Uncertainty, Learning about Productivity, and Human Capital Acquisition: A Reassessment of Sorting*

Áureo de Paula† Cristina Gualdani‡ Elena Pastorino§ 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' ability, and firm monopsony power. We allow workers' ability and human capital, acquired before and after entry in the labor market, to be general across firms to varying degrees. Such a framework nests and extends known models of job turnover, occupational choice, wage differentials across occupations, firms, and industries, and wage inequality 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 observables and unobservables. By exploiting our identification argument, we develop a constructive estimator of the model primitives that relies on simple extremal quantile regression methods commonly used for static selection models. Through the lens of the framework we propose, we investigate the ability of standard empirical measures of the assortativeness of matching to detect the degree of sorting in the labor market. We show that typical measures of sorting severely understate its importance because they ignore the option value of acquired human capital and information about ability for future sorting.

Keywords: Sorting, Matching, Inequality, Wage Variance, Identification, Estimation

*We are grateful to Jerome Adda, 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).

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

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

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

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

1 Introduction

Matching models of the labor market have been extensively used in both the labor and macro economics literature to study a wide range of phenomena, including workers' occupational choice and turnover across firms, 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: workers' *acquisition of new human capital* and the gradual *learning* of workers' true productivity, which may be unknown to both workers and firms when workers 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 based on workers' accumulation of new human capital and information about ability encompasses many known 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 the variability of wages across individuals and over time due to learning (Farber and Gibbons, 19996; Altonji and Pierret, 2001), and many others that nest or extend these models (Jovanovic and Nyarko (1997), Gibbons and Waldman (1999a,b), Gibbons et al. (2005), Gibbons and Waldman (2006), Lange (2007), Nagypál (2007), Antonovics and Golan (2012), Kahn and Lange (2014), and Pastorino (2024)). For reviews of the literature emphasizing the central role of uncertainty and learning about workers' ability in accounting for the dispersion of wages across workers and over the life cycle, see Gibbons and Waldman (1999a) and Rubinstein and Weiss (2006).

Despite the widespread use of these models to measure the determinants of job mobility and wage inequality, their empirical content is difficult to establish for three well-understood reasons. First, workers' career paths result from a complex process of dynamic selection based on multiple dimensions of unobservables. As a consequence, wages depend on worker, firm, and job characteristics that are typically hard to measure, serially correlated, and may endogenously evolve over time as new human capital and information about ability are acquired through employment. Second, firm, occupation, and industry choices are by their very nature discrete, leading to the standard identification challenge of dynamic discrete choice models with unobserved state variables, which are known to be nonparametrically underidentified. 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 interactive fixed-effect models inapplicable (Bonhomme et al., 2019; Freyberger, 2018). For instance, a worker of high ability with low human capital or whose ability is uncertain may prefer employment at a job at which the worker may not be very productive but that allows the worker to acquire more human capital or more information about ability, which will lead to higher wages. Then, by arbitrage, equilibrium wages in a competitive labor market depend on the relative *option value* of alternative employment possibilities that provide different human capital and information prospects. But this option value is typically highest at intermediate levels of human capital and information, at which additional capital or information may induce a worker and a firm to make different employment, hiring, or assignment decisions—hence the general nonmonotonicity, and thus nonlinearity, of wages in unobservables. In these settings, the inference about the sources of inequality is further complicated by firms' monopsony power, which has been documented to be large (Seegmiller, 2021; Lamadon et al., 2022), causing sizable systematic deviations of wages from workers' productivity.

In this paper, we establish a novel result on the identification of this general class of dynamic models with ex-ante heterogeneous firms and workers, symmetric uncertainty and learning about workers' ability, and firm monopsony power using data on only workers' jobs and wages. Our argument relies on simple conditions that accommodate arbitrary patterns of selection on endogenously time-varying unobservables, are easy to verify, and naturally lead to constructive estimators of model primitives that are straightforward to implement using common methods for static selection models. Finally, we estimate a general version of our model on U.S. data and find that it helps reconcile a key empirical puzzle: why measured sorting is typically very low despite the high degree of observed wage inequality—an outcome that matching models indeed attribute to sorting.

Formally, we study a broad class of non-stationary dynamic matching models in which a finite number of heterogeneous firms Bertrand compete for a large pool of workers in each period over a discrete time horizon of either finite or infinite length.[1] Firms differ along three dimensions observed by all: their *output* technology (how labor produces output), *human capital* technology (how on-the-job experience generates more skills), and *information* technology (how output provides information about a worker's unobserved ability). For example, a low-wage "stepping-stone" job at which a worker is not that productive in terms of current output may allow a worker to acquire much new human capital or information about ability. Conversely, a "star" job at which a worker is very

[1] Bertrand price competition provides an appealing modeling approach for markets with differentiated labor inputs since it places the bargaining power on the "long side" of the market as in any auction-like mechanism, thus allowing for a nontrivial and flexible sharing of the surplus arising from matches between firms and workers, without the need for any of the additional parameters that typical bargaining setups require, such as bargaining weights and haggling costs.

productive may not provide much human capital or information about ability. Workers also differ in three dimensions: their initial and acquired *human capital* (both observed by the model's agents, with only the initial component observed by the econometrician), *efficiency* (a latent time-invariant characteristic observed to model agents but unobserved to the econometrician), and *ability* (a latent time-invariant characteristic that is initially unobserved to both model agents and the econometrician and is gradually learned by model agents as experience accumulates). As standard in the literature, since human capital, efficiency, and ability stochastically map into output, which is publicly observed, output ("performance") provides a noisy signal that firms and workers use to update their beliefs about a worker's ability.

We characterize the set of Markov perfect equilibria in this setting and 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 the dynamic value of the foregone opportunity of human capital and information acquisition at the second-best firm in that period—a *compensating differential*. We pin down equilibrium allocations as the solution to a pseudo-planning problem in which the planner chooses each period a job for a worker among the set of each firm's preferred assignment.

Econometrically, this framework amounts to a dynamic generalized equilibrium Roy model with selection on *unobservables*, namely, the idiosyncratic match-specific productivity shocks affecting output (time varying and serially uncorrelated); a worker's efficiency (time invariant "type"); and the common beliefs about a worker's ability (time varying, serially correlated, and endogenously evolving with a worker's past job choices). A worker's efficiency and beliefs about ability determine both a worker's expected output and the compensating-differential component of wages. As argued, the latter affect wages *nonadditively* and potentially *nonmonotonically* because it captures the difference in wage returns from tomorrow onward between accepting a job today at the employing firm and at the second-best firm. As a difference in future values, it represents an endogenous dynamic payoff that generally depends on all observable and unobservable characteristics of firms and workers.

