al., 2014). (See Appendix H of the Online Supplementary Material for the extension of our identification arguments to search models.) This wide applicability of our approach stems from the extremal quantile step not relying on the specific mechanism governing job choices, $D_{n,t}$. It can thus accommodate any arbitrary dependence between those choices and the unobserved shocks $\epsilon_{n,t}$ in any model with a wage equation of the form in (10) or (11).

Productivity Shocks. So far, we have shown how several primitives of interest—such as the law of motion of the state including the learning process, the CCPs, and the deterministic component $\varphi(\cdot)$ of potential wages—can be identified with essentially no parametric assumptions through the mixture step. From this step, as discussed, it is also easy to identify the distribution of $\varphi(d, s_{n,t})+\epsilon_{n,t}(d)$ *conditional* on a worker's job choice $D_{n,t} = d$ and state $s_{n,t}$. With $\varphi(d, s_{n,t})$ identified via the extremal quantile step, we can then pin down the distribution of $\epsilon_{n,t}(d)$ conditional on $(D_{n,t} = d,\ s_{n,t})$.[32] To identify the unconditional distribution of $\epsilon_{n,t}(d)$, which is of interest to recover $y(\cdot)$, measure the importance of the output complementarities captured by $y(\cdot)$ relative to shocks for sorting, and quantify the determinants of the wage markdown, we proceed as follows.

By (15), the extremal right-tail probabilities of observed wages at any job are equal to the extremal right-tail probabilities of unobserved potential wages up to the nonzero finite factor $c_t(d, s_{n,t})$. This factor $c_t(d, s_{n,t})$ cancels out when we take *ratios of survival functions* of observed wages far out in the right tail. Since wages are the sum of $\varphi(d, s_{n,t})$ and $\epsilon_{n,t}(d)$, and $\varphi(d, s_{n,t})$ is identified by the extremal quantile step, it follows that we also nonparametrically identify the ratios of the survival functions of $\epsilon_{n,t}(d)$ far out in the right tail. From this asymptotic tail information alone, we cannot nonparametrically recover the full unconditional marginal distribution of $\epsilon_{n,t}(d)$. However, for any parametric family for $\epsilon_{n,t}(d)$ such that sufficiently far out in the right tail, the ratios of survival functions are injective in the parameters of the family—we term such families *tail-ratio identifiable*—these ratios yield a finite system of equations whose unique solution identifies the parameters of the unconditional marginal distribution of $\epsilon_{n,t}(d)$.[33] In the scalar-parameter case, this property of tail-ratio identifiability reduces to the condition that the elements of the family be locally—that is, far in the right tail—orderable in the monotone likelihood-ratio sense. In Appendix E, we prove this

[32]The mixture step yields the distribution of $\epsilon_{n,t}(d)$ only conditional on the time-t *observed* job choice $D_{n,t} = d$ and $s_{n,t}$, which can substantially differ from the unconditional one except in special cases; see footnote 24. Recovering the *unconditional* distribution of $\epsilon_{n,t}(d)$ would be possible, by the law of total probability, if the mixture step pinned down the distribution of $\epsilon_{n,t}(d)$ conditional on $D_{n,t} = \tilde{d}$ for *every* $\tilde{d} \in \mathcal{D}$ (factual and counterfactual) and $s_{n,t}$.

[33]For related identification results from tail restrictions on the error distribution, see Abbring and Ridder (2015).

identification result and show that the property of tail-ratio identifiability is shared by many families of distributions common in applied work, such as the Normal, Lognormal, Pareto, and Exponential.[34]

Expected Output and Compensating Differential. When equilibrium wages do not feature a compensating differential term or the wage equation, as in search or wage-bargaining models, has a particular structure nested by our model (see Appendices G.3 and H), the extreme quantile step also recovers expected output $y(\cdot)$, since either $\varphi(\cdot)$ reduces to $y(\cdot)$ or $y(\cdot)$ can be immediately obtained from it, for each $s_{n,t}$. In the general case of our model, we identify the expected output function $y(\cdot)$ as follows. As shown in Appendix F.2 of the Online Supplementary Material, the market-wide allocation of workers to firms can be represented as the solution to a pseudo-planning problem—a collection of pair-wise match surplus maximization problems between workers and their employing firms. Thus, once the state law of motion, CCPs, and the distribution of the vector of productivity shocks $\epsilon_{n,t}$ are recovered, $y(\cdot)$ is identified by standard arguments for dynamic discrete choice models (see, for instance, Magnac and Thesmar, 2002) and so is $\Psi(\cdot)$ from $\varphi(\cdot)$.

Human Capital Process. From the expected output function $y(\cdot)$, we can recover the human capital function $h_d(\cdot)$ as follows. Let $Z_{n,t} := (H_{n,1}, I_n^{t-1}, e_n)$ so that $s_{n,t} = (Z_{n,t}, P_{n,t})$. Given $\mathcal{N} = \{\bar{\eta}, \underline{\eta}\}$, define $\underline{h}_{d,t}(Z_{n,t}) := h_d(Z_{n,t}, \underline{\eta})$ and $\bar{h}_{d,t}(Z_{n,t}) := h_d(Z_{n,t}, \bar{\eta})$. Then, expected output can be expressed as $y(d, Z_{n,t}, P_{n,t}) = \underline{h}_{d,t}(Z_{n,t}) + \left[\bar{h}_{d,t}(Z_{n,t}) - \underline{h}_{d,t}(Z_{n,t})\right] P_{n,t}$, where $P_{n,t}$ is known, since the learning process has been identified. The identification of $\underline{h}_{d,t}(Z_{n,t})$ and $\bar{h}_{d,t}(Z_{n,t})$ is guaranteed by a simple rank condition. Specifically, there must exist two realizations, p_1 and p_2, of $P_{n,t}$ such that $p_1 \neq p_2$. Under this condition, the two equations obtained by evaluating $y(d, Z_{n,t}, P_{n,t})$ at $P_{n,t} = p_1$ and $P_{n,t} = p_2$ provide a non-degenerate linear system in the two unknowns $\underline{h}_{d,t}(Z_{n,t})$ and $\bar{h}_{d,t}(Z_{n,t})$ that has a unique solution. The argument naturally extends to $\mathcal{N}$ of any finite cardinality.

3.5 Second-Best Firm

Our wage equation in (10) depends on both a worker's first-best (or employing) firm $D_{n,t}$ and second-best firm $D'_{n,t}$. Like $D_{n,t}$, $D'_{n,t}$ is determined by a worker's ranking of the wage offers received in period t, which is a function of $s_{n,t}$ and $\epsilon_{n,t}$. Unlike $D_{n,t}$, however, under Assumption 2, the econometrician does not observe the distribution of $D'_{n,t}$. In general, such a distribution cannot be

[34] In competing-risks models without covariates, the *joint* distribution of the latent risks cannot be nonparametrically identified (Tsiatis, 1975). Heckman and Honoré (1989) establish that under sufficiently rich covariate variation—that is, with at least as many continuous covariates as causes of failure among other conditions—this joint distribution can be nonparametrically identified. In our setting, this requirement of rich covariate variation amounts to the availability of at least as many continuous regressors as firms—either observed in the data or components of the state $s_{n,t}$ identified in the mixture step. Absent those, as discussed in the context of the recovery of the deterministic component of wages, the joint distribution of $\epsilon_{n,t}$ can only be point identified, say, via an explicit independence assumption or parametric copula, or partially identified, say, via the Fréchet-Höffding bounds or the tighter bounds characterized by Petersen (1976).

identified from the joint distribution of $\{w_{n,t}, H_{n,1}, D_{n,t}\}_{t=1}^{T}$. This difficulty is analogous to that in a second-price auction in which the econometrician only observes the winning bidder's identity and the transaction price. The latter reveals the value of the second-order statistic of the submitted bids—the second-highest bid—yet it does not reveal the *identity* of the second-highest bidder. Absent additional structure, the distribution of the second-highest bidder's identity need not be uniquely pinned down by the joint distribution of the winning bidder and the transaction price.

In principle, we can identify the conditional distribution of $D'_{n,t}$ by incorporating it into the wage-mixture representation in (12) as an additional latent component in that

$$\begin{aligned}\Pr\big(w_{n,t} \leq w \big| H_{n,1}, D_n^t\big) = \sum\nolimits_{(d',e,\eta^{t-1}) \in \mathcal{D}\times\mathcal{E}\times\mathcal{N}^{t-1}} &\Pr\big(D'_{n,t} = d', e_n = e, \eta_n^{t-1} = \eta^{t-1} \big| H_{n,1}, D_n^t\big) \\ &\times \Pr\big(w_{n,t} \leq w \big| H_{n,1}, D_n^t, D'_{n,t} = d', e_n = e, \eta_n^{t-1} = \eta^{t-1}\big). \qquad (17)\end{aligned}$$

This augmented finite mixture can be identified once it is cast as a clusterable one (Aragam et al., 2020). In particular, identification of this augmented mixture yields the conditional distribution of $D'_{n,t}$ given the observables and the other latent components (e_n, η_n^{t-1}) indexing the mixture. The arguments in Sections 3.3 and 3.4 then proceed analogously.

In addition to this approach, we also propose an empirically motivated strategy that is consistent with common settings of incumbent-poacher competition (see, for instance, Bagger et al., 2014 for a related application and the references therein). In the U.S. LEHD data we use in our empirical application, worker mobility is predominantly local, with most job-to-job transitions occurring within narrowly defined industries and counties. Based on this pattern, we define a worker's labor market by the worker's industry (NAICS3) and county of employment. Markets constructed at this NAICS3 $\times$ county level capture the bulk of observed mobility: about 80% of workers remain in the same such market between any two consecutive years. This definition of a labor market leads to an easy recovery of each worker's second-best firm $D'_{n,t}$. Indeed, given this definition, employment in most labor markets is concentrated among a small number of firms. In more than 70% of these markets, although the median number of firms per market is five, only two firms have each at least 10% of employment in the market (see Bardoczy et al., 2025). Together, those two firms on average account for 78% of total market employment. For any such market, we cap the number of firms at two so that once the employing firm is observed, the second-best firm is identified as the remaining one. For the residual 30% of markets in which more than two firms have each at least 10% of employment, we proceed along the lines of Sorkin (2018) and infer a worker's second-best firm from the pattern of worker turnover across firms in the relevant market. Assumption 3 formalizes this approach.

Assumption 3 (Second-Best Firm)**.** *Let $\mathcal{D}(X_{n,t}) \subseteq \mathcal{D}$ denote worker n's labor market in period t observed by the econometrician and constructed from covariates $X_{n,t}$ external to the model. If $|\mathcal{D}(X_{n,t})| = 2$, then $D'_{n,t}$ is the other firm in $\mathcal{D}(X_{n,t})$. If $|\mathcal{D}(X_{n,t})| > 2$, then there exists a known mapping m_t such that $D'_{n,t} = m_t(D_{n,t}, s_{n,t})$.*

Here, $X_{n,t}$ can be industry and county of employment; m_t can be determined from the transitions of "otherwise identical" workers across firms.[35] In Appendix B, we extend the identification arguments outlined in Sections 3.3 and 3.4 to our general wage equation in (10) under Assumption 3.

4 Estimation

Our estimation approach mirrors our identification argument—a mixture step followed by an extremal quantile step. We just introduce a few parametric assumptions and simplifications for tractability. As mentioned in Section 3.1, our identification strategy does not require the econometrician to observe beliefs $P_{n,t}$, performance signals $\{\eta_{n,t}(d)\}_{d\in\mathcal{D}}$, or output $\{Y_{n,t}(d)\}_{d\in\mathcal{D}}$ nor does it require proxies for these objects. When additional information on worker performance is available—or when credible proxies for it can be constructed—estimation can nonetheless be simplified by recovering the learning process in a preliminary step prior to the wage-mixture estimation.

We illustrate this possibility by constructing a proxy for the performance signal about ability using the LEHD data (see Ganong et al., 2025 for a related construction). Specifically, we focus on worker-firm pairs with at least five quarters of employment. For any given quarter t, we first calculate the average quarterly labor earnings from the preceding three quarters $t-3$ to $t-1$. We then define an observation of *high performance pay* in quarter t as the event that earnings are more than 50% higher than lagged average earnings. Accordingly, we interpret this event of high pay as an instance of a *high performance signal*, $\eta_{n,t} = \bar{\eta}$. Given these constructed proxy signals, we estimate the signal distribution conditional on θ_n—namely, $\alpha(H_{n,1}, D_{n,t}, e_n)$ and $\beta(H_{n,1}, D_{n,t}, e_n)$—and the prior belief $P_{n,1}$ via the binomial mixture representation discussed in Appendix B.2. This step yields posterior beliefs $P_{n,t}$ for each worker n and, by recursively updating $P_{n,1}$ according to (3), for every $t \geq 2$.

We determine each worker's second-best firm in every period, $D'_{n,t}$, as outlined in Section 3.5, with labor markets defined at the NAICS3 $\times$ county level. Having recovered the learning process

[35] In each market and for each group of workers defined by common observed characteristics such as gender, education, and occupation, we define the second-best firm as the "destination" firm that attracts a sufficiently large share of workers who separate from a given firm. For the average firm in our sample, about 75% of workers who separate from their employer (90% when weighted by wages) switch to a single firm within the same NAICS3 $\times$ county market. Note that given our market definition, no dimension-reducing aggregation of firms into types (Bonhomme et al., 2019) is needed.

and workers' second-best firms, the observables used in the mixture and extremal quantile steps are $(D_{n,t}, D'_{n,t}, H_{n,t}, I_n^{t-1}, P_{n,t})$. In most of our applications, we specify the mixture distribution of wages as a finite mixture of Normals for simplicity, by assuming that the productivity shock $\epsilon_{n,t}(D'_{n,t})$ in the wage equation in (10) is normally distributed conditional on $(D_{n,t}, D'_{n,t}, H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$, with mean and variance allowed to flexibly depend on these variables to account for worker selection into $D_{n,t}$ and $D'_{n,t}$ based on $\epsilon_{n,t}$. Then, the conditional wage distribution in t is the finite mixture

$$\begin{aligned}
&\Pr(w_{n,t} \leq w | D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, I_n^{t-1} = \iota, P_{n,t} = p) \\
&\quad = \sum\nolimits_{e \in \mathcal{E}} \Pr(e_n = e | D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, I_n^{t-1} = \iota, P_{n,t} = p) \\
&\quad \times \Pr(w_{n,t} \leq w | D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, I_n^{t-1} = \iota, P_{n,t} = p, e_n = e),
\end{aligned} \tag{18}$$

where each component, $w_{n,t} | D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, I_n^{t-1} = \iota, P_{n,t} = p, e_n = e$, is normally distributed, $\mathcal{N}(\varphi(d, d', h, \iota, p, e) + \mu(d, d', h, \iota, p, e), \sigma^2(d, d', h, \iota, p, e))$, with the unknown conditional mean and variance of $\epsilon_{n,t}(d', e)$ given by $\mu(d, d', h, \iota, p, e)$ and $\sigma^2(d, d', h, \iota, p, e)$. We examine the robustness of our results to alternative mixtures of Pareto and Pareto-lognormal distributions.

For each latent class $e_n = e \in \mathcal{E}$ of the wage mixture in (18), we parameterize the deterministic wage component $\varphi(d, d', h, \iota, p, e) := y(d', h, \iota, p, e) + \Psi(d, d', h, \iota, p, e)$ with the finite-dimensional vector of parameters $b(d, d', e) := (b_y(d', e), b_\Psi(d, d', e))$. As discussed in Section 2 after equation (4), our model implies that the expected output function net of productivity shocks is affine in $P_{n,t}$, and so is fully characterized by an intercept and a slope with respect to $P_{n,t}$. Accordingly, the expected output function (net of productivity shocks) takes the form

$$\begin{aligned}
y(d, h, \iota, p, e) :=& \big[b_{y,0}(d, e) + b_{y,1}(d, e)\, h + b_{y,2}(d, e)\, \iota + b_{y,3}(d, e)\, h\, \iota\big] \\
&+ \big[b_{y,4}(d, e) + b_{y,5}(d, e)\, h + b_{y,6}(d, e)\, \iota + b_{y,7}(d, e)\, h\, \iota\big]\, p.
\end{aligned}$$

We approximate the compensating differential $\Psi(d, d', h, \iota, p, e)$ with a flexible quartic polynomial in (h, ι, p) with parameter vector $b_\Psi(d, d', e)$, allowing for rich nonlinearities and interaction terms while keeping the number of coefficients manageable to ensure a fast and numerically stable implementation. We estimate the parameters $\{b(d, d', e)\}_{(d,d') \in \mathcal{D}^2}$ by adapting the extremal quantile regression approach of D'Haultfoeuille et al. (2018), who extend the extremal quantile regression methods of Chernozhukov (2005) to allow for selection. This inner extremal quantile step is nested within an outer maximum likelihood (ML) step, in which we fit a finite mixture with $|\mathcal{E}|$ components and mixture weights $\{\Pr(e_n = e | D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, I_n^{t-1} = \iota, P_{n,t} = p)\}_{e \in \mathcal{E}}$ as in (18).[36]

[36]We estimate an intercept for wages by centering the unconditional extremal quantile of the productivity shock at zero. Since expected output is estimated from the wage mixture under our parametric specification, without relying on

Remark 4 (Estimator and Large-Sample Properties)**.** *Let the conditional wage distribution admit the finite-mixture representation in* (18)*, with wages parameterized by* $b(d, d', e)$ *for* $(d, d', e) \in \mathcal{D}^2 \times \mathcal{E}$*. Our estimator* $\widehat{b}$ *is obtained by (i) fitting the* $|\mathcal{E}|$*-component mixture in* (18) *by maximum likelihood; and (ii) estimating* $\{b(d, d', e)\}$ *via extremal quantile regression. Under standard regularity conditions ,* $\widehat{b}$ *is consistent and asymptotically normal.*

In the next section, we show that this estimator behaves well in finite samples.

5 Wage Inequality: The Role of Sorting and Monopsony Power

We now rely on our framework and econometric approach to investigate labor market sorting and firm monopsony power using the U.S. LEHD data.

5.1 Simulation Exercises

Here, we illustrate key features of our framework and the properties of our estimator through representative simulation exercises. In what follows, we describe how we simulate the model, how we estimate its parameters on the simulated data and the quality of the resulting estimates, and what these estimates imply for the sources of labor market inequality and their appropriate measurement.

Model Simulation. We simulate a finite-horizon version of our model with two worker-efficiency types (or "discrete fixed effects", following the terminology of Bonhomme et al., 2019), $e_n \in \mathcal{E} := \{\bar{e}, \underline{e}\}$, high and low, and two firms, $|\mathcal{D}| = 2$. Each firm consists of two jobs, $j \in \mathcal{J} := \{H, L\}$, high and low, ordered by their complementarity with respect to worker ability θ_n and efficiency e_n.[37] The two jobs provide identical opportunities for human capital acquisition and learning about ability, but the jobs of different firms are differentially informative about ability. Thus, jobs differ only in how worker ability θ_n and efficiency e_n map into expected output. We maintain that High-ability ($\theta_n = \bar{\theta}$) and high-efficiency ($e_n = \bar{e}$) workers produce more, on average, than low-ability ($\theta_n = \underline{\theta}$) and low-efficiency ($e_n = \underline{e}$) workers in the high-efficiency job $j = H$, whereas low-ability ($\theta_n = \underline{\theta}$) and low-efficiency ($e_n = \underline{e}$) workers produce more, on average, than high-ability ($\theta_n = \bar{\theta}$) and high-efficiency ($e_n = \bar{e}$) workers in the low-efficiency job $j = L$.

We solve for the equilibrium by backward induction exploiting that in this two-firm, two-job setting with symmetric human capital and information acquisition across jobs, the equilibrium is efficient. The allocation of workers to firms and jobs can then be characterized as the solution to

methods for dynamic discrete choice models, no additional normalization is needed. See Appendix B for details.

[37] As explained in Section 2, our model and econometric results immediately extend to settings in which firms' production is organized into multiple jobs and, accordingly, firms' offers specify both a wage and a job assignment for a worker in a period. See Appendix F.2 of the Online Supplementary Material.

the planner's problem of maximizing the expected present discounted value of (log) output in each period; see Appendix F of the Online Supplementary Material. Such *social welfare* value is

$$S(s_{n,t}, \epsilon_{n,t}) = \max_{(d,j) \in \mathcal{D} \times \mathcal{J}} \Big\{ y(d, j, s_{n,t}) + \epsilon_{n,t}(d, j) \\ + \delta \left[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j)\right] \int \mathbb{E}[S(s_{n,t+1}, \epsilon_{n,t+1}) \mid s_{n,t}, d, j] \; dF_{\epsilon_{n,t+1}} \Big\} \tag{19}$$

for worker n at state $(s_{n,t}, \epsilon_{n,t})$. We start from the terminal period $T = 40$ and set workers' continuation values to zero beyond it.[38] We determine equilibrium wages in period T at each state using equation (10), with the compensating differential term set to zero. For each remaining period $t \in \{1, \ldots, 39\}$, we solve the planner's problem in (19) taking as given the continuation social welfare values at each state from period $t + 1$ onward. In any such period, we determine both the allocation of workers to firms and jobs and workers' wages, which now include the compensating differential term $\Psi(d, d', j, j', s_{n,t})$—note that the compensating differential is now a function of job j at the first-best firm d and job j' at the second-best firm d'. We calculate this term using a worker's continuation match surplus value when matched with the worker's current second-best firm d' from period $t + 1$ onward. Interpreting one period as one year, the simulated panel consists of 1 million high-ability workers ($\theta_n = \bar{\theta}$) and 1 million low-ability workers ($\theta_n = \underline{\theta}$), who are followed from age 20 ($t = 1$) to age 60 ($t = T$), thereby capturing most of their working life, and are assigned a sequence of firm- and job-specific productivity shocks, θ_n-specific probabilities of generating a high signal $\eta_{n,t} = \bar{\eta}$, and an e_n-specific prior $P_{n,1}$ of being of high ability. This prior is updated over time based on realized output as in (3).

We choose the model's wage parameters—namely, the slope and intercept of expected output, the law of motion of human capital, the learning process, and the variance of productivity shocks—so that the simulated data reproduce a rich set of moments from the U.S. distributions of (log labor) earnings and (log) earnings changes from a PSID sample of heads of household aged 20 to 60 over the period between 1970 and 2019. Table 1 reports the model's fit to the targeted moments (left panel) and a host of untargeted ones (right panel). The targeted moments describe the life-cycle profiles of the mean and standard deviation of earnings. The untargeted moments consist of concentration measures and AKM-type moments as well as cross-sectional moments of the distribution of earnings changes over 1-, 3-, and 5-year horizons. As the table shows, in addition to fitting the targeted moments, our model matches well the untargeted ones, including the share of the (residual) variance

[38] Whereas the terminal period $T = 40$ in this simulation and the panel length in the datasets coincide, this needs not necessarily be the case as noted in our previous discussion of Assumption 2.

Table 1: Model Fit to Targeted and Untargeted Moments from PSID

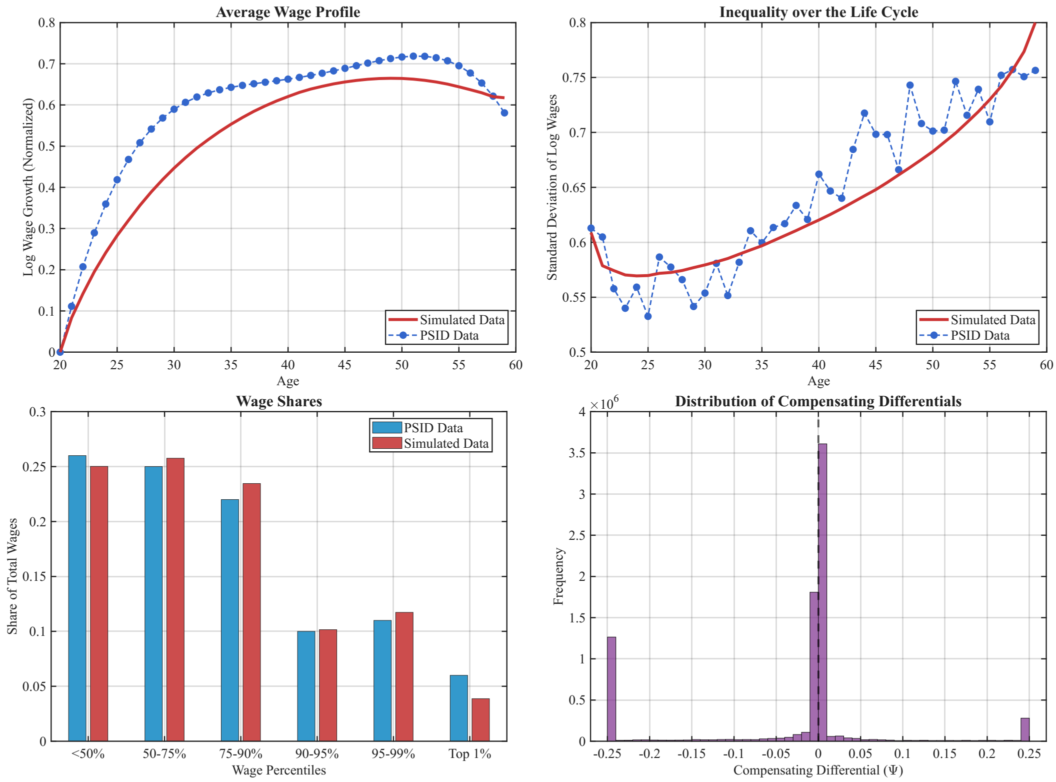
Targeted Moments			Untargeted Moments		
Life-Cycle Properties	Model	Data	Cross-Sectional Properties	Model	Data
Wage Growth (Life)	0.69	0.70	Top 10% Share	0.26	0.27
Wage Growth [20-30]	0.50	0.60	Top 1% Share	0.04	0.06
Wage Growth [20-40]	0.67	0.68	AKM Worker Share	0.53	0.52
Wage Growth [20-50]	0.71	0.70	AKM Firm Share	0.09	0.09
Wage Growth [20-60]	0.69	0.59	AKM Sorting	0.15	0.12
Std. Dev. (Life)	0.72	0.65	Std. Dev. (d_1)	0.19	0.35
Std. Dev. (20)	0.61	0.62	Std. Dev. (d_3)	0.26	0.45
Std. Dev. (30)	0.67	0.55	Std. Dev. (d_5)	0.32	0.50
Std. Dev. (40)	0.62	0.66	Skewness (d_1)	0.04	0.42
Std. Dev. (50)	0.68	0.70	Skewness (d_3)	-0.04	-0.03
Std. Dev. (60)	0.80	0.76	Skewness (d_5)	-0.13	-0.16

Note: The table shows moments of the life-cycle and cross-sectional distributions of wages calculated using the model-generated data ("Model" column) and the PSID data ("Data" column). "Life" refers to the dispersion of log wages over the life cycle across all ages; the numbers in parentheses denote, for the relevant moments, the ages over which they are calculated, for instance, Wage Growth [20-30] is the average (log) wage growth between ages 20 and 30; growth moments are calculated in the data up to 1997. Std. Dev. (20) is the standard deviation of log wages at age 20. The left panel reports model fit in terms of the life-cycle profiles of log wages (targeted moments). The right panel reports model fit in terms of the concentration of wages and the log wage growth distribution (untargeted moments).

of log wages accounted for by firm and worker fixed effects, obtained by estimating a standard AKM model as in (22) below on our simulated data. Consistent with Song et al. (2019), we find that the shares of the variance of log wages attributable to worker and firm fixed effects are about 51% and 9%, respectively, whereas the share attributable to the covariance between worker and firm fixed effects is close to 15%. These estimates are also close to those in Bonhomme et al. (2023). Finally, workers receive as wages about 55% of their output, which is in the low range of the estimates for the U.S. manufacturing sector (Yeh et al., 2022).

To better appreciate how the model compares with the PSID data, Figure 2 shows additional moments for our simulated economy and, whenever possible, for the data. Starting from the top-left panel, the simulated data (red solid line) match an increase in average earnings of about 0.6 log points (82%) for workers between 20 and 60 years of age, as observed in the PSID (dotted blue line). Similarly, the red solid line in the top-right panel, which reports the life-cycle profile of the standard deviation of (log) wages implied by the model, overall increases by 20 percentage points in line with the increase observed in the PSID over the same horizon (dotted blue line). Importantly, the bottom-left panel shows that the distribution of wages generated by our model (red bars) closely resembles the distribution of earnings in the PSID (blue bars). Finally, the bottom-right panel displays the distribution of the compensating differential across workers, which is on average negative and equal to -0.32, corresponding to about 22% of wages and 4% of expected output (in absolute value) for the average worker. Hence, the compensating differential in our simulated economy on average reduces wages relative to workers' expected output at the second-best firm, as workers trade off human capital accumulation and learning opportunities for lower current wages.

Figure 2: Wages in Simulated Data and PSID Data



Note: The figure shows the life-cycle profile of mean (log) labor earnings (top left), the life-cycle profile of the standard deviation of log labor earnings (top right), and the cross-sectional distribution of wages (bottom left) calculated using the model-generated data ("Simulated Data") and the PSID data ("PSID Data") as well as the cross-sectional distribution of the compensating differential calculated using the model-generated data (bottom right).

We note that the difference between a worker's expected output at the first-best and second-best firms is on average only about 35% of the expected output at the first-best firm—for more than a quarter of all workers, this difference is at most 25%—and relatively stable over the life cycle. Thus, the gap is moderate, which justifies focusing the discussion that follows about the measurement of sorting and firm monopsony power on the role of the compensating differential in wages.

Estimation. We now show how the estimation procedure outlined in Section 4 performs on the simulated data. To do so, we consider the (log) wage equation

$$\begin{aligned} w_{n,t} = &\sum\nolimits_{(d,j,d',j')\in\mathcal{D}^2\times\mathcal{J}^2} \mathbb{1}\{D_{n,t}=d,\, J_{n,t}=j,\, D'_{n,t}=d',\, J'_{n,t}=j'\} \\ &\times \left[y(d',j',s_{n,t}) + \Psi(d,d',j,j',s_{n,t}) + \epsilon_{n,t}(j',d')\right], \end{aligned} \tag{20}$$

where $J_{n,t}$ and $J'_{n,t}$ denote the first- and second-best firms' job offers, and expected (log) output, compensating differential, and productivity shocks are firm- and job-specific. In estimation, we treat $D'_{n,t}$, $J'_{n,t}$, and $P_{n,t}$ as observed together with $(w_{n,t}, D_{n,t}, J_{n,t}, H_{n,1}, I_n^{t-1})$. As discussed in Section 4, we parameterize expected output (net of productivity shocks) at firm d and job j as

$$y(d,j,s_{n,t}) := \underbrace{b_{y,0}(d,j,e_n) + b_{y,1}(d,j,e_n)\,H_{n,1} + b_{y,2}(d,j,e_n)\,I_n^{t-1} + b_{y,3}(d,j,e_n)\,H_{n,1}\,I_n^{t-1}}_{\text{"intercept"}}$$

$$+\underbrace{\left[b_{y,4}(d,j,e_n)+b_{y,5}(d,j,e_n)\,H_{n,1}+b_{y,6}(d,j,e_n)\,I_n^{t-1}+b_{y,7}(d,j,e_n)\,H_{n,1}\,I_n^{t-1}\right]}_{\text{“slope”}}\,P_{n,t}, \tag{21}$$

where $H_{n,1}$ is a dummy variable taking value 1 if a worker's education is high (say, college) and I_n^{t-1} denotes labor market experience or age. We define the "intercept" of expected output as the term on the first line of (21) and its "slope" as the term in squared brackets on the second line of (21) multiplying $P_{n,t}$. The intercept term is the expected output of worker n when of low ability, whereas the slope term is the difference in expected output between worker n when of high ability and when of low ability.These intercept and slope terms are of primitive interest because they describe the patterns of workers' absolute and comparative advantages across firms and jobs. In particular, they capture the gains from sorting based on workers' ability θ_n and efficiency e_n. A higher slope, for instance, due to higher initial or acquired human capital, amounts to a greater difference in the expected output of a high-ability relative to a low-ability worker, and so implies larger gains from sorting based on ability. Similarly, that output parameters $b_y(d,j,e_n)$ increase with e_n implies larger gains from sorting based on efficiency e_n.[39] We return to this point below. Lastly, we approximate the compensating differential $\Psi(d,d',j,j',s_{n,t})$ using a flexible quartic polynomial in $(H_{n,1}, I_n^{t-1}, P_{n,t})$.[40]

Table 2 compares the values of the intercept and slope terms in (21) estimated by our approach (columns "Our") and OLS (columns "OLS") together with their true values used to generate the simulated data (columns "True"). For each worker-efficiency and firm type (panels A to D), the table reports these values averaged across jobs, separately for workers with low education (left panel) and high education (right panel) at ages 25, 35, and 55. Clearly, the selection-corrected estimates obtained with our procedure are much closer to the true values than the OLS estimates; see Table 4 in Appendix A for the corresponding standard errors. Table 3 reports cross-sectional moments of the average compensating-differential-to-wage ratio $\Psi(d',d',j,j',s_{n,t})/w_{n,t}$, calculated on the simulated data (row "True") and estimated by our approach (row "Our") and OLS (row "OLS"), across worker-

[39] We parameterize the probabilities $\alpha(H_{n,1},d,j,e_n)$ and $\beta(H_{n,1},d,j,e_n)$ of high output for a high- and a low-ability worker, respectively, as $\alpha(d)$ and $\beta(d)$. As discussed in Section 4, like in our empirical application, we obtain $\alpha(d)$ and $\beta(d)$ in a preliminary step, prior to the wage-mixture estimation, by constructing proxies for performance signals using earnings data. Details on how the second-best firm $D'_{n,t}$ and job $J'_{n,t}$ are recovered are provided in Section 3.5. In the exercises based on the simulated data, we use the values of $P_{n,t}$, $D'_{n,t}$, and $J'_{n,t}$ obtained from the simulation.

[40] With $\overline{h}_{n,t}(d,j) = h_{d,j}(H_{n,1}, I_n^{t-1}, e_n, \overline{\eta})$ and $\underline{h}_{n,t}(d,j) = h_{d,j}(H_{n,1}, I_n^{t-1}, e_n, \underline{\eta})$, output in (2) takes values $\overline{Y}_{n,t}(d,j) = \overline{h}_{n,t}(d,j) + \epsilon_{n,t}(d,j)$ and $\underline{Y}_{n,t}(d,j) = \underline{h}_{n,t}(d,j) + \epsilon_{n,t}(d,j)$ under Assumption 1. Thus, net of productivity shocks or in expectation over them, the expected output of a high-ability worker is $\underline{h}_{n,t}(d,j) + \alpha(H_{n,1},d,j,e_n)[\overline{h}_{n,t}(d,j)-\underline{h}_{n,t}(d,j)]$ and of a low-ability worker is $\underline{h}_{n,t}(d,j)+\beta(H_{n,1},d,j,e_n)[\overline{h}_{n,t}(d,j)-\underline{h}_{n,t}(d,j)]$. Thus, expected output $y(d,j,s_{n,t})$ equals the sum of $P_{n,t}\{\underline{h}_{n,t}(d,j)+\alpha(H_{n,1},d,j,e_n)[\overline{h}_{n,t}(d,j)-\underline{h}_{n,t}(d,j)]\}$ and $(1-P_{n,t})\{\underline{h}_{n,t}(d,j)+\beta(H_{n,1},d,j,e_n)[\overline{h}_{n,t}(d,j)-\underline{h}_{n,t}(d,j)]\}$, which reduces to $\underline{h}_{n,t}(d,j)+\beta(H_{n,1},d,j,e_n)[\overline{h}_{n,t}(d,j)-\underline{h}_{n,t}(d,j)]+[\alpha(H_{n,1},d,j,e_n)-\beta(H_{n,1},d,j,e_n)][\overline{h}_{n,t}(d,j)-\underline{h}_{n,t}(d,j)]P_{n,t}$. So, the intercept term is the expected output of a low-ability worker and the slope term is the difference in expected output between a high- and low-ability worker.

Table 2: Parameter Values of Expected Output Equation

Age	Low Education						High Education					
	Intercept			Slope			Intercept			Slope		
	True	Our	OLS	True	Our	OLS	True	Our	OLS	True	Our	OLS
Panel A $(\underline{e}, \underline{d})$												
25	1.32	1.41	1.51	0.35	0.34	0.57	1.99	1.97	2.26	0.53	0.52	0.82
35	1.71	1.70	1.95	0.46	0.44	0.69	2.91	2.88	3.34	0.77	0.76	1.16
55	1.57	1.60	1.82	0.42	0.40	0.59	3.83	3.81	4.52	1.02	0.98	1.40
Panel B $(\underline{e}, \overline{d})$												
25	1.26	1.35	1.44	0.28	0.27	0.50	1.92	1.90	2.19	0.46	0.45	0.76
35	1.64	1.63	1.89	0.39	0.37	0.62	2.84	2.81	3.27	0.71	0.69	1.09
55	1.51	1.53	1.75	0.35	0.33	0.52	3.76	3.75	4.46	0.95	0.92	1.33
Panel C $(\overline{e}, \underline{d})$												
25	1.22	1.31	1.39	0.26	0.24	0.46	1.89	1.87	2.16	0.42	0.42	0.73
35	1.61	1.60	1.84	0.36	0.34	0.59	2.81	2.77	3.24	0.68	0.65	1.05
55	1.47	1.48	1.71	0.32	0.30	0.48	3.73	3.71	4.42	0.92	0.88	1.30
Panel D $(\overline{e}, \overline{d})$												
25	1.39	1.48	1.57	0.42	0.41	0.64	2.06	2.03	2.33	0.60	0.59	0.89
35	1.78	1.76	2.02	0.52	0.51	0.76	2.97	2.94	3.41	0.84	0.82	1.22
55	1.64	1.67	1.88	0.49	0.47	0.66	3.90	3.88	4.59	1.09	1.05	1.46

Note: The table reports the intercept and slope parameters of expected output in (21) used to generate the simulated data (columns "True"), those estimated using our approach (columns "Our"), and those estimated by OLS (columns "OLS") for each worker-efficiency type and firm type and for workers with high and low education at ages 25, 35, and 55 averaged across jobs.

efficiency types, firms, jobs, education groups, and ages. Again, the selection-corrected average implied by our estimates is much closer to the true average than that implied by the OLS estimates.

Table 3: Statistics on Compensating Differential

	Variable	Mean	Std. Deviation	Skewness
True	$\Psi(\cdot)/w_{n,t}$	-0.18	0.60	-5.17
Our	$\widehat{\Psi}(\cdot)/\widehat{w}_{n,t}$	-0.16	0.74	-5.34
OLS	$\widehat{\Psi}(\cdot)/\widehat{w}_{n,t}$	-0.04	0.12	2.10

Note: The table reports cross-sectional moments of the average compensating-differential-to-wage ratio computed on the simulated data ("True") as well as the corresponding moments estimated by our approach ("Our") and by OLS ("OLS"). These moments are obtained by averaging across worker-efficiency types, firm types, jobs, education groups, ages, and time periods.

Monte-Carlo Exercises. We provide further evidence on the performance of our estimator based on a Monte Carlo exercise where we estimate the parameters of the wage equation in (21) on 200 random samples of one-million individuals each ($|\mathcal{N}| = 1,000,000$), generated as our simulated data. Results are displayed in Table 5 in Appendix A, which reports the mean, median, and standard deviation of the empirical distribution of the estimates across samples. We note that this empirical distribution is closely centered on the true values of the wage parameters and displays little dispersion around them, suggesting that our estimator is well-behaved in finite samples. In each sample, the estimated parametric wage equation well approximates the true one implied by our model.