The econometric literature on the static Roy model provides methods to account for worker selection on idiosyncratic shocks that affect the wage equation additively or, more generally, monotonically, since this type of selection already arises in the static Roy setting. However, accounting for selection on the other two classes of unobservables—worker efficiency and evolving beliefs about worker ability—in the environments we consider requires a different identification strategy. Our strategy augments the standard *quantile* approach for static Roy models with a *mixture* approach, which

accounts for the multiple dimensions of unobservables in our problem. Namely, we first represent the wage distribution at any firm and time period, conditional on workers' job history and other observables, as a mixture over latent worker classes indexed by worker efficiency and by all possible histories of signals about ability. We then recover the wage distribution of each latent class from the corresponding mixture component, which is determined by the distribution of the idiosyncratic shock that governs the *selected* job and firm. Similarly, we recover the probabilities of these latent classes from the mixture weights. We can do so under the mild condition that the wage distribution can be expressed as a finite mixture whose components are (potentially continuous) Gaussian mixtures. We refer to this class of distributions as a *generalized finite mixture*, since finite mixtures of continuous Gaussian mixtures are known to approximate any distribution arbitrarily well (Bruni and Koch, 1985; Nguyen and McLachlan, 2019; Aragam et al., 2020). Hence, these mixtures are especially suited to describe general distributions contaminated by selection that do not admit a regular parametric shape, as firms' equilibrium wage distributions are in our settings.

As experience accumulates, the weights of such a mixture distribution capture the probabilities of the employment histories of workers with different observed and unobserved characteristics, including their histories of output signals. By concatenating these weights over time, we can recover from them not only the initial distribution of key unobserved states—namely, worker efficiency and the evolving beliefs about worker ability—together with their laws of motion, but also conditional choice probabilities. With the initial distributions and laws of motion of efficiency and beliefs in hand, we adapt standard quantile methods from the static Roy literature to recover the distribution of the "potential wage" at each firm, job, and time period from the identified components of the mixture described, which are contaminated by selection.

Observe that in the class of models we study, exclusion restrictions do not naturally arise—no state variable that shifts one component of the model leaves others unaffected. In fact, by interpreting the wage distribution at each firm and point in time as a mixture—the *mixture step* of our argument—it is easy to see that in general any variable that affects mixture components also affects mixture weights and vice versa. As a consequence, we cannot apply well-known identification strategies for mixture models that rely on excluded variables (see Henry et al., 2014; Compiani and Kitamura, 2016; Jochmans et al., 2017). In the second *quantile step* of our argument that addresses selection on productivity shocks, we face the same difficulty that every state variable affects both wages at all jobs and job choices. Classical identification arguments for the static Roy model instead leverage excluded regressors with rich support that shift wages only in one job, so that for some workers,

their choice of job is independent of their characteristics. Since such regressors are unavailable in our context, we adapt this "at-infinity" logic by exploiting the rich support of wages, our continuous outcome of interest. Intuitively, at extreme wage quantiles, idiosyncratic match-specific productivity shocks dominate the deterministic component of wages in governing a worker's job assignment. Namely, for very high values of a job-specific productivity shock, the corresponding job becomes by far a worker's best assignment. Hence, conditioning on working in a particular job is essentially equivalent to conditioning on a tail event of the job-specific shock. By moving from tail probabilities to quantiles and evaluating quantiles across groups of workers at suitably matched high quantiles, the contribution of productivity shocks to wages cancels out and we can recover the deterministic wage component as desired. From it, each firm's output technology, the compensating differential in wages, and the distribution of productivity shocks are immediate to back out. All these objects are key to measuring the impact of sorting on inequality, which is the focus of our empirical exercise.

To this end, the two-step identification approach described yields a natural estimator that integrates standard finite mixture-model methods, such as the `fmm` routine in Stata, and extreme-quantile regression methods, such as the `eqregsel` routine in Stata (D'Haultfoeuille et al., 2020), which we implement in our empirical exercise. We note that our approach does not require monotonicity restrictions on endogenous variables, which are common in the dynamic discrete choice literature when unobserved states are persistent, or assumptions about the dynamics of states, choices, or outcomes such as "sufficient mobility", which are common in the empirical literature on sorting.

We use the econometric approach described to measure how labor market sorting affects U.S. wage inequality. The most widely used empirical framework for this exercise is that of Abowd et al. (1999)—hereafter, AKM—which decomposes wages into worker and firm fixed effects, observable covariates, and random shocks. The impact of sorting on wage inequality is then gauged by the fraction of the total variance of wages attributable to the covariance between worker and firm effects. Empirical findings based on this framework often suggest a negligible role for sorting, as implied by the weak correlations between worker and firm effects; see, for instance, Song et al. (2019) and Card et al. (2013).[2] Building on the insights offered by the class of models we study, we argue that typical AKM estimates of the correlation between firm and worker effects tend to understate it, as they omit two key forces. First, the compensating-differential term in the wage equation dampens the direct impact of worker and firm characteristics on wages, as it compensates workers for the foregone future wage returns associated with the human capital and information they could have

[2]Bonhomme et al. (2023) show that once AKM estimates are corrected for biases stemming from workers' limited job mobility, the correlation between worker and firm effects in general increases.

acquired by accepting offers from competing firms rather than their chosen firms. For example, with persistent uncertainty about ability, a high-type worker in a low-output but steep-learning job, which offers rich training or informative feedback about how well-suited the worker is for the job, can be paid less than a low-type worker in a high-output but flat-learning job. This force tends to equalize wages across very different jobs. Second, *endogenous matching frictions*—such as the gradual resolution of uncertainty about ability—prevent high-type workers from immediately joining the most productive firms. For example, workers might temporarily choose less-productive firms that offer better training opportunities or better prospects to learn about their productivity. But this is at odds with the presumption that on average workers sort into the most productive matches given their *true time-invariant* characteristics—their fixed effect—especially since workers who turn over the most, who are key for identification, are less experienced ones so most likely to be mismatched.