Labor Market Sorting Based on AKM. As discussed in Section 2.2, statistical measures of sorting such as those proposed by Abowd et al. (1999) may be affected by econometric bias when applied to our class of models, as they ignore the compensating differential in wages (Remark 2). We now consider an exercise that illustrates such a bias. We proceed by first generating multiple datasets from economies that are parametrized to resemble the U.S. one and differ *only* in the level of the compensating differential. We then consider an econometrician who estimates an AKM-type (log) wage equation on each such dataset, which does not feature any terms that capture the compensating differential, so as to assess the misspecification implied by a wage equation that ignores it. As the exercise makes clear, we find bias in the resulting AKM estimates, which increases as the compensating differential embedded in the generated data increases in absolute value.

More precisely, we calculate in our *baseline* simulated data *(i)* the average compensating-differential-to-wage ratio, which is around -20%; *(ii)* the estimated correlation between worker and firm fixed effects, $\widehat{\rho}_{\text{AKM}} := \text{corr}(\widehat{e}_n, \widehat{\psi}(D_{n,t}))$ obtained from the AKM-type (log) wage equation

$$w_{n,t} = e_n + \psi(D_{n,t}) + \beta^\top (H_{n,1}, I_n^{t-1}, P_{n,t}) + \varepsilon_{n,t}, \tag{22}$$

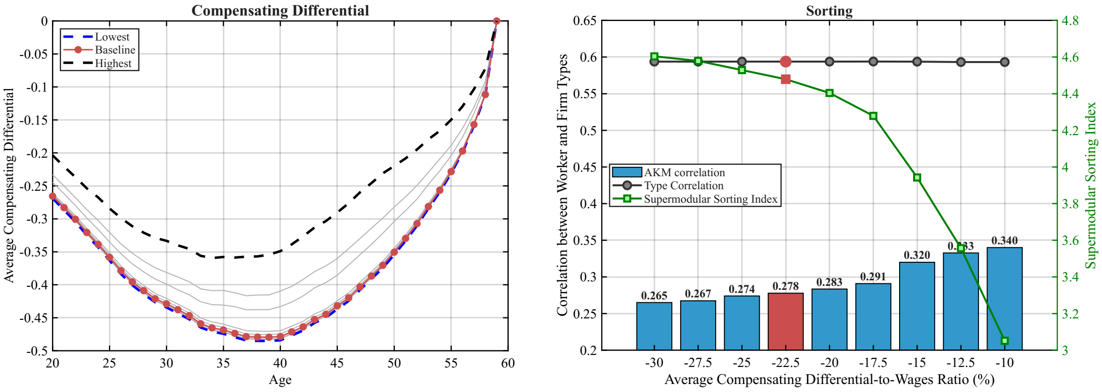
on the simulated data; and *(iii)* the true correlation between workers' efficiency types and firm types, $\rho := \text{corr}(e_n, D_{n,t})$.[41] We then calculate these three objects on additional data constructed as follows. We re-fit the model's primitives to the targeted moments on the left side of Table 1 with the additional constraint of a *progressively lower* average compensating-differential-to-wage ratio. That is, instead of freely choosing all wage parameters—including those governing the compensating differential—by only targeting the moments on the left side of Table 1 as we did in the baseline, we restrict the moment-matching procedure to choose values for the wage parameters that also deliver a specific level of the average compensating-differential-to-wage ratio. Given the values of the wage parameters resulting from this constrained-fit exercise, we then compute the new equilibrium matching patterns and wages. We next use these *constrained-fit data* to calculate ρ and $\widehat{\rho}_{\text{AKM}}$.

The left panel of Figure 3 reports the average compensating differential (in level) over the life cycle in our baseline simulated data (red line) and constrained-fit data (grey lines), with dashed blue and black lines denoting the cases in which we fit the most negative ("Lowest") and the least negative ("Highest") average compensating-differential-to-wage ratios, respectively.[42] These profiles are U-shaped in all cases. Intuitively, early in the life cycle, greater experience improves expected output and matching. This is because, as it turns out, experience increases both the intercept and the slope of expected output in (21). As discussed after equation (21), a higher slope amounts to a greater dif-

[41]When estimating the wage equation in (22), we allow each worker to have a distinct type, as is standard in AKM regressions. Moreover, we do not correct the AKM estimates for potential low-mobility bias, since our simulated data include only two firms. In all simulations, we consider two firms with distinct technologies (types) so firm identities and firm types coincide. Recall that the correlation coefficient is scale invariant so it is appropriate to compare $\widehat{\rho}_{\text{AKM}}$ and ρ.

[42]We obtain analogous results but of *opposite* sign when the average compensating differential is instead positive.

Figure 3: Wages and Sorting



Notes: The left panel reports the average compensating differential over the life cycle in our baseline simulated data (red line) and in the alternative data generated via the constrained-fit procedure (grey lines), with dashed blue and black lines for the cases with the most negative ("Lowest", -30%, blue) and least negative ("Highest", -10%, black) average compensating-differential-to-wage ratios. The right panel (left-side y-axis) reports $\widehat{\rho}_{\text{AKM}}$ in the baseline simulated data (red bar), $\widehat{\rho}_{\text{AKM}}$ in the constrained-fit data (blue bars), and ρ in both the baseline and constrained-fit data (solid black line with dots; the red dot corresponds to the baseline simulated data). The right panel (right-side y-axis) also reports the supermodular index $S_{\mathbf{x}}$ evaluated at the average $\mathbf{x}$ in both the baseline and the constrained-fit data (solid green line with dots; the red dot corresponds to the baseline simulated data).

ference in the expected output of a high-ability relative to a low-ability worker and so implies larger gains from sorting based on θ_n. Hence, experience increases the gains from sorting based on *true productivity*, $(H_{n,1}, I_n^{t-1}, e_n, \theta_n)$. Also, early on, uncertainty about ability is relatively high so information about it acquired through employment greatly reduces it, which improves matching by basing it more and more closely on true productivity. As a result, the human capital and information value of current employment in terms of higher future wages makes the compensating differential initially more negative. Later on, human capital grows more slowly, beliefs about ability become more precise, and the horizon over which greater human capital and information about ability improve output and matching is shorter, thus reducing the worth of additional human capital and information. The compensating differential then becomes smaller and smaller in absolute value.

In the right panel of Figure 3, focusing on the left-side y-axis, the red bar reports $\widehat{\rho}_{\text{AKM}}$ in the baseline simulated data, the blue bars report $\widehat{\rho}_{\text{AKM}}$ in the constrained-fit data, and the solid black line with dots reports ρ in both the baseline simulated data (red dot)—the data that best resemble U.S. data—and constrained-fit data (black dots). The figure shows that as the compensating differential becomes more negative, $\widehat{\rho}_{\text{AKM}}$ decreases and the gap between ρ and $\widehat{\rho}_{\text{AKM}}$—which is primarily driven by the bias due to ignoring the compensating differential—increases. That is, if the econometrician estimates an AKM-type wage equation as in (22) that features a progressively smaller compensating differential, then the estimator of the correlation between worker-efficiency types and firm types is increasingly biased, as the magnitude of the omitted compensating differential (in absolute value) increases. This pattern suggests that AKM estimates of worker-firm complementarity may severely *understate* the degree of sorting in a labor market in which wages feature a large negative compensating differential, as is likely early in workers' careers when workers place a high value on acquiring new human capital and information about ability. These results then help to reconcile the recurrent empirical finding that AKM-measured sorting is often low, even in labor markets that exhibit substantial wage dispersion, because AKM *indirect wage measures* omit a key component of wages—the compensating differential—thereby inducing a bias that compresses the estimated correlation between worker and firm types. As we argue next, a more appropriate statistic for sorting in this case is one based on *direct output measures* of worker-firm complementarity.

Labor Market Sorting Based on Output Complementarities. In Section 2.2, we argue that besides any potential econometric bias, AKM-type sorting measures—and thus ρ and $\widehat{\rho}_{\text{AKM}}$ above—may be conceptually ill-suited to the class of models we consider (Remark 1). As an alternative, we consider a classic sorting measure based on the supermodularity of output in workers' and firms' latent characteristics, a notion widely used in the transferable-utility matching literature to capture complementarities between matched agents' attributes (Graham, 2011).

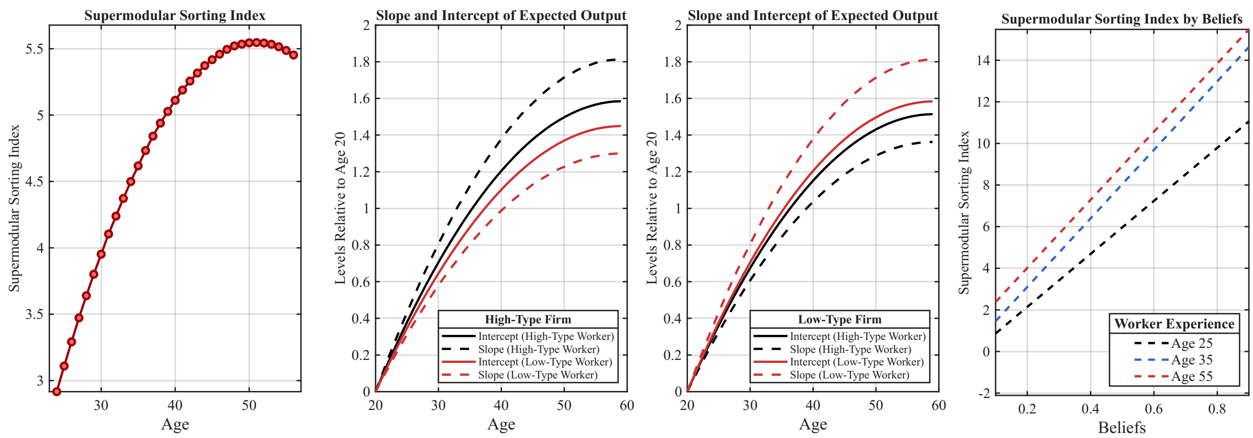
To elaborate, in our simulated economy, worker-efficiency types and firms are ordered from low to high by their contribution to expected output—we relabel firms in $\mathcal{D}$ as $\underline{d}$ and $\bar{d}$.[43] Fix now the remaining components of the state $s_{n,t}$, namely $(H_{n,1}, I_n^{t-1}, P_{n,t})$, and a worker's job $J_{n,t}$ at reference values collected in a vector $\mathbf{x}$. Denote by $y_{\mathbf{x}}(d, e)$ the expected (log) output (net of productivity shocks) in (21) when a worker has efficiency type $e \in \{\bar{e}, \underline{e}\}$ and is employed by firm $d \in \{\bar{d}, \underline{d}\}$, holding $\mathbf{x}$ fixed. If the incremental value of a *higher* worker-efficiency type is larger at a *higher* firm type in that $y_{\mathbf{x}}(\bar{d}, \bar{e}) - y_{\mathbf{x}}(\bar{d}, \underline{e}) > y_{\mathbf{x}}(\underline{d}, \bar{e}) - y_{\mathbf{x}}(\underline{d}, \underline{e})$, then expected output is supermodular in d and e—that is, worker-efficiency types and firm types are complementary inputs to producing (expected) match output. Define the index $S_{\mathbf{x}}$ of the supermodularity of expected output as $S_{\mathbf{x}} := y_{\mathbf{x}}(\underline{d}, \underline{e}) + y_{\mathbf{x}}(\bar{d}, \bar{e}) - y_{\mathbf{x}}(\underline{d}, \bar{e}) - y_{\mathbf{x}}(\bar{d}, \underline{e})$. So, there exist gains from positive sorting in the labor market if $S_{\mathbf{x}} > 0$, such gains are stronger the larger $S_{\mathbf{x}}$ is, and they increase over the life cycle if $S_{\mathbf{x}}$ increases with labor market experience.[44] We report the index $S_{\mathbf{x}}$, evaluated at the average value of $\mathbf{x}$ across workers and time periods, in the right panel of Figure 3 on the right-side y-axis, shown as the solid green line across the different constructed datasets. The red dot represents its value in the baseline simulated data, and the green dots in the constrained-fit data, obtained as discussed.

Note that the index is always positive and large, suggesting substantial output gains to posi-

[43] Since e_n represents worker efficiency, higher e_n naturally corresponds to higher expected output. Likewise, ordering firms by their productivity implies that higher d corresponds to higher expected output. For the reasons discussed in Section 2.2, no analogous monotonicity holds for the compensating differential.

[44] This notion immediately generalizes to any square array with $|\mathcal{D}| = |\mathcal{E}|$. It also extends to multi-dimensional attributes, rectangular arrays with $|\mathcal{D}| \neq |\mathcal{E}|$, and, under standard regularity conditions, continuous type spaces.

Figure 4: Supermodular Index of Sorting and Expected Output



Note: The leftmost panel shows how $S_{\mathbf{x}}$, evaluated at the average $\mathbf{x}$ and calculated on our baseline simulated data, varies over the life cycle. The two middle panels show how the intercept and the slope of expected output in (21) in a high-type firm $\overline{d}$ (center-left panel) and in a low-type firm $\underline{d}$ (center-right panel) vary over the life cycle. The rightmost panel shows how $S_{\mathbf{x}}$ varies with beliefs at age 25, 35, and 55.

tive sorting in the labor market. To understand why $S_{\mathbf{x}}$ monotonically decreases as the average compensating-differential-to-wage ratio gets closer and closer to zero, recall that in calculating this index, we use the parameters of expected output that generate the baseline simulated data (red dot) and those inferred through the constrained-fit procedure described (green dots). As this panel illustrates, intuitively, as the level of the average compensating-differential-to-wage ratio imposed in the constrained-fit procedure gets closer to zero—and so smaller and smaller in absolute value than the level implied by the baselined simulated data—the expected output implied by the constrained-fit procedure must increase, for a given variance of productivity shocks, to keep wage levels in line with the targeted moments on the left side of Table 1. But if so, then the expected output implied by the constrained-fit procedure must absorb part of the *nonmonotone* dependence of wages on worker-efficiency types and firm types that the pre-specified average compensating-differential-to-wage ratio is *not* free to capture. (Recall that the deterministic component of wages is the sum of expected output at the second-best firm and the compensating differential, the former of which is monotone with worker and firm types unlike the latter; see footnote 43). As a result, the supermodular index calculated on the expected output obtained through the constrained-fit procedure implies progressively weaker complementarity between workers and firms.

The left panel of Figure 4 further shows how $S_{\mathbf{x}}$ in our baseline simulated data varies over the life cycle, that is, over I_n^{t-1}, with $\mathbf{x}$ given by the average value of $(H_{n,1}, P_{n,t}, J_{n,t})$ across workers. The right panel instead shows how $S_{\mathbf{x}}$ varies with the belief $P_{n,t}$ that a worker is of high ability, with $\mathbf{x}$ given by the average value of $(H_{n,1}, J_{n,t})$ across workers and time periods and I_n^{t-1} fixed at 25 (black line), 35 (blue line), and 55 (red line). The left panel shows that the output gains from positive sorting ($S_{\mathbf{x}} > 0$) increase over the life cycle, as workers accumulate labor market experience.

In particular, acquired human capital not only improves expected output—both the intercept and the slope of expected output in (21) increase with experience as illustrated in the two middle panels of the figure—buy also strengthens the complementarity in production between workers and firms. Since the returns to acquiring new human capital decrease with experience, they lead to an overall concave profile for $S_{\mathbf{x}}$ (left panel)—see also the two middle panels showing that the intercept and the slope of expected output increase at a decreasing rate. Then, early in a career, additional experience generates large increases in expected output and magnifies the benefit of assigning workers to the firms that best fit their productive characteristics. Later on, as the growth in expected output decelerates, the gains from improved matching also rise at a decreasing rate, and so does $S_{\mathbf{x}}$.

Note that this finding that experience raises the intercept and the slope of expected output implies that both the expected output of a low-ability worker—the intercept term—and the difference in expected output between a high- and a low-ability worker—the slope term—increase with experience. Hence, both high- and low-ability workers become more productive over time, but high-ability workers more so. As the middle panels of figure 4 show, the slope term increases more for a high-efficiency-type (resp., low-efficiency-type) worker than for a low-efficiency-type (resp., high-efficiency-type) worker in a high-type (resp. low-type) firm. So, in a high-type firm, the difference in expected output between a high- and a low-ability worker increases with experience *more* for an $\bar{e}$-type than an $\underline{e}$-type worker. That is, the complementarity in production between workers with high efficiency $\bar{e}$ and high ability $\bar{\theta}$ and firms of type $\bar{d}$ intensifies over time. By contrast, in a low-type firm, the difference in expected output between a high- and a low-ability worker increases with experience *more* for an $\underline{e}$-type than an $\bar{e}$-type worker. That is, the complementarity in production between workers with low efficiency $\underline{e}$ and high ability $\bar{\theta}$ and firms of type $\underline{d}$ intensifies over time.

The right panel of Figure 4 illustrates that the gains from sorting also increase (approximately linearly) with the belief $P_{n,t}$ that a worker is of high ability, $\theta_n = \bar{\theta}$. Intuitively, as $P_{n,t}$ increases, uncertainty about a worker's true ability being high declines. Expected output increases, since a high-ability worker is more likely to produce high output, and sorting strengthens, as matching becomes increasingly based on true productivity due to the lower uncertainty about $\theta_n = \bar{\theta}$. As the panel shows, $S_{\mathbf{x}}$ increases with beliefs more steeply for older workers, that is, the increase in productivity and in the degree of sorting are more pronounced for workers with greater experience. In this sense, human capital and information have complementary effects on worker productivity and labor market sorting. Indeed, for a given value of $P_{n,t}$, sorting becomes more marked with experience—namely, human capital raises the output complementarity between workers and firms—consistent with the three stacked curves for $S_{\mathbf{x}}$ in the right panel for ages 25 (lowest), 35, and 55 (highest).

To evaluate the magnitude of the sorting measured by $S_{\mathbf{x}}$ and its contribution to wage inequality, we compare the value of $S_{\mathbf{x}}$ in the baseline simulated data with its value in counterfactually simulated data under the assumption that workers randomly match with firms (see Bonhomme et al., 2019 for a similar exercise). When we do so, we find that the index $S_{\mathbf{x}}$ evaluated at the average value of $\mathbf{x}$ decreases from the benchmark value of 4.5 to 3.7, namely, it is 19% lower. The gap between the counterfactual and baseline values of $S_{\mathbf{x}}$ increases with workers' experience: at age 30, the index is 15% lower than in the baseline, and by age 60, it is 27% lower. This random-matching assignment substantially lowers not only labor market sorting but also wage inequality. The wage share of workers at the top 10% of the wage distribution decreases from about 26% in the baseline to a little less than 24%, whereas the wage share of workers at the top 1% of the wage distribution decreases from about 4% in the baseline to 3.5%. The variance of (log) wages decreases by 45%. Hence, the contribution of labor market sorting to wage inequality is *more than twice* as large as the contribution predicted by the AKM wage statistic once appropriately measured (45% vs. 15%).

Labor Market Sorting without Uncertainty. So far, we have proposed measuring the gains from sorting in our framework through the index $S_{\mathbf{x}}$, which primarily captures complementarities between worker-efficiency types e_n and firm types d. In our setting, ability θ_n is another worker productivity attribute unobserved by the econometrician that is relevant for sorting, since it determines the distribution of output in any period and eventually governs the matching of workers to firms, as uncertainty about ability is progressively resolved. Unlike e_n, though, θ_n is initially *unknown* to both workers and firms and is only gradually learned over time. As a result, workers and firms cannot sort directly on θ_n. This informational friction hampers matching on true productivity: workers are unable to immediately match with the firms at which they are most productive given their ability θ_n.

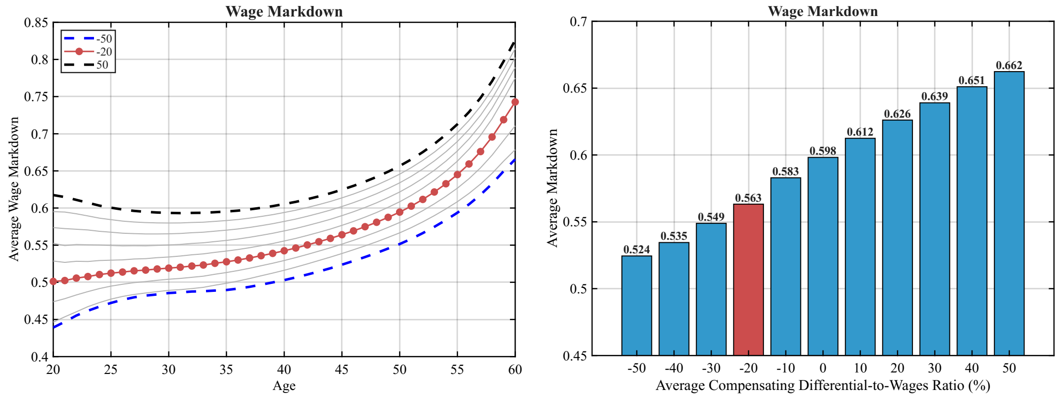
To measure the importance of this effect, we now turn to examining how sorting would change if uncertainty about θ_n were decreased—that is, a scenario in which workers and firms are more certain about θ_n so are closer to being able to sort on both e_n and θ_n. To this purpose, we simulate a version of our model in which uncertainty about ability is reduced. Specifically, we endow high-ability workers ($\theta_n = \bar{\theta}$) with an initial belief $P_{n,1}$ of 0.99 that their ability is high and low-ability workers ($\theta_n = \underline{\theta}$) with an initial belief $P_{n,1}$ of 0.01 that their ability is high—in the baseline, $P_{n,1}$ for each worker n is randomly drawn from a uniform distribution over $[0, 1]$. Once we calculate $S_{\mathbf{x}}$ at the average value of $\mathbf{x}$ using the resulting simulated data, we find that $S_{\mathbf{x}}$ increases to the value of 6 from the value of 4.5 in the baseline. Uncertainty about θ_n compresses not only labor market sorting but also wage inequality. The wage share of workers at the top 10% of the wage distribution increases from about 26% in the baseline to over 28% in this counterfactual scenario—the wage share of workers at the top 1% of the wage distribution slightly increases from about 4% in the baseline to a little over 4%. Accordingly, the variance of wages is 23% higher, that is, much larger than in the baseline.

Firm Monopsony Power. As discussed in Section 2.2, the compensating differential in wages not only blurs traditional AKM-type measures of sorting, but also calls into question the usual interpretation of wage markdowns as evidence of firm monopsony power (Remark 3). To elaborate, a low wage markdown—namely, a low wage-to-output ratio—is often interpreted as evidence of a substantial degree of monopsony power. In our framework, however, workers dynamically acquire new skills and learn about their abilities when employed. Whenever firms differ in the human capital and information prospects they offer, this process of human capital and information acquisition generates a natural force that depresses current wages—namely, an option value in terms of future high wages that is priced into current wages through the compensating differential, lowering them.

Since the wage markdown changes with experience as the compensating differential itself varies, a more appropriate measure of firm monopsony power is an *intertemporal* one that takes into account the entire time path along which the returns to the investments in human capital and information materialize. To shed light on these features, we perform the following exercise. We first estimate our wage equation in (20) on the baseline simulated data and calculate the resulting estimated average wage markdown—given the good performance of our estimator, it is close to the true average wage markdown. We then re-estimate (20) on the *same* baseline simulated *data*, but impose an additional constraint that enforces a *progressively lower or higher* average compensating-differential-to-wage ratio. That is, rather than freely estimating all wage parameters, we restrict the estimation procedure to select values for the wage parameters that deliver a given target level of the average compensating-differential-to-wage ratio.The idea is to estimate a progressively more "misspecified" wage equation in which the compensating differential is counterfactually reduced or increased, mirroring what an econometrician would obtain from a model that abstracts from, or magnifies, the human capital and informational option value of employment at different firms.

The left panel of Figure 5 reports the estimated average wage markdown over the life cycle under the *unconstrained* estimation procedure (red line) and under the *constrained*, misspecified one (grey lines). The dashed blue and black lines correspond to the cases in which we impose the most negative ("-50") and the most positive ("50") average compensating-differential-to-wage ratios, respectively. As the panel shows, even for relatively low levels of the *average* wage markdown—as in the data

Figure 5: Wages and Monopsony Power



Note: The left panel shows the estimated average wage markdown over the life cycle under the unconstrained estimation procedure (red line) and the constrained, misspecified procedure (grey lines). The dashed blue and black lines correspond to the cases in which we impose the most negative ("-50", blue) and the most positive ("50", black) average compensating-differential-to-wage ratios. The right panel shows the estimated average wage markdown under the unconstrained estimation procedure (red bar) and under the constrained, misspecified procedure (blue bars).

(red line)—the markdown increases sharply over the life cycle. Indeed, early on, firms discount in current wages the value of the higher future wages that workers will receive as compensation for the skills and information they acquire while employed at them—firms anticipate that future competition for workers' improved labor services will bid up wages and reward workers for their current investments. As the end of the life cycle nears, investments in human capital or information become less valuable, workers engage less in these activities, and so firms front-load the values they offer to workers. Across the levels of the targeted compensating differential, the implied increase in the share of output that workers appropriate as wages are on the order of 20 to 25 percentage points.

In the right panel of Figure 5, the red bar reports the estimated average wage markdown (averaged across ages) under the unconstrained estimation procedure, whereas the blue bars report the corresponding estimates under the constrained, misspecified procedure. The panel shows that an econometrician estimating a wage equation that imposes *progressively lower values* of the compensating differential (in absolute value) on our baseline simulated data would incorrectly interpret the *same* labor market as exhibiting *progressively lower degrees* of firm competition. Intuitively, the lower the compensating differential, the lower the wage and the wage markdown, and so the higher the inferred degree of firm monopsony power. Likewise, an econometrician estimating a wage equation that imposes *progressively higher values* of the compensating differential would incorrectly interpret the same labor market as exhibiting *progressively higher degrees* of firm competition. Hence, ignoring the wage markdown can grossly misstate the true degree of competition among firms.

5.2 Empirical Application

We now turn to estimate our wage equation in (10) using U.S. Census employer-employee match panel data (LEHD). This rich dataset provides information on quarterly labor earnings for all employed workers across 21 U.S. states from the mid-1990s to 2022, covering about 95% of private-sector jobs and most state/local employment. In these data, we observe each worker's employing firm, or firms if a worker is employed by more than one, labor earnings, gender, education, age, and occupation. Similarly to our simulation exercise in Section 5.1, we select for our analysis a sample of workers who are between 20 and 60 years of age, have at least four quarters of non-missing and non-zero labor earnings—for our results not to be affected by short employment spells—and have gender, education, age, and occupation information, as well as firm identifier, non-missing. Since in every quarter a worker can be employed by multiple firms, we define a worker's "main" employer as the firm providing the largest labor earnings in a quarter.[45] We construct accordingly a measure of annual labor earnings for each individual. Our empirical results—obtained through the estimation procedure outlined in the previous section and currently awaiting approval for disclosure from the Census Bureau—are quantitatively and qualitatively consistent with the results from the simulation exercises discussed. In particular, they corroborate the importance of accounting for workers' life-cycle investments in their productivity and in the process of discovery of their best matches across firms and jobs for assessing the degree of labor market sorting and firm monopsony power.

6 Conclusion

In this paper, we examine the empirical content of a large class of dynamic matching models of the labor market with ex-ante heterogeneous firms and workers, human capital acquisition, symmetric uncertainty and learning about workers' ability, and firm monopsony power. This class nests and extends known models that have been widely applied to study worker turnover across firms, occupational choice, wage growth, wage differentials across occupations, firms, and industries, and wage inequality across workers and over the life cycle. We provide a novel argument to establish that these models are identified solely from data on job choices and wages under intuitive conditions, which allow for arbitrary patterns of selection, are easy to verify, impose minimal data requirements, and yield a simple constructive estimator of the model primitives, as shown in our empirical application.

Through this framework, we revisit an outstanding puzzle about the role of labor market sorting for wage inequality. We demonstrate that ignoring the dynamics of the matching process between firms and workers induced by worker human capital acquisition and learning about worker productivity—and how it is priced into wages when firms differ in the human capital and information opportunities they offer—can lead to systematically understating the importance of sorting for

[45] The measure of labor earnings includes gross wages and salaries, bonuses, stock options, tips and other gratuities, meals and, when part of compensation, lodging and employer contributions to deferred compensation plans.

wages. By the same logic, firms' monopsony power in the labor market can be largely overstated.

References

ABBRING, J. H. AND G. RIDDER (2015): "Regular Variation And The Identification Of Generalized Accelerated Failure-Time Models," *Econometric Theory*, 31, 1229–1248.

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.

ALTONJI, J. G. AND C. R. PIERRET (2001): "Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 313–350.

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.

BANKS, J. S. AND R. K. SUNDARAM (1992): "Denumerable-Armed Bandits," *Econometrica*, 60, 1071–1096.

BARDOCZY, B., G. BORNSTEIN, AND S. SALGADO (2025): "Monopsony Power and the Transmission of Monetary Policy," *Working Paper*.

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.

——— (2006): "Dynamic Price cCompetition," *Journal of Economic THheory*, 127, 232–263.

BERGER, D., K. HERKENHOFF, AND S. MONGEY (2026): "Labor Market Power: From Micro Evidence to Macro Consequences," *Journal of Economic Perspectives*, 40, 93–114.

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

BLACKWELL, D. (1951): "The Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley: University of California Press, 93–102.

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.

BOROVICKA, K. AND R. SHIMER (2024): "Assortative Matching and Wages: The Role of Selection," *Working Paper*.

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.

CHERNOZHUKOV, V. (2005): "Extremal Quantile Regression," *The Annals of Statistics*, 33, 806–839.

CHING, A. (2010): "A Dynamic Oligopoly Structural Model for the Prescription Drug Market after Patent Expiration," *International Economic Review*, 51, 1175–1207.

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.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

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

EECKHOUT, J. AND P. KIRCHER (2011): "Identifying Sorting—In Theory," *Review of Economic Studies*, 78, 872—-906.

FARBER, H. S. AND R. GIBBONS (1996): "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.

GANONG, P., P. NOEL, C. PATTERSON, J. VAVRA, AND A. WEINBERG (2025): "Earnings Instability," *Chicago mimeo*, University of Chicago and NBER.

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.

GRAHAM, B. S. (2011): "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," In *Handbook of Social Economics*, 1B, J. Benhabib, M.O. Jackson, and A. Bisin, North–Holland, 965–1052.

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.

HECKMAN, J. J. AND B. E. HONORÉ (1989): "The Identifiability of the Competing Risks Model," *Biometrika*, 76, 325–330.

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*.

HONORÉ, B. E. AND L. HU (2020): "Selection Without Exclusion," *Econometrica*, 88, 1007–1029.

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 Conference 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.

KEHOE, P., B. LARSEN, C. MALGIERI, AND E. PASTORINO (2024): "Dynamic Competition in the Era of Big Data," *Stanford mimeo*.

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.

KRISTENSEN, D., L. NESHEIM, AND A. DA PAULA (2015): "CCP and the Estimation of Nonseparable Dynamic Models," *UCL mimeo*.

LAMADON, T., J. LISE, C. MEGHIR, AND J.-M. ROBIN (2026): "Labor Market Matching, Wages, and Amenities," *Working Paper*.

LAMADON, T., M. MOGSTAD, AND B. SETZLER (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112, 169–212.

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.

LOW, H., C. MEGHIR, AND L. PISTAFERRI (2010): "Wage Risk and Employment Risk over the Life Cycle," *American Economic Review*, 100, 1432–1467.

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.

NGUYEN, H. D. AND G. MCLACHLAN (2019): "On Approximations via Convolution-Defined Mixture Models," *Communications in Statistics - Theory and Methods*, 48, 3945–3955.

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

PETERSEN, A. V. (1976): "Bounds for a Joint Distribution Function with Fixed Sub-Distribution Functions: Application to Competing Risks," *Proceedings of the National Academy of Sciences of the United States of America*, 73, 11–13.

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.

RUST, J. (1994): "Structural Estimation of Markov Decision Processes," In *Handbook of Econometrics*, 4, Robert F. Engle and Daniel L. McFadden, Eds., Amsterdam, North–Holland, 3081–3143.

SASAKI, Y. AND Y. WANG (2024): "On Uniform Confidence Intervals for the Tail Index and the Extreme Quantile," *Journal of Econometrics*, 244, 105865.

——— (2025): "Extreme Quantile Treatment Effects under Endogeneity," *Forthcoming at Journal of Business & Economic Statistics*.

SEEGMILLER, B. (2021): "Valuing Labor Market Power: The Role of Productivity Advantages," Tech. rep., MIT Sloan Working Paper.

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.

SORKIN, I. (2018): "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 33, 1331–1393.

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

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

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

TIROLE, J. (1988): *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.

TSIATIS, A. (1975): "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proc. Nat. Acad. Sci. USA*, 72, 20–22.

WALDMAN, M. (2012): "Theory and Evidence in Internal Labor Markets," in *Handbook of Organizational Economics*, ed. by R. Gibbons and J. Roberts, Princeton, NJ: Princeton University Press, 520–572.

YEH, C., C. MACALUSO, AND B. HERSHBEIN (2022): "Monopsony in the US Labor Market," *American Economic Review*, 112, 2099–2138.

Appendix

A Omitted Tables and Figures

Table 4: Parameter Values of Expected Output Equation in Baseline Simulated Data

	Low Education						High Education					
Age	Intercept			Slope			Intercept			Slope		
	True	Our	OLS	True	Our	OLS	True	Our	OLS	True	Our	OLS
Panel A $(\underline{e}, \underline{d})$												
25	1.32	1.41	1.51	0.35	0.34	0.57	1.99	1.97	2.26	0.53	0.52	0.82
		(0.001)	(0.018)		(0.001)	(0.025)		(0.001)	(0.027)		(0.003)	(0.088)
35	1.71	1.70	1.95	0.46	0.44	0.69	2.91	2.88	3.34	0.77	0.76	1.16
		(0.000)	(0.011)		(0.001)	(0.017)		(0.000)	(0.018)		(0.010)	(0.063)
55	1.57	1.60	1.82	0.42	0.40	0.59	3.83	3.81	4.52	1.02	0.98	1.40
		(0.000)	(0.012)		(0.001)	(0.020)		(0.001)	(0.020)		(0.019)	(0.074)
Panel B $(\underline{e}, \overline{d})$												
25	1.26	1.35	1.44	0.28	0.27	0.50	1.92	1.90	2.19	0.46	0.45	0.76
		(0.000)	(0.004)		(0.001)	(0.015)		(0.000)	(0.017)		(0.001)	(0.022)
35	1.64	1.63	1.89	0.39	0.37	0.62	2.84	2.81	3.27	0.71	0.69	1.09
		(0.000)	(0.006)		(0.001)	(0.006)		(0.000)	(0.003)		(0.002)	(0.011)
55	1.51	1.53	1.75	0.35	0.33	0.52	3.76	3.75	4.46	0.95	0.92	1.33
		(0.000)	(0.009)		(0.001)	(0.007)		(0.000)	(0.019)		(0.005)	(0.072)
Panel C $(\overline{e}, \underline{d})$												
25	1.22	1.31	1.39	0.26	0.24	0.46	1.89	1.87	2.16	0.42	0.42	0.73
		(0.002)	(0.010)		(0.014)	(0.038)		(0.003)	(0.013)		(0.114)	(0.169)
35	1.61	1.60	1.84	0.36	0.34	0.59	2.81	2.77	3.24	0.68	0.65	1.05
		(0.003)	(0.006)		(0.011)	(0.027)		(0.003)	(0.008)		(0.072)	(0.121)
55	1.47	1.48	1.71	0.32	0.30	0.48	3.73	3.71	4.42	0.92	0.88	1.30
		(0.004)	(0.006)		(0.020)	(0.035)		(0.013)	(0.008)		(0.283)	(0.153)
Panel D $(\overline{e}, \overline{d})$												
25	1.39	1.48	1.57	0.42	0.41	0.64	2.06	2.03	2.33	0.60	0.59	0.89
		(0.099)	(0.425)		(0.054)	(0.365)		(0.255)	(0.603)		(0.319)	(0.901)
35	1.78	1.76	2.02	0.52	0.51	0.76	2.97	2.94	3.41	0.84	0.82	1.22
		(0.137)	(0.314)		(0.087)	(0.270)		(0.237)	(0.416)		(0.340)	(0.633)
55	1.64	1.67	1.88	0.49	0.47	0.66	3.90	3.88	4.59	1.09	1.05	1.46
		(0.213)	(0.393)		(0.121)	(0.337)		(0.805)	(0.573)		(1.127)	(0.865)

Note: The table reports the intercept and slope parameters of expected output in (21) used to generate the simulated data (columns "True"), those estimated using our approach (columns "Our"), and those estimated by OLS (columns "OLS") for each worker-efficiency type and firm type and for workers with high and low education at ages 25, 35, and 55 averaged across jobs, together with their standard errors (in parentheses).

Table 5: Parameter Estimates for Expected Output Equation Across Monte Carlo Samples

Age	Low Education						High Education					
	Intercept			Slope			Intercept			Slope		
	Mean	Std. Dev.	p50	Mean	Std. Dev.	p50	Mean	Std. Dev.	p50	Mean	Std. Dev.	p50
Panel A $(\underline{e}, \underline{d})$												
25	1.32	0.05	1.32	0.36	0.02	0.36	1.66	0.30	1.69	0.54	0.04	0.54
35	1.72	0.04	1.72	0.43	0.02	0.43	3.06	0.31	3.08	0.72	0.04	0.72
55	1.60	0.03	1.60	0.41	0.02	0.41	5.07	0.29	5.08	1.03	0.05	1.03
Panel B $(\underline{e}, \overline{d})$												
25	1.25	0.06	1.26	0.29	0.02	0.29	1.59	0.32	1.62	0.47	0.05	0.47
35	1.65	0.05	1.65	0.36	0.04	0.36	2.99	0.32	3.02	0.65	0.05	0.65
55	1.53	0.05	1.53	0.34	0.02	0.34	5.00	0.31	5.01	0.97	0.06	0.97
Panel C $(\overline{e}, \underline{d})$												
25	1.25	0.05	1.25	0.29	0.04	0.30	1.59	0.32	1.61	0.47	0.05	0.47
35	1.65	0.07	1.65	0.36	0.02	0.36	2.99	0.31	3.02	0.65	0.05	0.65
55	1.53	0.05	1.53	0.34	0.02	0.35	4.99	0.30	5.01	0.96	0.06	0.96
Panel D $(\overline{e}, \overline{d})$												
25	1.39	0.06	1.39	0.43	0.07	0.43	1.74	0.34	1.77	0.61	0.07	0.61
35	1.78	0.09	1.78	0.49	0.03	0.49	3.12	0.31	3.15	0.79	0.05	0.79
55	1.66	0.07	1.67	0.48	0.02	0.48	5.12	0.30	5.14	1.10	0.07	1.10

Note: The table reports features of the distribution of the estimated expected-output parameters in (21) across 200 replication samples, specifically, the mean (columns "Mean"), standard deviation (columns "Std. Dev."), and median (columns "p50"), for each firm and worker-efficiency type, workers with high and low education, at ages 25, 35, and 55, averaged across jobs.

B Formal Identification Argument

This section formalizes the identification approach previewed in Section 3 for our general class of models. Appendices B.1, B.2, and B.3 develop the mixture step of Section 3.3 to identify the learning process, the law of motion for the state, and the CCPs. Appendices B.4, B.5, and B.6 develop the extremal quantile step of Section 3.4 to identify the deterministic component of potential wages, the shock distribution, the expected output, the compensating differential, and the human capital function. Assumptions 1, 2, and 3 are assumed to hold throughout. In particular, we present results and proofs for the less trivial case covered by Assumption 3, namely when worker n's labor market $\mathcal{D}(X_{n,t}) \subseteq \mathcal{D}$ contains more than two firms and there exists a known mapping m_t such that $D'_{n,t} = m_t(D_{n,t}, s_{n,t})$. To keep notation simple, we assume without loss of generality that $\mathcal{D}(X_{n,t}) = \mathcal{D}$.

B.1 Wage Mixture

Using the law of total probability, we can represent the wage distribution at time t, conditional on worker n's observed initial human capital $H_{n,1}$, job history $D_{n,t}^{t-1}$, and current employing firm $D_{n,t}$, as a mixture over latent classes indexed by the efficiency type e_n and by the history of noisy performance signals η_n^{t-1}. Namely,

$$\Pr\big(w_{n,t} \le w \mid H_{n,1},\ D_n^t\big) = \sum_{(e,\eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}} \Pr\big(w_{n,t} \le w \mid H_{n,1},\ D_n^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1}\big) \times \Pr\big(e_n = e,\ \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1},\ D_n^t\big). \tag{23}$$

In what follows, we show identification of the mixture components and weights.

We identify the mixture components and weights in (23) under the condition that the wage distribution admits a *clusterable finite mixture* representation; that is, it can be written as a finite mixture whose components are sufficiently distinct—and hence *clusterable*, in the sense formalised by Aragam et al. (2020, Condition (19))—to be identifiable. A canonical and widely studied example is a finite mixture of possibly uncountable Gaussian mixtures whose means and variances are sufficiently separated across components, as shown by Bruni and Koch (1985, p.1354) and Aragam et al. (2020, p.31). Below, to keep the exposition streamlined and to avoid unnecessary technicalities, we present the identification argument in this canonical setting. We emphasise, however, that the underlying logic is more general and extends beyond finite mixtures of possibly uncountable Gaussian mixtures.

Assumption 4 (Finite mixture of possibly uncountable Gaussian mixtures)**.** *For each* $t \in \{1, \ldots, T\}$ *and* $(h, d^t) \in \mathcal{H} \times \mathcal{D}^t$*, assume:*

(i) (Mixture of Normals.) For each $(e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$ *and conditional on* $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1})$*, the productivity shock of the time-t second-best firm* $D'_{n,t} = d' \in \mathcal{D}$*,* $\epsilon_{n,t}(d')$*, is distributed as a mixture of a (possibly uncountable) family of Gaussian distributions. Formally, let* $f_{\epsilon_{n,t}(d')}(\cdot \mid H_{n,1} = h, D_n^t = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1})$ *denote the density of* $\epsilon_{n,t}(d')$ *conditional on* $(H_{n,1} = h, D_n^t = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1})$*. Then, for each* $r \in \mathbb{R}$*,*

$$f_{\epsilon_{n,t}(d')}(r|H_{n,1}{=}h, D_n^t{=}d^t, e_n{=}e, \eta_n^{t-1}{=}\eta^{t-1}) = \int_{(\mu,\sigma^2)\in\mathcal{G}_{h,d^t,e,\eta^{t-1}}} \phi(r;\mu,\sigma^2)d\pi(\mu,\sigma^2;h,d^t,e,\eta^{t-1}),$$

where $\phi(\cdot;\mu,\sigma^2)$ *is the Gaussian density with mean* μ *and variance* σ^2*;* $\mathcal{G}_{h,d^t,e,\eta^{t-1}} \subset \mathbb{R} \times (0,\infty)$ *is the (possibly unknown) support of the Gaussian parameters* (μ, σ^2) *conditional on* $(H_{n,1} = h, D_n^t = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1})$*; and* $\pi(\cdot; h, d^t, e, \eta^{t-1})$ *is a probability measure on* $\mathcal{G}_{h,d^t,e,\eta^{t-1}}$*, representing the distribution of* (μ, σ^2) *conditional on* $(H_{n,1} = h, D_n^t = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1})$*.*

(ii) (Supports.) The supports $\mathcal{E}$ *of* e_n *and* $\mathcal{N}$ *of* $\eta_{n,t}$ *are finite sets with known cardinality.*

(iii) (Compactness.) For each $(e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$*, the set* $\mathcal{G}_{h,d^t,e,\eta^{t-1}}$ *is a compact subset of* $\mathbb{R} \times (0,\infty)$*.*

(iv) (Non-Overlap.) There exists a Borel subset $G_{h,d^t,e,\eta^{t-1}} \subseteq \mathcal{G}_{h,d^t,e,\eta^{t-1}} \subset \mathbb{R} \times (0,\infty)$ *such that* $\pi(G_{h,d^t,e,\eta^{t-1}}; h, d^t, e, \eta^{t-1}) = 1$ *for each* $(e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$*. Moreover,* $G_{h,d^t,e,\eta^{t-1}} \cap G_{h,d^t,\tilde{e},\tilde{a}^{t-1}} = \emptyset$ *for each* $(e, \eta^{t-1}) \neq (\tilde{e}, \tilde{\eta}^{t-1})$ *with* $(e, \eta^{t-1}), (\tilde{e}, \tilde{\eta}^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$*.*

Assumption 4(i) imposes that, conditional on $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1})$, the productivity shock of the second-best firm d' at time t, $\epsilon_{n,t}(d')$, is distributed as a mixture of a possibly uncountable family of Gaussian distributions. Combined with Assumptions 3 and 4(ii), this implies that the the distribution of $w_{n,t}$ conditional on $(H_{n,1} = h,\ D_n^t = d^t)$ is a *finite mixture whose components are possibly uncountable Gaussian mixtures*.[46]

[46] We remark that the wage mixture in (23) is *not* assumed to be a finite mixture of Gaussians, but a finite mixture of possibly uncountable Gaussian mixtures, which is a much more general specification.

To see why, recall: *(a)* Our wage equation in (10) is a function of $D_{n,t}$, $D'_{n,t}$, $s_{n,t}$, and $\epsilon_{n,t}(D'_{n,t})$. Thus, the distribution of $w_{n,t} \mid (D_{n,t}, D'_{n,t}, s_{n,t})$ is the distribution of $\epsilon_{n,t}(D'_{n,t}) \mid (D_{n,t}, D'_{n,t}, s_{n,t})$ shifted by the unknown potential-wage component $\varphi(\cdot)$. *(b)* Under Assumption 3, $D'_{n,t}$ is a known deterministic function of $(D_{n,t}, s_{n,t})$. *(c)* The state $s_{n,t}$ evolves deterministically as a function of $(H_{n,1}, D_n^{t-1}, e_n, \eta_n^{t-1})$ (see footnote 22). Therefore, fixing $(H_{n,1},\ D_n^t,\ e_n,\ \eta_n^{t-1})$ also fixes the value of $s_{n,t}$. Moreover, since $D_{n,t}$ is the t-th element of D_n^t, its value is fixed as well. Taken together, *(b)* and *(c)* imply that, conditional on $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1})$, $D'_{n,t}$ takes a known value $d' \in \mathcal{D}$. Thus, by *(a)*, each mixture component in (23) satisfies

$$\begin{aligned}
&\Pr\Big(w_{n,t} \leq w \,\Big|\, H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1}\Big) \\
&\qquad = \Pr\bigg(\epsilon_{n,t}(d') \leq w - \varphi(d, d', s) \,\bigg|\, H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1}\bigg).
\end{aligned}$$

That is, each mixture component is governed by the distribution of the productivity shock $\epsilon_{n,t}(d')$ conditional $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1})$, shifted by the unknown potential-wage component $\varphi(\cdot)$. By Assumption 4(i), this conditional distribution of $\epsilon_{n,t}(d')$ is a mixture of a possibly uncountable family of Gaussian distributions. Since e_n and η_n^{t-1} can take only finitely many values under Assumption 4(ii), it follows that the distribution of $w_{n,t}$ conditional on $(H_{n,1} = h,\ D_n^t = d^t)$ is a finite mixture whose components are possibly uncountable Gaussian mixtures.

Assumption 4(ii) posits that efficiency e_n and the signal $\eta_{n,t}$ have finite supports, $\mathcal{E}$ and $\mathcal{N}$, with known cardinalities. Given the finiteness of $\mathcal{E}$, e_n can be viewed as a worker "finite fixed effect," following the terminology of Bonhomme et al. (2019). In the illustrative model in Section 2, we assumed for simplicity that each signal $\eta_{n,t}$ takes only two values. Assumption 4(ii) instead allows for an arbitrary (but finite) cardinality. It is also straightforward to relax this assumption by requiring the researcher to know only upper bounds on the cardinalities of $\mathcal{E}$ and $\mathcal{N}$; in that case, the result in Bruni and Koch (1985) can be used to recover the cardinalities of $\mathcal{E}$ and $\mathcal{N}$. Appendix D discusses an extension to the case in which e_n and $\eta_{n,t}$ are continuous, multidimensional random vectors. See also footnote 23 regarding the supports of e_n and $\eta_{n,t}$.

Assumptions 4(iii) is a regularity condition and requires that all Gaussian means and variances (μ, σ^2) that may arise lie in a bounded rectangle, with variances bounded away from 0 and ∞.

Assumption 4(iv) is a standard separation condition in the identification of mixture models and requires that the mixing distributions $\{\pi(\cdot; h, d^t, e, \eta^{t-1}) : (e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}\}$ place all their mass on pairwise disjoint sets $G_{h,d^t,e,\eta^{t-1}} \subseteq \mathcal{G}_{h,d^t,e,\eta^{t-1}}$ in the (μ, σ^2)-space. Otherwise, they may not be distinguished. Importantly, it does not require the densities $\{f_{\epsilon_{n,t}(d')}(r \mid H_{n,1} = h, D_n^t = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1}) : (e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}\}$, and so the wage mixture components, to have disjoint supports, and allows them to overlap arbitrarily. This corresponds to the clusterability condition of Aragam et al. (2020, Condition (19)) for the case of finite mixtures of possibly uncountable Gaussian mixtures.

Proposition B.1 formalizes the identification result under Assumption 4.

Proposition B.1 (Wage Mixture)**.** *Let Assumption 4 holds. Then, for each $t \in \{1, \ldots, T\}$ and $(h, d^t) \in \mathcal{H} \times \mathcal{D}^t$:*

(i) The probability $\Pr\big(e_n = e,\ \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h,\ D_n^t = d^t\big)$ *is identified for each* $(e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$.

(ii) The probability $\Pr\big(w_{n,t} \leq w \mid H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ \eta_n^{t-1} = \eta^{t-1}\big)$ *is identified for each* $(e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$ *and* $w \in \mathbb{R}$.

The proof of Proposition B.1 is omitted because it follows directly from a straightforward application of Bruni and Koch (1985). Specifically, the wage mixture (23) corresponds to the mixture model discussed in Section 4.c of Bruni and Koch (1985) and is shown to be identified under Assumption 4.[47] As is well known and as mentioned in Section 3.3, the wage mixture (23) is identified only up to labeling of its components and weights. As we proceed, we clarify when fixing a labeling is necessary and how we do so.

Lastly, as an immediate implication of Proposition B.1, we identify the joint distribution of histories of signals by combining the wage mixture weights across periods. We report below the identification of two specifications of signal histories that will be useful for the arguments that follow.

Corollary B.1 (Signal Distribution)**.** *(a) Assume that the wage mixture weights in* (23) *are identified and labeled at times* t *and* $t+1$, *with* $t \in \{1, \ldots, T-1\}$. *(See Proposition B.1 for sufficient conditions.) Then, the conditional signal distribution*

$$\Pr\big(\eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^t = d^t,\ e_n = e\big),$$

is identified for each $(\eta^t, h, d^t, e) \in \mathcal{N}^t \times \mathcal{H} \times \mathcal{D}^t \times \mathcal{E}$.
(b) Assume that the wage mixture weights in (23) *are identified and labeled at times* $t+2$ *and* $t+3$, *with* $t \in \{1, \ldots, T-3\}$. *(See Proposition B.1 for sufficient conditions.) Then, the conditional distribution of three consecutive signals at firm* d,

$$\Pr\big(\eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} | H_{n,1} = h, D_{n,t} = d, D_{n,t+1} = d, D_{n,t+2} = d,\ e_n = e\big),$$

is identified for each $(\eta_t, \eta_{t+1}, \eta_{t+2}, h, d, e) \in \mathcal{N}^3 \times \mathcal{H} \times \mathcal{D} \times \mathcal{E}$.

B.2 Learning Process

In this section, we use the wage mixture weights in (23), identified by Proposition B.1, to identify the learning process about a worker's ability θ_n, which is characterized by the prior belief $(P_{n,1}, 1-P_{n,1})$ about θ_n being high ($\bar{\theta}$) or low ($\underline{\theta}$) and by the distribution of signals conditional on θ_n—that is, the probabilities $\alpha(H_{n,1}, D_{n,t}, e_n)$ and $\beta(H_{n,1}, D_{n,t}, e_n)$ defined in Section 2. To this end, we first introduce an assumption that disciplines the distribution of signals conditional on θ_n.

Assumption 5 (Signal Distribution Conditional on Ability)**.** *(i)* $\mathcal{N} := \{\bar{\eta}, \underline{\eta}\}$ *and* $\Theta := \{\bar{\theta}, \underline{\theta}\}$.

[47] If Assumption 4(ii) is relaxed so that the researcher need only know upper bounds on the cardinalities of $\mathcal{E}$ and $\mathcal{N}^{t-1}$ (as mentioned above), then Proposition B.1 also identifies the cardinalities of $\mathcal{E}$ and $\mathcal{N}^{t-1}$.

(ii) Signals are conditionally independent over time: for each $t \in \{1, \ldots, T-k\}$ and integer $k > 0$,

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

(iii) The distribution of $\eta_{n,t}$ conditional on $(H_{n,1}, D_{n,t}, e_n, \theta_n)$ is time-invariant, with $\alpha(h,d,e) := \Pr\big(\eta_{n,t} = \bar{\eta} \mid H_{n,1} = h,\ D_{n,t} = d,\ e_n = e,\ \theta_n = \bar{\theta}\big)$ and $\beta(h,d,e) := \Pr\big(\eta_{n,t} = \bar{\eta} \mid H_{n,1} = h,\ D_{n,t} = d,\ e_n = e,\ \theta_n = \underline{\theta}\big)$, for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$, and $\alpha(h,d,e) > \beta(h,d,e)$.

Assumption 5(i) imposes that the signal $\eta_{n,t}$ and the latent ability θ_n have finite supports of cardinality two.[48] We already adopted this assumption in Section 2 to simplify the description of the learning process; see, for instance, equation (3). It is therefore convenient to maintain it when identifying this learning process. In Appendix D, we discuss how this assumption can be relaxed to allow for supports of cardinality greater than two, as well as for continuous and multidimensional $\eta_{n,t}$ and θ_n. Assumption 5(ii) imposes that signals are conditionally independent over time, as is standard in learning models. Assumption 5(iii) requires that the distribution of $\eta_{n,t}$ conditional on the chosen job and θ_n is time-invariant and described by the parameters $\alpha(h,d,e)$ and $\beta(h,d,e)$. The latter is an unavoidable requirement for identifying the learning process: if that distribution varied over time, we could not recover belief dynamics solely from observing workers who remain employed in the same job for sufficiently many periods. Assumption 5(iii) also imposes $\alpha(h,d,e) > \beta(h,d,e)$, which is a natural restriction since high-ability types are more likely to generate high signals.

In addition to Assumption 5, the identification of the learning process also builds on Corollary B.1. In particular, it requires the identification of the conditional distribution of three consecutive signals at firm d to be identified, for which sufficient conditions are provided by Corollary B.1(b), and of the conditional distribution of the initial signal at firm d, for which sufficient conditions are provided by Corollary B.1(a). These sufficient conditions essentially amount to the identification of certain wage mixture weights in (23) (see more details at the end of the section).

Proposition B.2 (Learning Process)**.** *Suppose that:*

(i) Assumption 5 hold.

(ii) For each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ and for some $t \in \{1, \ldots, T-3\}$, the conditional distribution of three consecutive signals at firm d,

$$\Pr\big(\eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1},\ \eta_{n,t+2} = \eta_{t+2} | H_{n,1} = h, D_{n,t} = d, D_{n,t+1} = d, D_{n,t+2} = d, e_n = e\big),$$

is identified for each $(\eta_t, \eta_{t+1}, \eta_{t+2}) \in \mathcal{N}^3$, with t possibly varying across (h,d,e), and the conditional distribution of the initial signal at firm d,

$$\Pr(\eta_{n,1} = \eta \mid H_{n,1} = h,\ D_{n,1} = d,\ e_n = e),$$

is identified for each $\eta \in \mathcal{N}$. See Corollary B.1 for sufficient conditions.

[48] The mixture identification in Proposition B.1 assumes that the signal $\eta_{n,t}$ has a finite support with any cardinality.

Then, $\alpha(h,d,e)$*,* $\beta(h,d,e)$*, the prior belief* $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$*, and the set of realizations of the posterior beliefs* $\{P_{n,t}\}_{t=2}^{T}$ *are identified for each* $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$*.*

The proof of Proposition B.2 is straightforward. Under Assumption 5 and condition (ii) from Proposition B.2, we can represent the *identified* signal distribution $\Pr\big(\eta_{n,t} = \eta_t,\ \eta_{n,t+1} = \eta_{t+1},\ \eta_{n,t+2} = \eta_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d,\ e_n = e\big)$ as a *binomial* mixture over θ_n with two components, characterized by $\alpha(h,d,e)$ and $\beta(h,d,e)$, and three trials. The components of this mixture can be identified using the results in Blischke (1964, 1978) for binomial mixtures (without any labelling indeterminacy by Assumption 5(iii)). This explains why condition (ii) of Proposition B.2 requires observing workers employed at job d for three consecutive periods: by Blischke (1964, 1978), at least three trials are needed to identify two binomial mixture components. Still using Assumption 5 and condition (ii) from Proposition B.2, we can represent the *identified* initial signal distribution $\Pr(\eta_{n,1} = \eta \mid H_{n,1} = h,\ D_{n,1} = d,\ e_n = e)$ as a Bernoulli mixture over θ_n with two components, again characterized by $\alpha(h,d,e)$ and $\beta(h,d,e)$, and with mixture weights given by the prior belief, $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e)$ and $1 - \Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e)$. Since $\alpha(h,d,e)$ and $\beta(h,d,e)$ are already identified, we can readily identify $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e)$. In turn, the set of realizations of the posterior beliefs $\{P_{n,t}\}_{t=2}^{T}$ is identified, since each $P_{n,t}$ can be computed recursively as in equation (3) using $\alpha(h,d,e)$, $\beta(h,d,e)$, and $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$. The formal proof of Proposition B.2 is in Appendix E.

Note that, to establish Proposition B.2, we need a minimum panel length of four periods ($T \geq 4$). This is due to two requirements: *(I)* By Corollary B.1(b), identifying $\Pr\big(\eta_{n,t}, \eta_{n,t+1}, \eta_{n,t+2} \mid H_{n,1},\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d,\ e_n\big)$—as in condition (ii) of Proposition B.2—relies on recovering the mixture weights at times $t+2$ and $t+3$ for some t, which is possible only when at least four periods are observed. *(II)* By Corollary B.1(a), identifying $\Pr\big(\eta_{n,1} \mid H_{n,1},\ D_{n,1} = d,\ e_n\big)$—as in condition (ii) of Proposition B.2—relies on recovering the mixture weights at times $t = 1$ and $t = 2$. Therefore, if there is sufficient variation in workers' job and signal histories, the first four periods are sufficient to satisfy condition (ii) of Proposition B.2 for each firm d.

Finally, as the proof of Corollary B.1(a) highlights, identifying $\Pr\big(\eta_{n,t}, \eta_{n,t+1}, \eta_{n,t+2} \mid H_{n,1},\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d,\ e_n\big)$, relies on combining the recovered mixture weights at times $t+2$ and $t+3$ for some t. Similarly, identifying $\Pr\big(\eta_{n,1} \mid H_{n,1},\ D_{n,1} = d,\ e_n\big)$ relies on combining the recovered mixture weights at times $t = 1$ and $t = 2$. This necessitates consistent labeling of the mixture weights across those periods. To do so, we use the variances of the mixture components to order the weights with respect to e_n—as is standard in the literature—and means of the mixture components to order the weights with respect to η_n^{t-1}. The latter follows from natural monotonicity conditions characterizing the class of models we study: higher-ability workers are more likely to generate higher signals, which implies that the expected output component of wages are monotone in beliefs for any given e_n. See Section 3.3 and footnote 26 for details. Once the learning process is identified, the remaining primitives—namely the law of motion of the state, the CCPs, the deterministic wage component, the shock distribution, the expected output, the compensating differential, and the human capital function—can be recovered without combining mixture weights and components across time, and hence without any further labeling step.

B.3 Law of Motion of the State and Conditional Choice Probabilities

In this section, we use the learning process identified by Proposition B.2, to recover the law of motion of the state and the CCPs. As no additional assumptions are required, we state the formal results directly and provide the proofs in Appendix E. Before doing so, note that, having identified the learning process by Proposition B.2, the map from realizations of $(H_{n,1}, D_{n,t}^{t-1}, e_n, \eta_n^{t-1})$ to realizations of $s_{n,t} := (H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$ is identified. Denote this map by g_t, that is, $(H_{n,1}, D_{n,t}^{t-1}, e_n, \eta_n^{t-1}) \mapsto s_{n,t} = g_t(H_{n,1}, D_{n,t}^{t-1}, e_n, \eta_n^{t-1})$. Denote by $\mathcal{S}_t$ the image set of g_t, which is also identified. Hereafter, refer to $\mathcal{S}_t$ as the support of $s_{n,t}$.

Proposition B.3 (Law of Motion of the State)**.** *Assume that $\alpha(h,d,e)$, $\beta(h,d,e)$, and the prior belief $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$ are identified for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$. See Proposition B.2 for sufficient conditions. Then, the law of motion of the state, $\Pr(s_{n,t} = s \mid D_{n,t-1} = d,\ s_{n,t-1} = \tilde{s})$, is identified for each $s \in \mathcal{S}_t$, $d \in \mathcal{D}$, $\tilde{s} \in \mathcal{S}_{t-1}$, and $t \in \{2, \dots, T\}$.*

Proposition B.4 (Conditional Choice Probabilities)**.** *Assume that, for $t \in \{1, \dots, T\}$, the wage mixture weights in* (23) *are identified at time t, together with $\alpha(h,d,e)$, $\beta(h,d,e)$, and the prior belief $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$ for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$. See Propositions B.1 and B.2 for sufficient conditions. Then, the conditional choice probability $\Pr(D_{n,t} = d | s_{n,t} = s)$ is identified for each $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$.*

B.4 Deterministic Wage Component

In this section, we identify the deterministic potential-wage component $\varphi(\cdot) := y(\cdot) + \Psi(\cdot)$. We begin by introducing additional notation. Recall that, under Assumption 3, $D'_{n,t}$ is a deterministic function m_t of $(D_{n,t}, s_{n,t})$. Hence, fixing $(D_{n,t}, s_{n,t})$ also pins down the value of $D'_{n,t}$. For any $(d, d') \in \mathcal{D}^2$, define $\mathcal{S}_t(d, d') \subseteq \mathcal{S}_t$ as the set of realizations s of $s_{n,t}$ such that $m_t(d, s) = d'$. Because $m_t(\cdot)$ is known under Assumption 3 and $\mathcal{S}_t$ is identified, the set $\mathcal{S}_t(d, d')$ is identified for each $(d, d') \in \mathcal{D}^2$.

We then establish identification of the distribution of wages $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$, which will be a key ingredient for recovering $\varphi(\cdot)$.

Proposition B.5 (Conditional Wage Distribution)**.** *Assume that, for $t \in \{1, \dots, T\}$, the wage mixture weights in* (23) *are identified at time t, together with $\alpha(h,d,e)$, $\beta(h,d,e)$, and the prior belief $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$ for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$. See Propositions B.1 and B.2 for sufficient conditions. Then, the conditional wage distribution $\Pr\big(w_{n,t} \le w \mid D_{n,t} = d,\ s_{n,t} = s\big)$ is identified for each $w \in \mathbb{R}$, $d \in \mathcal{D}$, and $s \in \mathcal{S}_t$.*

We next introduce additional assumptions.

Assumption 6 (Supports)**.** *Let $t \in \{1, \dots, T\}$, $(d, d') \in \mathcal{D}^2$, and $s \in \mathcal{S}_t(d, d')$. Then, $\omega_t(d, s) := \inf\{w : \Pr(w_{n,t}(d, d') \le w \mid s_{n,t} = s) > 0\} = -\infty$ and $\omega_t^{\text{obs}}(d, s) := \inf\{w : \Pr(w_{n,t} \le w \mid D_{n,t} = d, s_{n,t} = s) > 0\} = -\infty$, where $w_{n,t}(d, d') := \varphi(d, d', s_{n,t}) + \epsilon_{n,t}(d')$ denotes worker n's potential wage in period t when the first- and second-best firms are d and d', respectively.*

Assumption 7 (Tail Limit)**.** *Let* $t \in \{1, \ldots, T\}$*,* $(d, d') \in \mathcal{D}^2$*, and* $s \in \mathcal{S}_t(d, d')$*. There exists an (unknown) constant* $q_t(d, d') \in (0, 1]$ *such that* $\lim_{w \to -\infty} \Pr\big(D_{n,t} = d \mid s_{n,t} = s,\ w_{n,t}(d, d') < w\big) = q_t(d, d')$*.*

Assumption 8 (Tail Regularity)**.** *Let* $t \in \{1, \ldots, T\}$*,* $(d, d') \in \mathcal{D}^2$*, and* $s \in \mathcal{S}_t(d, d')$*. There exist (unknown) thresholds* $a_t(d, s) > -\infty$ *and* $a_t^{\text{obs}}(d, s) > -\infty$ *such that the cumulative distribution functions* $F_{w_{n,t}(d,d')|s_{n,t}=s}$ *and* $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ *are continuous and strictly increasing on* $(-\infty, a_t(d, s))$ *and* $(-\infty, a_t^{\text{obs}}(d, s))$*, respectively.*

Assumption 9 (Normalization)**.** *Let* $t \in \{1, \ldots, T\}$*,* $(d, d') \in \mathcal{D}^2$*, and* $s \in \mathcal{S}_t(d, d')$*. There exists a known* $\overline{s} \in \mathcal{S}_t(d, d')$ *such that* $\varphi(d, d', \overline{s}) = 0$*.*

Assumption 6 requires that both the distribution of potential wages $w_{n,t}(d, d')|s_{n,t} = s$ (unobserved) and the distribution of observed wages $w_{n,t}|(D_{n,t} = d, s_{n,t} = s)$ have unbounded left support. This requirement is inessential, however, and it serves only to simplify the argument.[49]

Assumption 7 imposes a common limit $q_t(d, d') \in (0, 1]$, independent of $s \in \mathcal{S}_t(d, d')$, on $\Pr(D_{n,t} = d \mid s_{n,t} = s, w_{n,t}(d, d') < w)$ as the potential wage $w_{n,t}(d, d')$ becomes large. The assumption has two parts, with the second strengthening the first. First, it requires that the limit $\lim_{w \to -\infty} \Pr(D_{n,t} = d \mid s_{n,t} = s, w_{n,t}(d, d') < w)$ exists and is strictly positive. Second, it requires that this limit is invariant across $s \in \mathcal{S}_t(d, d')$. See Appendix G.1 in the Online Supplementary Material for a microfoundation of Assumption 7 and related discussion in Section 3.4.

Assumption 8 is a tail-regularity condition: continuity and strict monotonicity of the left tails of the relevant wage CDFs ensure a one-to-one mapping between right tail probabilities and quantiles. This condition is satisfied by many parametric families of distributions, including both thin-tailed and fat-tailed ones.

Finally, Assumption 9 is a location normalization. As in standard Roy models, wages are identified only up to an additive constant, here $\varphi(d, d', \bar{s})$. Alternatively, the error term can be normalised to have zero unconditional mean or median (French and Taber, 2011).

We now use Assumptions 6 to 9 to establish the identification of $\varphi(\cdot)$.

Proposition B.6 (Deterministic Wage)**.** *Let* $t \in \{1, \ldots, T\}$*. Assume:*

(i) The conditional wage distribution $\Pr\big(w_{n,t} \le w \,\big|\, D_{n,t} = d,\ s_{n,t} = s\big)$ *is identified for each* $w \in \mathbb{R}$*,* $d \in \mathcal{D}$*, and* $s \in \mathcal{S}_t$*. See Proposition B.5 for sufficient conditions. The conditional choice probability* $\Pr(D_{n,t} = d \mid s_{n,t} = s)$ *is identified for each* $d \in \mathcal{D}$ *and* $s \in \mathcal{S}_t$*. See Proposition B.4 for sufficient conditions.*

[49] Exactly one of the following cases obtains: (a) $\omega_t(d, s) = \omega_t^{\text{obs}}(d, s) = -\infty$; (b) $\omega_t(d, s) = \omega_t^{\text{obs}}(d, s) > -\infty$; (c) $\omega_t^{\text{obs}}(d, s) > \omega_t(d, s) \ge -\infty$. We ignore $\omega_t^{\text{obs}}(d, s) < \omega_t(d, s)$ as $\text{supp}(w_{n,t}|D_{n,t} = d, s_{n,t} = s) \subseteq \text{supp}(w_{n,t}(d, d')|s_{n,t} = s)$ implies $\omega_t^{\text{obs}}(d, s) \ge \omega_t(d, s)$. Case (a) is covered by Assumption 6. Under (b), Proposition B.6 and its proof go through with minimal edits—replace limits as $w \to -\infty$ with limits as $w \to \omega_t(d, s)$. Under (c)—where the left endpoint of the distribution of observed wages $w_{n,t}|(D_{n,t} = d, s_{n,t} = s)$ can differ from, and be finite relative to, the left endpoint of the distribution of potential wages $w_{n,t}(d, d')|s_{n,t} = s$ so that selection affects the shape *and support* of the distribution of wages—the identification result retains the spirit of Proposition B.6, but limits need to be taken appropriately because the two endpoints differ. In Appendix G.2 of the Online Supplementary Material, we treat case (c) and further show that, when finite, the endpoints of the potential wage distribution $w_{n,t}(d, d')|s_{n,t} = s$ and of the shock $\epsilon_{n,t}(d')$ can be nonparametrically identified.

(ii) Assumptions 6 to 9 hold.

For $(d, d') \in \mathcal{D}^2$ and $s \in \mathcal{S}_t(d, d')$, let $\{\tau^{(k)}_{d,\bar{s},t}\}_{k\geq 1} \subset (0,1)$ be any sequence with $\tau^{(k)}_{d,\bar{s},t} \to 1$ as $k \to \infty$. Define

$$\tau^{(k)}_{d,s,t} \;:=\; 1 - \frac{\Pr(D_{n,t} = d \mid s_{n,t} = s)}{\Pr(D_{n,t} = d \mid s_{n,t} = \bar{s})}\left(1 - \tau^{(k)}_{d,\bar{s},t}\right).$$

Then,

$$\lim_{k\to\infty}\left[Q_{w_{n,t}\,|\,D_{n,t}=d, s_{n,t}=s}\big(\tau^{(k)}_{d,s,t}\big) - Q_{w_{n,t}\,|\,D_{n,t}=d, s_{n,t}=\bar{s}}\big(\tau^{(k)}_{d,\bar{s},t}\big)\right] = \varphi(d, d', s). \tag{24}$$

Hence, $\varphi(d, d', s)$ is identified.

We provide a proof sketch of Proposition B.6 in Section 3.4 for the simplified wage equation (11), and the formal proof of Proposition B.6 in Appendix E.

One minor difference between Proposition B.6 and the proof sketch in Section 3.4 is worth highlighting. In Assumptions 6–9, we focus on the *left* extremal tails of the wage distributions, whereas the proof sketch in Section 3.4 considers the right extremal tails. The argument is symmetric, but focusing on the left tails ensures that the quantity $q_t(d, d')$ remains strictly positive. By contrast, if we let $w \to \infty$ (rather than $w \to -\infty$), then $\Pr\big(D_{n,t} = d \,\big|\, s_{n,t} = s,\ w_{n,t}(d, d') > w\big)$ would tend to zero because the equilibrium pricing mechanism in our model mirrors a second-price auction. Namely, in the class of models we study, the equilibrium wage for job d equals the expected output at the second-best firm d', $y(d', s_{n,t}) + \epsilon_{n,t}(d')$, plus a compensating differential $\Psi(d, d', s_{n,t})$. Letting the wage of job d go to infinity while holding $s_{n,t}$ and the wages of all other jobs fixed effectively corresponds to letting the second-best firm's productivity shock $\epsilon_{n,t}(d')$ go to infinity *while holding fixed the productivity shocks of all other firms*, including that of the first-best firm d. In this scenario, firm d would no longer be the worker's best option, driving $q_t(d, d')$ to zero. One could avoid this by letting the productivity shocks of both the first- and second-best firms jointly go to infinity, but doing so would substantially complicate the identification argument. Focusing on left tails, combined with the fact that in our setting labour markets are typically small with two dominant firms, sidesteps this issue by preserving the equilibrium ranking between the first- and second-best firms.

B.5 Distribution of Productivity Shocks

In Section 3.4, we discuss at length the challenges in identifying the unconditional joint distribution of the productivity-shock vector $\epsilon_{n,t}$, as well as the marginal distribution of each component $\epsilon_{n,t}(d)$, and we motivate our approach. In this section, we formalize the result for our general class of models. Before doing so, we introduce some notation and an additional assumption.

Assumption 10 (Tail-Ratio Identifiable Distribution)**.** *Let $t \in \{1, \dots, T\}$. For each $d \in \mathcal{D}$, $\epsilon_{n,t}(d)$ belongs to a parametric family, indexed by the $p_d \times 1$ vector of parameters μ_d, that is tail-ratio identifiable. Namely, for any $p_d + 1$ distinct sufficiently small thresholds $e_0 < e_1 < \cdots < e_{p_d} < 0$, the map*

$$\mu_d \mapsto \left(\frac{F_{\epsilon_{n,t}(d)}(e_1; \mu_d)}{F_{\epsilon_{n,t}(d)}(e_0; \mu_d)}, \dots, \frac{F_{\epsilon_{n,t}(d)}(e_{p_d}; \mu_d)}{F_{\epsilon_{n,t}(d)}(e_0; \mu_d)}\right),$$

is injective.

Proposition B.7 (Identification of the Shock Distribution)**.** *Let $t \in \{1, \dots, T\}$. Assume:*

(i) The conditional wage distribution $\Pr\big(w_{n,t} \le w \,\big|\, D_{n,t} = d,\ s_{n,t} = s\big)$ is identified for each $w \in \mathbb{R}$, $d \in \mathcal{D}$, and $s \in \mathcal{S}_t$. See Proposition B.5 for sufficient conditions.

(ii) Assumptions 6 to 9 hold, implying that $\varphi(d, d', s)$ is identified for each $(d, d') \in \mathcal{D}^2$ and $s \in \mathcal{S}_t(d, d')$ (Proposition B.6).

(iii) Assumption 10 holds.

Then, the parameter μ_d is identified for each $d \in \mathcal{D}$. Moreover, if the shocks $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$ are mutually independent across $d \in \mathcal{D}$, then the joint distribution of $\epsilon_{n,t}$ is identified as the product of the identified marginals. Alternatively, if a copula C_μ is specified so that

$$F_{\epsilon_{n,t}}(v_1, \dots, v_{|\mathcal{D}|}) \;=\; C_\mu\big(F_{\epsilon_{n,t}(1)}(v_1; \mu_1), \dots, F_{\epsilon_{n,t}(|\mathcal{D}|)}(v_{|\mathcal{D}|}; \mu_{|\mathcal{D}|})\big) \quad \forall (v_1, \dots, v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|},$$

and the copula parameter μ is known, then the joint distribution of $\epsilon_{n,t}$ is identified from the identified marginals and C_μ. Without further restrictions on the dependence among $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$, the joint distribution of $\epsilon_{n,t}$ is partially identified by the sharp Fréchet–Höffding bounds in that for all $(v_1, \dots, v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|}$,

$$\max\Big\{\sum_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d; \mu_d) - (|\mathcal{D}| - 1),\ 0\Big\} \;\le\; F_{\epsilon_{n,t}}(v_1, \dots, v_{|\mathcal{D}|}) \;\le\; \min_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d; \mu_d).$$

B.6 Expected Output, Compensating Differential, and Human Capital

In this section, we first show how to identify the expected output (net of productivity shocks) $y(\cdot)$, as defined in equation (4), using standard arguments for dynamic discrete choice models. Given $y(\cdot)$, the compensating differential $\Psi(\cdot)$, as defined in Proposition 1, can be obtained by solving for the model's equilibrium, and the human capital functions $\{h_d(\cdot)\}_{d\in\mathcal{D}}$, as defined in equation (1), can then be obtained directly from $y(\cdot)$.

We discuss the identification of $y(\cdot)$ in the simplest case in which the model's equilibrium is efficient; see Appendix F.2 of the Online Supplementary Material for the general inefficient case. When the equilibrium is efficient, job choices maximize the expected present discounted value of output, not just the surplus generated by the match between a worker and a firm. Hence, a firm's job offer and a worker's job choice solves a planning problem. In other words, the market-wide equilibrium allocation of workers to firms reduces to a single-agent dynamic decision problem of maximizing the expected present discounted value of market-wide output. In this sense, standard identification arguments for dynamic discrete choice models can be applied to identify $y(\cdot)$ from observed job choices.[50]

To elaborate, let $S(s_{n,t}, \epsilon_{n,t})$ denote the expected present discounted value of the output produced by worker n or, equivalently, the expected present discounted *social welfare* at state $(s_{n,t}, \epsilon_{n,t})$. Then,

[50] As known, in the finite-horizon case, identification is up to the match surplus values in the final period.

$S(s_{n,t}, \epsilon_{n,t})$ is the value of the following planning problem

$$S(s_{n,t}, \epsilon_{n,t}) = \max_{d \in \mathcal{D}_t(s_{n,t})} \Big\{ y(d, s_{n,t}) + \epsilon_{n,t}(d) \\ + \delta \left[1 - \varsigma(H_{n,1}, I_n^{t-1}, d)\right] \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[S(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, d \Big] dF_{\epsilon_{n,t+1}} \Big\}.$$

By Propositions B.3 and B.4, the law of motion of $s_{n,t}$, as well as the CCPs, are identified. The joint distribution $F_{\epsilon_{n,t}}$ of the shock vector $\epsilon_{n,t}$ is identified by Proposition B.7. The exogenous separation rate $\varsigma(H_{n,1}, I_n^{t-1}, d)$ is nonparametrically identified by the fraction of employed workers at firm d with given $(H_{n,1}, I_n^{t-1})$ who exit the labor market considered at the end of the period. Therefore, $y\big(d, s_{n,t}\big)$ is identified following Magnac and Thesmar (2002) under standard normalizations in dynamic discrete choice models. We now state the formal normalization conditions and the identification result.