To empirically validate these theoretical predictions, we provide both simulation-based and empirical evidence. In a Monte Carlo exercise, we simulate an economy that reproduces key features of our class of models. We choose model parameters so as to match the distribution of wages from the Panel Study of Income Dynamics (PSID), a representative survey of U.S. households dating back to 1968, and AKM-type moments from Song et al. (2019) estimated from Social Security Administration (SSA) data. Much like in a setting with standard omitted-variable bias, our findings suggest that when the compensating differential is *negative* under the true data-generating process—so that workers match with firms offering jobs with *more* valuable prospects for human-capital and informational gains than their competitors—the AKM estimates understate the impact of sorting on wage inequality, since the omitted compensating differential *attenuates* the measured output complementarities between firm and worker characteristics. Conversely, when the compensating differential is *positive*—so that workers match with firms offering jobs with *less* valuable prospects for human capital and information gains than their competitors—the AKM estimates *overstate* the impact of sorting, since the omitted compensating differential amplifies firm-worker complementarities.

Next, we estimate the wage equation implied by our model using U.S. matched employer-employee data from the Longitudinal Employer-Household Dynamics (LEHD) dataset, which provides quarterly earnings across 21 U.S. states from the mid 1990s to 2022. Our empirical results corroborate the findings from our simulations. In particular, the AKM estimates of the impact of sorting on wage inequality are much lower than the estimates implied by our model, which helps resolve the sorting puzzle. To further support this key finding, we conduct an exercise designed to capture the global importance of sorting. Specifically, the AKM framework measures sorting solely with

respect to a worker's *fixed* characteristic—the time-invariant efficiency type in our framework. By contrast, our setting allows workers to sort based on multiple *time-varying* characteristics—namely, their beliefs about ability and their accumulated human capital. To measure their importance, we estimate our model primitives and perform a number of random reallocation exercises that compare the observed wage distribution to counterfactual ones arising when workers and firms match at random with and without uncertainty about ability, learning, and human capital acquisition. Intuitively, if sorting matters, than these counterfactual wage distributions should exhibit markedly less dispersion and concentration at the top whenever workers and firms are not choosing the best matches. Our findings are consistent with this conjecture, which supports the view that the mechanisms we consider, especially uncertainty and learning about ability, attenuate standard measures of sorting, thus playing a potentially important role in explaining the typical findings of AKM exercises.

Literature Review. Our paper is related to a large literature on the estimation of human capital and learning models, including Heckman (1976), Cunha and Heckman (2008), Buchinsky et al. (2010), Bagger et al. (2014), and Lamadon et al. (2024); see Gibbons and Waldman (1999a), Rubinstein and Weiss (2006), and Keane et al. (2017) for reviews. Our work is the first to provide formal identification arguments for dynamic matching models in which firms are heterogeneous in their output, human capital, and information technologies and have monopsony power, whereas workers differ in both observed and unobserved (to model agents and the econometrician) persistent characteristics.

A large literature has also 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 extreme quantile regression models (Chernozhukov, 2005; Sasaki and Wang, 2024, 2025) to account for selection on unobservables in dynamic generalized equilibrium Roy models without excluded covariates. D'Haultfoeuille and Maurel (2013) propose an identification procedure for static Roy models with 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 is important for plausibly modeling the U.S. wage distribution. For sample selection in quantile regression models, see Arellano and Bonhomme (2017).

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 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 the extensive literature on sorting that builds and extends the AKM framework. This literature includes works such as Card et al. (2013), Card et al. (2018), Bonhomme et al. (2019), and Song et al. (2019), as well as studies that highlight 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).

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 in that unobservables enter in a potentially nonlinear, nonmonotone, and nonmultiplicative way, which renders the use of interactive fixed-effect methods infeasible. Moreover, unlike those papers, we allow for dynamic selection on multiple unobservables, namely, idiosyncratic productivity shocks (time-varying and serially uncorrelated), worker efficiency (time invariant), and workers and firms' beliefs about worker ability (time-varying, serially correlated, and endogenously evolving with past job choices).

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 presents the formal identification argument and derives our estimator for the model primitives. Section 5 discusses a Monte Carlo exercise that illustrates its performance and our empirical application. Appendix A examines extensions to our framework. Proofs are collected in Appendix D; Appendices E and F offer additional details.

2 Setup

We consider a canonical and broad class of non-stationary dynamic matching models in which a finite number of heterogeneous firms Bertrand compete for a large pool of workers in each period over a discrete time horizon of either finite or infinite length. Firms are heterogeneous along three dimensions, observed by both sides: their output technology (how labor produces output), human capital technology (how on-the-job experience generates more skills), and information technology (how output provides information about a worker's unobserved ability). Workers are also hetero-

geneous in three dimensions: their initial and acquired human capital (both observed by firms and workers, with only the initial component observed by the econometrician), their efficiency (a time-invariant characteristic observed by firms and workers but unobserved by the econometrician), and their ability (a time-invariant characteristic that is initially unobserved by firms, workers, and the econometrician, and is gradually learned by firms and workers as experience accumulates). Once matched, the firm–worker pair produces output and human capital accumulates. As standard in the literature, since human capital, efficiency, and ability stochastically map into output, which is publicly observed, output ("performance") provides a noisy signal that firms and workers use to update their beliefs about a worker's ability. In the following period, firms post wages anticipating these dynamics, workers choose among offers given their updated state, and the cycle repeats.

This class of models nests many existing frameworks used in both the labor and macroeconomic literature to study the determinants of occupational choice, worker turnover, firm-worker sorting, wage growth, and wage inequality. See Section 1 for key references.

Some Notational Guidance. Subscript n indexes a worker. A symbol with subscript n (for instance, X_n) denotes a random variable or vector; the corresponding symbol without the subscript and typically in lowercase (for instance, x) denotes a realization of that random object. When convenient, we make functional dependencies explicit—for example, $X_n(Z_n, W_n)$. Retaining the subscript n on $X_n(\cdot)$ indicates a random function: even after fixing realizations $Z_n = z$ and $W_n = w$, the object $X_n(z, w)$ remains stochastic due to other latent sources of randomness, which we suppress in the notation to maintain readability.

Firms. There is a finite number of heterogeneous firms, indexed by $d \in \mathcal{D} \subset \mathbb{N}$, where $2 \leq |\mathcal{D}| < \infty$. Firms produce a homogeneous good sold in a perfectly competitive market at a price normalized to 1. Each firm $d \in \mathcal{D}$ operates under a constant-returns-to-scale technology in workers' labor as the only input. Firms compete for workers by offering them wages each period for their employment during that period. The model and econometric results extend to settings in which firms comprise multiple jobs—the case we consider in our empirical application—where offers specify both a wage and a job assignment. As we proceed, we highlight features of the multi-job case that warrant special attention.