Assumption 11 (Normalization)**.** *Let $t \in \{1, \ldots, T\}$, $d \in \mathcal{D}$, and $s \in \mathcal{S}_t$. There exists a known $\tilde{d} \in \mathcal{D}$ such that $y(\tilde{d}, s) = 0$.*

Assumption 11 normalizes $y(\cdot)$ to zero at one firm for each state. As mentioned, this normalization is common and needed in dynamic discrete choice models, which are otherwise not identified; relaxing this assumption leads to a partially identified framework, as discussed, for example, by Kalouptsidi et al. (2021) and Kalouptsidi et al. (2024). Relative to the normalizations of certain values of the sum of expected output and the compensating differential imposed in Assumption 9, which we maintain for the identification of the deterministic wage component, Assumption 11 simply additionally normalizes certain values of expected output.

Proposition B.8 (Expected Output)**.** *Let $t \in \{1, \ldots, T\}$. Suppose that:*

(i) The law of motion of the state, $\Pr(s_{n,t} | D_{n,t-1}, s_{n,t-1})$*, and the CCPs,* $\Pr(D_{n,t} | s_{n,t})$*, are identified. See Propositions B.3 and B.4 for sufficient conditions.*

(ii) The joint distribution $F_{\epsilon_{n,t}}$ of the shock vector $\epsilon_{n,t}$ is identified. See PropositionB.7 for sufficient conditions.

(iii) The discount factor δ is known.

(iv) The separation rates $\{\varsigma(H_{n,1}, I_n^{t-1}, d)\}_{d \in \mathcal{D}}$ are identified (immediate consequence of Assumption 2).

(v) Assumption 11 holds.

Then, the expected output $y(d, s)$ is identified for each $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$.

With $y(\cdot)$ known, the compensating differential $\Psi(\cdot)$ can be obtained by solving for the model's equilibrium and computing the workers' continuation match surplus values.

Proposition B.9 (Compensating Differential)**.** *Let $t \in \{1, \ldots, T\}$. Assume that the expected output $y(d, s)$ is identified for each $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$. See Proposition B.8 for sufficient conditions. Then, the compensating differential $\Psi(d, d', s)$ is identified for each $(d, d') \in \mathcal{D}^2$ and $s \in \mathcal{S}_t(d, d')$.*

Analogously, the human capital functions $\{h_d(\cdot)\}_{d\in\mathcal{D}}$ can be obtained from $y(\cdot)$.

Proposition B.10. *Let $t \in \{1,\ldots,T-1\}$. Assume that:*

(i) $\alpha(h,d,e)$, $\beta(h,d,e)$, and the prior belief $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$ are identified for each $(h,d,e) \in \mathcal{H}\times\mathcal{D}\times\mathcal{E}$. See Proposition B.2 for sufficient conditions.

(ii) The expected output $y(d,s)$ is identified for each $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$. See Proposition B.8 for sufficient conditions.

(iii) $\mathcal{N} := \{\bar{\eta}, \underline{\eta}\}$. Let $Z_{n,t} := (H_{n,1}, I_n^{t-1}, e_n)$ so that $s_{n,t} = (Z_{n,t}, P_{n,t})$. Let $\mathcal{S}_t^{Z_{n,t}}$ denote the projection of $\mathcal{S}_t$ on $Z_{n,t}$. For each $z \in \mathcal{S}_t^{Z_{n,t}}$, there exists (z,p_1) and (z,p_2) in $\mathcal{S}_{n,t}$ with $p_1 \neq p_2$.

Then, the human capital function $h_d(z,\eta)$ is identified for each $d \in \mathcal{D}$, $z \in \mathcal{S}_t^{Z_{n,t}}$, and $\eta \in \mathcal{N}$.

The proofs of Propositions B.8 and B.9 are omitted because they follow directly from a straightforward application of Proposition 2 and Corollary 3 in Magnac and Thesmar (2002). The proof of Proposition B.10 is provided in Appendix E.

C Bias in AKM Estimators of Worker and Firm Fixed Effects

In this section, we provide analytical expressions for the bias in the AKM estimators of worker and firm effects and, in turn, in the AKM estimators of sorting that arises when the compensating differential is nonzero. Hereafter, we restrict attention to the largest connected set of the worker–firm mobility graph and impose the usual AKM-type normalisations; we do not formalise these conditions, and treat them as implicit assumptions throughout. Moreover, we abstract from bias due to limited worker mobility across firms, which is well understood in the literature.

Before presenting the formal results, we introduce some useful notation. Let $\mathcal{I} := \{(n,t) :$ worker n is observed in period $t\}$ be the set of observed worker–period pairs. Let $\mathcal{N}$ be the set of workers. For each worker $n \in \mathcal{N}$, let $T_n := \big|\{t : (n,t) \in \mathcal{I}\}\big|$ be the number of observed periods for worker n. For each firm $d \in \mathcal{D}$, let $N_d := \big|\{(n,t) \in \mathcal{I} : D_{n,t} = d\}\big|$ be the number of worker–period observations at firm d.

Assumption 12 (True versus Estimated Wage Equation—Compensating Differential is Nonzero)**.** *The true wage equation is*

$$w_{n,t} = e_n + \varphi(D_{n,t}) + \Psi\big(e_n, D_{n,t}, X_{n,t}; \gamma\big) + u_{n,t}, \tag{25}$$

with $\mathbb{E}[u_{n,t} \mid \boldsymbol{Z}] = 0$, where $\boldsymbol{Z} := \big\{(e_n)_{n\in\mathcal{N}},\ (D_{n,t})_{(n,t)\in\mathcal{I}},\ (X_{n,t})_{(n,t)\in\mathcal{I}}\big\}$. The researcher estimates by OLS the misspecified (AKM-type) equation

$$w_{n,t} = \alpha_n + \psi(D_{n,t}) + g(X_{n,t}; \beta) + \varepsilon_{n,t}. \tag{26}$$

Equation (25) is a simplified parametric version of our wage equation (8). Relative to our general class of models: *(i)* dependence on the second-best firm $D'_{n,t}$ is eliminated; *(ii)* the state vector

$s_{n,t} := (H_{n,1}, I_n^{t-1}, P_{n,t}, e_n)$ is absorbed into $(X_{n,t}, e_n)$, where $X_{n,t}$ is observed in the data and collects $H_{n,1}$, I_n^{t-1}, and additional covariates that may measure beliefs $(P_{n,t}, 1 - P_{n,t})$ about worker ability $\theta_n \in \{\overline{\theta}, \underline{\theta}\}$;[51] *(iii)* the expected output function $y(\cdot)$ is log-linear in worker and firm effects and independent of $X_{n,t}$, and is therefore fully characterised by the sum $e_n + \varphi(D_{n,t})$, where e_n denotes the worker effect and $\varphi(D_{n,t})$ the firm effect; *(iv)* the compensating differential is represented by the parametric function $\Psi(e_n, D_{n,t}, X_{n,t}; \gamma)$, with parameters γ, which we allow to be nonlinear and nonseparable in e_n and $D_{n,t}$, for the reasons discussed in Section 2.2; and *(v)* the productivity shock is not firm-specific, but only worker-specific, denoted by $u_{n,t}$ and assumed conditionally mean-independent of $D_{n,t}$.

Equation (26) is a misspecified version of equation (25) because, while it correctly represents the log-linear output technology in worker and firm effects through the term $\alpha_n + \psi(D_{n,t})$, it replaces $\Psi(e_n, D_{n,t}, X_{n,t}; \gamma)$—which, as noted above, can depend on worker and firm effects in a nonlinear and nonseparable way—with $g(X_{n,t}; \beta)$ often taken to be linear in $X_{n,t}$, which is independent of e_n and $D_{n,t}$. This misspecification induces bias in the estimators of worker and firm effects and, in turn, in $\widehat{S}_{\mathrm{AKM}}$ relative to S_{AKM}, which we now characterize analytically. Define

$$m_{n,t} \;:=\; \Psi\big(e_n, D_{n,t}, X_{n,t}; \gamma\big) - g(X_{n,t}; \beta), \tag{27}$$

which represents the *omitted component*, that is, the part of the compensating differential $\Psi\big(e_n, D_{n,t}, X_{n,t}; \gamma\big)$ that is *not* captured by the function $g(X_{n,t}; \beta)$ and is therefore absorbed into the regression error $\varepsilon_{n,t}$ in (26). Using (27), the true wage equation (25) can be rewritten as

$$w_{n,t} = e_n + \varphi(D_{n,t}) + g(X_{n,t}; \beta) + m_{n,t} + u_{n,t}. \tag{28}$$

Proposition C.1 (Bias in the AKM Worker and Firm Effects)**.** *Under Assumption 12, for each worker $n \in \mathcal{N}$ and firm $d \in \mathcal{D}$, the conditional biases of the OLS estimators of worker and firm effects obtained from* (26) *satisfy*

$$\mathbb{E}[\hat{\alpha}_n \mid \boldsymbol{Z}] - e_n = \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} \Big[m_{n,t} - \big(\mathbb{E}[\hat{\psi}(D_{n,t}) \mid \boldsymbol{Z}] - \varphi(D_{n,t})\big) - \big(\mathbb{E}[g(X_{n,t}; \hat{\beta}) \mid \boldsymbol{Z}] - g(X_{n,t}; \beta)\big)\Big], \tag{29}$$

$$\mathbb{E}[\hat{\psi}(d) \mid \boldsymbol{Z}] - \varphi(d) = \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} \Big[m_{n,t} - \big(\mathbb{E}[\hat{\alpha}_n \mid \boldsymbol{Z}] - e_n\big) - \big(\mathbb{E}[g(X_{n,t}; \hat{\beta}) \mid \boldsymbol{Z}] - g(X_{n,t}; \beta)\big)\Big]. \tag{30}$$

Equation 29 shows that the bias of the OLS estimator of worker effects obtained from (26) is given by the average, along worker n's observed job history, of three components: *(i)* the omitted term $m_{n,t}$, which captures the part of the compensating differential $\Psi\big(e_n, D_{n,t}, X_{n,t}; \gamma\big)$ not absorbed by $g(X_{n,t}; \beta)$; *(ii)* the bias in the estimators of firm effects at the firms where worker n is employed,

[51] For simplicity of exposition, we assume—for the purposes of this section only—that the researcher can observe, or accurately proxy for, $P_{n,t}$. Under this assumption, $P_{n,t}$ can be interpreted as a standard covariate. See also footnote 18.

$\mathbb{E}[\hat{\psi}(D_{n,t}) \mid \boldsymbol{Z}] - \varphi(D_{n,t})$; and *(iii)* the bias in the estimator of the misspecified function $g(X_{n,t}; \hat{\beta})$, $\mathbb{E}[g(X_{n,t}; \hat{\beta}) \mid \boldsymbol{Z}] - g(X_{n,t}; \beta)$. To isolate the role of the omitted term $m_{n,t}$, hold fixed the contributions of *(ii)* and *(iii)*. The worker-effect estimator bias in (29) then increases in magnitude with the worker-specific average omitted portion of the compensating differential, $\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t}$. In particular, $\left|\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t}\right|$—and hence the worker-effect estimator bias—will be large if the worker spends most periods at firms where the compensating differential $\Psi(e_n, d, X_{n,t}; \gamma)$ is systematically above $g(X_{n,t}; \beta)$ (so $g(X_{n,t}; \beta)$ persistently understates compensation) or systematically below $g(X_{n,t}; \beta)$ (so $g(X_{n,t}; \beta)$ persistently overstates compensation). Conversely, $\left|\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t}\right|$ will be small when the worker's career mixes firms where the compensating differential $\Psi(e_n, d, X_{n,t}; \gamma)$ lies above $g(X_{n,t}; \beta)$ with firms where it lies below, so that $g(X_{n,t}; \beta)$ is approximately correct on average along the worker's path.

This mechanism is especially transparent in markets where compensating differentials are highly heterogeneous across *firms*—that is, where firms differ sharply in the human-capital development and information opportunities they offer. In such markets, for a given worker type e_n and covariates $X_{n,t}$, $\Psi(e_n, d, X_{n,t}; \gamma)$ varies substantially with d, so there are firms for which the discrepancy between $\Psi(e_n, d, X_{n,t}; \gamma)$ and the fixed benchmark $g(X_{n,t}; \beta)$ is large. When workers are concentrated in such firms—as is likely early in their careers, when they most value acquiring new information and human capital—and the discrepancies are predominantly of the same sign, $\left|\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t}\right|$ will be large, leading to a large bias in the estimated worker effects.

Components *(ii)* and *(iii)* can attenuate or exacerbate this pattern. Say, holding fixed the contributions of *(i)* and *(iii)*, the worker-effect estimator bias decreases as the average firm-effect bias $\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} (\mathbb{E}[\hat{\psi}(D_{n,t}) \mid \boldsymbol{Z}] - \varphi(D_{n,t}))$ increases, as it enters (29) with a minus sign. Specifically, if $\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t}$ and $\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} (\mathbb{E}[\hat{\psi}(D_{n,t}) \mid \boldsymbol{Z}] - \varphi(D_{n,t}))$ have the same sign, component *(ii)* offsets part of the bias generated by *(i)*; if they have opposite signs, component *(ii)* reinforces it.

Equation 30 provides the firm-side analogue of (29). It shows that the bias in the estimator of firm effect at d is an average, over the workers and periods in which firm d is observed, of three terms: the omitted component $m_{n,t}$; the worker-effect estimator bias, $\mathbb{E}[\hat{\alpha}_n \mid \boldsymbol{Z}] - e_n$; and the bias in the estimator of the misspecified function $g(X_{n,t}; \hat{\beta})$, $\mathbb{E}[g(X_{n,t}; \hat{\beta}) \mid \boldsymbol{Z}] - g(X_{n,t}; \beta)$.

As above, to isolate the role of the omitted term $m_{n,t}$, consider markets in which compensating differentials are highly heterogeneous across *workers*—that is, where workers differ sharply in their processes of human-capital accumulation and information acquisition. In such markets, for a given firm d and covariates $X_{n,t}$, $\Psi(e_n, d, X_{n,t}; \gamma)$ can vary substantially across worker types e_n, so there exist workers for whom the discrepancy between $\Psi(e_n, d, X_{n,t}; \gamma)$ and the fixed benchmark $g(X_{n,t}; \beta)$ is large. As a result, firms that disproportionately employ workers for which $\Psi(e_n, d, X_{n,t}; \gamma)$ lies systematically above $g(X_{n,t}; \beta)$ (or systematically below it) will have a large firm-specific average omitted component $\frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} m_{n,t}$ and, all else equal, a large bias in $\hat{\psi}(d)$. The sign and magnitude of the firm-effect estimator bias ultimately depend on how this firm-specific average omitted component interacts with the composition of worker-effect estimator biases among the workers employed at d and with any bias in $g(X_{n,t}; \hat{\beta})$.

We now show how the bias characterised in Proposition C.1 propagates to the standard AKM estimator of sorting, defined as

$$\widehat{S}_{\mathrm{AKM}} \;:=\; \frac{\mathrm{Cov}\big(\hat{\alpha}_n, \hat{\psi}(D_{n,t})\big)}{\mathrm{Var}(w_{n,t})},$$

where $\mathrm{Cov}(\cdot,\cdot)$ and $\mathrm{Var}(\cdot)$ denote the sample covariance and variance computed over the worker–period observations $(n,t) \in \mathcal{I}$. Let the "true" AKM sorting measure be

$$S_{\mathrm{AKM}} := \frac{\mathrm{Cov}(e_n, \varphi(D_{n,t}))}{\mathrm{Var}(w_{n,t})}.$$

For each worker $n \in \mathcal{N}$ and firm $d \in \mathcal{D}$, define the fixed-effect estimation errors

$$\Delta\alpha_n \;:=\; \hat{\alpha}_n - e_n \qquad \text{and} \qquad \Delta\psi(d) \;:=\; \hat{\psi}(d) - \varphi(d).$$

Proposition C.2 (Bias in the AKM Sorting Measure)**.** *Under Assumption 12, the bias in $\widehat{S}_{\mathrm{AKM}}$ relative to S_{AKM} is*

$$\mathbb{E}\big[\widehat{S}_{\mathrm{AKM}} \mid \boldsymbol{Z}\big] - S_{\mathrm{AKM}} = \mathbb{E}\left[\frac{Cov\big(e_n, \Delta\psi(D_{n,t})\big) + Cov\big(\Delta\alpha_n, \varphi(D_{n,t})\big) + Cov\big(\Delta\alpha_n, \Delta\psi(D_{n,t})\big)}{Var(w_{n,t})} \;\middle|\; \boldsymbol{Z}\right]. \tag{31}$$

Proposition C.2 shows that the gap between $\widehat{S}_{\mathrm{AKM}}$ and S_{AKM} is driven by three covariance terms: *(i)* $\mathrm{Cov}\big(e_n, \Delta\psi(D_{n,t})\big)$; *(ii)* $\mathrm{Cov}\big(\Delta\alpha_n, \varphi(D_{n,t})\big)$; and *(iii)* $\mathrm{Cov}\big(\Delta\alpha_n, \Delta\psi(D_{n,t})\big)$. In particular, holding fixed the contributions of *(i)* and *(ii)*, the bias in $\widehat{S}_{\mathrm{AKM}}$ is increasing in $\mathrm{Cov}\big(\Delta\alpha_n, \Delta\psi(D_{n,t})\big)$.

To build intuition for term *(iii)*, consider a market in which some workers have upward-biased worker-effect estimation error (i.e., $\hat{\alpha}_n - e_n > 0$) and others have downward-biased ones. Similarly, some firms have upward-biased firm-effect estimation error (i.e., $\hat{\psi}(d) - \varphi(d) > 0$) and others have downward-biased ones. Such dispersion in worker- and firm-effect estimation errors can arise, for instance, when compensating differentials are highly heterogeneous across firms and workers, so that the worker-specific average omitted component $\frac{1}{T_n}\sum_{t:(n,t)\in\mathcal{I}} m_{n,t}$ and the firm-specific average omitted component $\frac{1}{N_d}\sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} m_{n,t}$ are sizable for some workers and some firms (as discussed above). Whether term *(iii)* increases or decreases $\widehat{S}_{\mathrm{AKM}}$ depends on the extent to which worker- and firm-effect estimation errors co-move along observed matches. If workers with upward-biased worker-effect estimation error tend to be employed at firms with upward-biased firm-effect estimation error, and likewise for downward biases, then $\mathrm{Cov}\big(\Delta\alpha_n, \Delta\psi(D_{n,t})\big)$ is positive and component *(iii)* increases $\widehat{S}_{\mathrm{AKM}}$ relative to S_{AKM}. If, instead, upward-biased workers tend to match with downward-biased firms (and vice versa), then $\mathrm{Cov}\big(\Delta\alpha_n, \Delta\psi(D_{n,t})\big)$ is negative and component *(iii)* decreases $\widehat{S}_{\mathrm{AKM}}$ relative to S_{AKM}. Finally, if there is little systematic co-movement between the two estimation errors, then $\mathrm{Cov}\big(\Delta\alpha_n, \Delta\psi(D_{n,t})\big)$ is close to zero and component *(iii)* contributes little to the gap between $\widehat{S}_{\mathrm{AKM}}$ and S_{AKM}. The proof of Proposition C.2 is omitted, as it amounts to a straightforward variance–covariance decomposition. We now provide the proof of Proposition C.1.

Proof of Proposition C.1: Let the fitted residual from (26) be

$$\hat{\varepsilon}_{n,t} \coloneqq w_{n,t} - \hat{\alpha}_n - \hat{\psi}(D_{n,t}) - g(X_{n,t}; \hat{\beta}).$$

We derive the worker and firm bias expressions in turn.

Worker Effects. The OLS normal equation for α_n is

$$\sum_{t:(n,t)\in\mathcal{I}} \hat{\varepsilon}_{n,t} = 0. \tag{32}$$

Dividing (32) by T_n gives the identity

$$\hat{\alpha}_n = \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} w_{n,t} - \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} \hat{\psi}(D_{n,t}) - \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} g(X_{n,t}; \hat{\beta}). \tag{33}$$

Taking worker-n averages of (28) yields

$$\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} w_{n,t} = e_n + \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} \varphi(D_{n,t}) + \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} g(X_{n,t}; \beta) + \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t} + \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} u_{n,t}. \tag{34}$$

Substituting (34) into (33) and rearranging gives

$$\begin{aligned}\hat{\alpha}_n - e_n = \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} m_{n,t} + \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} u_{n,t} + \Big(\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} \varphi(D_{n,t}) - \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} \hat{\psi}(D_{n,t})\Big) \\ + \Big(\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} g(X_{n,t}; \beta) - \frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} g(X_{n,t}; \hat{\beta})\Big).\end{aligned} \tag{35}$$

Taking the conditional expectations given $\boldsymbol{Z}$ of (35) and using the conditional exogeneity of the error term, which implies $\mathbb{E}[\frac{1}{T_n} \sum_{t:(n,t)\in\mathcal{I}} u_{n,t} \mid \boldsymbol{Z}] = 0$, yields (29).

Firm Effects. The OLS normal equation for $\psi(d)$ is

$$\sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} \hat{\varepsilon}_{n,t} = 0. \tag{36}$$

Dividing (36) by N_d gives the identity

$$\hat{\psi}(d) = \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} w_{n,t} - \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} \hat{\alpha}_n - \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} g(X_{n,t}; \hat{\beta}). \tag{37}$$

Taking firm-d averages of (28) yields

$$\begin{aligned}\frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} w_{n,t} = \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} e_n + \varphi(d) + \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} g(X_{n,t}; \beta) \\ + \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} m_{n,t} + \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} u_{n,t}.\end{aligned} \tag{38}$$

Substituting (38) into (37) and rearranging gives

$$\hat{\psi}(d) - \varphi(d) = \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} m_{n,t} + \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} u_{n,t} + \Big(\frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} e_n - \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} \hat{\alpha}_n\Big)$$
$$+ \Big(\frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} g(X_{n,t};\beta) - \frac{1}{N_d} \sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} g(X_{n,t};\hat{\beta})\Big). \tag{39}$$

Taking the conditional expectations given $\boldsymbol{Z}$ of (39) and using the conditional exogeneity of the error term, which implies $\mathbb{E}[\frac{1}{N_d}\sum_{(n,t)\in\mathcal{I}:\, D_{n,t}=d} u_{n,t} \mid \boldsymbol{Z}] = 0$, yields (30). □

D Supports of Worker Efficiency, Signals, and Worker Ability

We discuss here the supports of worker efficiency, performance signals, and worker ability.

Support $\mathcal{E}$ of Efficiency e_n. To identify the wage mixture (23), Assumption 4(ii) imposes that $\mathcal{E}$ is finite. If e_n is continuous (and potentially multidimensional), then the wage mixture (23) becomes a *continuous* mixture model, making identification more challenging. In particular, the identification result of Aragam et al. (2020) does not extend to this class of mixtures. This impasse can be easily resolved by assuming away selection into $D_{n,t}$ based on $\epsilon_{n,t}$; that is, assuming that $D_{n,t}$ is independent of $\epsilon_{n,t}$ In that case, Assumption 4(i) 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 (23) is a continuous mixture of Normals, whose identification is established by Bruni and Koch (1985)'s Theorem 1. See Bunting et al. (2024) for an application of Bruni and Koch (1985)'s Theorem 1 to interactive fixed-effect panel-data models in the absence of selection on unobservables.

Support $\mathcal{N}$ of Signal $\eta_{n,t}$. As for $\mathcal{E}$, Proposition B.1 can also be easily adapted to cases where the performance signal $\eta_{n,t}$ about ability is continuous (and potentially multidimensional) in the absence of selection into $D_{n,t}$ based on $\epsilon_{n,t}$. To identify the learning process in Proposition B.2, Assumption 5(i) imposes that $\mathcal{N}$ has a cardinality of two. As explained in Section B.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 and multidimensional $\eta_{n,t}$, provided that the signal distribution remains an identifiable mixture. For instance, if $\eta_{n,t}$ 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 B.2, Assumption 5(i) requires that Θ has cardinality two. As explained in Section B.2, 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 $\mathcal{N}$ 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, 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 the proof of Proposition B.2 in Appendix E for more details). Keeping $\mathcal{N}$ of cardinality two, Assumption 6(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 $\eta_{n,t}$ 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.

E Omitted Proofs

Proof of Proposition 1 The proof proceeds in three steps. First, observe that firms' individual rationality implies that the worker must be indifferent between working at the first-best firm, $D_{n,t} := D(s_{n,t}, \epsilon_{n,t})$ and working at the second-best firm, $D'_{n,t} := D'(s_{n,t}, \epsilon_{n,t})$. The logic of this intermediate result is by contradiction. Since the worker must weakly prefer working at firm $D_{n,t}$ to working at firm $D'_{n,t}$ for firm $D_{n,t}$ to be the employing firm in equilibrium, if the worker strictly preferred firm $D_{n,t}$ to firm $D'_{n,t}$, then firm $D_{n,t}$ would be able to marginally lower its wage offer and still attract the worker. Hence, firm $D_{n,t}$ would not be maximizing profits, which is a contradiction.

Formally, first, the worker's indifference between firms $D_{n,t}$ and $D'_{n,t}$ can be expressed as

$$w_{n,t,D_{n,t}} + \delta[1-\varsigma(I_n^{t-1}, D_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, D_{n,t}\right] dF_{\epsilon_{n,t+1}}$$

$$= w_{n,t,D'_{n,t}} + \delta[1-\varsigma(I_n^{t-1}, D'_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, D'_{n,t}\right] dF_{\epsilon_{n,t+1}},$$

where $w_{n,t,D_{n,t}}$ and $w_{n,t,D'_{n,t}}$ are the equilibrium wage offers of firms $D_{n,t}$ and $D'_{n,t}$, respectively. By straightforward algebra,

$$\begin{aligned} w_{n,t,D_{n,t}} =& w_{n,t,D'_{n,t}} + \delta[1-\varsigma(I_n^{t-1}, D'_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, D'_{n,t}\right] dF_{\epsilon_{n,t+1}} \\ & - \delta[1-\varsigma(I_n^{t-1}, D_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, D_{n,t}\right] dF_{\epsilon_{n,t+1}}. \end{aligned} \tag{40}$$

Second, recall that by condition (*iv*) of equilibrium, each non-employing firm must be indifferent between employing and not-employing the worker. Hence, by (7), the equilibrium wage offer $w_{n,t,D'_{n,t}}$ of the second-best firm $D'_{n,t}$ must satisfy

$$\begin{aligned} w_{n,t,D'_{n,t}} =& y(D'_{n,t}, s_{n,t}, \epsilon_{n,t}(D'_{n,t})) + \delta[1-\varsigma(I_n^{t-1}, D'_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_{D'_{n,t}}(\cdot)|s_{n,t}, D'_{n,t}] dF_{\epsilon_{n,t+1}} \\ & - \delta[1-\varsigma(I_n^{t-1}, D_{n,t})] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_{d'}(\cdot)|s_{n,t}, D_{n,t}] dF_{\epsilon_{n,t+1}}. \end{aligned} \tag{41}$$

Third, by substituting for $w_{n,t,D'_{n,t}}$ in (40) using (41), we obtain

$$\begin{aligned}w_{n,t,D_{n,t}} =& y(D'_{n,t}, s_{n,t}, \epsilon_{n,t}(D'_{n,t}))+\delta[1-\varsigma(I_n^{t-1}, D'_{n,t})]\int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_{D'_{n,t}}(\cdot)|s_{n,t}, D'_{n,t}]dF_{\epsilon_{n,t+1}}\\ &- \delta[1-\varsigma(I_n^{t-1}, D_{n,t})]\int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_{D'_{n,t}}(\cdot)|s_{n,t}, D_{n,t}]dF_{\epsilon_{n,t+1}}\\ &+ \delta[1-\varsigma(I_n^{t-1}, D'_{n,t})]\int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(\cdot)|s_{n,t}, D'_{n,t}\right]dF_{\epsilon_{n,t+1}}\\ &- \delta[1-\varsigma(I_n^{t-1}, D_{n,t})]\int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(\cdot)|s_{n,t}, D_{n,t}\right]dF_{\epsilon_{n,t+1}}.\end{aligned}$$

Rearranging the terms of this expression then gives

$$\begin{aligned}w_{n,t,D_{n,t}} =& y(D'_{n,t}, s_{n,t}, \epsilon_{n,t}(D'_{n,t}))\\ &+\delta[1-\varsigma(I_n^{t-1}, D'_{n,t})]\int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(s_{n,t+1}, \epsilon_{n,t+1})+\Pi_{D'_{n,t}}(\cdot)|s_{n,t}, D'_{n,t}\right]dF_{\epsilon_{n,t+1}}\\ &- \delta[1-\varsigma(I_n^{t-1}, D_{n,t})]\int_{\epsilon_{n,t+1}} \mathbb{E}\left[W(s_{n,t+1}, \epsilon_{n,t+1})+\Pi_{D'_{n,t}}(\cdot)|s_{n,t}, D_{n,t}\right]dF_{\epsilon_{n,t+1}}.\end{aligned}$$

Using this expression, the definition of match surplus value $V_{D'_{n,t}}(s_{n,t}, \epsilon_{n,t})$, and of the compensating differential $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$, (8) immediately follows, as desired. □

Proof of Proposition 2 The result is an immediate implication of the fact that if the human capital process is sufficiently similar across firms, then the state $s_{n,t}$ evolves differently when worker n is employed by the first-best ($D_{n,t} := D(s_{n,t}, \epsilon_{n,t})$) and the second-best ($D'_{n,t} := D'(s_{n,t}, \epsilon_{n,t})$) firm only because beliefs about ability are updated differently based on the worker's realized output at the two firms. Now, as apparent from (3), when $P_{n,t}$ is arbitrarily close to zero or one, the updated value $P_{n,t+1}$ of the belief $P_{n,t}$ that worker n is of high ability is arbitrarily close to $P_{n,t}$ regardless of whether the worker is employed by firm $D_{n,t}$ or $D'_{n,t}$. Hence, the distribution of the state in $t+1$ conditional on its value $s_{n,t}$ in t does not depend on whether the worker is employed by the first-best or (off path) by the second-best firm in t. Then, we have that

$$\int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, D'_{n,t}]dF_{\epsilon_{n,t+1}} - \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, D_{n,t}]dF_{\epsilon_{n,t+1}} = 0.$$

When the risk of exogenous separation is sufficiently similar across firms, it therefore follows that $\Psi(D_{n,t}, D'_{n,t}, s_{n,t}) = 0$ for any possible $D_{n,t}$ and $D'_{n,t}$. That for given $D_{n,t}$ and $D'_{n,t}$, $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$ is strictly positive for some value of $P_{n,t}$ is an immediate implication of Proposition 3. Hence, $\Psi(D_{n,t}, D'_{n,t}, s_{n,t})$ must be non-monotone in $P_{n,t}$, which yields the desired claim. □

Proof of Proposition 3 We begin by abstracting from the risk of exogenous separation and the possibility for workers to accumulate human capital when employed. As some of the details of the logic are tedious yet straightforward, we keep the exposition purposely informal whenever possible.

As shown in the proof of Proposition 1 in Pastorino (2024), the match surplus value for the first-best firm $D_{n,t} := D(s_{n,t}, \epsilon_{n,t})$ reduces to the value of output when $|\mathcal{D}| = 2$. Namely, $V_{D_{n,t}}(s_{n,t}, \epsilon_{n,t})$ can be shown to solve

$$V_{D_{n,t}}(s_{n,t}, \epsilon_{n,t}) = y(D_{n,t}, s_{n,t}, \epsilon_{n,t}(D_{n,t})) + \delta \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D_{n,t}}(\cdot)|s_{n,t}, D_{n,t}]dF_{\epsilon_{n,t+1}}. \tag{42}$$

In a slight abuse of notation, let now $V_{D_{n,t}}$ be any function convex in the prior $P_{n,t}$ in t that worker n is of high ability. Define the operator T as $TV_{D_{n,t}}(s_{n,t}, \epsilon_{n,t}) = y(D_{n,t}, s_{n,t}, \epsilon_{n,t}(D_{n,t})) + \delta \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D_{n,t}}(\cdot)|$-$s_{n,t}, d]dF_{\epsilon_{n,t+1}}$, with $V_{D_{n,t}}(\cdot) = V_{D_{n,t}}(s_{n,t+1}, \epsilon_{n,t+1})$. Note that expected output $y(D_{n,t}, s_{n,t}, \epsilon_{n,t}(D_{n,t}))$ is an affine function of $P_{n,t}$. Hence, expected output is a convex function of $P_{n,t}$.

As for the third term on the right side of (42), one can show that $\mathbb{E}[V_{D_{n,t}}(s_{n,t+1}, \epsilon_{n,t+1})|s_{n,t}, d]$ is an affine, hence convex, function of the prior $P_{n,t}$ by using the law of iterated expectations, the fact that the distribution of signals under any prior is a mixture of the distributions of signals conditional on the worker being of high and of low ability, with weights given by the corresponding prior probabilities $P_{n,t}$ and $1 - P_{n,t}$, and that the average value of $P_{n,t+1}$ conditional on $P_{n,t}$ is $P_{n,t}$.[52] Now, given that $\int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D_{n,t}}(\cdot)|s_{n,t}, D_{n,t}]dF_{\epsilon_{n,t+1}}$ is a convex combination—a probability-weighted average—of convex functions of $P_{n,t}$, it is also a convex function. Thus, $TV_{D_{n,t}}$ is a convex function of $P_{n,t}$, which implies that the operator T maps convex functions into convex functions. As T is a contraction mapping, it admits a unique fixed point. Given that the space of convex functions is closed, the fixed point of T must be a convex function. Hence, $V_{D_{n,t}}(s_{n,t}, \epsilon_{n,t})$ is a convex function of $P_{n,t}$. Since $V_{D_{n,t}}(s_{n,t}, \epsilon_{n,t})$ in the multi-job case is the maximum of job-specific match surplus values that are convex functions of $P_{n,t}$, an analogous argument applies.

The convexity of $V_{D_{n,t}}(s_{n,t}, \epsilon_{n,t})$ with respect to $P_{n,t}$ implies that mean-preserving spreads of posterior beliefs increase match surplus value. Recall now that if the job of firm $D_{n,t}$ is *more* informative than that of the second-best firm $D'_{n,t} := D'(s_{n,t}, \epsilon_{n,t})$ in the Blackwell sense, then the distribution of posterior beliefs reached at the end of t conditional on the worker's employment at firm $D_{n,t}$ in t is a mean-preserving spread of the distribution of posterior beliefs reached at the end of t conditional on the worker's employment at firm $D'_{n,t}$ in t. It is then immediate that $\int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(\cdot)|s_{n,t}, D'_{n,t}]dF_{\epsilon_{n,t+1}} \leq \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(\cdot)|s_{n,t}, D_{n,t}]dF_{\epsilon_{n,t+1}}$ by the convexity of $V_{D'_{n,t}}$ and the property that the law of motion of the state differs across firms only in terms of the evolution of $P_{n,t}$. By contrast, it follows that $\int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(\cdot)|s_{n,t}, D'_{n,t}]dF_{\epsilon_{n,t+1}} \geq \int_{\epsilon_{n,t+1}} \mathbb{E}[V_{D'_{n,t}}(\cdot)|s_{n,t}, D_{n,t}]$-$dF_{\epsilon_{n,t+1}}$ if the job of firm $D_{n,t}$ is *less* informative than that of firm $D'_{n,t}$ in the Blackwell sense.

When the risk of exogenous separation and the human capital process are sufficiently similar across firms, an analogous argument applies, which concludes the proof of the proposition. □

Proof of Corollary B.1(a). We show how to identify the conditional signal distribution

$$\Pr\left(\eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^t = d^t, e_n = e\right), \tag{43}$$

[52] This result extends Lemma 3.1 in Banks and Sundaram (1992) to the case in which the relevant state of nature, here a worker's ability, is correlated across firms, which are here the "arms" of the market-wide bandit problem of the worker's employment at one of the firms competing for the worker's services.

for each $t \in \{1, \dots, T-1\}$ and $(\eta^t, h, d^t, e) \in \mathcal{N}^t \times \mathcal{H} \times \mathcal{D}^t \times \mathcal{E}$, where $\eta^t := (\eta_1, \dots, \eta_t)$ and $d^t := (d_1, \dots, d_t)$. By Proposition B.1(i) at time $t+1$,

$$\Pr\big(e_n = e,\ \eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^{t+1} = d^{t+1}\big),$$

is identified for each $(e, \eta^t) \in \mathcal{E} \times \mathcal{N}^t$ and $(h, d^{t+1}) \in \mathcal{H} \times \mathcal{D}^{t+1}$, where $d^{t+1} := (d_1, d_2, \dots, d_t, d_{t+1}) = (d^t, d_{t+1})$. Using the law of total probability,

$$\begin{aligned}&\Pr\big(e_n = e,\ \eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^t = d^t\big)\\ &= \sum_{d_{t+1}} \Pr\big(e_n = e,\ \eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^{t+1} = (d^t, d_{t+1})\big) \times \Pr\big(D_{n,t+1} = d_{t+1} \mid H_{n,1} = h,\ D_n^t = d^t\big).\end{aligned}$$

Therefore

$$\Pr\big(e_n = e,\ \eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^t = d^t\big), \tag{44}$$

is identified.

By Proposition B.1(i) at time t,

$$\Pr\big(e_n = e,\ \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h,\ D_n^t = d^t\big),$$

is identified for each $(e, \eta^{t-1}) \in \mathcal{E} \times \mathcal{N}^{t-1}$ and $(h, d^t) \in \mathcal{H} \times \mathcal{D}^t$, where $d^t := (d_1, \dots, d_t)$ and $\eta^{t-1} := (\eta_1, \dots, \eta_{t-1})$. Therefore,

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_n^t = d^t\big), \tag{45}$$

is identified from

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_n^t = d^t\big) = \sum_{a^{t-1}} \Pr\big(e_n = e,\ \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h,\ D_n^t = d^t\big).$$

By combining (44) and (45), the conditional distribution in (43) is identified via the ratio

$$\frac{\Pr\big(e_n = e,\ \eta_n^t = \eta^t \mid H_{n,1} = h,\ D_n^t = d^t\big)}{\Pr\big(e_n = e \mid H_{n,1} = h,\ D_n^t = d^t\big)},$$

for any $e \in \mathcal{E}$. Note that, to combine (44) and (45), the mixture weights at t and $t+1$ must be consistently labeled, which we accomplish as explained in Appendix B.3 by using the means and variances of the mixture components.

Proof of Corollary B.1(b). We show how to identify the conditional signal distribution

$$\Pr\big(\eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} | H_{n,1} = h, D_{n,t} = d_t, D_{n,t+1} = d_{t+1}, D_{n,t+2} = d_{t+2}, e_n = e\big), \tag{46}$$

for each $t \in \{1, \dots, T-3\}$ and $(\eta_t, \eta_{t+1}, \eta_{t+2}, h, d_t, d_{t+1}, d_{t+2}, e) \in \mathcal{N}^3 \times \mathcal{H} \times \mathcal{D}^3 \times \mathcal{E}$.