Workers. There is a large pool of workers, index by $n \in \mathbb{N}$. Upon entering the labor market, each worker n is endowed with time-invariant characteristics denoted by $H_{n,1}$, with support $\mathcal{H}$, which are observed by workers, firms, and the econometrician. These may include attributes such as gender, race, and education, that capture worker n's initial human capital. For expositional simplicity,

we assume $\mathcal{H}$ is finite; all arguments extend to continuous $H_{n,1}$, with the usual care in handling conditional probabilities and densities. Worker n has also other skills that are unobserved by the econometrician and can be distinguished into two components: e_n with support $\mathcal{E}$, which denotes worker n's efficiency (a time-invariant productivity multiplier) observed by workers and firms; and θ_n with support Θ, which denotes worker n's ability (a time-invariant skill type), initially unknown to workers and firms but gradually and symmetrically learned by all based on worker n's realized output through a process described in detail later. Both e_n and θ_n are general traits that can influence worker n's performance when employed at any firm d.[3] In the model, e_n may be scalar or multidimensional, discrete or continuous. In the econometric analysis, we assume that its support $\mathcal{E}$ is finite, thereby accommodating multidimensional types while restricting them to finitely many realisations.[4] Hereafter, we let θ_n take values in $\Theta := \{\bar{\theta}, \underline{\theta}\}$, referred to as high $(\bar{\theta})$ and low $(\underline{\theta})$ ability. This binary assumption simplifies the exposition of the learning process. We maintain the same assumption on θ_n in the econometric section. Extensions to non-finite supports (including continuous multidimensional e_n and θ_n) are provided in Appendix A.[5]

Human Capital. Hereafter, we use the letter t to denote a time period, which does not represent calendar time but rather a worker's experience in the labor market. Hence, $t = 1$ denotes the first period of worker n in the labor market.[6] As standard in the literature, worker n accumulates human capital over time through a process that depends on the initial characteristics $(H_{n,1}, e_n, \theta_n)$ and on the employment history $D_n^{t-1} := (D_{n,1}, \ldots, D_{n,t-1})$, where $D_{n,t}$ is a random variable representing the firm employing worker n in period t, with support $\mathcal{D}$. Formally, if employed by firm $d \in \mathcal{D}$ in period t, worker n with efficiency $e_n = e \in \mathcal{E}$ has accumulated a human capital $H_{n,t}(d, e)$ at the *end* of the period, given by

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

In (1), $H_{n,t}(d, e)$ is determined by two components: the *labor-input* $\ell(H_{n,1}, \kappa_{n,t}; d, e) + \epsilon_{n,t}(d, e)$

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

[4] It is straightforward to accommodate a discrete bivariate e_n in which one dimension is as described and the other is a kth-order Markov process; see Low et al. (2010) for a similar formulation.

[5] To preview Appendix A, we could allow e_n and θ_n to be continuous and multidimensional—for instance, to capture settings in which ability and output signals are conjugate normal distributions.

[6] Some workers may first appear in the dataset several years after their initial entry into the labor market. In the class of models we study, this affects only the identification of the initial distribution of beliefs about ability—the initial prior. Specifically, if workers are observed only *after* their labor market entry, our methodology recovers the prior belief about worker n's ability, θ_n, as of worker n's first appearance in the data, rather than as of the worker's labor-market entry.

and the *total factor productivity* (TFP) $a_{n,t}(d,e)$. In the labor-input component, $\kappa_{n,t} \coloneqq \kappa(H_{n,1}, D_n^{t-1})$ is a deterministic function—known to workers, firms, and the econometrician—of worker n's initial human capital $H_{n,1}$ and employment history D_n^{t-1} that captures, for example, market experience and firm-specific tenure. We denote by $\mathcal{K}_t$ the support of $\kappa_{n,t}$. $\ell(H_{n,1}, \kappa_{n,t}; d, e)$ is a (d,e)-specific function of $(H_{n,1}, \kappa_{n,t})$, known to workers and firms but unknown to the econometrician. $\epsilon_{n,t}(d,e)$ is an idiosyncratic, (d,e)-specific productivity shock (or amenity), unobserved by the econometrician and known to workers and firms.

The TFP term $a_{n,t}(d,e)$ is a (d,e)-specific random variable whose distribution may depend on $(H_{n,1}, \theta_n)$ and can vary across (d,e). Accordingly, θ_n affects $H_{n,t}(d,e)$ via the distribution of $a_{n,t}(d,e)$. Importantly, the dependence of $a_{n,t}(d,e)$ on θ_n is stochastic rather than deterministic: different realizations of $a_{n,t}(d,e)$ may arise even for the same θ_n. Therefore, once realized and observed by workers and firms, $a_{n,t}(d,e)$ serves as a *noisy* signal of ability—informative about θ_n but not perfectly revealing. We detail how this signal updates beliefs later in this section.

In most employer-employee match datasets, $a_{n,t}(d,e)$ is unobserved; thus, this will be the canonical case considered in the econometric analysis. Henceforth, we assume that $a_{n,t}(d,e) \in \mathcal{A} \coloneqq \{\bar{a}, \underline{a}\}$, interpreted as a high ($\bar{a}$) and low ($\underline{a}$) signal. As with θ_n, this binary specification is adopted for expositional simplicity. We maintain the same assumption in the econometric section; extensions to non-finite supports (including continuous multidimensional $a_{n,t}(d,e)$) are provided in Appendix A.[7]

Output Technology. Normalizing labor supply to one, (1) represents the (potential) output $Y_{n,t}(d,e)$ produced by worker n with efficiency $e_n = e \in \mathcal{E}$ at the *end* of t when employed by firm $d \in \mathcal{D}$,

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

Because the firm index d enters the function $\ell(\cdot; d, e)$ and the distributions (and realizations) of the random components $a_{n,t}(d,e)$ and $\epsilon_{n,t}(d,e)$, firms are ex-ante differentiated by their output (and human-capital) technologies. For instance, a startup may exhibit higher baseline output and steeper human-capital accumulation than a back-office operation. Moreover, as is typically the case