By Proposition B.1(i) at time $t+3$,

$$\Pr\bigl(e_n = e,\ \eta_n^{t+2} = \eta^{t+2} \mid H_{n,1} = h,\ D_n^{t+3} = d^{t+3}\bigr),$$

is identified for each $(e, \eta^{t+2}) \in \mathcal{E} \times \mathcal{N}^{t+2}$ and $(h, d^{t+3}) \in \mathcal{H} \times \mathcal{D}^{t+3}$, where $d^{t+3} := (d_1, d_2, \ldots, d_{t-1}, d_t, d_{t+1}, d_{t+2}, d_{t+3}) = (d^{t-1}, d_t, d_{t+1}, d_{t+2}, d_{t+3})$ and $\eta^{t+2} := (\eta_1, \eta_2, \ldots, \eta_{t-1}, \eta_t, \eta_{t+1}, \eta_{t+2}) = (\eta^{t-1}, \eta_t, \eta_{t+1}, \eta_{t+2})$. Using the law of total probability,

$$\begin{aligned}
&\Pr\bigl(e_n = e, \eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} \mid H_{n,1} = h, D_{n,t} = d_t, D_{n,t+1} = d_{t+1}, D_{n,t+2} = d_{t+2}\bigr)\\
&= \sum_{\eta^{t-1}} \sum_{d^{t-1}, d_{t+3}} \Pr\bigl(e_n = e, \eta_n^{t+2} = (\eta^{t-1}, \eta_t, \eta_{t+1}, \eta_{t+2}) \mid H_{n,1} = h, D_n^{t+3} = (d^{t-1}, d_t, d_{t+1}, d_{t+2}, d_{t+3})\bigr)\\
&\times \Pr\bigl(D_n^{t-1} = d^{t-1}, D_{n,t+3} = d_{t+3} \mid H_{n,1} = h, D_{n,t} = d_t, D_{n,t+1} = d_{t+1}, D_{n,t+2} = d_{t+2}\bigr).
\end{aligned}$$

Therefore,

$$\Pr\bigl(e_n = e, \eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} | H_{n,1} = h, D_{n,t} = d_t, D_{n,t+1} = d_{t+1}, D_{n,t+2} = d_{t+2}\bigr), \tag{47}$$

is identified.

By Proposition B.1(i) at time $t+2$,

$$\Pr\bigl(e_n = e,\ \eta_n^{t+1} = \eta^{t+1} \mid H_{n,1} = h,\ D_n^{t+2} = d^{t+2}\bigr),$$

is identified for each $(e, \eta^{t+1}) \in \mathcal{E} \times \mathcal{N}^{t+1}$ and $(h, d^{t+2}) \in \mathcal{H} \times \mathcal{D}^{t+2}$, where $d^{t+2} := (d_1, d_2, \ldots, d_t,$-$d_{t+1}, d_{t+2}) = (d^{t-1}, d_t, d_{t+1}, d_{t+2})$ and $\eta^{t+1} := (\eta_1, \eta_2, \ldots, \eta_t, \eta_{t+1}) = (\eta^{t-1}, \eta_t, \eta_{t+1})$. By the law of total probability,

$$\begin{aligned}
&\Pr\bigl(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\bigr)\\
&= \sum_{\eta^{t+1}} \sum_{d^{t-1}} \Pr\bigl(e_n = e,\ \eta_n^{t+1} = \eta^{t+1} \mid H_{n,1} = h,\ D_n^{t+2} = (d^{t-1}, d_t, d_{t+1}, d_{t+2}))\bigr)\\
&\times \Pr\bigl(D_n^{t-1} = d^{t-1} \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\bigr).
\end{aligned}$$

Thus,

$$\Pr\bigl(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\bigr), \tag{48}$$

is identified.

By combining (47) and (48), the conditional distribution in (46) is identified via the ratio

$$\frac{\Pr\bigl(e_n = e, \eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} \mid H_{n,1} = h, D_{n,t} = d_t, D_{n,t+1} = d_{t+1}, D_{n,t+2} = d_{t+2}\bigr)}{\Pr\bigl(e_n = e \mid H_{n,1} = h, D_{n,t} = d_t, D_{n,t+1} = d_{t+1}, D_{n,t+2} = d_{t+2}\bigr)},$$

for any $e \in \mathcal{E}$.

Note that, to combine (47) and (48), the mixture weights at $t+2$ and $t+3$ must be consistently labeled, which we accomplish as explained in Appendix B.3 by using the means and variances of the mixture components.

Proof of Proposition B.2. *Step 1: Identification of* $\alpha(h,d,e)$ *and* $\beta(h,d,e)$. In this step, we identify the conditional probabilities

$$\alpha(h,d,e) \qquad \text{and} \qquad \beta(h,d,e), \tag{49}$$

for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ such that, for some $t \in \{1,\ldots,T-3\}$, $\Pr\big(\eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} \mid H_{n,1} = h, D_{n,t} = d, D_{n,t+1} = d, D_{n,t+2} = d, e_n = e\big)$ is identified for each $(\eta_t, \eta_{t+1}, \eta_{t+3}) \in \mathcal{N}^3$.

Proof: Fix $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ and $t \in \{1,\ldots,T-3\}$ such that, for some $t \in \{1,\ldots,T-3\}$, $\Pr\big(\eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} \mid H_{n,1} = h, D_{n,t} = d, D_{n,t+1} = d, D_{n,t+2} = d, e_n = e\big)$ is identified for each $(\eta_t, \eta_{t+1}, \eta_{t+3}) \in \mathcal{N}^3$. See remark below for sufficient conditions based on Corollary B.1(b).

For any $(\eta_t, \eta_{t+1}, \eta_{t+2}) \in \mathcal{N}^3$, using the law of total probability and Assumption 5, we can write

$$\begin{aligned}
&\Pr\big(\eta_{n,t} = \eta_t, \eta_{n,t+1} = \eta_{t+1}, \eta_{n,t+2} = \eta_{t+2} \mid H_{n,1} = h, D_{n,t} = d, D_{n,t+1} = d, D_{n,t+2} = d, e_n = e\big) \\
&\quad = \alpha(h,d,e)^{\sum_{\ell=0}^{2} \mathbb{1}\{\eta_{t+\ell} = \bar{\eta}\}} \big(1 - \alpha(h,d,e)\big)^{3 - \sum_{\ell=0}^{2} \mathbb{1}\{\eta_{t+\ell} = \bar{\eta}\}}\, q(h,d,e) \\
&\qquad + \beta(h,d,e)^{\sum_{\ell=0}^{2} \mathbb{1}\{\eta_{t+\ell} = \bar{\eta}\}} \big(1 - \beta(h,d,e)\big)^{3 - \sum_{\ell=0}^{2} \mathbb{1}\{\eta_{t+\ell} = \bar{\eta}\}} \big(1 - q(h,d,e)\big),
\end{aligned} \tag{50}$$

where $q(h,d,e) := \Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ D_{n,t} = d, D_{n,t+1} = d, D_{n,t+2} = d,\ e_n = e)$.

Equation (50) is a binomial mixture with two components and three trials. The left-hand side of (50) is identified by assumption. Following Blischke (1964, 1978), the weights and components of the binomial mixture in (50), $\{\alpha(h,d,e), \beta(h,d,e), q(h,d,e)\}$, are identified if the number of trials is greater than or equal to $2r-1$, where r is the number of mixture components. In our case, $r = 2$. Therefore, we need to observe workers who remain in job d for at least $2r-1 = 3$ periods, which motivates our focus on periods $t, t+1, t+2$ in (50). In particular, $\alpha(h,d,e)$ and $\beta(h,d,e)$ are identified without any labeling indeterminacy with respect to θ_n, using the restriction $\alpha(h,d,e) > \beta(h,d,e)$ imposed by Assumption 5(iii). □

Step 2: Identification of the Prior and Posterior Beliefs. In the proof below, we identify the prior

$$\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e), \tag{51}$$

for each $(h,e) \in \mathcal{H} \times \mathcal{E}$ such that, for some $d \in \mathcal{D}$, $\Pr\big(\eta_{n,1} = a \mid H_{n,1} = h, D_{n,1} = d, e_n = e\big)$ is identified, and $\alpha(h,d,e)$ and $\beta(h,d,e)$ are identified. In turn, the set of realizations of the posterior beliefs $\{P_{n,t}\}_{t=2}^{T}$ is identified, since each $P_{n,t}$ can be computed recursively as in equation (3) using $\{\alpha(h,d,e),\ \beta(h,d,e),\ \Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e)\}_{(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}}$.

Proof: Fix $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ such that $\Pr\big(\eta_{n,1} = \eta \mid H_{n,1} = h, D_{n,1} = d, e_n = e\big)$ is identified and $\alpha(h,d,e)$ and $\beta(h,d,e)$ are identified. See remark below for sufficient conditions based on

Corollary B.1(a). Using the law of total probability and Assumption 5(i) and (iii), we can write

$$\begin{aligned}\Pr\big(\eta_{n,1}=\eta|H_{n,1}=h,D_{n,1}=d,e_n=e\big) &= \alpha(h,d,e)^{\mathbb{1}\{\eta=\bar{\eta}\}}\big(1-\alpha(h,d,e)\big)^{\mathbb{1}\{a=\underline{\eta}\}}p_1(h,e)\\ &+\beta(h,d,e)^{\mathbb{1}\{\eta=\bar{\eta}\}}\big(1-\beta(h,d,e)\big)^{\mathbb{1}\{\eta=\underline{\eta}\}}\big(1-p_1(h,e)\big),\end{aligned}$$

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

$$p_1(h,e)=\begin{cases}\frac{\Pr(\eta_{n,1}=\eta|H_{n,1}=h,D_{n,1}=d,e_n=e)-\beta(h,d,e)}{\alpha(h,d,e)-\beta(h,d,e)} & \text{if } \eta=\bar{\eta},\\ \frac{\Pr(a_{n,1}=a|H_{n,1}=h,D_{n,1}=d,e_n=e)-(1-\beta(h,d,e))}{\beta(h,d,e)-\alpha(h,d,e)} & \text{if } \eta=\underline{\eta}.\end{cases} \tag{52}$$

Therefore, $p_1(h,e)$ is identified if $\Pr(\eta_{n,1}=\eta\mid H_{n,1}=h,D_{n,1}=d,e_n=e)$ is identified, $\alpha(h,d,e)$ and $\beta(h,d,e)$ are identified, and $\alpha(h,d,e)\neq\beta(h,d,e)$ by Assumption 5(iii).

Remark. Let $t\in\{1,\dots,T-3\}$ and $(h,d,e)\in\mathcal{H}\times\mathcal{D}\times\mathcal{E}$. By Corollary B.1(b),

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

is identified for each $(\eta_t,\eta_{t+1},\eta_{t+2})\in\mathcal{N}^3$ if the wage mixture weights in (23) are identified at times $t+2$ and $t+3$. See Proposition B.1 for sufficient conditions.

By Corollary B.1,

$$\Pr\big(\eta_{n,1}=\eta\mid H_{n,1}=h,D_{n,1}=d,e_n=e\big),$$

is identified for each $\eta\in\mathcal{N}$ if the wage mixture weights in (23) are identified at times 1 and 2. See Proposition B.1 for sufficient conditions.

Proof of Proposition B.3. Let $2\le t\le T$, $s\coloneqq(h,\iota,p,e)\in\mathcal{S}_t$, $d\in\mathcal{D}$, and $\tilde{s}\coloneqq(\tilde{h},\tilde{\iota},\tilde{p},\tilde{e})\in\mathcal{S}_{t-1}$. We have

$$\begin{aligned}&\Pr(s_{n,t}=s\mid D_{n,t-1}=d,s_{n,t-1}=\tilde{s})\\ &\quad=\Pr(H_{n,1}=h,I_n^{t-1}=\iota,e_n=e\mid D_{n,t-1}=d,H_{n,1}=\tilde{h},I_n^{t-2}=\tilde{\iota},P_{n,t-1}=\tilde{p},e_n=\tilde{e})\\ &\qquad\times\Pr(P_{n,t}=p\mid D_{n,t-1}=d,H_{n,1}=\tilde{h},I_n^{t-2}=\tilde{\iota},P_{n,t-1}=\tilde{p},e_n=\tilde{e}).\end{aligned}$$

For $(h,e)\neq(\tilde{h},\tilde{e})$, $\Pr(s_{n,t}=s\mid D_{n,t-1}=d,s_{n,t-1}=\tilde{s})=0$. For $(h,e)=(\tilde{h},\tilde{e})$,

$$\begin{aligned}&\Pr(s_{n,t}=s|D_{n,t-1}=d,s_{n,t-1}=\tilde{s})\\ &\quad=\Pr(I_n^{t-1}=\iota|D_{n,t-1}=d,I_n^{t-2}=\tilde{\iota})\\ &\qquad\times\Pr(P_{n,t}=p|D_{n,t-1}=d,H_{n,1}=h,P_{n,t-1}=\tilde{p},e_n=e).\end{aligned}\tag{53}$$

In (53), $\Pr(I_n^{t-1}=\iota|D_{n,t-1}=d,I_n^{t-2}=\tilde{\iota})$ is known because I_n^{t-1} is a known function of D_n^{t-1} and I_n^{t-2} is a known function of D_n^{t-2}. By (3), p can take two values, $\{\bar{p},\underline{p}\}$, depending on whether $\eta_{n,t-1}=\bar{\eta}$ or $\eta_{n,t-1}=\underline{\eta}$. Thus, $\Pr(P_{n,t}=p|D_{n,t-1}=d,H_{n,1}=h,P_{n,t-1}=\tilde{p},e_n=e)$ in (53) can be

$$\begin{aligned}&\Pr(P_{n,t}=\bar{p}|D_{n,t-1}=d,H_{n,1}=h,P_{n,t-1}=\tilde{p},e_n=e)\\ &\quad=\Pr(\eta_{n,t-1}=\bar{\eta}|D_{n,t-1}=d,H_{n,1}=h,P_{n,t-1}=\tilde{p},e_n=e),\end{aligned}$$

or

$$
\begin{aligned}
&\Pr(P_{n,t} = \underline{p} | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) \\
&\qquad\qquad = \Pr(\eta_{n,t-1} = \underline{\eta} | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e),
\end{aligned}
$$

which are identified by Proposition B.2. Indeed,

$$
\begin{aligned}
&\Pr(\eta_{n,t-1} = \bar{\eta} | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) \\
&= \sum_{\theta \in \Theta} \Pr(\eta_{n,t-1} = \bar{\eta} | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e, \theta_n = \theta) \\
&\qquad\qquad \times \Pr(\theta_n = \theta | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) \\
&= \sum_{\theta \in \Theta} \Pr(\eta_{n,t-1} = \bar{\eta} | D_{n,t-1} = d, H_{n,1} = h, e_n = e, \theta_n = \theta) \\
&\qquad\qquad \times \Pr(\theta_n = \theta | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) \\
&= \alpha(h, d, e)\tilde{p} + \beta(h, d, e)(1 - \tilde{p}).
\end{aligned}
$$

Similarly,

$$
\Pr(\eta_{n,t-1} = \underline{\eta} | D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) = (1 - \alpha(h, d, e))\tilde{p} + (1 - \beta(h, d, e))(1 - \tilde{p}).
$$

Therefore, $\Pr(s_{n,t} = s \mid D_{n,t-1} = d, s_{n,t-1} = \tilde{s})$ is identified. $\square$

Proof of Proposition B.4. Let $2 \leq t \leq T$. The conditional probability

$$
\Pr(D_{n,t} = d \mid H_{n,1} = h, D_n^{t-1} = d^{t-1}, e_n = e, \eta_n^{t-1} = \eta^{t-1}), \tag{54}
$$

is identified for each $(d, h, d^{t-1}, e, \eta^{t-1}) \in \mathcal{D} \times \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{N}^{t-1}$, where $d^{t-1} := (d_1, \ldots, d_{t-1})$ and $\eta^{t-1} := (\eta_1, \ldots, \eta_{t-1})$. This is because, by Bayes' rule,

$$
\begin{aligned}
&\Pr(D_{n,t} = d \mid H_{n,1} = h, D_n^{t-1} = d^{t-1}, e_n = e, \eta_n^{t-1} = \eta^{t-1}) \\
&\quad = \frac{\Pr(e_n = e, \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h, D_n^t = (d^{t-1}, d)) \Pr(D_{n,t} = d \mid H_{n,1} = h, D_n^{t-1} = d^{t-1})}{\Pr(e_n = e, \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h, D_n^{t-1} = d^{t-1})},
\end{aligned}
$$

where:

- $\Pr(e_n = e, \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h, D_n^t = (d^{t-1}, d))$ is identified by Proposition B.1(i) at time t for each $(e, \eta^{t-1}, h, d^{t-1}, d) \in \mathcal{E} \times \mathcal{N}^{t-1} \times \mathcal{H} \times \mathcal{D}^t$.
- $\Pr(D_{n,t} = d \mid H_{n,1} = h, D_n^{t-1} = d^{t-1})$ is known from the data for each $(d, h, d^{t-1}) \in \mathcal{D} \times \mathcal{H} \times \mathcal{D}^{t-1}$.
- $\Pr(e_n = e, \eta_n^{t-1} = \eta^{t-1} \mid H_{n,1} = h, D_n^{t-1} = d^{t-1})$ is identified, as shown in (44), from Proposition B.1(i) at time t for each $(h, d^{t-1}, e, \eta^{t-1}) \in \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{N}^{t-1}$.

The joint distribution

$$
\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}, e_n = e, \eta_n^{t-1} = \eta^{t-1}), \tag{55}
$$

is identified from Proposition B.1(i) at time t for each $(h, d^{t-1}, e, \eta^{t-1}) \in \mathcal{H}\times\mathcal{D}^{t-1}\times\mathcal{E}\times\mathcal{N}^{t-1}$. Given (54), (55), and knowledge of the map g_t from realizations of $(H_{n,1}, D_n^{t-1}, e_n, \eta_n^{t-1})$ to realizations of $s_{n,t}$—which follow from Proposition B.2—we identify $\Pr(D_{n,t} = d \mid s_{n,t} = s)$ for all $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$.

For $t = 1$, the same steps apply, with the obvious modification that $D_{n,t-1}$ and $\eta_{n,t-1}$ are not present in the derivations. □

Proof of Proposition B.5. Let $2 \le t \le T$. From Proposition B.1(ii) at time t, we identify

$$\Pr\big(w_{n,t} \le w \mid H_{n,1} = h, D_{n,t}^{t} = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1}\big), \tag{56}$$

for each $(h, d^t, e, \eta^{t-1}) \in \mathcal{H} \times \mathcal{D}^t \times \mathcal{E} \times \mathcal{N}^{t-1}$, where $d^t := (d_1, \dots, d_t) = (d^{t-1}, d_t)$ and $\eta^{t-1} := (\eta_1, \dots, \eta_{t-1})$.

From Proposition B.1(i) at time t, we identify

$$\Pr\big(H_{n,1} = h, D_{n,t}^{t} = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1}\big), \tag{57}$$

for each $(h, d^t, e, \eta^{t-1}) \in \mathcal{H} \times \mathcal{D}^t \times \mathcal{E} \times \mathcal{N}^{t-1}$. From Proposition B.1(i) at time t, we identify

$$\Pr\big(H_{n,1} = h, D_{n,t}^{t-1} = d^{t-1}, e_n = e, \eta_n^{t-1} = \eta^{t-1}\big), \tag{58}$$

for each $(h, d^{t-1}, e, \eta^{t-1}) \in \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{N}^{t-1}$.

By taking the ratio between (57) and (58), we identify

$$\begin{aligned}\Pr\big(D_{n,t} = d_t \mid H_{n,1} = h, D_{n,t}^{t-1} = d^{t-1}, e_n = e, \eta_n^{t-1} = \eta^{t-1}\big)\\ = \frac{\Pr(H_{n,1} = h, D_{n,t}^{t} = d^t, e_n = e, \eta_n^{t-1} = \eta^{t-1})}{\Pr(H_{n,1} = h, D_{n,t}^{t-1} = d^{t-1}, e_n = e, \eta_n^{t-1} = \eta^{t-1})},\end{aligned} \tag{59}$$

for each $(h, d^t, e, \eta^{t-1}) \in \mathcal{H} \times \mathcal{D}^t \times \mathcal{E} \times \mathcal{N}^{t-1}$.

Let $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$. Using Bayes' rule, for each $w \in \mathbb{R}$, we can write

$$\begin{aligned}&\Pr(w_{n,t} \le w | D_{n,t} = d, s_{n,t} = s) = \\ &\quad \sum_{\substack{(h,d^{t-1},e,\eta^{t-1}):\\ g(h,d^{t-1},e,\eta^{t-1})=s}} \Pr(w_{n,t} \le w | H_{n,1} = h, D_{n,t}^{t-1} = d^{t-1}, D_{n,t} = d, e_n = e, \eta_n^{t-1} = \eta^{t-1}) \\ &\quad \times \frac{\Pr(H_{n,1} = h, D_{n,t}^{t-1} = d^{t-1}, D_{n,t} = d, e_n = e, \eta_n^{t-1} = \eta^{t-1})}{\sum\limits_{\substack{(h,d^{t-1},e,\eta^{t-1}):\\ g(h,d^{t-1},e,\eta^{t-1})=s}} \Pr(H_{n,1} = h, D_{n,t}^{t-1} = d^{t-1}, D_{n,t} = d, e_n = e, \eta_n^{t-1} = \eta^{t-1})}.\end{aligned} \tag{60}$$

All the components on the right-hand side of (60) are identified by (56) to (59). Therefore, $\Pr(w_{n,t} \le w \mid D_{n,t} = d, s_{n,t} = s)$ is identified.

For $t = 1$, the same steps apply, with the obvious modification that $\eta_{n,t-1}$ and D_n^{t-1} are not present in the derivations. □

Proof of Proposition B.6. To keep the notation and exposition lean, we provide the proof of Proposition B.6 for the wage equation (11), which omits dependence on the second-best firm $D'_{n,t}$—Proposition E.1 below. The proof for the general equation (10) follows the same steps, with $D'_{n,t}$ entering the relevant expressions where needed.

Proposition E.1 (Deterministic Wage—Simplified Wage Equation (11))**.** *Let $t \in \{1, \ldots, T\}$ and $d \in \mathcal{D}$. Assume that the conditional wage distribution $\Pr\big(w_{n,t} \le w \,\big|\, D_{n,t} = d,\ s_{n,t} = s\big)$ is identified for each $w \in \mathbb{R}$ and $s \in \mathcal{S}_t$ (see Proposition B.5 for sufficient conditions), and that the conditional choice probability $\Pr(D_{n,t} = d \mid s_{n,t} = s)$ is identified for each $s \in \mathcal{S}_t$ (see Proposition B.4 for sufficient conditions). Moreover, assume:*

(i) (Supports.) For each $s \in \mathcal{S}_t$, $\omega_t(d, s) := \sup\{u : \Pr(w_{n,t}(d) \le u | s_{n,t} = s) < 1\} = \infty$ and $\omega_t^{\text{obs}}(d, s) := \sup\{u : \Pr(w_{n,t} \le u \mid D_{n,t} = d, s_{n,t} = s) < 1\} = \infty$.

(ii) (Tail Limit.) There exists an unknown constant $q_t(d) \in (0, 1]$ such that for each $s \in \mathcal{S}_t$, $\lim_{w\to\infty} \Pr\big(D_{n,t} = d | s_{n,t} = s,\ w_{n,t}(d) > w\big) = q_t(d)$.

(iii) (Tail Regularity.) For each $s \in \mathcal{S}_t(d)$, there exist unknown thresholds $a_t(d, s) < \infty$ and $a_t^{\text{obs}}(d, s) < \infty$ such that the cumulative distribution functions of wages $F_{w_{n,t}(d)|s_{n,t}=s}$ and $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ are continuous and strictly increasing on $(a_t(d, s), \infty)$ and $(a_t^{\text{obs}}(d, s), \infty)$, respectively.

(iv) (Normalization.) There exists a known $\bar{s} \in \mathcal{S}_t$ such that $\varphi(d, \bar{s}) = 0$.

Let $\{\tau^{(k)}_{d,\bar{s},t}\}_{k\ge1} \subset (0, 1)$ be any sequence with $\tau^{(k)}_{d,\bar{s},t} \to 1$ as $k \to \infty$. For each $s \in \mathcal{S}_t$, define the sequence $\{\tau^{(k)}_{d,s,t}\}_{k\ge1} \subset (0, 1)$ with $\tau^{(k)}_{d,s,t} := 1 - \frac{\Pr(D_{n,t}=d|s_{n,t}=\bar{s})}{\Pr(D_{n,t}=d|s_{n,t}=s)} (1 - \tau^{(k)}_{d,\bar{s},t})$. Then,

$$\lim_{k\to\infty} \Big[\, Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=s}\big(\tau^{(k)}_{d,s,t}\big) \;-\; Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{d,\bar{s},t}\big) \Big] \;=\; \varphi(d, s). \tag{61}$$

Hence, $\varphi(d, s)$ is identified for each $s \in \mathcal{S}_t$.

Proof: Let $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$. For any threshold w, Bayes' rule and the identity $\Pr(w_{n,t} > w \mid D_{n,t} = d, s_{n,t} = s) = \Pr(w_{n,t}(d) > w \mid D_{n,t} = d, s_{n,t} = s)$ give

$$\Pr(w_{n,t} > w | D_{n,t} = d, s_{n,t} = s) = \frac{\Pr(w_{n,t}(d) > w | s_{n,t} = s)\ \Pr(D_{n,t} = d | s_{n,t} = s,\ w_{n,t}(d) > w)}{\Pr(D_{n,t} = d | s_{n,t} = s)}.$$

Using Assumptions 6 and 7,

$$\Pr(w_{n,t} > w \mid D_{n,t} = d, s_{n,t} = s) \;\sim\; c_t(d, s)\, \Pr\big(w_{n,t}(d) > w \mid s_{n,t} = s\big) \qquad (w \to \infty), \tag{62}$$

where $c_t(d, s) := \frac{q_t(d)}{\Pr(D_{n,t}=d|s_{n,t}=s)} \in (0, \infty)$ and "$\sim$" denotes that the ratio of the two sides converges to 1.

Write $S_{d,s,t}(w) := S_{w_{n,t}(d)|s_{n,t}=s}(w)$ and $S^{\text{obs}}_{d,s,t}(w) := S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w)$ to denote the survival functions of the potential wages $w_{n,t}(d)|s_{n,t} = s$ and observed wages $w_{n,t}|(D_{n,t} = d, s_{n,t} = s)$. Then,

(62) reads as

$$S^{\text{obs}}_{d,s,t}(w) \;\sim\; c_t(d,s)\, S_{d,s,t}(w) \qquad (w \to \infty). \tag{63}$$

By Assumption 6, both right wage endpoints are ∞. By Assumption 8, the upper-tail CDFs $F_{w_{n,t}(d)|s_{n,t}=s}$ and $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ are continuous and strictly increasing beyond finite thresholds, so their tail quantile maps are the ordinary inverses on the corresponding index ranges near 1. Hence, by Lemma E.1,

$$Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=s}(\tau) \;=\; Q_{w_{n,t}(d)\,|\,s_{n,t}=s}\Big(1-\frac{1-\tau}{c_t(d,s)} \;+\; o_s(1-\tau)\Big) \qquad (\tau \to 1), \tag{64}$$

where $o_s(1-\tau)/(1-\tau) \to 0$ as $\tau \to 1$.
From $w_{n,t}(d) = \varphi(d,s) + \epsilon_{n,t}(d)$, for all $u \in (0,1)$,

$$Q_{w_{n,t}(d)\,|\,s_{n,t}=s}(u) \;=\; \varphi(d,s) + Q_{\epsilon_{n,t}(d)}(u).$$

Plugging into (64) gives

$$Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=s}(\tau) \;=\; \varphi(d,s) \;+\; Q_{\epsilon_{n,t}(d)}\Big(1-\frac{1-\tau}{c_t(d,s)} \;+\; o_s(1-\tau)\Big) \qquad (\tau \to 1). \tag{65}$$

Let $\{\tau^{(k)}_{d,\bar{s},t}\}_{k\geq 1} \subset (0,1)$ with $\tau^{(k)}_{d,\bar{s},t} \to 1$. Define $\tau^{(k)}_{d,s,t}$ as

$$\tau^{(k)}_{d,s,t} \;\coloneqq\; 1-\frac{c_t(d,s)}{c_t(d,\bar{s})}\big(1-\tau^{(k)}_{d,\bar{s},t}\big). \tag{66}$$

Since $c_t(d,s), c_t(d,\bar{s}) \in (0,\infty)$, we have $\tau^{(k)}_{d,s,t} \in (0,1)$ for all large k and $\tau^{(k)}_{d,s,t} \to 1$ as $k \to \infty$. Note also that by (66), $1-\tau^{(k)}_{d,s,t} = (c_t(d,s)/c_t(d,\bar{s}))(1-\tau^{(k)}_{d,\bar{s},t})$, so $1-\tau^{(k)}_{d,s,t}$ and $1-\tau^{(k)}_{d,\bar{s},t}$ are of the same order.
Evaluate (65) at $\tau = \tau^{(k)}_{d,s,t}$ and, with $s=\bar{s}$, at $\tau = \tau^{(k)}_{d,\bar{s},t}$:

$$\begin{aligned} Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=s}\big(\tau^{(k)}_{d,s,t}\big) &= \varphi(d,s) + Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{d,s,t}}{c_t(d,s)} + o_s(1-\tau^{(k)}_{d,s,t})\Big), \\ Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{d,\bar{s},t}\big) &= \varphi(d,\bar{s}) + Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{d,\bar{s},t}}{c_t(d,\bar{s})} + o_{\bar{s}}(1-\tau^{(k)}_{d,\bar{s},t})\Big), \end{aligned} \qquad (k\to\infty). \tag{67}$$

By construction (66),

$$1-\frac{1-\tau^{(k)}_{d,s,t}}{c_t(d,s)} \;=\; 1-\frac{1-\tau^{(k)}_{d,\bar{s},t}}{c_t(d,\bar{s})}.$$

Therefore, (67) can be written as

$$\begin{aligned} Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=s}\big(\tau^{(k)}_{d,s,t}\big) &= \varphi(d,s) + Q_{\epsilon_{n,t}(d)}\Big(1-\tfrac{1-\tau^{(k)}_{d,\bar{s},t}}{c_t(d,\bar{s})} + o_s(1-\tau^{(k)}_{d,s,t})\Big), \\ Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{d,\bar{s},t}\big) &= \varphi(d,\bar{s}) + Q_{\epsilon_{n,t}(d)}\Big(1-\tfrac{1-\tau^{(k)}_{d,\bar{s},t}}{c_t(d,\bar{s})} + o_{\bar{s}}(1-\tau^{(k)}_{d,\bar{s},t})\Big), \end{aligned} \qquad (k\to\infty). \tag{68}$$

Also note that $o_s(1-\tau_{d,s,t}^{(k)}) \to 0$ and $o_{\bar{s}}(1-\tau_{d,\bar{s},t}^{(k)}) \to 0$ as $k \to \infty$. Therefore, by continuity of $Q_{\epsilon_{n,t}(d)}$ near 1 under Assumption 8,

$$Q_{\epsilon_{n,t}(d)}\Big(u+o_s(1-\tau_{d,s,t}^{(k)})\Big)-Q_{\epsilon_{n,t}(d)}\Big(u+o_{\bar{s}}(1-\tau_{d,\bar{s},t}^{(k)})\Big) = o(1), \qquad u := 1-\frac{1-\tau_{d,\bar{s},t}^{(k)}}{c_t(d,\bar{s})} \qquad (k \to \infty).$$

Subtracting the two equations in (68) and using the normalization $\varphi(d,\bar{s}) = 0$ from Assumption 9:

$$\lim_{k\to\infty} \Big[\, Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=s}\big(\tau_{d,s,t}^{(k)}\big) \;-\; Q_{w_{n,t}\,|\,D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tau_{d,\bar{s},t}^{(k)}\big)\,\Big] \;=\; \varphi(d,s).$$

This proves the claim.

Note that, while $\{\tau_{d,\bar{s},t}^{(k)}\}_{k\geq 1}$ can be chosen arbitrarily as any sequence with $\tau_{d,\bar{s},t}^{(k)} \in (0,1)$ and $\tau_{d,\bar{s},t}^{(k)} \to 1$, the sequence $\{\tau_{d,s,t}^{(k)}\}_{k\geq 1}$ is defined in (66) as

$$\tau_{d,s,t}^{(k)} \;:=\; 1-\frac{c_t(d,s)}{c_t(d,\bar{s})}\big(1-\tau_{d,\bar{s},t}^{(k)}\big).$$

Using Assumption 7—in particular, the invariance across s of $\lim_{w\to\infty}\Pr(D_{n,t} = d \mid s_{n,t} = s,\ w_{n,t}(d) > w)$—we have that

$$\frac{c_t(d,s)}{c_t(d,\bar{s})} = \frac{q_t(d)/\Pr(D_{n,t}=d \mid s_{n,t}=s)}{q_t(d)/\Pr(D_{n,t}=d \mid s_{n,t}=\bar{s})} = \frac{\Pr(D_{n,t}=d \mid s_{n,t}=\bar{s})}{\Pr(D_{n,t}=d \mid s_{n,t}=s)}.$$

Therefore,

$$\tau_{d,s,t}^{(k)} = 1-\frac{\Pr(D_{n,t}=d \mid s_{n,t}=\bar{s})}{\Pr(D_{n,t}=d \mid s_{n,t}=s)}\,\big(1-\tau_{d,\bar{s},t}^{(k)}\big),$$

which can be calculated in the data for any given choice of $\{\tau_{d,\bar{s},t}^{(k)}\}_{k\geq 1}$. Hence, $\tau_{d,s,t}^{(k)}$ can also be calculated from the data. $\square$

Lemma E.1 (Survival-to-quantile index inversion—Simplified Wage Equation (11))**.** *Let $t \in \{1,\dots,T\}$, $d \in \mathcal{D}$, and $s \in \mathcal{S}_t$. Under Assumptions (i) and (iii) of Proposition E.1,*

$$S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) \;\sim\; c_t(d,s)\,S_{w_{n,t}(d)|s_{n,t}=s}(w) \qquad (w\to\infty), \tag{69}$$

for some $c_t(d,s) \in (0,\infty)$, if and only if

$$Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau) \;=\; Q_{w_{n,t}(d)|s_{n,t}=s}\Big(1-\frac{1-\tau}{c_t(d,s)}+o_s(1-\tau)\Big) \qquad (\tau\to 1), \tag{70}$$

where $o_s(1-\tau)/(1-\tau) \to 0$ as $\tau \to 1$.

Proof: *($\Rightarrow$) Assume*

$$S_{d,s,t}^{\text{obs}}(w) \sim c_t(d,s)\,S_{d,s,t}(w) \qquad (w\to\infty). \tag{71}$$

Fix $\tau \to 1$ and define

$$w_\tau := Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau). \tag{72}$$

By (i), $\omega_t^{\text{obs}}(d,s)=\infty$, so $w_\tau\to\infty$ as $\tau\to 1$. For τ close enough to 1, w_τ lies in the tail region where (iii) applies; thus, by continuity on the tail and (72),

$$F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w_\tau)=\tau,$$

which is equivalent to

$$S^{\text{obs}}_{d,s,t}(w_\tau)=1-\tau. \tag{73}$$

From (71) evaluated at $w=w_\tau$ and (73), we get

$$S_{d,s,t}(w_\tau)=\frac{1-\tau}{c_t(d,s)}+o_s(1-\tau)\qquad(\tau\to 1), \tag{74}$$

where $o_s(1-\tau)/(1-\tau)\to 0$ as $\tau\to 1$.
Define

$$u_\tau:=F_{w_{n,t}(d)|s_{n,t}=s}(w_\tau)=1-S_{d,s,t}(w_\tau). \tag{75}$$

By (74)–(75),

$$u_\tau=1-\frac{1-\tau}{c_t(d,s)}+o_s(1-\tau)\qquad(\tau\to 1). \tag{76}$$

Since $u_\tau\to 1$, for τ close enough to 1 we have u_τ in the tail index range where $F_{w_{n,t}(d)|s_{n,t}=s}$ is invertible; combining this with (75),

$$w_\tau=\big(F_{w_{n,t}(d)|s_{n,t}=s}\big)^{-1}(u_\tau)=Q_{w_{n,t}(d)|s_{n,t}=s}(u_\tau)\qquad(\tau\to 1). \tag{77}$$

Substituting (76) into (77) yields

$$w_\tau=Q_{w_{n,t}(d)|s_{n,t}=s}\Big(1-\frac{1-\tau}{c_t(d,s)}+o_s(1-\tau)\Big)\qquad(\tau\to 1). \tag{78}$$

Combining (78) with (72) gives (70).

($\Leftarrow$) Conversely, assume

$$Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau)=Q_{w_{n,t}(d)|s_{n,t}=s}\Big(1-\frac{1-\tau}{c_t(d,s)}+o_s(1-\tau)\Big)\qquad(\tau\to 1). \tag{79}$$

Fix $\tau\to 1$ and set

$$u_\tau:=1-\frac{1-\tau}{c_t(d,s)}+o_s(1-\tau). \tag{80}$$

Then, (79) becomes

$$Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau)=Q_{w_{n,t}(d)|s_{n,t}=s}(u_\tau)\qquad(\tau\to 1). \tag{81}$$

Since $u_\tau\to 1$, it lies in the tail index range where (iii) yields invertibility, so

$$F_{w_{n,t}(d)|s_{n,t}=s}\big(Q_{w_{n,t}(d)|s_{n,t}=s}(u_\tau)\big)\;=\;u_\tau. \tag{82}$$

Applying $F_{w_{n,t}(d)|s_{n,t}=s}$ to both sides of (81) and using (80)-(82) gives

$$F_{w_{n,t}(d)|s_{n,t}=s}\big(Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau)\big) = 1 - \frac{1-\tau}{c_t(d,s)} + o_s(1-\tau) \qquad (\tau \to 1).$$

Equivalently, in survival notation,

$$S_{d,s,t}\big(Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau)\big) = \frac{1-\tau}{c_t(d,s)} + o_s(1-\tau) \qquad (\tau \to 1). \tag{83}$$

Moreover, by continuity on the tail under (iii),

$$S^{\text{obs}}_{d,s,t}\big(Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau)\big) \;=\; 1-\tau. \tag{84}$$

Define

$$w_\tau \;\coloneqq\; Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau). \tag{85}$$

Then, by tail continuity under (iii),

$$S^{\text{obs}}_{d,s,t}(w_\tau) = 1-\tau. \tag{86}$$

From (84), (85), and (86), with $r_{d,s,t}(\tau) \coloneqq o_s(1-\tau)/(1-\tau) \to 0$,

$$\frac{S^{\text{obs}}_{d,s,t}(w_\tau)}{S_{d,s,t}(w_\tau)} = \frac{1-\tau}{\frac{1-\tau}{c_t(d,s)}\big(1+c_t(d,s)r_{d,s,t}(\tau)\big)} = c_t(d,s)\,\frac{1}{1+c_t(d,s)r_{d,s,t}(\tau)} = c_t(d,s)\{1+o(1)\},$$

whence

$$S^{\text{obs}}_{d,s,t}(w_\tau) \;\sim\; c_t(d,s)\,S_{d,s,t}(w_\tau) \qquad (\tau \to 1).$$

Finally, since $\tau \mapsto w_\tau$ is increasing and unbounded, any sequence $w \to \infty$ can be written as w_{τ_k} with $\tau_k \to 1$,

$$S^{\text{obs}}_{d,s,t}(w_\tau) \;\sim\; c_t(d,s)\,S_{d,s,t}(w) \qquad (w \to \infty),$$

which is (69). □

Proof of Proposition B.7. To keep the notation and exposition lean, we provide the proof of Proposition B.7 for the wage equation (11), which omits dependence on the second-best firm $D'_{n,t}$—Proposition E.2 below. The proof for the general equation (10) follows the same steps, with $D'_{n,t}$ entering the relevant expressions where needed.

Assumption 13 (Tail-Ratio Identifiable Distribution—Simplified Wage Equation (11))**.** *Let* $t \in \{1,\dots,T\}$. *For each* $d \in \mathcal{D}$, $\epsilon_{n,t}(d)$ *belongs to a parametric family, indexed by the* $p_d \times 1$ *vector of parameters* μ_d, *that is tail-ratio identifiable. Namely, for any* p_d+1 *distinct large thresholds* $0 < e_0 < e_1 < \cdots < e_{p_d}$, *the map*

$$\mu_d \mapsto \left(\frac{S_{\epsilon_{n,t}(d)}(e_1;\mu_d)}{S_{\epsilon_{n,t}(d)}(e_0;\mu_d)}, \dots, \frac{S_{\epsilon_{n,t}(d)}(e_{p_d};\mu_d)}{S_{\epsilon_{n,t}(d)}(e_0;\mu_d)}\right),$$

is injective, where $S_{\epsilon_{n,t}(d)}$ denotes the survival function of $\epsilon_{n,t}(d)$.

Proposition E.2 (Shock Distribution—Simplified Wage Equation (11))**.** *Let $t \in \{1, \ldots, T\}$. Assume:*

(i) The conditional wage distribution $\Pr\big(w_{n,t} \leq w \,\big|\, D_{n,t} = d,\ s_{n,t} = s\big)$ is identified for each $w \in \mathbb{R}$, $d \in \mathcal{D}$, and $s \in \mathcal{S}_t$. See Proposition B.5 for sufficient conditions.

(ii) Assumptions (i) to (iv) of Proposition E.1 hold, implying that $\varphi(d, s)$ is identified for each $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$.

(iii) Assumption 13 holds.

Then, the parameter μ_d is identified for each $d \in \mathcal{D}$. Moreover, if the shocks $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$ are mutually independent across $d \in \mathcal{D}$, then the joint distribution of $\epsilon_{n,t}$ is identified as the product of the identified marginals. Alternatively, if a copula C_μ is specified so that

$$F_{\epsilon_{n,t}}(v_1, \ldots, v_{|\mathcal{D}|}) \;=\; C_\mu\big(F_{\epsilon_{n,t}(1)}(v_1; \mu_1), \ldots, F_{\epsilon_{n,t}(|\mathcal{D}|)}(v_{|\mathcal{D}|}; \mu_{|\mathcal{D}|})\big) \quad \forall (v_1, \ldots, v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|},$$

and the copula parameter μ is known, then the joint distribution of $\epsilon_{n,t}$ is identified from the identified marginals and C_μ. Without further restrictions on the dependence among $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$, the joint distribution of $\epsilon_{n,t}$ is partially identified by the sharp Fréchet–Höffding bounds in that for all $(v_1, \ldots, v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|}$,

$$\max\Big\{\sum_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d; \mu_d) - (|\mathcal{D}| - 1),\ 0\Big\} \;\leq\; F_{\epsilon_{n,t}}(v_1, \ldots, v_{|\mathcal{D}|}) \;\leq\; \min_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d; \mu_d).$$

Proof: Let $t \in \{1, \ldots, T\}$ and $d \in \mathcal{D}$. We prove that μ_d is identified. Let $s \in \mathcal{S}_t$. By Bayes' rule and the identity $\Pr(w_{n,t} > w \mid D_{n,t} = d, s_{n,t} = s) = \Pr(w_{n,t}(d) > w \mid D_{n,t} = d, s_{n,t} = s)$, for any $w \in \mathbb{R}$,

$$S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) = S_{w_{n,t}(d)|s_{n,t}=s}(w)\, \frac{\Pr\big(D_{n,t} = d \mid s_{n,t} = s,\ w_{n,t}(d) > w\big)}{\Pr(D_{n,t} = d \mid s_{n,t} = s)}. \tag{87}$$

Under Assumption (i) and (ii) of Proposition E.1, if follows from (87) that

$$S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) \;\sim\; c_t(d, s)\, S_{w_{n,t}(d)|s_{n,t}=s}(w) \qquad (w \to \infty), \tag{88}$$

where $c_t(d, s) := q_t(d)/\Pr(D_{n,t} = d \mid s_{n,t} = s) \in (0, \infty)$.

Take two thresholds $w_1, w_2 > 0$ and let $\min\{w_1, w_2\} \to \infty$. Dividing (88) at $w = w_1$ and $w = w_2$ gives

$$\lim_{\min\{w_1,w_2\}\to\infty} \frac{S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w_1)}{S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w_2)} = \frac{S_{w_{n,t}(d)|s_{n,t}=s}(w_1)}{S_{w_{n,t}(d)|s_{n,t}=s}(w_2)}. \tag{89}$$

By exogeneity of $\epsilon_{n,t}$,

$$S_{w_{n,t}(d)|s_{n,t}=s}(w) = \Pr\big(\epsilon_{n,t}(d) > w - \varphi(d, s)\big) = S_{\epsilon_{n,t}(d)}\big(w - \varphi(d, s); \mu_d\big). \tag{90}$$

Substituting (90) into (89) yields

$$\lim_{\min\{w_1,w_2\}\to\infty} \frac{S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(w_1\big)}{S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(w_2\big)} \;=\; \frac{S_{\epsilon_{n,t}(d)}\big(w_1-\varphi(d,s);\mu_d\big)}{S_{\epsilon_{n,t}(d)}\big(w_2-\varphi(d,s);\mu_d\big)}. \tag{91}$$

Now choose p_d+1 distinct large thresholds $0<w_0<w_1<\cdots<w_{p_d}$ and form the p_d ratios

$$R_j(s) \;\coloneqq\; \frac{S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(w_j\big)}{S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(w_0\big)}, \qquad j=1,\ldots,p_d.$$

Applying (91) with $(w_1,w_2)=(w_j,w_0)$ for each $j=1,\ldots,p_d$ and letting all thresholds be large gives

$$\lim_{\min\{w_j,w_0\}\to\infty} R_j(s) \;=\frac{S_{\epsilon_{n,t}(d)}\big(w_j-\varphi(d,s);\mu_d\big)}{S_{\epsilon_{n,t}(d)}\big(w_0-\varphi(d,s);\mu_d\big)}\,, \qquad j=1,\ldots,p_d, \tag{92}$$

(92) tells us that the observed vector

$$\big(R_1(s),\ldots,R_{p_d}(s)\big),$$

converges to the vector

$$\Big(\frac{S_{\epsilon_{n,t}(d)}\big(w_1-\varphi(d,s);\mu_d\big)}{S_{\epsilon_{n,t}(d)}\big(w_0-\varphi(d,s);\mu_d\big)},\ldots,\frac{S_{\epsilon_{n,t}(d)}\big(w_{p_d}-\varphi(d,s);\mu_d\big)}{S_{\epsilon_{n,t}(d)}\big(w_0-\varphi(d,s);\mu_d\big)}\Big).$$

By Assumption 10, the map

$$\mu_d \mapsto \Big(\frac{S_{\epsilon_{n,t}(d)}\big(w_1-\varphi(d,s);\mu_d\big)}{S_{\epsilon_{n,t}(d)}\big(w_0-\varphi(d,s);\mu_d\big)},\ldots,\frac{S_{\epsilon_{n,t}(d)}\big(w_{p_d}-\varphi(d,s);\mu_d\big)}{S_{\epsilon_{n,t}(d)}\big(w_0-\varphi(d,s);\mu_d\big)}\Big),$$

is injective. Therefore, the system of equations 92 identifies μ_d. □

Examples of Tail-Ratio Identifiable Distributions. Consider a random variable ϵ belonging to a parametric family, indexed by the $p\times 1$ vector of parameters β, that is *tail-ratio identifiable*. Namely, for any any $p+1$ distinct large thresholds $0<e_0<e_1<\cdots<e_p$, the map

$$\beta \mapsto \left(\frac{S_\epsilon(e_1;\beta)}{S_\epsilon(e_0;\beta)},\ldots,\frac{S_\epsilon(e_p;\beta)}{S_\epsilon(e_0;\beta)}\right),$$

is injective. We now provide examples of parametric families that are, or are not, tail-ratio identifiable.

Normal. Suppose $\epsilon\sim\mathcal{N}(\mu,\sigma^2)$, with parameters (μ,σ), so $p=2$. The survival function is

$$S_\epsilon(e;\mu,\sigma)=\Pr(\epsilon>e)=1-F\left(\frac{e-\mu}{\sigma}\right),$$

where $F(\cdot)$ denotes the standard normal CDF. Fix any $0 < e_0 < e_1 < e_2$ and define

$$z_0 := \frac{e_0 - \mu}{\sigma}, \qquad \Delta_j := e_j - e_0 > 0 \;\; (j = 1, 2).$$

Then

$$\frac{S_\epsilon(e_j; \mu, \sigma)}{S_\epsilon(e_0; \mu, \sigma)} = \frac{1 - F\left(\frac{e_j - \mu}{\sigma}\right)}{1 - F\left(\frac{e_0 - \mu}{\sigma}\right)} = \frac{1 - F(z_0 + \Delta_j/\sigma)}{1 - F(z_0)} := Q_j(z_0, \sigma).$$

For fixed σ and Δ_j, $Q_j(z_0, \sigma)$ is strictly decreasing in z_0. For fixed z_0 and Δ_j, $Q_j(z_0, \sigma)$ is strictly decreasing in $1/\sigma$ (equivalently, strictly increasing in σ), since $\Delta_j > 0$ and $1 - F(\cdot)$ is strictly decreasing. Hence, the map

$$(\mu, \sigma) \longmapsto \left(\frac{S_\epsilon(e_1; \mu, \sigma)}{S_\epsilon(e_0; \mu, \sigma)}, \frac{S_\epsilon(e_2; \mu, \sigma)}{S_\epsilon(e_0; \mu, \sigma)} \right),$$

is injective, i.e., the normal family is tail-ratio identifiable.

Exponential. Suppose $\epsilon \sim \mathrm{Exp}(\lambda)$, with parameter $\lambda > 0$, so $p = 1$. The survival function is

$$S_\epsilon(e; \lambda) = \Pr(\epsilon > e) = \exp(-\lambda e).$$

Fix any $0 < e_0 < e_1$ and define

$$\Delta_1 := e_1 - e_0 > 0.$$

Then

$$\frac{S_\epsilon(e_1; \lambda)}{S_\epsilon(e_0; \lambda)} = \frac{\exp(-\lambda e_1)}{\exp(-\lambda e_0)} = \exp\left(-\lambda \Delta_1\right) := Q_1(\lambda).$$

For fixed $\Delta_1 > 0$, $Q_1(\lambda)$ is strictly decreasing in λ, since the exponential function $\exp(-\lambda\Delta_1)$ decreases as λ increases. Hence, the map

$$\lambda \longmapsto \frac{S_\epsilon(e_1; \lambda)}{S_\epsilon(e_0; \lambda)},$$

is injective, i.e., the exponential family is tail-ratio identifiable.

Lognormal. Suppose $\epsilon \sim \mathrm{LN}(m, \sigma^2)$, with parameters (m, σ), so $p = 2$. The survival function is

$$S_\epsilon(e; m, \sigma) = 1 - F\left(\frac{\log(e) - m}{\sigma}\right),$$

where $F(\cdot)$ is the standard normal CDF. Fix any $0 < e_0 < e_1 < e_2$ and define

$$z_0 := \frac{\log(e_0) - m}{\sigma}, \qquad \Delta_j^{\log} := \log\left(\frac{e_j}{e_0}\right) > 0 \;\; (j = 1, 2).$$

Then

$$\frac{S_\epsilon(e_j;m,\sigma)}{S_\epsilon(e_0;m,\sigma)} = \frac{1-F\left(\frac{\log(e_j)-m}{\sigma}\right)}{1-F\left(\frac{\log(e_0)-m}{\sigma}\right)} = \frac{1-F\left(z_0+\frac{\Delta_j^{\log}}{\sigma}\right)}{1-F(z_0)} := Q_j(z_0,\sigma).$$

For fixed σ and $\Delta_j^{\log}$, $Q_j(z_0,\sigma)$ is strictly decreasing in z_0. For fixed z_0 and $\Delta_j^{\log}$, $Q_j(z_0,\sigma)$ is strictly decreasing in $1/\sigma$ (equivalently, strictly increasing in σ), since $\Delta_j^{\log} > 0$ and $1-F(\cdot)$ is strictly decreasing. Hence, the map

$$(m,\sigma) \longmapsto \left(\frac{S_\epsilon(e_1;m,\sigma)}{S_\epsilon(e_0;m,\sigma)}, \frac{S_\epsilon(e_2;m,\sigma)}{S_\epsilon(e_0;m,\sigma)}\right)$$

is injective, i.e., the lognormal family is tail-ratio identifiable.

Pareto. Suppose $\epsilon \sim \mathrm{Par}(\alpha,k)$, with parameters (α,k), so $p=2$. The survival function for $t>k$ is

$$S_\epsilon(t;\alpha,k) = \left(\frac{k}{t}\right)^\alpha.$$

Fix any $0<e_0<e_1$ and define

$$\Delta_j := e_j - e_0 > 0 \ \ (j=1,2).$$

Then the ratio of survival functions is

$$\frac{S_\epsilon(e_j;\alpha,k)}{S_\epsilon(e_0;\alpha,k)} = \left(\frac{e_0}{e_j}\right)^\alpha = \left(\frac{e_0}{e_0+\Delta_j}\right)^\alpha := Q_j(\alpha).$$

For fixed α and Δ_j, $Q_j(\alpha)$ is strictly decreasing in α, since $\Delta_j > 0$. Hence, the map

$$\alpha \longmapsto \left(\frac{S_\epsilon(e_1;\alpha,k)}{S_\epsilon(e_0;\alpha,k)}, \frac{S_\epsilon(e_2;\alpha,k)}{S_\epsilon(e_0;\alpha,k)}\right)$$

is injective in α, but not in k, since the scale parameter k cancels out in the ratio. Thus, the pareto family is tail-ratio identifiable in α but not in k.

Mixture of Normals. Suppose ϵ is a mixture of two normal distributions with parameters (μ_1,σ_1^2) and (μ_2,σ_2^2), with mixture weight $\pi \in [0,1]$, so $p=5$. Specifically,

$$f_\epsilon(e) = \pi \cdot \phi(e;\mu_1,\sigma_1^2) + (1-\pi)\cdot\phi(e;\mu_2,\sigma_2^2),$$

where $\phi(e;\mu,\sigma^2)$ is the normal density function with mean μ and variance σ^2.

The survival function is the weighted sum of the survival functions of the two normal distributions,

$$S_\epsilon(e;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2) = \pi\cdot S_{\mathcal{N}}(e;\mu_1,\sigma_1) + (1-\pi)\cdot S_{\mathcal{N}}(e;\mu_2,\sigma_2),$$

where

$$S_{\mathcal{N}}(e;\mu,\sigma) = 1 - F\left(\frac{e-\mu}{\sigma}\right),$$

and $F(\cdot)$ is the standard normal CDF. Fix any $0 < e_0 < e_1 < \cdots < e_5$ and define

$$\Delta_j := e_j - e_0 > 0 \ \ (j = 1, \ldots, 5).$$

The ratio of survival functions is

$$\frac{S_\epsilon(e_j;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2)}{S_\epsilon(e_0;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2)} = \frac{\pi \cdot S_{\mathcal{N}}(e_j;\mu_1,\sigma_1) + (1-\pi)\cdot S_{\mathcal{N}}(e_j;\mu_2,\sigma_2)}{\pi \cdot S_{\mathcal{N}}(e_0;\mu_1,\sigma_1) + (1-\pi)\cdot S_{\mathcal{N}}(e_0;\mu_2,\sigma_2)}.$$

For fixed $\mu_1, \sigma_1, \mu_2, \sigma_2$, and Δ_j, the above ratio depends on the mixture weight π. Hence, the map

$$\pi \longmapsto \left(\frac{S_\epsilon(e_1;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2)}{S_\epsilon(e_0;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2)}, \frac{S_\epsilon(e_2;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2)}{S_\epsilon(e_0;\pi,\mu_1,\sigma_1,\mu_2,\sigma_2)}\right)$$

is injective in π, but not in $\mu_1, \sigma_1, \mu_2, \sigma_2$. Thus, the mixture of Normals is tail-ratio identifiable in π but not in $\mu_1, \sigma_1, \mu_2, \sigma_2$.

Proof of Proposition B.10. Let $t \in \{1, \ldots, T-1\}$ and $d \in \mathcal{D}$. Define the two human capital functions $\underline{h}_d(Z_{n,t}) := h_d(Z_{n,t}, \underline{\eta})$ and $\bar{h}_d(Z_{n,t}) := h_d(Z_{n,t}, \bar{\eta})$, corresponding to a low and high realized performance signals, respectively. Then, writing the expectation operator explicitly, we have

$$y(d, Z_{n,t}, P_{n,t}) = \underline{h}_d(Z_{n,t}) + [\bar{h}_d(Z_{n,t}) - \underline{h}_d(Z_{n,t})]P_{n,t}. \tag{93}$$

Now evaluate (93) at two realizations (z, p_1) and (z, p_2) of $(Z_{n,t}, P_{n,t})$ with $p_1 \neq p_2$:

$$\begin{aligned} y(d, z, p_1) &= \underline{h}_d(z) + [\bar{h}_d(z) - \underline{h}_d(z)]p_1, \\ y(d, z, p_2) &= \underline{h}_d(z) + [\bar{h}_d(z) - \underline{h}_d(z)]p_2. \end{aligned}$$

Subtracting the second equation from the first yields

$$y(d, z, p_1) - y(d, z, p_2) = [\bar{h}_d(z) - \underline{h}_d(z)] \times [p_1 - p_2].$$

We can solve for the difference $\Delta_{d,t}(z) := \bar{h}_d(z) - \underline{h}_d(z)$ as

$$\Delta_{d,t}(z) = \frac{y(d, z, p_1) - y(d, z, p_2)}{p_1 - p_2}. \tag{94}$$

Substituting (94) into (93) evaluated at $P_{n,t} = p_1$ gives

$$\underline{h}_d(z) = y(d, z, p_1) - \Delta_{d,t}(z)p_1. \tag{95}$$

On the right side of (95), $y(d, z, p_1)$ and $y(d, z, p_2)$ are identified by Assumption (i). Thus, $\Delta_{d,t}(z)$ is also known, so $\underline{h}_d(z)$ is identified. Finally, $\bar{h}_d(z) = \underline{h}_d(z) + \Delta_{d,t}(z)$ is identified.

Online Supplementary Material

F Omitted Model Details

We present here omitted details and proofs about the characterization of the model equilibrium.

F.1 Efficiency of Equilibrium in the One-Job Case

Although our model is one of imperfect competition in the labor market as firms are differentiated in their technologies, the equilibrium does not need to be inefficient. Intuitively, wages are determined by a second-price auction mechanism for a worker's services, which, as any such mechanism, has desirable efficiency properties. However, inefficiencies can arise because through employment, firms allow workers to acquire human capital and generate information about their ability, but neither human capital nor information about ability can be priced separately from a worker's labor services. Put differently, the pricing mechanism is not rich enough to align a worker and *all* firms' incentives to produce output, human capital, and information.

Yet, we are able to establish that equilibrium is efficient provided that the process of accumulation of human capital and information and the risk of exogenous separation are similar enough across firms that workers are able to internalize through their employment decisions the willingness of the third-best firm, fourth-best firm, and so on to employ them.[53] This result extends to the case in which firms consist each of many jobs, as shown in Appendix F.2 of the Online Supplementary Material.

Proposition F.1 (Efficiency)**.** *For any $\overline{D}$ such that $|\mathcal{D}| \leq \overline{D}$, whenever the process of human capital acquisition, the probabilities $\alpha(H_{n,1}, d, e_n)$ and $\beta(H_{n,1}, d, e_n)$ of high output for a high-ability worker and a low-ability worker, respectively, and the probability of exogenous separation $\varsigma(I_n^{t-1}, d)$ are sufficiently similar between any two firms d and d' for each realization of $H_{n,1}$, e_n, and I_n^{t-1}, where the required degree of similarity depends on $\overline{D}$, the equilibrium is efficient.*

Proof: Consider first the case in which $|\mathcal{D}| = 2$. The proof of the result in this case is an extension of Theorem 2 in Bergemann and Välimaki (1996). Although their result applies to a setting in which a worker's ability, unlike our case, is independent across firms—namely, a consumer's taste in their framework of two firms competing for a consumer is unknown to all and independent across firms—no steps of the argument is reliant on this assumption. We note that for this case, the assumption that the human capital process is sufficiently similar across firms can be dispensed with. When $|\mathcal{D}| > 2$, the assumptions of the proposition ensure that firms ranked third-best, fourth-best, and so on by the worker can always be chosen to have technologies that are sufficiently similar to that of the first- and second-best firms. Hence, the same argument as that under the case of $|\mathcal{D}| = 2$ applies. $\square$

F.2 Multi-Job-Firm Case: Equilibrium, Efficiency, and Identification

We now argue that the probability of a worker's employment at any firm in the market and allocation to any job at the employing firm admits a pseudo-programming characterization, which only requires

[53] It is easy to construct counterexamples to efficiency when the market consists of three or more firms or each firm consists of multiple jobs, if the human capital or informational process are unrestricted across firms.

knowledge of a worker's wage in equilibrium. For this result, we consider the general case of our model in which firms consist of finitely many jobs indexed by j—without loss, we assume the set of jobs to be the same across firms. In this section, we denote by f a generic firm, by d the equilibrium employing (first-best) firm, and by d' the equilibrium second-best firm.

In this more general case, the definition of equilibrium extends straightforwardly; we note here only the important differences. In particular, we denote by $(w_{n,t,f}, j_{n,t,f})$, with $w_{n,t,f} := w_f(s_{n,t}, \epsilon_{n,t})$ and $j_{n,t,f} := j_f(s_{n,t}, \epsilon_{n,t})$, the wage and job offer strategy of a generic firm f and by $\{w_{n,t,f}, j_{n,t,f}\}_{f\in\mathcal{D}}$ the collection of all offer strategies. We denote by $l_{n,t,f} := l_f(s_{n,t}, \epsilon_{n,t}, \{w_{n,t,f}, j_{n,t,f}\}_{f\in\mathcal{D}})$ the acceptance strategy of worker n for firm f's offer—an indicator function, taking value one if f is the employing firm and zero otherwise at a given state—and by $\{l_{n,t,f}\}_{f\in\mathcal{D}}$ the collection of all acceptance strategies. Given firms' strategies, worker n's strategy satisfies

$$\tilde{W}(s_{n,t}, \epsilon_{n,t}, \{w_{n,t,f}, j_{n,t,f}\}_{f\in\mathcal{D}}) = \max_{\{l_f\}_{f\in\mathcal{D}}} \sum_{f\in\mathcal{D}} l_f \times \Big\{ w_{n,t,f} + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j_{n,t,f})]$$

$$\times \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[\tilde{W}(s_{n,t+1}, \epsilon_{n,t+1}, \{w_{n,t+1,f}, j_{n,t+1,f}\}_{f\in\mathcal{D}}) | s_{n,t}, f, j_{n,t,f}\Big] dF_{\epsilon_{n,t+1}} \Big\}, \tag{96}$$

where $j_{n,t,f}$ affects the exogenous separation rate $\varsigma(\cdot)$ and conditions the law of motion of the state, as is apparent from the conditional expectation in the last line of (96). Given worker n's strategy and its competitors' strategies, firm f's strategy satisfies

$$\Pi_f(s_{n,t}, \epsilon_{n,t}) = \max_{w,j} \Big(l_{n,t,f} \times \Big\{ y(f, s_{n,t}, j, \epsilon_{n,t}(f,j)) - w$$

$$+ \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[\Pi_f(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, f, j\Big] dF_{\epsilon_{n,t+1}} \Big\} \tag{97}$$

$$+ \sum_{f'\in\mathcal{D}\setminus\{f\}} l_{n,t,f'} \Big\{ \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f', j_{n,t,f'})] \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[\Pi_d(s_{n,t+1}, \epsilon_{n,t+1}) | s_{n,t}, f', j_{n,t,f'}\Big] dF_{\epsilon_{n,t+1}} \Big\} \Big),$$

which shows how each firm now chooses a wage and a job for the worker, taking into account the wage and job offers of all other firms—see the last line of (97). The cautious equilibrium refinement, namely, condition (*iv*) of our equilibrium definition, requires that if firm $d \in \mathcal{D}$ employs worker n at state $(s_{n,t}, \epsilon_{n,t})$, then for any other firm $f \in \mathcal{D}$,

$$\delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j_{n,t,d})] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_f(\cdot) | s_{n,t}, d, j_{n,t,d}] dF_{\epsilon_{n,t+1}} = \max_{w,j} \Big\{ y(f, s_{n,t}, j, \epsilon_{n,t}(f,j))$$

$$- w + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_f(\cdot) | s_{n,t}, f, j] dF_{\epsilon_{n,t+1}} \Big\}. \tag{98}$$

Two immediate implications of equilibrium are as follows. First, worker n must be indifferent between the offers of the first- and second-best firms, as discussed in the proof of Proposition 1, and weakly prefer either of these two offers to any other offers. Second, for each firm f, maximizing profits is equivalent to maximizing the sum of its own value of profits and worker n's value of wages from accepting its offer. We denote this value by $\tilde{W}(s_{n,t}, \epsilon_{n,t}|f)$, which is the value $\tilde{W}(s_{n,t}, \epsilon_{n,t}, \{w_{n,t,f'}, j_{n,t,f'}\}_{f'\in\mathcal{D}})$ conditional on the worker accepting firm f's offer, after suppressing the notation for firms' offers since they are function of the state $(s_{n,t}, \epsilon_{n,t})$.

That maximizing profits is equivalent to maximizing match surplus follows from: *(i)* any firm f different from the first-best firm d being indifferent between *not employing* the worker—in which case its payoff is $\Pi_f(s_{n,t}, \epsilon_{n,t}|d)$, the left side of (98)—and *employing* the worker—in which case its payoff is $\Pi_f(s_{n,t}, \epsilon_{n,t}|f)$, the right side of (98); and *(ii)* worker n weakly preferring employment at the first-best firm d to employment at any other firm f, ($\tilde{W}(s_{n,t}, \epsilon_{n,t}|d) \geq \tilde{W}(s_{n,t}, \epsilon_{n,t}|f)$). Summing the two conditions $\Pi_f(s_{n,t}, \epsilon_{n,t}|d) = \Pi_f(s_{n,t}, \epsilon_{n,t}|f)$ and $\tilde{W}(s_{n,t}, \epsilon_{n,t}|d) \geq \tilde{W}(s_{n,t}, \epsilon_{n,t}|f)$ term by term indeed gives that

$$\Pi_f(s_{n,t}, \epsilon_{n,t}|d) + \tilde{W}(s_{n,t}, \epsilon_{n,t}|d) \geq \Pi_f(s_{n,t}, \epsilon_{n,t}|f) + \tilde{W}(s_{n,t}, \epsilon_{n,t}|f),$$

so maximizing profits is equivalent to maximizing match surplus for any firm f different from the first-best firm d. As for the first-best firm d, since this firm must prefer employing to not employing the worker, that is, $\Pi_d(s_{n,t}, \epsilon_{n,t}|d) \geq \Pi_d(s_{n,t}, \epsilon_{n,t}|f)$, an analogous argument applies.

Thus, at any state at which a firm f is the employing firm d, we can rewrite the firm problem as

$$V_d(s_{n,t}, \epsilon_{n,t}) = \max_j \left\{ y(d, s_{n,t}, j, \epsilon_{n,t}(d,j)) + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_d(\cdot)|s_{n,t}, d, j] dF_{\epsilon_{n,t+1}} \right\}, \tag{99}$$

whereas at any state at which firm f is *not* the employing firm d, we can rewrite the firm problem as

$$\begin{aligned} V_f(s_{n,t}, \epsilon_{n,t}) = \max \Bigg(& w_{n,t,d} + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j_{n,t,d})] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_f(\cdot)|s_{n,t}, d, j_{n,t,d}] dF_{\epsilon_{n,t+1}}, \\ & \max_j \left\{ y(f, s_{n,t}, j, \epsilon_{n,t}(f,j)) + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_f(\cdot)|s_{n,t}, f, j] dF_{\epsilon_{n,t+1}} \right\} \Bigg), \end{aligned} \tag{100}$$

where $w_{n,t,d}$ is worker n's wage at the first-best firm d. At each state, the solution to the subproblem in (99) determines worker n's probability of job assignment at the employing firm (firm d) conditional on employment at it, whereas the solution to the subproblem in (100) determines worker n's probability of employment at firms different from the first-best firm (any firm $f \neq d$) together with their job offer. By (99) and (100), the best-response problem of any firm f can be represented as

$$V_f(s_{n,t}, \epsilon_{n,t}) = \begin{cases} \max_j \Big\{ y(f, s_{n,t}, j, \epsilon_{n,t}(f,j)) \\ \quad + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_f(\cdot)|s_{n,t}, f, j] dF_{\epsilon_{n,t+1}} \Big\} \text{ if } f = d \\ \max \Big(w_{n,t,d} + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j_{n,t,d})] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_f(\cdot)|s_{n,t}, d, j_{n,t,d}] dF_{\epsilon_{n,t+1}} \;, \\ \quad \max_j \Big\{ y(f, s_{n,t}, j, \epsilon_{n,t}(f,j)) \\ \quad + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[V_f(\cdot)|s_{n,t}, f, j] dF_{\epsilon_{n,t+1}} \Big\} \Big) \text{ if } f \neq d \end{cases}$$

which is a pseudo-planning problem up to the one-period match surplus value of firm f when it is not the employing firm, given by the wage $w_{n,t,d}$ paid to worker n by the first-best firm. The collection of all such match surplus values for all firms determines worker n employing firm and assigned job.

By the discrete choice logic in Kristensen et al. (2015), it is easy to see that up to δ and knowledge of $y(f, s_{n,t}, j, \epsilon_{n,t}(f,j))$ for one j, say $\tilde{j}$, the remaining $y(f, s_{n,t}, j), \epsilon_{n,t}(f,j)$, $j \neq \tilde{j}$, are identified once the remaining primitives are known—see Proposition B.8 for the formal statement of this result.

In terms of the efficiency of equilibrium, recall that in equilibrium, worker n must be indifferent between the first-best (d) and second-best (d') firm and weakly prefer their offers to any other; the employing firm must prefer employing the worker to not employing the worker; and each non-employing firm must prefer not employing the worker to employing the worker. Respectively, in equilibrium, for worker n to be employed at firm d, it must be that

$$\begin{aligned} w_{n,t,d} + &\delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j_{n,t,d})] \\ &\times \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[W(s_{n,t+1}, \epsilon_{n,t+1}, \{w_{n,t+1,f'}, j_{n,t+1,f'}\}_{f' \in \mathcal{D}}) | s_{n,t}, d, j_{n,t,d}\Big] dF_{\epsilon_{n,t+1}} \\ &\geq w_{n,t,f} + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j_{n,t,f})] \\ &\times \int_{\epsilon_{n,t+1}} \mathbb{E}\Big[W(s_{n,t+1}, \epsilon_{n,t+1}, \{w_{n,t+1,f'}, j_{n,t+1,f'}\}_{f' \in \mathcal{D}}) | s_{n,t}, f, j_{n,t,f}\Big] dF_{\epsilon_{n,t+1}}, \end{aligned}$$

with strict equality when f is the second-best firm d'; for the employing firm d, it must be that

$$\begin{aligned} &\max_{w,j} \Big\{ y(d, s_{n,t}, j, \epsilon_{n,t}(d,j)) - w + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_d(\cdot) | s_{n,t}, d, j] dF_{\epsilon_{n,t+1}} \Big\} \\ &\geq \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j_{n,t,f})] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_d(\cdot) | s_{n,t}, f, j_{n,t,f}] dF_{\epsilon_{n,t+1}}; \end{aligned}$$

and for any other firm $f \neq d$, it must be that

$$\begin{aligned} &\delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, d, j_{n,t,d})] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_f(\cdot) | s_{n,t}, d, j_{n,t,d}] dF_{\epsilon_{n,t+1}}, \\ &\geq \max_{w,j} \Big\{ y(f, s_{n,t}, j, \epsilon_{n,t}(f,j)) - w + \delta[1 - \varsigma(H_{n,1}, I_n^{t-1}, f, j)] \int_{\epsilon_{n,t+1}} \mathbb{E}[\Pi_f(\cdot) | s_{n,t}, f, j] dF_{\epsilon_{n,t+1}} \Big\}. \end{aligned}$$

By summing term by term these three sets of inequalities and recalling the definition of (social) welfare $S(s_{n,t}, \epsilon_{n,t})$ as the sum of values of worker n and all firms, namely, $S(\cdot) = W(\cdot) + \sum_f \Pi_f(\cdot)$, it is immediate that if the law of motion of the state and the risk of exogenous separation are similar enough across firms and jobs or firms' continuation profits are independent enough of the state (as under perfect competition when they are zero), then any MPE is efficient. Intuitively, in this case, the values of all players can be "passed through" the relevant maximization operators so that maximizing match surplus is equivalent to maximizing total surplus. We note that this argument applies to all MPEs, even those that do not satisfy the cautious requirement. In the general case, however, this argument does not apply, and it is easy to see that the equilibrium need not be efficient (see also Bergemann and Välimaki, 2006). □

G Additional Identification Results

G.1 Micro-Foundation of Assumption (ii) of Proposition E.1

Lemma G.1 shows that if the productivity shocks are "sufficiently independent," then Assumption (ii) of Proposition E.1 holds. This corresponds to Corollary 4.1 of D'Haultfoeuille and Maurel (2013).

Lemma G.1 (Moderate Dependence—Simplified Wage Equation (11))**.** *Let Assumption (i) of Proposition E.1 hold and, without loss,* $\mathcal{D} \coloneqq \{0, 1\}$*. For some* $q_t(1) \in (0, 1]$*, let*

$$\lim_{u\to\infty} \Pr\big(\epsilon_{n,t}(0) \leq \epsilon_{n,t}(1) + a \,\big|\, \epsilon_{n,t}(1) \geq u\big) = q_t(1) \qquad \textit{for all } a \in \mathbb{R}. \tag{101}$$

Then, Assumption (ii) of Proposition E.1 holds in that $\lim_{w\to\infty} \Pr\big(D_{n,t} = 1 \mid s_{n,t} = s,\ w_{n,t}(1) \geq w\big) = q_t(1)$ *for every* $s \in \mathcal{S}_t$*. Moreover, if* $\epsilon_{n,t}(0)$ *and* $\epsilon_{n,t}(1)$ *are independent, then* $q_t(1) = 1$*.*

Proof: For simplicity, assume that the equilibrium is efficient. All the steps can be generalized to the non-efficient case by leveraging the pseudo-planner problem representation in Appendix F.2 of the Online Supplementary Material. Then. for any $s \in \mathcal{S}_t$ and $w \in \mathbb{R}$,

$$\begin{aligned}
&\Pr(D_{n,t} = 1 \,\big|\, s_{n,t} = s,\ w_{n,t}(1) \geq w) \\
&= \Pr\big(\Upsilon(1, s) + \epsilon_{n,t}(1) \ \geq\ \Upsilon(0, s) + \epsilon_{n,t}(0) \,\big|\, s_{n,t} = s,\ \varphi(1, s) + \epsilon_{n,t}(1) \geq w\big) \\
&= \Pr\big(\epsilon_{n,t}(0) \leq \epsilon_{n,t}(1) + \Upsilon(1, s) - \Upsilon(0, s) \,\big|\, \epsilon_{n,t}(1) \geq w - \varphi(1, s)\big),
\end{aligned} \tag{102}$$

where $\Upsilon(d, s) + \epsilon_{n,t}(d)$ is the expected present discounted value of output for firm d in state s after productivity shocks have realised.

Set $u \coloneqq w - \varphi(1, s)$, so $w \to \infty$ iff $u \to \infty$. Applying (101) with $a = \Upsilon(1, s) - \Upsilon(0, s)$ gives

$$\lim_{w\to\infty} \Pr\big(\epsilon_{n,t}(0) \leq \epsilon_{n,t}(1) + a(x) \,\big|\, \epsilon_{n,t}(1) \geq w - \varphi(1, s)\big) = q_t(1). \tag{103}$$

By (102), (103) is precisely

$$\lim_{w\to\infty} \Pr(D_{n,t} = 1 \mid s_{n,t} = s,\ w_{n,t}(1) \geq w) = q_t(1),$$

which is Assumption (ii) of Proposition E.1.

Now, suppose $\epsilon_{n,t}(1)$ and $\epsilon_{n,t}(0)$ are independent. Then, for any $a \in \mathbb{R}$ and $u \in \mathbb{R}$,

$$\begin{aligned}
\Pr(\epsilon_{n,t}(0) \leq \epsilon_{n,t}(1) + a \mid \epsilon_{n,t}(1) \geq u) &= \mathbb{E}\Big[\Pr(\epsilon_{n,t}(0) \leq \epsilon_{n,t}(1) + a \mid \epsilon_{n,t}(1)) \,\Big|\, \epsilon_{n,t}(1) \geq u\Big] \\
&= \mathbb{E}\Big[F_{\epsilon_{n,t}(0)}\big(\epsilon_{n,t}(1) + a\big) \,\Big|\, \epsilon_{n,t}(1) \geq u\Big],
\end{aligned} \tag{104}$$

where $F_{\epsilon_{n,t}(0)}$ is the CDF of $\epsilon_{n,t}(0)$ and the second line uses independence between $\epsilon_{n,t}(0)$ and $\epsilon_{n,t}(1)$. Since $F_{\epsilon_{n,t}(0)}$ is nondecreasing,

$$F_0(u + a) \ \leq\ F_{\epsilon_{n,t}(0)}\big(\epsilon_{n,t}(1) + a\big) \ \leq\ 1 \quad \text{on the event } \{\epsilon_{n,t}(1) \geq u\}.$$

Taking conditional expectations yields the bounds

$$F_{\epsilon_{n,t}(0)}(u+a) \;\leq\; \mathbb{E}\Big[F_{\epsilon_{n,t}(0)}\big(\epsilon_{n,t}(1)+a\big)\,\Big|\,\epsilon_{n,t}(1)\geq u\Big] \;\leq\; 1.$$

Letting $u \to \infty$ and using $\lim_{\tau\to\infty} F_{\epsilon_{n,t}(0)}(\tau) = 1$, we conclude that

$$\lim_{u\to\infty} \Pr(\epsilon_{n,t}(0) \leq \epsilon_{n,t}(1) + a \mid \epsilon_{n,t}(1) \geq u) = 1,$$

so $q_t(1) = 1$. $\square$

Remark. Suppose that $\epsilon_{n,t}(1)$ and $\epsilon_{n,t}(0)$ are jointly normal—or lognormal. If $\mathrm{cov}(\epsilon_{n,t}(1), \epsilon_{n,t}(0)) < \mathrm{Var}(\epsilon_{n,t}(1))$—or if $\mathrm{cov}(\log(\epsilon_{n,t}(1)), \log(\epsilon_{n,t}(0))) < \mathrm{Var}(\log(\epsilon_{n,t}(1)))$—then (101) holds. Similar "sufficient independence" conditions can be given for many other parametric families, including both thin-tailed (for instance, Normal, Exponential, Gamma, Logistic, Gumbel) and fat-tailed (for instance, Pareto, Cauchy, Burr, Fréchet, log-logistic, and lognormal) distributions.

G.2 Proposition E.1 with Bounded Support

Proposition G.1 establishes identification of the deterministic wage components entering our simplified wage equation (11) in the case where the potential wages $w_{n,t}(d) \mid s_{n,t} = s$ and the observed, selected wages $w_{n,t} \mid (D_{n,t} = d, s_{n,t} = s)$ have different right endpoints. The identification result retains the spirit of Proposition E.1, but extra care is needed in taking limits because the two endpoints differ. We further show that, when finite, the lower and upper endpoints of the potential wages $w_{n,t}(d) \mid s_{n,t} = s$ and shock $\epsilon_{n,t}(d)$ are nonparametrically identified (Corollary G.1).