[7] Restricting the dependence on θ_n to operate through the distribution of $a_{n,t}(d,e)$ (and not also through $\ell(H_{n,1}, \kappa_{n,t}; d, e)$) is for expositional simplicity. More general specifications are admissible—for example, a nonseparable term $\ell(H_{n,1}, \kappa_{n,t}, a_{n,t}(d,e); d, e)$. For the purposes of equilibrium characterisation, it suffices that: (i) $H_{n,t}(d,e)$ is strictly monotone in whichever component(s) are allowed to depend on θ_n; and (ii) the dependence of $H_{n,t}(d,e)$ on θ_n is stochastic rather than deterministic. Condition (i) ensures that, upon observing the output $Y_{n,t}(d,e)$ in equation (2), workers and firms can invert the mapping and recover the *unique* realisation of the component(s) through which θ_n affects $Y_{n,t}(d,e)$—so the signal about θ_n is *well defined*—which is key for the learning process. Condition (ii) ensures that learning is *nontrivial*, that is, θ_n is not revealed after a single observation of output. Lastly, the additive separability of $H_{n,t}(d,e)$ with respect to the idiosyncratic component $\epsilon_{n,t}(d,e)$ is common in the literature on dynamic discrete choice models, and it is exploited both for the equilibrium characterization and for our identification proof.

in matching frameworks, such output (and human-capital) technologies are allowed to be tailored to worker n's characteristics and job history, as reflected in the dependence of the components in equation (2) on $H_{n,1}$, e_n, θ_n, and $\kappa_{n,t}$—see the discussion of equation (1). We present next how firms differ in their technology of information generation.

Information Technology (Learning Process). At the beginning of each period t, firms make wage offers and workers express acceptance decisions to maximize, respectively, the expected present discounted value of profits (output minus wages) and wages. Before making these decisions, firms and workers with efficiency $e_n = e \in \mathcal{E}$ observe initial human capital $H_{n,1}$, tenure and experience summarized by $\kappa_{n,t}$, and the productivity shocks $\{\epsilon_{n,t}(d,e)\}_{d\in\mathcal{D}}$ at each potential firm $d \in \mathcal{D}$. However, the TFP components $\{a_{n,t}(d,e)\}_{d\in\mathcal{D}}$ are observed only at the end of period t, after production. As a result, firms and workers do not know the potential outputs $\{Y_{n,t}(d,e)\}_{d\in\mathcal{D}}$ ex-ante and therefore make decisions based on their expectations about $\{a_{n,t}(d,e)\}_{d\in\mathcal{D}}$ and, in turn, about $\{Y_{n,t}(d,e)\}_{d\in\mathcal{D}}$. Since the distribution of each $a_{n,t}(d,e)$ depends on θ_n, these expectations depend on beliefs about θ_n. The next paragraph describes how these beliefs are formed.

Firms and workers with efficiency $e_n = e \in \mathcal{E}$ learn about θ_n based on the common observations of $Y_{n,t}(d,e)$, and so $a_{n,t}(d,e)$, at the end of each period t at the employing firm $d \in \mathcal{D}$. In this precise sense, $a_{n,t}(d,e)$ represents the public noisy signal about worker n's ability θ_n that firms and workers extract from realized output $Y_{n,t}(d,e)$. (Recall that $a_{n,t}(d,e)$ is a random function of θ_n. If $a_{n,t}(d,e)$ was a deterministic function of θ_n, then the value of θ_n could be learned in one period after observing $a_{n,t}(d,e)$, and thus learning would become trivial.) As standard in the models we nest, we focus on symmetric learning: *all* firms and workers share a common belief about θ_n in each period t. Formally, at the beginning of period $t = 1$, firms and workers with efficiency $e_n = e \in \mathcal{E}$ have a common prior belief of $\theta_n = \bar{\theta}$, $P_{n,1}(e) \coloneqq \Pr(\theta_n = \bar{\theta} \mid H_{n,1}, e_n = e)$. This prior need not coincide with the true conditional distribution of θ_n and may incorporate any learning about θ_n that has taken place before entry into the labor market, for instance, during schooling. At the end of period $t \geq 1$, firms and workers observe $Y_{n,t}(d,e)$ at the employing firm $d \in \mathcal{D}$, and thus extract the signal $a_{n,t}(d,e)$ about worker n's ability θ_n. At the beginning of period $t+1$, firms and workers update their belief about θ_n based on $a_{n,t}(d,e)$ using Bayes' rule. Assuming that the performance signals are conditionally independent over time, the updated belief of $\theta_n = \bar{\theta}$ can be defined *recursively* as

$$P_{n,t+1}(d,e) = \begin{cases} \frac{\alpha(H_{n,1},d,e)P_{n,t}(D_{n,t-1},e)}{\alpha(H_{n,1},d,e)P_{n,t}(D_{n,t-1},e)+\beta(H_{n,1},d,e)(1-P_{n,t}(D_{n,t-1},e))} & \text{if } a_{n,t}(d,e) = \bar{a}, \\ \frac{(1-\alpha(H_{n,1},d,e))P_{n,t}(D_{n,t-1},e)}{(1-\alpha(H_{n,1},d,e))P_{n,t}(D_{n,t-1},e)+(1-\beta(H_{n,1},d,e))(1-P_{n,t}(D_{n,t-1},e))} & \text{if } a_{n,t}(d,e) = \underline{a}, \end{cases} \tag{3}$$

where $P_{n,t}(D_{n,t-1}, e)$ is the belief at the start of period t with $D_{n,t-1}$ denoting worker n's employment choice at $t-1$ (the realisation of $D_{n,t-1}$ is left unspecified in the notation), and

$$\alpha(H_{n,1}, d, e) \coloneqq \Pr\big(a_{n,t}(d,e) = \bar{a} \,\big|\, H_{n,1}, D_{n,t} = d, e_n = e, \theta_n = \bar{\theta}\big),$$
$$\beta(H_{n,1}, d, e) \coloneqq \Pr\big(a_{n,t}(d,e) = \bar{a} \,\big|\, H_{n,1}, D_{n,t} = d, e_n = e, \theta_n = \underline{\theta}\big).$$

Importantly, because the terms $\alpha(H_{n,1}, d, e)$ and $\beta(H_{n,1}, d, e)$ may vary across firms d, jobs can differ in their informativeness about θ_n. Hence, firms are ex-ante differentiated not only by their output (and human-capital) technology but also by their information technology. For example, observing the same high signal $\bar{a}$ in a problem-solving role (e.g., troubleshooting unexpected issues) may raise the posterior probability that a worker is high type $\bar{\theta}$ more than observing $\bar{a}$ in a highly standardized role (e.g., processing routine transactions); the former technology is more informative. Consequently, belief updating—and thus the speed of learning about θ_n—depends on the *entire* history of jobs undertaken by worker n, as well as worker n's characteristics, $H_{n,1}$, e_n, and θ_n.