Proposition G.1 (Deterministic Wage—Simplified Wage Equation (11) with Bounded Supports)**.** *Let $t \in \{1,\dots,T\}$ and $d \in \mathcal{D}$. Assume that the conditional wage distribution $\Pr\big(w_{n,t} \leq w \,\big|\, D_{n,t} = d,\ s_{n,t} = s\big)$ is identified for each $w \in \mathbb{R}$ and $s \in \mathcal{S}_t$ (see Proposition B.5 for sufficient conditions), and that the conditional choice probability $\Pr(D_{n,t} = d \mid s_{n,t} = s)$ is identified for each $s \in \mathcal{S}_t$ (see Proposition B.4 for sufficient conditions). Moreover, assume:*

- *(i) (Supports.) For each $s \in \mathcal{S}_t$, $\omega_t(d,s) := \sup\{u : \Pr(w_{n,t}(d) \leq u \mid s_{n,t} = s) < 1\} \leq \infty$ and $\omega_t^{\mathrm{obs}}(d,s) := \sup\{u : \Pr(w_{n,t} \leq u \mid D_{n,t} = d, s_{n,t} = s) < 1\} < \omega_t(d,s)$, with $\omega_t(d,s)$ and $\omega_t^{\mathrm{obs}}(d,s)$ potentially unknown.*
- *(ii) (Relative Tail Decay.) For each $s \in \mathcal{S}_t$, define $r_{d,s,t}(u) := \Pr\big(D_{n,t} = 1 \mid s_{n,t} = s, w_{n,t}(d) > Q_{w_{n,t}(d)|s_{n,t}=s}(u)\big)$, $u \in (0,1)$. There exists an (unknown) constant $q_t(d) \in (0,\infty)$ such that, for every $s \in \mathcal{S}_t$ and a fixed reference $\bar{s} \in \mathcal{S}_t$, $\lim_{u\to 1} \frac{r_{d,s,t}(u)}{r_{d,\bar{s},t}(u)} = q_t(d)$.*
- *(iii) (Tail Regularity.) For each $s \in \mathcal{S}_t$, there exist (unknown) thresholds $a_t(d,s) < \infty$ and $a_t^{\mathrm{obs}}(d,s) < \omega_t^{\mathrm{obs}}(d,s)$ such that $F_{w_{n,t}(d)|s_{n,t}=s}$ and $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ are* continuous and strictly increasing *on $(a_t(d,s),\infty)$ and $(a_t^{\mathrm{obs}}(d,s), \omega_t^{\mathrm{obs}}(d,s))$, respectively. Moreover, $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ is* continuous at the endpoint*:* $\lim_{w\to\omega_t^{\mathrm{obs}}(d,s)} F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) = 1$.
- *(iv) (Normalization.) There exists a known $\bar{s} \in \mathcal{S}_t$ with $\varphi(d,\bar{s}) = 0$.*

For each $s \in \mathcal{S}_t$, let $\{\tau^{(k)}_{d,\bar{s},t}\}_{k\geq 1} \subset (0,1)$ be any sequence with $\tau^{(k)}_{d,\bar{s},t} \to 1$ as $k \to \infty$ and define the sequence $\{\tau^{(k)}_{d,s,t}\}_{k\geq 1} \subset (0,1)$ where

$$\tau^{(k)}_{d,s,t} \coloneqq 1 - \frac{\Pr(D_{n,t} = d \mid s_{n,t} = \bar{s})}{\Pr(D_{n,t} = d \mid s_{n,t} = s)} \left(1 - \tau^{(k)}_{d,\bar{s},t}\right).$$

Then,

$$\lim_{k\to\infty} \left[Q_{w_{n,t} \mid D_{n,t}=d, s_{n,t}=s}\left(\tau^{(k)}_{d,s,t}\right) \; - \; Q_{w_{n,t} \mid D_{n,t}=d, s_{n,t}=\bar{s}}\left(\tau^{(k)}_{d,\bar{s},t}\right) \right] \; = \varphi(d,s). \tag{105}$$

Hence, $\varphi(d,s)$ is identified for each $s \in \mathcal{S}_t$.

Proof: To facilitate reading, we divide the proof into steps. Without loss, consider firm $d = 1$.
Step 1: Bayes rule. Let $s \in \mathcal{S}_t$. For any real w, Bayes' rule gives

$$\begin{aligned} S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(w) &= \frac{\Pr\big(w_{n,t}(1) > w \mid s_{n,t} = s\big) \; \Pr\big(D_{n,t} = 1 \mid s_{n,t} = s, \; w_{n,t}(1) > w\big)}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)} \\ &= S_{w_{n,t}(1)|s_{n,t}=s}(w) \; \frac{\Pr\big(D_{n,t} = 1 \mid s_{n,t} = s, \; w_{n,t}(1) > w\big)}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)}. \end{aligned} \tag{106}$$

Define

$$r_{1,s,t}(u) \; \coloneqq \; \Pr\big(D_{n,t} = 1 \mid s_{n,t} = s, \; w_{n,t}(1) > Q_{w_{n,t}(1)|s_{n,t}=s}(u)\big), \qquad u \in (0,1).$$

Evaluating (106) at $w = Q_{w_{n,t}(1)|s_{n,t}=s}(u)$ yields

$$S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}\big(Q_{w_{n,t}(1)|s_{n,t}=s}(u)\big) \; = \; (1-u) \, \frac{r_{1,s,t}(u)}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)}, \qquad u \in (0,1). \tag{107}$$

Step 2: Behavior of the composed survival near the observed endpoint. By Assumption (iii), there exist thresholds $a_t(1,s) < \omega_t(1,s)$ and $a^{\text{obs}}_t(1,s) < \omega^{\text{obs}}_t(1,s)$ such that $F_{w_{n,t}(d)|s_{n,t}=s}$ is continuous and strictly increasing on $(a_t(1,s), \omega_t(1,s))$, and $F_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}$ is continuous and strictly increasing on $(a^{\text{obs}}_t(1,s), \omega^{\text{obs}}_t(1,s))$. Define

$$u^*_{1,s,t} \coloneqq F_{w_{n,t}(1)|s_{n,t}=s}\big(a_t(1,s)\big), \qquad \tau^*_{1,s,t} \coloneqq F_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}\big(a^{\text{obs}}_t(1,s)\big),$$

so $Q_{w_{n,t}(1)|s_{n,t}=s} : (u^*_{1,s,t}, 1) \to (a_t(1,s), \omega_t(1,s))$ and $Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s} : (\tau^*_{1,s,t}, 1) \to (a^{\text{obs}}_t(1,s), \omega^{\text{obs}}_t(1,s))$ are strictly increasing. Because $\omega^{\text{obs}}_t(1,s) < \omega_t(1,s)$, set

$$\bar{u}_{1,s,t} \coloneqq \sup\left\{u \in (u^*_{1,s,t}, 1) : \; Q_{w_{n,t}(1)|s_{n,t}=s}(u) < \omega^{\text{obs}}_t(1,s)\right\} \in (u^*_{1,s,t}, 1).$$

Then, $Q_{w_{n,t}(1)|s_{n,t}=s}(u) \to \omega^{\text{obs}}_t(1,s)$ as $u \to \bar{u}_{1,s,t}$. Since $Q_{w_{n,t}(1)|s_{n,t}=s}(u)$ is increasing and $\omega^{\text{obs}}_t(1,s)$ is finite, there exists $\tilde{u}_{1,s,t} \in (u^*_{1,s,t}, \bar{u}_{1,s,t})$ such that $Q_{w_{n,t}(1)|s_{n,t}=s}(u) \in (a^{\text{obs}}_t(1,s), \omega^{\text{obs}}_t(1,s))$ for all $u \in (\tilde{u}_{1,s,t}, \bar{u}_{1,s,t})$. On that interval, the map

$$u \; \mapsto \; S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}\big(Q_{w_{n,t}(1)|s_{n,t}=s}(u)\big),$$

is a composition of a continuous, strictly increasing function (the potential quantile) with a continuous, strictly decreasing function (the observed survival on its tail), hence it is continuous and strictly decreasing on $(\tilde{u}_{1,s,t}, \bar{u}_{1,s,t})$. By the endpoint continuity in Assumption (iii),

$$\lim_{w \to \omega_t^{\text{obs}}(1,s)} F_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(w) = 1,$$

and therefore

$$\lim_{u \to \bar{u}_{1,s,t}} S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}\big(Q_{w_{n,t}(1)|s_{n,t}=s}(u)\big) = 0. \tag{108}$$

Step 3: Exact tail matching. By the continuity and strict decrease of $u \mapsto S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(Q_{w_{n,t}(1)|s_{n,t}=s}(u))$ on $(\tilde{u}_{1,s,t}, \bar{u}_{1,s,t})$ and its limit 0 as $u \to \bar{u}_{1,s,t}$, there exists $\tilde{\tau}_{1,s,t} \in (\tau^*_{1,s,t}, 1)$ such that for every $\tau \in (\tilde{\tau}_{1,s,t}, 1)$ there is a unique $u_{1,s,t}(\tau) \in (\tilde{u}_{1,s,t}, \bar{u}_{1,s,t})$ solving

$$S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}\big(Q_{w_{n,t}(1)|s_{n,t}=s}(u_{1,s,t}(\tau))\big) = 1 - \tau.$$

Combining this with $S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(\tau)) = 1 - \tau$ for all $\tau \in (\tau^*_{1,s,t}, 1)$ and the strict decrease of $w \mapsto S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(w)$ on $(a_t^{\text{obs}}(1,s), \omega_t^{\text{obs}}(1,s))$ yields

$$Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(\tau) = Q_{w_{n,t}(1)|s_{n,t}=s}(u_{1,s,t}(\tau)) \qquad \text{for all } \tau \in (\tilde{\tau}_{1,s,t}, 1). \tag{109}$$

Moreover,

$$\lim_{\tau \to 1} u_{1,s,t}(\tau) = \bar{u}_{1,s,t}. \tag{110}$$

Step 4: Cross-x τ-alignment and the product identity. Fix $\bar{s} \in \mathcal{S}_t$ satisfying Assumption (iv). Let $\{\tau^{(k)}_{1,\bar{s},t}\}_{k \geq 1} \subset (0,1)$ be any sequence with $\tau^{(k)}_{1,\bar{s},t} \to 1$ as $k \to \infty$ and define the sequence $\{\tau^{(k)}_{1,s,t}\}_{k \geq 1} \subset (0,1)$ where

$$\tau^{(k)}_{1,s,t} \coloneqq 1 - \frac{\Pr(D_{n,t} = 1 \mid s_{n,t} = \bar{s})}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)} \left(1 - \tau^{(k)}_{1,\bar{s},t}\right). \tag{111}$$

Let $u^{(k)}_{1,s,t} \coloneqq u_{1,s,t}(\tau^{(k)}_{1,s,t})$ and $u^{(k)}_{1,\bar{s},t} \coloneqq u_{1,\bar{s},t}(\tau^{(k)}_{1,\bar{s},t})$. Using (107) at $u = u^{(k)}_{1,s,t}$ and $u = u^{(k)}_{1,\bar{s},1}$,

$$1 - \tau^{(k)}_{1,s,t} = (1 - u^{(k)}_{1,s,t}) \frac{r_{1,s,t}(u^{(k)}_{1,s,t})}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)}, \qquad 1 - \tau^{(k)}_{1,\bar{s},t} = (1 - u^{(k)}_{1,\bar{s},t}) \frac{r_{1,\bar{s},t}(u^{(k)}_{1,\bar{s},t})}{\Pr(D_{n,t} = 1 \mid s_{n,t} = \bar{s})}.$$

Divide the two equalities and use (111) to obtain

$$\frac{(1 - u^{(k)}_{1,s,t})\, r_{1,s,t}(u^{(k)}_{1,s,t})}{(1 - u^{(k)}_{1,\bar{s},t})\, r_{1,\bar{s},t}(u^{(k)}_{1,\bar{s},t})} = 1. \tag{112}$$

Step 5: Aligning tail probabilities across states. Under Assumption (ii),

$$\lim_{k \to \infty} \frac{r_{1,s,t}(u^{(k)}_{1,s,t})}{r_{1,\bar{s},t}(u^{(k)}_{1,\bar{s},t})} = q_t(1) \in (0, \infty).$$

By (110), $u^{(k)}_{1,\bar{s},t} \to \bar{u}_{1,\bar{s},t}$ and $u^{(k)}_{1,s,t} \to \bar{u}_{1,s,t}$. Since $\bar{u}_{1,\bar{s},t}, \bar{u}_{1,s,t} < 1$ and $r_{1,\bar{s},t}(\cdot), r_{1,s,t}(\cdot)$ are continuous near those limits (by Assumption (iii)), (112) implies

$$\lim_{k\to\infty} \frac{1-u_{1,s,t}^{(k)}}{1-u_{1,\bar{s},t}^{(k)}} = 1,$$

and therefore

$$\lim_{k\to\infty} \left(u_{1,s,t}^{(k)} - u_{1,\bar{s},t}^{(k)}\right) = 0. \tag{113}$$

Step 6: Identification by differencing. By exogeneity of $\epsilon_{n,t}(1)$,

$$Q_{w_{n,t}(1)|s_{n,t}=s}(u) \;=\; \varphi(1,s) \;+\; Q_{\epsilon_{n,t}(1)}(u) \qquad \text{for all } u \in (0,1).$$

Apply (109) at $\tau = \tau_{1,s,t}^{(k)}$ and at $\tau = \tau_{1,\bar{s},t}^{(k)}$ to obtain

$$\begin{aligned} Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}\big(\tau_{1,s,t}^{(k)}\big) &= \varphi(1,s) + Q_{\epsilon_{n,t}(1)}(u_{1,s,t}^{(k)}), \\ Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=\bar{s}}\big(\tau_{1,\bar{s},t}^{(k)}\big) &= y(1,\bar{s}) + Q_{\epsilon_{n,t}(1)}(u_{1,\bar{s},t}^{(k)}). \end{aligned} \tag{114}$$

By (113) and continuity of $Q_{\epsilon_{n,t}(1)}$ near the upper tail,

$$\lim_{k\to\infty} \Big(Q_{\epsilon_{n,t}(1)}(u_{1,s,t}^{(k)}) - Q_{\epsilon_{n,t}(1)}(u_{1,\bar{s},t}^{(k)})\Big) = 0.$$

Subtracting the equations in (114) and using $\varphi(1,\bar{s}) = 0$ (Assumption (iv)) yields the identification result:

$$\lim_{k\to\infty} \Big[\, Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tau_{1,s,t}^{(k)}\big) \;-\; Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=\bar{s}}\big(\tau_{1,\bar{s},t}^{(k)}\big) \,\Big] \;=\varphi(1,s). \quad \square$$

Remark. The only substantive difference between Proposition E.1 and Proposition G.1—apart from the support restrictions in Assumption (i)—is that Assumption (ii) in the unbounded case is replaced, in the bounded case, by a *relative tail decay* condition. For reference, Assumption (ii) of Proposition E.1 posits that there exists an (unknown) constant $q_t(d) \in (0,1]$ such that, for each $s \in \mathcal{S}_t$,

$$\lim_{w\to\omega_t(d,s)} \Pr\big(D_{n,t} = d \mid s_{n,t} = s, w_{n,t}(d) > w\big) = q_t(d). \tag{115}$$

The requirement (115) is too strong—and in fact necessarily violated—under a strict support gap $\omega_t^{\text{obs}}(d,s) < \omega_t(d,s)$ (Assumption (i) of Proposition G.1). To see this, Bayes' rule (Step 1 of the proof) implies, for any $s \in \mathcal{S}_t$ and any w,

$$S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) = \; S_{w_{n,t}(d)|s_{n,t}=s}(w)\, \frac{\Pr(D_{n,t} = d \mid s_{n,t} = s,\; w_{n,t}(d) > w)}{\Pr(D_{n,t} = d \mid s_{n,t} = s)}.$$

For any $w \in (\omega_t^{\text{obs}}(d,s), \omega_t(d,s))$ we have $S_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) = 0$ while $S_{w_{n,t}(d)|s_{n,t}=s}(w) > 0$

and $\Pr(D_{n,t} = d \mid s_{n,t} = s) > 0$, hence

$$\Pr(D_{n,t} = d \mid s_{n,t} = s,\ w_{n,t}(d) > w) = 0 \quad \text{for all } w \in (\omega_t^{\text{obs}}(d,s), \omega_t(d,s)).$$

Therefore, the tail selection probability collapses to zero as $w \to \omega_t(d,s)$, forcing $q_t(d) = 0$ in (115). A positive limit could arise only in the case $\omega_t^{\text{obs}}(d,s) = \omega_t(d,s)$, which is excluded by Assumption (i). This is why we adopt a *relative* tail condition in place of (115), which governs the *rate* at which tail probabilities vanish across s (via ratios) rather than imposing a common nonzero limit that cannot hold under a support gap.

G.2.1 Identification of Support Endpoints

Corollary G.1 shows that, when finite, the lower and upper endpoints of the potential wages $w_{n,t}(d) \mid s_{n,t} = s$ and shock $\epsilon_{n,t}(d)$ are nonparametrically identified. Intuitively, for each s, the lower and upper endpoints of the observed, selected wage distribution, (denoted by $\underline{\omega}_t^{\text{obs}}(d,s)$ and $\omega_t^{\text{obs}}(d,s)$, respectively), are read directly from extremal quantiles: very low quantiles approach the lower endpoint and very high quantiles approach the upper endpoint. Because the deterministic part of wages $\varphi(d,s)$ is already known by Proposition G.1, we can "shift" the observed $\underline{\omega}_t^{\text{obs}}(d,s)$ and $\omega_t^{\text{obs}}(d,s)$ to learn about the latent lower and upper endpoints of both the shock $\epsilon_{n,t}(d)$ (denoted by $\underline{\omega}_\epsilon(d)$ and $\omega_\epsilon(d)$, respectively; time-invariant for simplicity) and the potential wage $w_{n,t}(d) = \varphi(d,s) + \epsilon_{n,t}(d)$ (denoted by $\underline{\omega}_t(d,s)$ and $\omega_t(d,s)$, respectively). Namely, selection trims the extremals, so the observed support sits inside the latent one: $\underline{\omega}_t^{\text{obs}}(d,s) \geq \underline{\omega}_t(d,s)$ and $\omega_t^{\text{obs}}(d,s) \leq \omega_t(d,s)$, with $\omega_t(d,s) = \varphi(d,s) + \omega_\epsilon(d)$ and $\underline{\omega}_t(d,s) = \varphi(d,s) + \underline{\omega}_\epsilon(d)$. Taking the best (tightest) such shifts across s gives bounds:

$$\sup_s\{\omega_t^{\text{obs}}(d,s) - \varphi(d,s)\} \leq \omega_\epsilon(d), \qquad \underline{\omega}_\epsilon(d) \leq \inf_s\{\underline{\omega}_t^{\text{obs}}(d,s) - \varphi(d,s)\},$$

and adding back $\varphi(d,s)$ yields corresponding tighest bounds for $\omega_t(d,s)$ and $\underline{\omega}_t(d,s)$. Moreover, if there exists a state value $s^\star$ where selection does *not* truncate the top ($\omega_t^{\text{obs}}(d,s^\star) = \omega_t(d,s^\star)$), the upper latent endpoint is revealed by the extremal quantile at $s^\star$:

$$\omega_\epsilon(d) = \lim_{\tau \to 1}\left\{Q_{w_{n,t}|D_{n,t}=d, s_{n,t}=s^\star}(\tau) - \varphi(d,x^\star)\right\},$$

and then $\omega_t(d,s) = \varphi(d,s) + \omega_\epsilon(d)$ for every s. A symmetric argument applies to the lower endpoint if selection does not truncate the bottom at some $s^\dagger$.

Corollary G.1 (Identification of finite right and left endpoints of $\epsilon_{n,t}(d)$ and $w_{n,t}(d)$—Simplified Wage Equation (11))**.** *Let $t \in \{1,\ldots,T\}$ and $d \in \mathcal{D}$. Maintain the assumptions of Proposition G.1, implying that $\varphi(d,s)$ is identified for each $s \in \mathcal{S}_t$. In addition, assume finite and distinct endpoints, with two-sided tail regularity: for each $s \in \mathcal{S}_t$,*

$$\underline{\omega}_t(d,s) := \inf\{u : \Pr(w_{n,t}(d) \leq u \mid s_{n,t} = s) > 0\} > -\infty,$$
$$\omega_t(d,s) := \sup\{u : \Pr(w_{n,t}(d) \leq u \mid s_{n,t} = s) < 1\} < \infty,$$

$$\underline{\omega}_t^{\mathrm{obs}}(d,s) \coloneqq \inf\{u : \Pr(w_{n,t} \le u \mid D_{n,t} = d, s_{n,t} = s) > 0\} > \underline{\omega}_t(d,s) > -\infty,$$
$$\omega_t^{\mathrm{obs}}(d,s) \coloneqq \sup\{u : \Pr(w_{n,t} \le u \mid D_{n,t} = d, s_{n,t} = s) < 1\} < \omega_t(d,s) < \infty,$$

with $F_{w_{n,t}(d)|s_{n,t}=s}$ continuous and strictly increasing on $(\underline{\omega}_t(d,s),\ a_t(d,s)) \cup (\tilde{a}_t(d,s),\ \omega_t(d,s))$ for some $a_t(d,s) < \tilde{a}_t(d,s)$, and $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ continuous and strictly increasing on $(\underline{\omega}_t^{\mathrm{obs}}(d,s),\ a_t^{\mathrm{obs}}(d,s)) \cup ((\tilde{a}_t^{\mathrm{obs}}(d,s)),\ \omega_t^{\mathrm{obs}}(d,s))$ for some $a_t^{\mathrm{obs}}(d,s) < (\tilde{a}_t^{\mathrm{obs}}(d,s))$, as well as continuous at both endpoints:

$$\lim_{w \to \underline{\omega}_t^{\mathrm{obs}}(d,s)} F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) = 0, \qquad \lim_{w \to \omega_t^{\mathrm{obs}}(d,s)} F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(w) = 1.$$

Define the shock $\epsilon_{n,t}(d)$ (finite) endpoints as:

$$\underline{\omega}_\epsilon(d) \coloneqq \inf\{u \in \mathbb{R} : F_{\epsilon_{n,t}(d)}(u) > 0\} > -\infty, \qquad \omega_\epsilon(d) \coloneqq \sup\{u \in \mathbb{R} : F_{\epsilon_{n,t}(d)}(u) < 1\} < \infty.$$

Then:

(a) (Observed wage endpoints are identified.) For every $s \in \mathcal{S}_t$, $\underline{\omega}_t^{\mathrm{obs}}(d,s)$ and $\omega_t^{\mathrm{obs}}(d,s)$ are identified:

$$\underline{\omega}_t^{\mathrm{obs}}(d,s) = \lim_{\tau \to 0} Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau), \qquad \omega_t^{\mathrm{obs}}(d,s) = \lim_{\tau \to 1} Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}(\tau).$$

(b) (Sharp bounds for latent endpoints.) For every $s \in \mathcal{S}_t$, a lower (resp. upper bound) bound for $\omega_\epsilon(d)$ (resp. $\underline{\omega}_\epsilon(d)$) and a lower (resp. upper bound) bound for $\omega_t(d,s)$ (resp. $\underline{\omega}_t(d,s)$) are identified:

$$L_\epsilon(d) \coloneqq \sup_s \left\{\omega_t^{\mathrm{obs}}(d,s) - \varphi(d,s)\right\} \ \le\ \omega_\epsilon(d),\ U_\epsilon(1) \coloneqq \inf_s \left\{\underline{\omega}_t^{\mathrm{obs}}(d,s) - \varphi(d,s)\right\} \ge \underline{\omega}_\epsilon(d),$$
$$\underline{\omega}_t(d,s) \le \min\left\{\ \underline{\omega}_t^{\mathrm{obs}}(d,s), \varphi(d,s) + U_\epsilon(d)\right\}, \omega_t(d,s) \ge \max\left\{\ \omega_t^{\mathrm{obs}}(d,s), \varphi(d,s) + L_\epsilon(d)\ \right\}.$$

Moreover, these bounds are sharp under the stated assumptions.

(c) (Upper endpoint point identification under no top truncation at some $s^\star$.) If there exists a known $s^\star \in \mathcal{S}_t$ with $\omega_t^{\mathrm{obs}}(d,s^\star) = \omega_t(d,s^\star)$ (i.e., the finite upper endpoint of the selected observed wages equals the finite lower endpoint of the potential wages; in other words, selection does not affect the rightmost support of wages at s^*), then, for every $s \in \mathcal{S}_t$, $\omega_\epsilon(d)$ and $\omega_t(d,s)$ are identified:

$$\omega_\epsilon(d) = \lim_{\tau \to 1} \Big[Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s^\star}(\tau) - \varphi(d,s^\star)\Big],$$
$$\omega_t(d,s) = \varphi(d,s) + \lim_{\tau \to 1} \Big[Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s^\star}(\tau) - \varphi(d,s^\star)\Big].$$

(d) (Lower endpoint point identification under no bottom truncation at some $s^\dagger$.) If there exists a known $s^\dagger \in \mathcal{S}_t$ with $\underline{\omega}_t^{\mathrm{obs}}(d,s^\dagger) = \underline{\omega}_t(d,s^\dagger)$ (i.e. the finite lower endpoint of the selected observed wages equals the finite left endpoint of the potential wages; in other words, selection does not affect the leftmost support of wages at $s^\dagger$), then, for every $s \in \mathcal{S}_t$, $\underline{\omega}_\epsilon(d)$ and $\underline{\omega}_t(d,s)$

are identified:

$$\underline{\omega}_\epsilon(d) = \lim_{\tau\to 0}\Big[Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s^\dagger}(\tau) - \varphi(d,s^\dagger)\Big],$$
$$\underline{\omega}_t(d,s) = \varphi(d,s) + \lim_{\tau\to 0}\Big[Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s^\dagger}(\tau) - \varphi(d,s^\dagger)\Big].$$

Proof: We present the proof for the upper endpoints; the argument for the lower endpoints is symmetric. Without loss, consider $d = 1$.

(a) Fix any $s \in \mathcal{S}_t$. By Assumption (iii) of Proposition G.1, $F_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}$ is continuous and strictly increasing on $(a_t^{\text{obs}}(1,s), \omega_t^{\text{obs}}(1,s))$ and continuous at the endpoint. Therefore, its upper quantiles converge to the endpoint, yielding *(a)*.

(b) Fix any $s \in \mathcal{S}_t$. It holds that

$$\omega_t(1,s) \;=\; \varphi(1,s) + \omega_\epsilon(1). \tag{116}$$

By Assumption (ii) of Proposition G.1, $\omega_t^{\text{obs}}(1,s) < \omega_t(1,s)$, so

$$\omega_t^{\text{obs}}(1,s) - \varphi(1,s) \;<\; \omega_t(1,s) - \varphi(1,s) \;=\; \omega_\epsilon(1).$$

Taking the supremum over s yields a lower bound for $\omega_\epsilon(1)$. Adding $\varphi(1,s)$ gives a lower bound for $\omega_t(1,s)$. These bounds are the best possible (sharp) without further restrictions.

(c) Fix any $s \in \mathcal{S}_t$. If there exists a known $s^\star \in \mathcal{S}_t$ with $\omega_t^{\text{obs}}(1,s^\star) = \omega_t(1,s^\star)$, then by *(a)*, $\omega_t(1,s^\star)$ is identified:

$$\omega_t(1,s^\star) = \lim_{\tau\to 1} Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s^\star}(\tau).$$

Using (116) written for s^* and recalling that $\varphi(1,s^\star)$ is identified by Proposition G.1 gives $\omega_\epsilon(1) = \omega_t(1,s^\star) - \varphi(1,s^\star)$. We plug this into (116) and complete the proof. □

G.3 Proposition E.1 with Location and Scale Parameters

Propositions E.1 and G.1 extend to wage specifications in which the shock $\epsilon_{n,t}(d)$ is multiplied by a scale parameter $\sigma(d, s_{n,t}) > 0$:

$$w_{n,t} = \sum_{d\in\{0,1\}} \mathbb{1}\{D_{n,t} = d\}\big[\varphi(d,s_{n,t}) + \sigma(d,s_{n,t})\epsilon_{n,t}(d)\big]. \tag{117}$$

Proposition G.2 (Identification of $\varphi(d,\cdot)$ and $\sigma(d,\cdot)$—Simplified Wage Equation (117))**.** *Let* $t \in \{1,\ldots,T\}$ *and* $d \in \mathcal{D}$*. Assume that the conditional wage distribution* $\Pr\big(w_{n,t} \le w \,\big|\, D_{n,t} = d,\ s_{n,t} = s\big)$ *is identified for each* $w \in \mathbb{R}$ *and* $s \in \mathcal{S}_t$ *(see Proposition B.5 for sufficient conditions), and that the conditional choice probability* $\Pr(D_{n,t} = d \mid s_{n,t} = s)$ *is identified for each* $s \in \mathcal{S}_t$ *(see Proposition B.4 for sufficient conditions). Moreover, assume:*

(i) (Supports.)[54] *For each* $s \in \mathcal{S}_t$, $\omega_t(d,s) := \sup\{u : \Pr(w_{n,t}(d) \le u \mid s_{n,t} = s) < 1\} = \infty$ *and*

[54] We focus on the case of unbounded supports. The bounded-support case follows analogously, with the technical modifications highlighted in Appendix G.2 of the Online Supplementary Material.

$\omega_t^{\text{obs}}(d,s) := \sup\{u : \Pr(w_{n,t} \le u \mid D_{n,t} = d, s_{n,t} = s) < 1\} = \infty.$

(ii) (Tail Limit.) There exists an (unknown) constant $q_t(d) \in (0,1]$ such that for every $s \in \mathcal{S}_t$, $\lim_{w\to\infty} \Pr\big(D_{n,t} = d \mid s_{n,t} = s, w_{n,t}(1) > w\big) = q_t(d)$.

(iii) (Tail Regularity.) For each $s \in \mathcal{S}_t$, there exist (unknown) thresholds $a_t(d,s) < \infty$ and $a_t^{\text{obs}}(d,s) < \infty$ such that the cumulative distribution functions $F_{w_{n,t}(d)|s_{n,t}=s}$ and $F_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}$ are continuous and strictly increasing on $(a_t(d,s),\infty)$ and $(a_t^{\text{obs}}(d,s),\infty)$, respectively.

(iv) (Location and Scale Normalizations.) There exists a known $\bar{s} \in \mathcal{S}_t$ with $\varphi(d,\bar{s}) = 0$ and $\sigma(d,\bar{x}) = 1$.

For each $s \in \mathcal{S}_t$, fix the following sequences

$$\tau_{d,\bar{s},t}^{(k)} := 1 - 2^{-k}, \qquad \tilde{\tau}_{d,\bar{s},t}^{(k)} := 1 - 3^{-k}, \qquad k = 1,2,\ldots$$

and let

$$\tau_{d,s,t}^{(k)} := 1 - \frac{\Pr(D_{n,t} = d \mid s_{n,t} = \bar{s})}{\Pr(D_{n,t} = d \mid s_{n,t} = s)}\big(1 - \tau_{d,\bar{s},t}^{(k)}\big), \qquad \tilde{\tau}_{d,s,t}^{(k)} := 1 - \frac{\Pr(D_{n,t} = d \mid s_{n,t} = \bar{s})}{\Pr(D_{n,t} = d \mid s_{n,t} = s)}\big(1 - \tilde{\tau}_{d,\bar{s},t}^{(k)}\big).$$

Then,

$$\sigma(d,s) = \lim_{k\to\infty} \frac{Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(\tau_{d,s,t}^{(k)}\big) - Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(\tilde{\tau}_{d,s,t}^{(k)}\big)}{Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tau_{d,\bar{s},t}^{(k)}\big) - Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tilde{\tau}_{d,\bar{s},t}^{(k)}\big)},$$

and

$$\varphi(d,s) = \lim_{k\to\infty} \Big[Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=s}\big(\tau_{d,s,t}^{(k)}\big) - \sigma(d,s)\, Q_{w_{n,t}|D_{n,t}=d,s_{n,t}=\bar{s}}\big(\tau_{d,\bar{s},t}^{(k)}\big)\Big].$$

Hence $\varphi(d,s)$ and $\sigma(d,s)$ are identified.

Proof: Let $s \in \mathcal{S}_t$, and without loss, consider firm $d = 1$. For any threshold w, Bayes' rule gives

$$\Pr(w_{n,t} > w \mid D_{n,t} = 1, s_{n,t} = s) = \frac{\Pr(w_{n,t}(1) > w \mid s_{n,t} = s)\ \Pr(D_{n,t} = 1 \mid s_{n,t} = s, w_{n,t}(1) > w)}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)}.$$

Letting $w \to \infty$ and using Assumption (ii),

$$\Pr(w_{n,t} > w \mid D_{n,t} = 1, s_{n,t} = s)\ \sim\ c_t(1,s)\ \Pr\big(w_{n,t}(1) > w \mid s_{n,t} = s\big) \qquad (w \to \infty), \quad (118)$$

where

$$c_t(1,s)\ :=\ \frac{q_t(1)}{\Pr(D_{n,t} = 1 \mid s_{n,t} = s)}\ \in (0,\infty),$$

and "$\sim$" denotes that the ratio of the two sides converges to 1.
Write $S_{1,s,t}(w) := S_{w_{n,t}(1)|s_{n,t}=s}(w)$ and $S_{1,s,t}^{\text{obs}}(w) := S_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(w)$. Then, (118) reads as

$$S_{1,s,t}^{\text{obs}}(w)\ \sim\ c_t(1,s)\, S_{1,s,t}(w) \qquad (w \to \infty). \quad (119)$$

By Assumption (i), both right wage endpoints are ∞; by Assumption (iii), the upper-tail CDFs $F_{w_{n,t}(1)|s_{n,t}=s}$ and $F_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}$ are continuous and strictly increasing beyond finite thresholds, so their tail quantile maps are the ordinary inverses on the corresponding index ranges near 1. Hence, by Lemma E.1 (stated and proved in Appendix E),

$$Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}(\tau) \;=\; Q_{w_{n,t}(1)\,|\,s_{n,t}=s}\Big(1-\frac{1-\tau}{c_t(1,s)} \;+\; o_s(1-\tau)\Big) \qquad (\tau\to 1), \tag{120}$$

where $o_s(1-\tau)/(1-\tau)\to 0$ as $\tau\to 1$.
From $w_{n,t}(1)=\varphi(1,s)+\sigma(1,s)\epsilon_{n,t}(1)$ and exogeneity of $\epsilon_{n,t}(1)$, for all $u\in(0,1)$,

$$Q_{w_{n,t}(1)\,|\,s_{n,t}=s}(u) \;=\; \varphi(1,s)+\sigma(1,s)Q_{\epsilon_{n,t}(1)}(u).$$

Plugging into (120) gives

$$Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}(\tau) \;=\; \varphi(1,s) \;+\; \sigma(1,s)Q_{\epsilon_{n,t}(1)}\Big(1-\frac{1-\tau}{c_t(1,s)} \;+\; o_s(1-\tau)\Big) \qquad (\tau\to 1). \tag{121}$$

Scale. Evaluate (121) at $\tau=\tau^{(k)}_{1,s,t}$ and $\tau=\tilde{\tau}^{(k)}_{1,s,t}$:

$$\left.\begin{aligned} Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tau^{(k)}_{1,s,t}\big) &= \varphi(1,s)+\sigma(1,s)Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{1,s,t}}{c_t(1,s)}+o_s(1-\tau^{(k)}_{1,s,t})\Big),\\ Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tilde{\tau}^{(k)}_{1,s,t}\big) &= \varphi(1,s)+\sigma(1,s)Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tilde{\tau}^{(k)}_{1,s,t}}{c_t(1,s)}+o_x(1-\tilde{\tau}^{(k)}_{1,s,t})\Big),\end{aligned}\right.\quad (k\to\infty). \tag{122}$$

Take the difference between the two equations in (122):

$$\begin{aligned}\Delta^{(k)}_{1,s,t} &:= Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(\tau^{(k)}_{1,s,t})-Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(\tilde{\tau}^{(k)}_{1,s,t})\\ &=\sigma(1,s)\Big[Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{1,s,t}}{c_t(1,s)}+o_s(1-\tau^{(k)}_{1,s,t})\Big)-Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tilde{\tau}^{(k)}_{1,s,t}}{c_t(1,s)}+o_x(1-\tilde{\tau}^{(k)}_{1,s,t})\Big)\Big]\ (k\to\infty).\end{aligned} \tag{123}$$

Repeat analogous steps for $\tau=\tau^{(k)}_{1,\bar{s},t}$ and $\tau=\tilde{\tau}^{(k)}_{1,\bar{s},t}$ and use the normalizations in Assumption (iv):

$$\begin{aligned}\Delta^{(k)}_{1,\bar{s},t} &:= Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=\bar{s}}(\tau^{(k)}_{1,\bar{s},t})-Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=\bar{s}}(\tilde{\tau}^{(k)}_{1,\bar{s},t})\\ &=\Big[Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}+o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big)-Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tilde{\tau}^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}+o_{\bar{s}}(1-\tilde{\tau}^{(k)}_{1,\bar{s},t})\Big)\Big]\qquad (k\to\infty).\end{aligned} \tag{124}$$

By the definition of $\tau^{(k)}_{1,s,t}$ and $\tilde{\tau}^{(k)}_{1,s,t}$,

$$1-\frac{1-\tau^{(k)}_{1,s,t}}{c_t(1,s)}=1-\frac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})},\qquad 1-\frac{1-\tilde{\tau}^{(k)}_{1,s,t}}{c_t(1,s)}=1-\frac{1-\tilde{\tau}^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}.$$

Thus, as $k\to\infty$, (123) and (124) can be written as

$$\left.\begin{aligned}\Delta^{(k)}_{1,s,t} &= \sigma(1,s)\Big[Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}+o_s(1-\tau^{(k)}_{1,s,t})\Big)-Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tilde{\tau}^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}+o_x(1-\tilde{\tau}^{(k)}_{1,s,t})\Big)\Big]\\ \Delta^{(k)}_{1,\bar{s},t} &= \Big[Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}+o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big)-Q_{\epsilon_{n,t}(1)}\Big(1-\tfrac{1-\tilde{\tau}^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}+o_{\bar{s}}(1-\tilde{\tau}^{(k)}_{1,\bar{s},t})\Big)\Big]\end{aligned}\right.. \tag{125}$$

Take the ratio between the two equations in (125). Assumption (iii) implies that $Q_{\epsilon_{n,t}(1)}$ is continuous and strictly increasing near 1. Since $\tau^{(k)}_{1,\bar{s},t} = 1 - 2^{-k}$ and $\tilde{\tau}^{(k)}_{1,\bar{s},t} = 1 - 3^{-k}$ are distinct for all k, the denominator of the ratio is nonzero for all large k. By continuity and $o_s(1-\tau^{(k)}_{1,s,t}), o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t}) \to 0$,

$$\lim_{k\to\infty} \frac{Q_{\epsilon_{n,t}(1)}\Big(1 - \frac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_s(1-\tau^{(k)}_{1,s,t})\Big) - Q_{\epsilon_{n,t}(1)}\Big(1 - \frac{1-\tilde{\tau}^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_x(1-\tilde{\tau}^{(k)}_{1,s,t})\Big)}{Q_{\epsilon_{n,t}(1)}\Big(1 - \frac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big) - Q_{\epsilon_{n,t}(1)}\Big(1 - \frac{1-\tilde{\tau}^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_{\bar{s}}(1-\tilde{\tau}^{(k)}_{1,\bar{s},t})\Big)}.$$