Expected Output. At the beginning of every period t—before making their decisions—firms and workers with efficiency $e_n = e \in \mathcal{D}$ calculate the *expected* output at firm $d \in \mathcal{D}$ as

$$\mathbb{E}\Big(Y_{n,t}(d,e)\Big|H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n = e, \epsilon_{n,t}\Big)$$
$$= \mathbb{E}(a_{n,t}(d,e) \mid s_{n,t}(e)) + \ell(H_{n,1}, \kappa_{n,t}; d, e) + \epsilon_{n,t}(d,e) \coloneqq y(d, s_{n,t}(e)) + \epsilon_{n,t}(d,e),$$

where the information available to firms and worker n is collected in $s_{n,t} \coloneqq (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ and $\epsilon_{n,t} \coloneqq \big(\epsilon_{n,t}(d,e) : d \in \mathcal{D},\, e \in \mathcal{E}\big)$; $P_{n,t}$ is *shorthand* for the belief $P_{n,t}(D_{n,t-1}, e_n)$ that $\theta_n = \bar{\theta}$ at the beginning of t, as recursively defined in equation (3); $s_{n,t}(e)$ denotes $s_{n,t}$ evaluated at $e_n = e$; and

$$y\big(d, s_{n,t}(e)\big) \coloneqq \mathbb{E}\big(a_{n,t}(d,e) \mid s_{n,t}(e)\big) + \ell(H_{n,1}, \kappa_{n,t}; d, e).$$

Equilibrium. Given the absence of complementarities in production among workers, to characterize the model's equilibrium, we can examine the competition of all firms for one worker at a time without any loss of generality. We adopt a refinement of the notion of Markov perfect equilibrium, which we term Robust Markov perfect equilibrium (RMPE). An RMPE consists of wage strategies for firms and an acceptance strategy for worker n, alongside a belief function such that: (i) the worker maximizes the expected present discounted value of wages; (ii) each firm maximizes the expected present discounted value of its profits; (iii) beliefs are consistently updated according to Bayes' rule; and (iv) non-employing firms are indifferent between not employing and employing the worker. Conditions (i) through (iii) define a standard MPE, under which multiple MPEs may exist.

Condition (iv) selects one of such equilibria and hence acts as a refinement condition. We provide further details on condition (iv) below. Under conditions (i)-(iv), an RMPE exists, is unique, and can be efficient (Bergemann and Välimaki, 1996).

More formally, the state that firms face at the time they make their wage offers to worker n consists of $(s_{n,t}, \epsilon_{n,t})$, and the state that worker n faces at the time they make their acceptance decisions includes $(s_{n,t}, \epsilon_{n,t})$ and the collection of all firms' wage offers that worker n receives. We denote by $w_{n,t,d} \coloneqq w_d(s_{n,t}, \epsilon_{n,t})$ the wage offer strategy of each generic firm d and by $\{w_{n,t,d}\}_{d\in\mathcal{D}}$ the collection of all wage offer strategies. We denote by $l_{n,t,d} \coloneqq l_d(s_{n,t}, \epsilon_{n,t}, \{w_{n,t,d}\}_{d\in\mathcal{D}})$ 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—and by $\{l_{n,t,d}\}_{d\in\mathcal{D}}$ the collection of all acceptance strategies.

Given firms' strategies, worker n's acceptance strategy when of efficiency type $e_n = e \in \mathcal{E}$ satisfies

$$\begin{aligned} W(s_{n,t}(e), \epsilon_{n,t}(e), \{w_{n,t,d}(e)\}_{d\in\mathcal{D}}) = \max_{\{l_{n,t,d}(e)\}_{d\in\mathcal{D}}} \sum_{d\in\mathcal{D}} l_{n,t,d}(e) \times \Big[w_{n,t,d}(e) \\ + \delta[1-\eta(\kappa_{n,t}, d)] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(W(s_{n,t+1}(e), \epsilon_{n,t+1}(e), \{w_{d,n,t+1}(e)\}_{d\in\mathcal{D}}) \mid s_{n,t}(e), d \Big) dF_e \Big]. \end{aligned} \tag{4}$$

In (4) $s_{n,t}(e)$ is the vector $s_{n,t}$ evaluated at $e_n = e \in \mathcal{E}$, $\epsilon_{n,t}(e) \coloneqq (\epsilon_{n,t}(d,e) : d \in \mathcal{D})$, $w_{n,t,d}(e) \coloneqq (w_d(s_{n,t}(e), \epsilon_{n,t}(e))$, $l_{n,t,d}(e) \coloneqq l_d(s_{n,t}(e), \epsilon_{n,t}(e), \{w_{n,t,d}(e)\}_{d\in\mathcal{D}})$, F_e is the cumulative distribution function of the vector of shocks $\epsilon_{n,t}(e)$, δ is the discount factor, and $\eta(\kappa_{n,t}, d)$ is the probability that worker n leaves the labor market at the end of period t, given the accumulated human capital investments $\kappa_{n,t}$ and the last employing firm $d \in \mathcal{D}$.[8] Note that, in (4), we assume that for each $e \in \mathcal{E}$, $\epsilon_{n,t}(e)$ is independent of $s_{n,t}(e)$, and that the vectors $\{\epsilon_{n,t}(e)\}_t$ are i.i.d. across periods t, as is standard in dynamic models. We maintain this assumption throughout. In our framework, time persistence in the state is generated through $\kappa_{n,t}$ (observed by the researcher) and $(P_{n,t}, e_n)$ (unobserved by the researcher).

Given worker n's strategy and its competitors' strategies, firm d's strategy satisfies

$$\begin{aligned} \Pi_d(s_{n,t}(e), \epsilon_{n,t}(e)) = \max_{w_{n,t,d}(e)} \Big(l_{n,t,d}(e) \times \Big[y(d, s_{n,t}(e)) + \epsilon_{n,t}(d,e) - w_{n,t,d}(e) \\ + \delta[1-\eta(\kappa_{n,t}, d)] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(\Pi_d(s_{n,t+1}(e), \epsilon_{n,t+1}(e)) \mid s_{n,t}(e), d \Big) dF_e \Big] \\ + \sum_{d'\in\mathcal{D}\setminus\{d\}} l_{d',n,t}(e) \Big\{ \delta[1-\eta(\kappa_{n,t}, d')] \int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Big(\Pi_d(s_{n,t+1}(e), \epsilon_{n,t+1}(e)) \mid s_{n,t}(e), d' \Big) dF_e \Big\} \Big). \end{aligned} \tag{5}$$

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

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

$$\begin{aligned}&\delta[1-\eta(\kappa_{n,t},d')]\int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Pi_d(\cdot|s_{n,t}(e),d')dF_e \\ &= \max_{w_{n,t,d}(e)} \Big\{ y(d,s_{n,t}(e)) + \epsilon_{n,t}(d,e) - w_{n,t,d}(e) + \delta[1-\eta(\kappa_{n,t},d)]\int_{\epsilon_{n,t+1}(e)} \mathbb{E}\Pi_d(\cdot|s_{n,t}(e),d)dF_e \Big\}.\end{aligned} \tag{6}$$

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 (6)—and employing the worker—in which case its payoff is the right side of (6). Importantly, under condition (iv), an employed worker's wage is uniquely determined—specifically, it equals the wage offered by the second-best firm plus a compensating differential, as shown in Proposition 1 below.

2.1 Equilibrium Wage

An intuition for how wages are determined can be gained by considering a static model with just two firms. Recall that in a static model of Bertrand price competition for a homogeneous product among two firms with heterogeneous output technology, the high productivity (low-cost) firm sells to a consumer at a price equal to the cost of the low-productivity (high-cost) firm, making the consumer indifferent between the two sellers. Analogously, in the static version of our model with two firms—where the two firms have heterogeneous output (and human-capital) technologies and there is no learning—worker n's wage in period t equals the worker's output *were* the worker hired by the competitor of the employing firm. Thus, in equilibrium, the worker is indifferent between employment at the employing firm and employment at its competitor. In the special case of perfect competition, where firms have identical output technology, the worker is paid their output, as the output at the non-employing firm is the same as at the employing firm.

In the dynamic version of our model with two firms differing in their output, human-capital,

and information technologies, the same indifference condition holds: in equilibrium, the worker is indifferent between the employing firm and its competitor. However, additional factors must be taken into account in this dynamic setting. Specifically, the human capital and information accumulated during employment at one firm lead to future returns for the worker. Therefore, a firm at which a worker can accumulate substantial human capital or information can afford to pay a lower wage while still employing the worker. Conversely, a firm that offers limited opportunities for accumulating human capital or information must offer a higher wage to attract the worker.

With more than two firms, a similar argument applies—in this case, the two firms competing for a worker are those offering the two highest expected present discounted values of wages. Specifically, we demonstrate that worker n's wage in period t equals the expected output the worker would produce if hired by the firm ranked as "second-best" in terms of the offered expected present discounted values of wages—akin to a *second-price auction*—plus a *compensating differential* term, which is either a *premium for the missed future returns* in terms of human capital and information acquisition that would have been gained by accepting a job at the second-best firm (and so is positive) or a *discount for the greater future returns* in terms of human capital and information acquisition that are gained by accepting a job at the first-best firm (and so is negative).

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

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

$$w_{n,t}(d, d', e) = y(d', s_{n,t}(e)) + \Psi(d, d', s_{n,t}(e)) + \epsilon_{n,t}(d', e), \quad \textit{with} \tag{7}$$

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

According to Proposition 1, a worker's wage is the sum of three terms: $y(d', s_{n,t}(e)) + \epsilon_{n,t}(d', e)$, which is the expected per-period output at d' (after the vector of productivity shocks $\epsilon_{n,t}$ is realised),

and $\Psi(d, d', s_{n,t}(e))$, which is a compensating differential. In particular, $\Psi(d, d', s_{n,t}(e))$ is the difference between two value functions: the first being the (counterfactual) future expected discounted match surplus value generated by worker n and firm d' had d' being chosen by n in period t, and the second being the future expected discounted match surplus value generated by worker n and firm d' when worker n chooses firm d in period t. Lastly, note that in the expression $w_{n,t}(d, d', e)$, the subscript (n, t) encapsulates not only the worker and time indices but also any dependence of the wage on the state $(s_{n,t}(e), \epsilon_{n,t}(e))$ that is worker- and time-specific.

An implication of Proposition 1 for multi-job firms—the case in our empirical application—is the following. When exogenous separation rates and the human-capital process are sufficiently similar across firms, if firm d's job is *more* informative than firm d''s in the Blackwell sense, the compensating differential is *negative*; if it is *less* informative, the compensating differential is *positive*. Intuitively, a firm pays less than static competition would predict when employment there delivers greater learning about ability (the worker enjoys an informational gain), and pays a premium when employment there entails forgoing such learning (an informational loss). We emphasise that we do not impose additional assumptions to guarantee this sign pattern; we note it simply as a qualitative implication that will help guide the interpretation of some of our empirical results.

Proposition 2 (Sign of Compensating Differential)**.** *When the differences $\eta(\kappa_{n,t}, d) - \eta(\kappa_{n,t}, d')$ are sufficiently small across any two firms d and d' for each $\kappa_{n,t}$ and the process of human capital acquisition is sufficiently similar across firms, the compensating differential is negative (respectively, positive) whenever performance signals at firm d are more (respectively, less) informative than performance signals at firm d'.*

2.2 Econometric Model

The model just described can be cast as a dynamic equilibrium generalised Roy model. In particular, the observed wage of worker n in period t is

$$\begin{aligned} w_{n,t} &= \sum_{(d,d')\in\mathcal{D}^2} \sum_{e\in\mathcal{E}} \mathbb{1}\{D_{n,t} = d, D'_{n,t} = d', e_n = e\} w_{n,t}(d, d', e) \\ &= \sum_{(d,d')\in\mathcal{D}^2} \sum_{e\in\mathcal{E}} \mathbb{1}\{D_{n,t} = d, D'_{n,t} = d', e_n = e\} \big[y(d', s_{n,t}(e)) + \Psi(d, d', s_{n,t}(e)) + \epsilon_{n,t}(d', e)\big], \end{aligned} \tag{8}$$

where $D_{n,t}$ denotes the employing (first-best) firm for worker n in period t, with generic realisation $d \in \mathcal{D}$; $D'_{n,t}$ denotes the second-best firm, with generic realisation $d' \in \mathcal{D}$; e_n denotes worker n's efficiency, with generic realisation $e \in \mathcal{E}$; $w_{n,t}(d, d', e)$ is the potential wage defined in equation (7); and $s_{n,t}(e)$ denotes the vector of state variables $s_{n,t} \coloneqq (H_{n,1},\, \kappa_{n,t},\, P_{n,t},\, e_n)$ evaluated at $e_n = e$.