Therefore,

$$\sigma(1,s) = \lim_{k\to\infty} \frac{Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(\tau^{(k)}_{1,s,t}) - Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=s}(\tilde{\tau}^{(k)}_{1,s,t})}{Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=\bar{s}}(\tau^{(k)}_{1,\bar{s},t}) - Q_{w_{n,t}|D_{n,t}=1,s_{n,t}=\bar{s}}(\tilde{\tau}^{(k)}_{1,\bar{s},t})}.$$

Location. Evaluate (121) at $\tau = \tau^{(k)}_{1,s,t}$ and, with $s = \bar{s}$, at $\tau = \tau^{(k)}_{1,\bar{s},t}$:

$$\begin{aligned} Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tau^{(k)}_{1,s,t}\big) &= \varphi(1,s) + \sigma(1,s) Q_{\epsilon_{n,t}(1)}\Big(1 - \tfrac{1-\tau^{(k)}_{1,s,t}}{c_t(1,s)} + o_s(1-\tau^{(k)}_{1,s,t})\Big), \\ Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{1,\bar{s},t}\big) &= Q_{\epsilon_{n,t}(1)}\Big(1 - \tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big), \end{aligned} \qquad (k\to\infty), \tag{126}$$

where we use the normalizations in Assumption (iv). By the definition of $\tau^{(k)}_{1,s,t}$,

$$1 - \frac{1-\tau^{(k)}_{1,s,t}}{c_t(1,s)} = 1 - \frac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})}.$$

Therefore, (126) can be written as

$$\begin{aligned} Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tau^{(k)}_{1,s,t}\big) &= \varphi(1,s) + \sigma(1,s) Q_{\epsilon_{n,t}(1)}\Big(1 - \tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_s(1-\tau^{(k)}_{1,s,t})\Big), \\ Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{1,\bar{s},t}\big) &= Q_{\epsilon_{n,t}(1)}\Big(1 - \tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big), \end{aligned} \qquad (k\to\infty). \tag{127}$$

As $k \to \infty$, subtracting the two equations in (127):

$$\begin{aligned} &Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tau^{(k)}_{1,s,t}\big) - \sigma(1,s) Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{1,\bar{s},t}\big) \\ &= \varphi(1,s) + \sigma(1,s)\Big[Q_{\epsilon_{n,t}(1)}\Big(1 - \tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_s(1-\tau^{(k)}_{1,s,t})\Big) - Q_{\epsilon_{n,t}(1)}\Big(1 - \tfrac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} + o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big)\Big]. \end{aligned}$$

Also note that $o_s(1-\tau^{(k)}_{1,s,t}) \to 0$ and $o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t}) \to 0$ as $k \to \infty$. Therefore, by continuity of $Q_{\epsilon_{n,t}(1)}$ near 1 under Assumption (iii),

$$Q_{\epsilon_{n,t}(1)}\Big(u + o_s(1-\tau^{(k)}_{1,s,t})\Big) - Q_{\epsilon_{n,t}(1)}\Big(u + o_{\bar{s}}(1-\tau^{(k)}_{1,\bar{s},t})\Big) = o(1), \qquad u := 1 - \frac{1-\tau^{(k)}_{1,\bar{s},t}}{c_t(1,\bar{s})} \qquad (k\to\infty).$$

Therefore, as desired,

$$\lim_{k\to\infty}\Big[\, Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=s}\big(\tau^{(k)}_{1,s,t}\big) \; - \; \sigma(1,s) Q_{w_{n,t}\,|\,D_{n,t}=1,s_{n,t}=\bar{s}}\big(\tau^{(k)}_{1,\bar{s},t}\big) \,\Big] = \varphi(1,s). \quad \square$$

H Extension to Search Models

The quantile approach in Proposition G.2 can be used to identify key parameters of equilibrium wage equations with inherent conditional heteroskedasticity as in standard search models. For example, in the spirit of Bagger et al. (2014), consider the potential wage of worker n at time t in firm $d \in \mathcal{D}$ is

$$w_{n,t}(d) = \omega\, \gamma_d^{\alpha_d}\, H_{n,t}^{1-\alpha_d}\, \epsilon_{n,t}(d) \;+\; (1-\omega)(1-\delta)\, U_d(H_{n,t}), \tag{128}$$

where $0 < \omega < 1$ is the worker's bargaining weight, $\gamma_d > 0$ is firm d's productivity, $\alpha_d \in [0,1)$ is the elasticity of wages with respect to γ_d, $\delta \in (0,1)$ is the discount factor, $H_{n,t}$ denotes a worker's human capital at time t with support $\mathcal{H}_t$, and $U_d(H_{n,t})$ is the value of unemployment,

$$U_d(H_{n,t}) \coloneqq z \;+ \delta\, \mathbb{E}_{\epsilon_{n,t} \sim F_{\epsilon_{n,t}}} \left[f\big(S\big(H_{n,t}, \epsilon_{n,t};\, \omega, \alpha_d, \gamma_d, \delta\big)\big) \right],$$

where z is the flow value of unemployment; $S(\cdot; \omega, \alpha_d, \gamma_d, \delta)$ is the match surplus; $f(\cdot)$ is a functional of the surplus; and the expectation is taken with respect to the shocks $\epsilon_{n,t} \coloneqq (\epsilon_{n,t}(d) : d \in \mathcal{D})$ with distribution $F_{\epsilon_{n,t}}$. The index d on U_d reflects the dependence of S on (α_d, γ_d). As is standard, we treat δ and ω as known. The parameters to be identified are then γ_d, α_d, and z. The functions f and S are known up to $(\gamma_d, \alpha_d, F_{\epsilon_{n,t}})$. We consider two cases: (1) $H_{n,t}$ is observed (or unobserved with known distribution and support); (2) $H_{n,t}$ is unobserved with unknown distribution and support. Hereafter, we denote by $F_{\epsilon_{n,t}(d)}$ the marginal CDF of $\epsilon_{n,t}(d)$ and by $S_{\epsilon_{n,t}(d)}$ the survival function of $\epsilon_{n,t}(d)$.

Remark. Hereafter, for simplicity, we focus on the case in which observed and potential wages, as well as shocks, have unbounded support. The bounded-support case follows analogously, with the technical modifications highlighted in Proposition G.1 of Appendix G.2 of the Online Supplementary Material. In that case, it is also possible to nonparametrically identify the finite lower and upper endpoints of potential wages and shocks, as shown in Corollary G.1.

Case 1: $H_{n,t}$ is Observed or Unobserved with Known Distribution and Support. As a preview, Proposition H.1, Proposition H.2, and Corollary H.1 stated below follow immediately from Proposition G.2 and Proposition E.2. For Proposition H.1, replace $s_{n,t}$ with $H_{n,t}$ and define

$$y(d, H_{n,t}) \coloneqq (1-\omega)(1-\delta)\, U_d(H_{n,t}) \;\text{ and }\; \sigma(d, H_{n,t}) \coloneqq \omega\, \gamma_d^{\alpha_d}\, H_{n,t}^{1-\alpha_d}.$$

Then, Proposition G.2 identifies $y(d, H_{n,t})$ and $\sigma(d, H_{n,t})$. As for Proposition H.2, once $y(d, H_{n,t})$ and $\sigma(d, H_{n,t})$ are identified, the joint distribution of the shock vector, $F_{\epsilon_{n,t}}$, is identified under the conditions of Proposition E.2. As for Corollary H.1, once $y(d, H_{n,t})$, $\sigma(d, H_{n,t})$, and $F_{\epsilon_{n,t}}$ are identified, the parameters α_d, γ_d, and z are straightforward to recover.

Proposition H.1 (Identification of $y(d, H_{n,t})$ and $\sigma(d, H_{n,t})$)**.** *For each firm $d \in \mathcal{D}$ and period $t \in \{1, \ldots, T\}$, assume:*

(i) (Supports.) For each $h \in \mathcal{H}_t$, $\sup\{u : \Pr(w_{n,t}(d) \le u \mid H_{n,t} = h) < 1\} = \infty$, $\sup\{u : \Pr(w_{n,t} \le u \mid D_{n,t} = d, H_{n,t} = h) < 1\} = \infty$, and $0 < \Pr(D_{n,t} = d \mid H_{n,t} = h) \le 1$.

(ii) (Tail Limit.) There exists a constant $q_t(d) \in (0,1]$ *such that for every* $h \in \mathcal{H}_t$, $\lim_{w\to\infty} \Pr\big(D_{n,t} = d \mid H_{n,t} = h, w_{n,t}(d) > w\big) = q_t(d)$.

(iii) (Tail Regularity.) For each $h \in \mathcal{H}_t$, *there exist thresholds* $w_t(d,h) < \infty$ *and* $w_t^{\mathrm{obs}}(d,h) < \infty$ *such that the cumulative distribution functions* $F_{w_{n,t}(d)|H_{n,t}=h}$ *and* $F_{w_{n,t}|D_{n,t}=d,H_{n,t}=h}$ *are continuous and strictly increasing on* $(w_t(d,h),\infty)$ *and* $(w_t^{\mathrm{obs}}(d,h),\infty)$, *respectively.*

(iv) (Normalization.) There exists a known $\bar{h} \in \mathcal{H}_t$ *with* $y(d,\bar{h}) = 0$ *and* $\sigma(d,\bar{h}) = 1$.

Then, $y(d,h)$ *and* $\sigma(d,h)$ *are identified for each* $d \in \mathcal{D}$, $h \in \mathcal{H}_t$, *and* $t \in \{1,\dots,T\}$.

Proposition H.2 (Identification of $F_{\epsilon_{n,t}}$**).** *Suppose that Assumptions (i) to (iv) of Proposition H.1 hold so that* $y(d,h)$ *and* $\sigma(d,h)$ *are identified for each* $d \in \mathcal{D}$, $h \in \mathcal{H}_t$, *and* $t \in \{1,\dots,T\}$. *Moreover, for each firm* $d \in \mathcal{D}$ *and period* $t \in \{1,\dots,T\}$, *assume* $\epsilon_{n,t}(d)$ *belongs to a parametric family indexed by the* $p_{t,d} \times 1$ *vector of parameters* $\mu_{t,d}$ *that is tail-ratio identifiable. Namely, for any* $p_{t,d}+1$ *distinct large thresholds* $0 < e_0 < e_1 < \cdots < e_{p_{t,d}+1}$, *the map*

$$\mu_{t,d} \mapsto \left(\frac{S_{\epsilon_{n,t}(d)}(e_1;\mu_{t,d})}{S_{\epsilon_{n,t}(d)}(e_0;\mu_{t,d})}, \dots, \frac{S_{\epsilon_{n,t}(d)}(e_{p_{t,d}};\mu_{t,d})}{S_{\epsilon_{n,t}(d)}(e_0;\mu_{t,d})} \right),$$

is injective. Under these assumptions, for each period $t \in \{1,\dots,T\}$*:*[55]

(a) (Marginal Identification.) For each firm $d \in \mathcal{D}$, *the parameter* $\mu_{t,d}$ *is identified.*

(b) (Joint Identification.) If the shocks $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$ *are mutually independent across* $d \in \mathcal{D}$, *then the joint distribution of* $\epsilon_{n,t}$ *is identified as the product of the identified marginals. Alternatively, if a copula* C_{μ_t} *is specified so that*

$$F_{\epsilon_{n,t}}(v_1,\dots,v_{|\mathcal{D}|}) \;=\; C_{\mu_t}\big(F_{\epsilon_{n,t}(1)}(v_1;\mu_{t,1}),\dots,F_{\epsilon_{n,t}(|\mathcal{D}|)}(v_{|\mathcal{D}|};\mu_{t,|\mathcal{D}|})\big) \quad \forall\,(v_1,\dots,v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|},$$

with μ_t *known, then the joint distribution is identified via the identified marginals and* C_{μ_t}. *Absent further restrictions on the dependence among* $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$, *the joint CDF is partially identified by the sharp Fréchet–Höffding bounds, namely, for all* $(v_1,\dots,v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|}$,

$$\max\Big\{ \sum_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d;\mu_{t,d}) - (|\mathcal{D}|-1),\; 0 \Big\} \;\le\; F_{\epsilon_{n,t}}(v_1,\dots,v_{|\mathcal{D}|}) \;\le\; \min_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d;\mu_{t,d}).$$

[55] Note that our identification result in Proposition H.2 is established period by period. Accordingly, the shock vectors $\{\epsilon_{n,t}\}_{t=1}^T$ are allowed to be correlated across periods, and we impose no restrictions on their time dependence. If one is willing to assume that $\epsilon_{n,t}$ follows a (covariance-)stationary VAR(p) process,

$$\epsilon_{n,t} = A_0 + A_1\epsilon_{n,t-1} + \cdots + A_p\epsilon_{n,t-p} + u_{n,t},$$

and that the (population) mean μ and autocovariances Γ_j for $j = 0,\dots,p$ are known, then the autoregressive coefficient matrices $A_1,\dots,A_p$ (and hence $A_0 = \big(I - \sum_{j=1}^p A_j\big)\mu$) are identified under standard conditions such as orthogonality/whiteness of $u_{n,t}$ with respect to past $\epsilon_{n,t-h}$ and the usual rank conditions. Note, however, that without parametric assumptions on the distribution of $u_{n,t}$, the marginal distribution of each $\epsilon_{n,t}$ is not identified even given knowledge of μ, Γ_j, and A_j for $j = 0,\dots,p$. To recover these marginal distributions, one can instead rely on Proposition H.2.

Corollary H.1 (Identification of α_d, γ_d, and z)**.** *Assume that $y(d,h)$ and $\sigma(d,h)$ are identified for each $d \in \mathcal{D}$, for every realization h of $H_{n,t}$, and for some period $t \geq 1$ (see Proposition H.1 for sufficient conditions). Assume also that the joint distribution of the shock vector $F_{\epsilon_{n,t}}$ is identified for the same period $t \geq 1$ (see Proposition H.2 for sufficient conditions). Then, the parameters α_d, γ_d, and z are identified for each $d \in \mathcal{D}$.*

Proof: The proof proceeds in three steps.
Step 1: Identification of α_d from $\sigma(d, H_{n,t})$. For any h, h' in $\mathcal{H}_t$,

$$\frac{\sigma(d,h)}{\sigma(d,h')} = \frac{\omega\, \gamma_d^{\alpha_d}\, h^{1-\alpha_d}}{\omega\, \gamma_d^{\alpha_d}\, (h')^{1-\alpha_d}} = \left(\frac{h}{h'}\right)^{1-\alpha_d}.$$

Taking logarithms and rearranging,

$$\alpha_d \; = 1 \; - \frac{\log\big(\sigma(d,h)/\sigma(d,h')\big)}{\log(h/h')}.$$

Note that α_d is identified *without* relying on the scale normalization $\sigma(d,\bar{h}) = 1$ in Assumption (iv) of Proposition H.1, because the ratio $\sigma(d,h)/\sigma(d,h')$ is identified without any such normalization, as shown in the proof of Proposition G.2. Also, to identify α_d, we do not need to know the distribution of $\epsilon_{n,t}$, namely, $F_{\epsilon_{n,t}}$.
Step 2: Identification of γ_d from $\sigma(d, H_{n,t})$. Recall that for any $h \in \mathcal{H}_t$, we have that $\sigma(d,h) = \omega\, \gamma_d^{\alpha_d}\, h^{1-\alpha_d}$. Solving for γ_d yields that

$$\gamma_d \; = \big(\sigma(d,h)\omega^{-1}\, h^{\alpha_d - 1}\big)^{1/\alpha_d}.$$

Note that, unlike α_d, the identification of γ_d relies on knowing the *level* $\sigma(d,h)$ and therefore depends on the scale normalization $\sigma(d,\bar{h}) = 1$ in Assumption (iv) of Proposition H.1. As is the case with α_d, to identify γ_d, we do not need to know $F_{\epsilon_{n,t}}$.
Step 3: Identification of z from $y(d, H_{n,t})$ and $F_{\epsilon_{n,t}}$. For any $h \in \mathcal{H}_t$, we have

$$y(d,h) \; = (1-\delta)(1-\omega)\Big[z \; + \delta\, \mathbb{E}_{\epsilon_{n,t} \sim F_{\epsilon_{n,t}}}\big[\, f\big(S\big(h, \epsilon_{n,t};\, \omega, \alpha_d, \gamma_d, \delta\big)\big)\big]\Big].$$

Solving for z yields that z is identified as

$$z \; = \frac{y(d,h)}{(1-\delta)(1-\omega)} - \; \delta\, \mathbb{E}_{\epsilon_{n,t} \sim F_{\epsilon_{n,t}}}\big[\, f\big(S\big(h, \epsilon_{n,t};\, \omega, \alpha_d, \gamma_d, \delta\big)\big)\big]\,.$$

The identification of z relies on knowing the *level* $y(d,h)$ and therefore depends on the location normalization $y(d,\bar{h}) = 0$ in Assumption (v) of Proposition H.1. Differently from α_d and γ_d, to identify z we need to know $F_{\epsilon_{n,t}}$. □

Case 1: Alternative Proof. Rather than relying on Proposition H.1, we can use a quantile-based approach that directly leverages the parametric structure of the wage equation in (128). We show how this approach works to identify α_d and γ_d in Proposition H.3. We discuss how the nonparametric approach illustrated so far and the parametric approach behind Proposition H.3 differ from each other after the proof of the proposition. We start by defining

$$y(d, H_{n,t}) := (1-\omega)(1-\delta)\, U_d(H_{n,t}) \quad \text{and} \quad M_{n,t}(d) := \omega\, \gamma_d^{\alpha_d} H_{n,t}^{1-\alpha_d}\, \epsilon_{n,t}(d).$$

Proposition H.3 (Identification of α_d and γ_d)**.** *For each firm $d \in \mathcal{D}$ and some period $t \in \{1, \dots, T\}$, assume:*

(i) (Unbounded Upper Tail of Human Capital.) The upper tail of human capital $H_{n,t}$ is unbounded, that is, $\lim_{p\to 1} Q_{\log H_{n,t}|D_{n,t}=d}(p) = \infty$.

(ii) (Negligible Quantile Reminder Relative to Human Capital.) For each $p \in (0,1)$, define the conditional quantile reminder

$$R_{t,d}(p) := Q_{\log w_{n,t}|D_{n,t}=d}(p) - \Big\{ \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \Big\}. \quad (129)$$

Then, the contribution of this reminder to the selected wage $w_{n,t}|D_{n,t} = d$ grows strictly more slowly than the contribution of human capital $H_{n,t}$ as the wage increases, that is,

$$\lim_{p\to 1} \frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} = 0. \quad (130)$$

(iii) (Normalization.) The upper tail of the remainder $R_{t,d}(p)$ has a known finite limit. That is,

$$\lim_{p\to 1} R_{t,d}(p) \;=\; L_{t,d}, \quad (131)$$

with $L_{t,d}$ known.

In addition, assume that $H_{n,t} > 0$, $w_{n,t} > 0$, and $\epsilon_{n,t}(d) > 0$ almost surely, so that all logarithms above are well defined. Then, for each $d \in \mathcal{D}$, the parameters α_d and γ_d are identified.

Proof: The proof is articulated in two steps.
Step 1: Identification of α_d. Fix a firm $d \in \mathcal{D}$. Using the structure of the wage equation,

$$\log w_{n,t}(d) = \log M_{n,t}(d) + \log\Big(1 + \tfrac{y(d,H_{n,t})}{M_{n,t}(d)}\Big). \quad (132)$$

Using the definition of $M_{n,t}(d)$,

$$\log M_{n,t}(d) = \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\log H_{n,t} + \log \epsilon_{n,t}(d). \quad (133)$$

Therefore, substituting (133) in (132),

$$\log w_{n,t}(d) = \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\log H_{n,t} + \log\epsilon_{n,t}(d) + \log\Big(1 + \tfrac{y(d,H_{n,t})}{M_{n,t}(d)}\Big). \quad (134)$$

Now condition on $D_{n,t} = d$ and apply the conditional quantile operator $Q_{\cdot|D_{n,t}=d}(p)$ to both sides of

(134). Using only that adding a constant shifts quantiles, we obtain that

$$Q_{\log w_{n,t}|D_{n,t}=d}(p) = \log\omega + \alpha_d \log\gamma_d + Q_{(1-\alpha_d)\log H_{n,t}+\log\epsilon_{n,t}(d)+\log\left(1+\frac{y(d,H_{n,t})}{M_{n,t}(d)}\right)|D_{n,t}=d}(p). \quad (135)$$

Define the conditional quantile remainder $R_{t,d}(p)$ as

$$R_{t,d}(p) := Q_{\log w_{n,t}|D_{n,t}=d}(p) - \Big[\log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p)\Big]. \quad (136)$$

Plugging (135) into (136) yields that

$$\begin{aligned}R_{t,d}(p) =& \log\omega + \alpha_d\log\gamma_d + Q_{(1-\alpha_d)\log H_{n,t}+\log\epsilon_{n,t}(d)+\log\left(1+\frac{y(d,H_{n,t})}{M_{n,t}(d)}\right)|D_{n,t}=d}(p)\\ &- \Big[\log\omega + \alpha_d\log\gamma_d + (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p)\Big],\end{aligned}$$

which implies that

$$Q_{(1-\alpha_d)\log H_{n,t}+\log\epsilon_{n,t}(d)+\log\left(1+\frac{y(d,H_{n,t})}{M_{n,t}(d)}\right)|D_{n,t}=d}(p) = (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) + R_{t,d}(p).$$

By substituting this expression into (135), we obtain that

$$Q_{\log w_{n,t}|D_{n,t}=d}(p) = \log\omega + \alpha_d\log\gamma_d + (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) + R_{t,d}(p). \quad (137)$$

Fix any $\bar{p} \in (0,1)$. Let $\Delta_W(p) := Q_{\log w_{n,t}|D_{n,t}=d}(p) - Q_{\log w_{n,t}|D_{n,t}=d}(\bar{p})$, $\Delta_H(p) := Q_{\log H_{n,t}|D_{n,t}=d}(p) - Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})$, and $\Delta_R(p) := R_{t,d}(p) - R_{t,d}(\bar{p})$. Subtracting (137) evaluated at p and at $\bar{p}$ yields that for all $p \in (0,1)$,

$$\Delta_W(p) = (1-\alpha_d)\,\Delta_H(p) + \Delta_R(p). \quad (138)$$

By Assumption (i), $\lim_{p\to 1} Q_{\log H_{n,t}|D_{n,t}=d}(p) = \infty$ so that

$$\lim_{p\to 1}\Delta_H(p) = \infty. \quad (139)$$

By combining (139) and Assumption (ii), we can show that

$$\lim_{p\to 1}\frac{\Delta_R(p)}{\Delta_H(p)} = 0. \quad (140)$$

Indeed, note that

$$\frac{\Delta_R(p)}{\Delta_H(p)} = \frac{R_{t,d}(p) - R_{t,d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p) - Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})} = \frac{\frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} - \frac{R_{t,d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)}}{1 - \frac{Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)}}. \quad (141)$$

As $p \to 1$, Assumptions (i) and (ii) imply that

$$\frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} \to 0, \qquad \frac{R_{t,d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} \to 0, \qquad \frac{Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} \to 0,$$

so that the numerator of (141) converges to 0 whereas the denominator of it to 1, which yields (140). Now divide both sides of (138) by $\Delta_H(p)$ to obtain

$$\frac{\Delta_W(p)}{\Delta_H(p)} = 1 - \alpha_d + \frac{\Delta_R(p)}{\Delta_H(p)}. \tag{142}$$

Taking limits of both sides of (142) as $p \to 1$ and using (140), we obtain

$$\lim_{p\to 1} \frac{\Delta_W(p)}{\Delta_H(p)} = \lim_{p\to 1} \left\{ (1-\alpha_d) + \frac{\Delta_R(p)}{\Delta_H(p)} \right\} = 1 - \alpha_d,$$

which identifies $1 - \alpha_d$ and hence α_d. Note that we have not used the normalization in Assumption (iv) to identify α_d. Assumption (iii) will be used below to identify γ_d.

Step 2: Identification of γ_d. Rearranging (137) gives

$$R_{t,d}(p) = Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) - (\log\omega + \alpha_d \log\gamma_d).$$

Taking limits as $p \to 1$ on both sides and using the tail normalization (131), we obtain

$$L_{t,d} = \lim_{p\to 1} R_{t,d}(p) = \lim_{p\to 1} \left\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \right\} - (\log\omega + \alpha_d \log\gamma_d)$$

and, by re-arranging terms,

$$\alpha_d \log\gamma_d = \lim_{p\to 1} \left\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \right\} - \log\omega - L_{t,d}.$$

Hence,

$$\gamma_d = \exp\left(\frac{1}{\alpha_d} \left[\lim_{p\to 1} \left\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1-\alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \right\} - \log\omega - L_{t,d} \right] \right).$$

The right side of this expression is identified from the conditional joint distribution of $(w_{n,t}, H_{n,t})$ given $D_{n,t} = d$, which determines the limit of $Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1-\alpha_d) Q_{\log H_{n,t}|D_{n,t}=d}(p)$ as $p \to 1$, the known bargaining parameter ω, the known constant $L_{t,d}$, and the already identified α_d. Thus, γ_d is identified. $\square$

Remarks. We now compare the approach of Proposition H.1, Proposition H.2, and Corollary H.1 (hereafter, the "*first approach*") with that of Proposition H.3 (hereafter, the "*second approach*") to recover α_d and γ_d. The *first approach* identifies the scale function $\sigma(d, H_{n,t})$ from the upper tail of the observed selected wage distribution $w_{n,t}$ conditional on $(D_{n,t}, H_{n,t})$. Given the structural relation $\sigma(d, H_{n,t}) := \omega\, \gamma_d^{\alpha_d} H_{n,t}^{1-\alpha_d}$, α_d is then identified from *ratios* of $\sigma(d, h)$ at different values of h, which do not depend on any normalization for $\sigma(d, \cdot)$. By contrast, γ_d is identified from the *level* of $\sigma(d, h)$ at some h and thus requires a level normalization, for example, that $\sigma(d, \bar{h}) = 1$ for some $\bar{h}$.

The *second approach* does not involve the intermediate identification of $\sigma(d, H_{n,t})$, but instead directly focuses on the log wage equation expressed as

$$\log w_{n,t}(d) = \log \omega + \alpha_d \log \gamma_d + (1 - \alpha_d) \log H_{n,t} + \text{error},$$

in particular on the equilibrium relationship between the upper quantiles of $\log w_{n,t}$ and $\log H_{n,t}$ given $D_{n,t} = d$. Under Assumptions (i) and (ii), we obtain that as $p \to 1$,

$$Q_{\log w_{n,t}|D_{n,t}=d}(p) = \log \omega + \alpha_d \log \gamma_d + (1 - \alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) + o\big(Q_{\log H_{n,t}|D_{n,t}=d}(p)\big),$$

so that ratios of differences in this expression across different values of p identify the *slope* term $1 - \alpha_d$ and so the parameter α_d without the need for any normalization. The parameter γ_d is then recovered from an *intercept*- type tail normalization of the composite error term encoded in Assumption (iii). In this sense, the *second approach* resembles an asymptotic linear quantile regression of $Q_{\log w_{n,t}|D_{n,t}}(p)$ on $Q_{\log H_{n,t}|D_{n,t}}(p)$ at high levels of p: the slope $1 - \alpha_d$ is identified from the limiting ratio of quantile differences, whereas the intercept $\log \omega + \alpha_d \log \gamma_d$ is pinned down through a normalization of the extremal tail of the composite error.

Thus, both approaches are fundamentally based on *upper-tail identification*. The first approach relies on the upper tail of $w_{n,t}$ conditional on $(H_{n,t}, D_{n,t})$, whereas the second approach relies on the upper tail of $\log w_{n,t}$ and $\log H_{n,t}$ conditional on $D_{n,t}$. Under both approaches, α_d is identified through a slope argument. Instead, the recovery of γ_d requires a normalization condition.

Under the first approach, the key restriction is a *tail limit condition* (Assumption (ii) of Proposition H.1). It requires that, conditional on human capital $H_{n,t}$, the probability of working at firm d upon receiving a very high *potential* wage $w_{n,t}(d)$ converges to a firm-specific constant $q_t(d)$, as the lower bound for the potential wage to be considered large grows arbitrarily large ($w \to \infty$). This stabilisation of selection in the upper tail allows the tail of the *potential* wage distribution to be recovered from observed wages. Under the second approach, the key restriction is a *dominance condition* (Assumptions (ii) of Proposition H.3) on the quantile reminder of observed wages, which leads to a relation between $Q_{\log w_{n,t}|D_{n,t}=d}(p)$ and $Q_{\log H_{n,t}|D_{n,t}=d}(p)$ that is asymptotically ($p \to 1$) affine. From the identified slope and intercept of this relation, respectively, α_d and γ_d can be recovered.

Case 2: $H_{n,t}$ is Unobserved with Unknown Distribution and Support. In this case, we proceed in two steps. First, in Proposition H.4, we account for the fact that $\mathcal{H}_t$ is unknown and work in the human-capital *rank space* by mapping $H_{n,t}$ to its quantile (percentile) index via its CDF, namely,

$$y^\circ(d, U_{n,t}) \;:=\; y\big(d, F^{-1}_{H_{n,t}}(U_{n,t})\big) \;\text{ and }\; \sigma^\circ(d, U_{n,t}) \;:=\; \sigma\big(d, F^{-1}_{H_{n,t}}(U_{n,t})\big),$$

defined on the support of $U_{n,t}$ rather than on the support of $H_{n,t}$. Second, in Proposition H.2, we assume that $F_{\epsilon_{n,t}}$ is known and that two values of $H_{n,t}$, h_a and h_b, corresponding to the values u_a and u_b of $U_{n,t}$, are known to identify α_d, γ_d, and z.

Proposition H.4 (Identification of $y^\circ(d, U_{n,t})$ and $\sigma^\circ(d, U_{n,t})$)**.** *Given $d \in \mathcal{D}$ and $t \in \{1, \ldots, T\}$, let $\mathcal{U}_{t,d} \subseteq (0, 1)$ be the set of realizations u of $U_{n,t}$ such that* $\Pr(D_{n,t} = d \mid U_{n,t} = u) > 0$. *For each firm $d \in \mathcal{D}$ and period $t \in \{1, \ldots, T\}$, assume:*

(i) (Supports.) For each $u \in \mathcal{U}_{t,d}$, $\sup\{w : \Pr(w_{n,t}(d) \leq w \mid U_{n,t} = u) < 1\} = \infty$ and $\sup\{w : \Pr(w_{n,t} \leq w \mid D_{n,t} = d, U_{n,t} = u) < 1\} = \infty$.

(ii) (Tail Limit.) There exists an (unknown) constant $q_t(d) \in (0, 1]$ such that for every $u \in \mathcal{U}_{t,d}$, $\lim_{w\to\infty} \Pr\big(D_{n,t} = d \mid U_{n,t} = u, w_{n,t}(d) > w\big) = q_t(d)$.

(iii) (Tail Regularity.) For each $u \in \mathcal{U}_{t,d}$, there exist (unknown) thresholds $w_{u,t,d} < \infty$ and $w^{\text{obs}}_{u,t,d} < \infty$ such that the cumulative distribution functions $F_{w_{n,t}(d)|U_{n,t}=u}$ and $F_{w_{n,t}|D_{n,t}=d,U_{n,t}=u}$ are continuous and strictly increasing on $(w_{u,t,d}, \infty)$ and $(w^{\text{obs}}_{u,t,d}, \infty)$, respectively.

(iv) (Normalization.) There exists a known $\bar{u} \in \mathcal{U}_{t,d}$ with $y^\circ(d, \bar{u}) = 0$ and $\sigma^\circ(d, \bar{u}) = 1$.

Then, the functions $y^\circ(d, u)$ and $\sigma^\circ(d, u)$ are identified for each $u \in \mathcal{U}_{t,d}$ and $d \in \mathcal{D}$.

Proof: The claim is an immediate consequence of Proposition G.2 after a change of conditioning variable from the latent value $H_{n,t}$ to its rank $U_{n,t} := F_{H_{n,t}}(H_{n,t})$. Note that this reparametrisation is without loss, because by the probability integral transform, $U_{n,t}$ is uniformly distributed on $(0, 1)$, and conditioning on $H_{n,t}$ is equivalent to conditioning on $U_{n,t}$. Since the support of $H_{n,t}$ is unknown, identification can only be stated for the *rank-indexed* objects $y^\circ(d, u)$ and $\sigma^\circ(d, u)$ rather than for the cardinal objects $y(d, h)$ and $\sigma(d, h)$ at the unknown levels h. $\square$

Proposition H.5 (Identification of $F_{\epsilon_{n,t}}$)**.** *Suppose that Assumptions (i) to (iv) of Proposition H.4 holds, which imply that $y^\circ(d, u)$ and $\sigma^\circ(d, u)$ are identified for each $d \in \mathcal{D}$, $u \in \mathcal{U}_{t,d}$, and $t \in \{1, \ldots, T\}$. Moreover, for each firm $d \in \mathcal{D}$ and period $t \in \{1, \ldots, T\}$, assume $\epsilon_{n,t}(d)$ belongs to a parametric family indexed by the $p_{t,d} \times 1$ vector of parameters $\mu_{t,d} \in M_{t,d} \subseteq \mathbb{R}^{p_{t,d}}$ that is tail-ratio identifiable. Namely, fix any $u \in \mathcal{U}_{t,d}$ and choose $p_{t,d}+1$ thresholds $0 < w_0 < w_1 < \cdots < w_{p_{t,d}}$. Define the function $\Phi_{t,d,u} : M_{t,d} \to \mathbb{R}^{p_{t,d}}$ as*

$$\Phi_{t,d,u}(\mu_{t,d}) := \left(\frac{S_{\epsilon_{n,t}(d)}\left(\frac{w_1 - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)}{S_{\epsilon_{n,t}(d)}\left(\frac{w_0 - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)}, \ \ldots, \ \frac{S_{\epsilon_{n,t}(d)}\left(\frac{w_{p_{t,d}} - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)}{S_{\epsilon_{n,t}(d)}\left(\frac{w_0 - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)} \right).$$

If $\Phi_{t,d,u}$ is injective, we say that $\epsilon_{n,t}(d)$ belongs to a parametric family that is tail-ratio identifiable. Under these assumptions, for each period $t \in \{1, \ldots, T\}$:

(a) (Marginal Identification.) The parameter $\mu_{t,d}$ is identified.

(b) (Joint Identification.) If the shocks $\{\epsilon_{n,t}(d)\}_{d\in\mathcal{D}}$ are mutually independent across $d \in \mathcal{D}$, then the joint distribution of $\epsilon_{n,t}(d)$ is identified as the product of the identified marginals. Alternatively, if a copula C_{μ_t} is specified so that

$$F_{\epsilon_{n,t}}(v_1, \ldots, v_{|\mathcal{D}|}) \ = \ C_{\mu_t}\big(F_{\epsilon_{n,t}(1)}(v_1; \mu_{t,1}), \ldots, F_{\epsilon_{n,t}(|\mathcal{D}|)}(v_{|\mathcal{D}|}; \mu_{t,|\mathcal{D}|})\big) \quad \forall\, (v_1, \ldots, v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|},$$

and the copula parameter μ_t is known, then the joint distribution is identified from the identified marginals and C_{μ_t}. The joint CDF is partially identified by the sharp Fréchet–Höffding *bounds, namely, for all* $(v_1,\dots,v_{|\mathcal{D}|})\in\mathbb{R}^{|\mathcal{D}|}$,

$$\max\Big\{\sum_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d;\mu_{t,d})-(|\mathcal{D}|-1),\ 0\Big\}\ \le\ F_{\epsilon_{n,t}}(v_1,\dots,v_{|\mathcal{D}|})\ \le\ \min_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d;\mu_{t,d}).$$

Corollary H.2 (Identification of α_d, γ_d, and z)**.** *For each firm* $d\in\mathcal{D}$ *and for some period* $t\in\{1,\dots,T\}$*, assume that:*

(i) $y^\circ(d,u)$ *and* $\sigma^\circ(d,u)$ *are identified for each* $u\in\mathcal{U}_{t,d}$ *(see Proposition H.4 for sufficient conditions).*

(ii) The distribution $F_{\epsilon_{n,t}}$ *of* $\epsilon_{n,t}$ *is identified (see Proposition H.5 for sufficient conditions).*

(iii) There exist two distinct ranks $u_a\neq u_b$ *in* $\mathcal{U}_{t,d}$ *such that the corresponding levels of human capital* $h_a:=F^{-1}_{H_{n,t}}(u_a)$ *and* $h_b:=F^{-1}_{H_{n,t}}(u_b)$ *are known to the researcher.*

Then, α_d*,* γ_d*, and* z *are identified for each* $d\in\mathcal{D}$*.*

Proof: The proof proceeds in three steps.
Step 1: Identification of α_d *from* $\sigma^\circ(d,U_{n,t})$. Recall that

$$\sigma^\circ(d,u)\ =\ \sigma\big(d,F^{-1}_{H_{n,t}}(u)\big)\ =\ \omega\,\gamma_d^{\alpha_d}\left(F^{-1}_{H_{n,t}}(u)\right)^{1-\alpha_d}.$$

Pick two ranks $u_a\neq u_b$ and the corresponding levels $h_a:=F^{-1}_{H_{n,t}}(u_a)$ and $h_b:=F^{-1}_{H_{n,t}}(u_b)$. Then

$$\frac{\sigma^\circ(d,u_a)}{\sigma^\circ(d,u_b)}\ =\ \frac{\omega\,\gamma_d^{\alpha_d}\,h_a^{1-\alpha_d}}{\omega\,\gamma_d^{\alpha_d}\,h_b^{1-\alpha_d}}\ =\ \left(\frac{h_a}{h_b}\right)^{1-\alpha_d}.$$

Taking logarithms and rearranging terms yields that

$$\alpha_d\ =\ 1\ -\ \frac{\log\big(\sigma^\circ(d,u_a)/\sigma^\circ(d,u_b)\big)}{\log(h_a/h_b)}.$$

Hence, given knowledge of the two ranks u_a,u_b and their corresponding levels h_a,h_b, α_d is identified.
Step 2: Identification of γ_d *from* $\sigma^\circ(d,U_{n,t})$. Using any anchored pair (u_*,h_*) with $*\in\{a,b\}$,

$$\sigma^\circ(d,u_*)=\omega\,\gamma_d^{\alpha_d}h_*^{1-\alpha_d}\quad\Longrightarrow\quad\gamma_d=\left(\tfrac{\sigma^\circ(d,u_*)}{\omega\,h_*^{1-\alpha_d}}\right)^{1/\alpha_d},$$

which identifies γ_d.
Step 3: Identify z *from* $y^\circ(d,U_{n,t})$ *and* $F_{\epsilon_{n,t}}$. Pick any anchored pair (u_*,h_*) with $*\in\{a,b\}$. Since $y^\circ(d,u_*)=y(d,h_*)$ is identified, (α_d,γ_d) are now known, and $F_{\epsilon_{n,t}}$ is known by assumption, z is identified as

$$z=\frac{y^\circ(d,u_*)}{(1-\delta)(1-\omega)}-\delta\,\mathbb{E}_{\epsilon_{n,t}\sim F_{\epsilon_{n,t}}}\Big[f\big(S(h_*,\epsilon_{n,t};\omega,\alpha_d,\gamma_d,\delta)\big)\Big],$$

which completes the proof. □