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

Assumption 1 describes the observation scheme maintained throughout. It requires the econometrician to have access to a panel of data on wages, initial attributes, and employment choices. We keep T finite and presume elsewhere that the number of workers grows arbitrarily large. We make *minimal* data requirements to accommodate the limited information typically available in standard employer-employee matched datasets. In particular, we do not rely on the availability of variables that can facilitate the identification of the learning process, such as proxies for beliefs or direct information on performance. In Section 4.8, we show how the availability of such additional data can simplify some estimation steps under extra assumptions. To simplify the notation, we assume that the panel is balanced; however, all econometric arguments remain valid even with an unbalanced panel. The occupation choice $D_{n,t}$ can depend on all the variables entering the equilibrium wage equation, some of which are not observed by the econometrician, namely e_n, $P_{n,t}$, and $\epsilon_{n,t}$. This dependency arises from the optimising behaviour of workers and firms, leading to dynamic selection on *unobservables*.

Primitives of Empirical Interest. We show below how to identify several primitives, which enable us to study the fundamental question of measuring the impact of sorting on earnings inequality. In particular, we identify the "deterministic" wage component $\varphi(d, d', s_{n,t}(e)) \coloneqq y(d', s_{n,t}(e)) + \Psi(d, d', s_{n,t}(e))$—defined as the sum of expected output (net of productivity shocks) and the compensating differential—and the distribution of the productivity-shock vector $\epsilon_{n,t}$. We then identify the output (and human-capital) technology $y(d', s_{n,t}(e))$ and thereby disentangle the compensating differential $\Psi(d, d', s_{n,t}(e))$ from $\varphi(d, d', s_{n,t}(e))$. Given $y(d', s_{n,t}(e))$, we recover its "deterministic" labor-input component $\ell(H_{n,1}, \kappa_{n,t}; d', e)$. We also identify the law of motion for the state $s_{n,t}$, including the information technology (learning process). Finally, we identify the distribution of job choices $D_{n,t}$ conditional on $s_{n,t}$ (conditional choice probabilities, or CCPs). Throughout, we take the discount factor δ as known, as is standard in dynamic models.

3 Overview of Identification

A key primitive for us is the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$, which we then use to separately identify the output (and human-capital) technology $y(\cdot)$ and the compensating differential $\Psi(\cdot)$. As previewed in Section 2.2, however, identification of $\varphi(\cdot)$ is complicated by selection of $D_{n,t}$ based on the following unobserved state variables: (i) idiosyncratic productivity shocks $\epsilon_{n,t}$

(time-varying and serially uncorrelated); (ii) worker efficiency e_n (time-invariant); and (iii) beliefs $P_{n,t}$ of workers and firms about worker ability θ_n (time-varying, serially correlated, and endogenously evolving with past occupational choices). The shocks $\epsilon_{n,t}$ enter *additively* into the output component $y(\cdot)$ of the wage equation. Worker efficiency e_n and beliefs $P_{n,t}$ enter *both* the output $y(\cdot)$ and the compensating-differential $\Psi(\cdot)$ components; in the latter, they enter *nonadditively* and, in general, *nonmonotonically*. This is because $\Psi(\cdot)$ captures the difference in wage returns—from $t+1$ onward—between accepting today's job at the employing firm and the second-best alternative. As a difference in future values, $\Psi(\cdot)$ is an endogenous dynamic payoff that generally depends on all observable and unobservable attributes of firms and workers in an *a priori* unknown manner.

The econometric literature on the static Roy model offers tools for handling selection on idiosyncratic shocks $\epsilon_{n,t}$ that enter the wage equation additively (or, more generally, monotonically)—since such selection arises in the standard Roy setting as well. By contrast, dealing with the other two classes of unobservables—e_n and $P_{n,t}$—in the environments we consider requires a different identification strategy. Our strategy augments the standard *quantile* approach for static Roy models with a *mixture* approach.

Namely, we first represent the cross-sectional wage distribution at time t—conditional on a worker's occupational history $D_n^t \coloneqq (D_{n,1}, \ldots, D_{n,t})$ and observables $H_{n,1}$—as a mixture over latent classes indexed by e_n and by the history of noisy performance signals $a_n^{t-1} \coloneqq (a_{n,1}, \ldots, a_{n,t-1})$ about θ_n. (*To simplify notation, we henceforth write* $a_{n,t}(D_{n,t}, e_n)$ *as* $a_{n,t}$.) Since the state $s_{n,t} \coloneqq (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ is a deterministic function of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$, it follows that each mixture component is determined by the distribution of $\epsilon_{n,t}$ conditional on the *selected* employing firm in period t, $D_{n,t}$. We identify the wage distribution of each latent class from the corresponding mixture component. Similarly, we identify the probabilities of the latent classes from the mixture weights. We recover such mixture components and weights under mild conditions, namely that the wage distribution admits a *generalised finite mixture* representation: a finite mixture whose components are (potentially continuous) Gaussian mixtures. We use the term *generalised finite mixture* because finite mixtures of continuous Gaussian mixtures can approximate any distribution arbitrarily well (Bruni and Koch, 1985; Nguyen and McLachlan, 2019; Aragam et al., 2020). This class is therefore well suited to model general, selection-contaminated distributions that need not follow a standard parametric form—as in our setting, where the mixture components are contaminated by the selection of $D_{n,t}$ based on $\epsilon_{n,t}$.

As workers accumulate experience, the weights of the wage mixture capture the probabilities of

the employment histories of workers with different observed and unobserved characteristics, including their histories of performance signals. By concatenating these weights over time, we identify not only the initial distributions of the key unobserved state variables—e_n and $P_{n,t}$—together with their laws of motion, but also the CCPs. Equipped with these initial distributions and transition laws, we then adapt standard quantile methods from the static Roy literature to recover, from the identified mixture components (which, recall, are contaminated by selection of $D_{n,t}$ based on $\epsilon_{n,t}$), the deterministic wage $\varphi(\cdot)$ at each firm and time period. Lastly, with $\varphi(\cdot)$ identified, we identify the distribution of $\epsilon_{n,t}$, the output technology $y(\cdot)$, the compensating differential $\Psi(\cdot)$, and the remaining components of the output (and human-capital) equation.

A challenging feature of the class of models we study is the absence of exclusion restrictions—that is, there is no state variable that shifts one component of the model while leaving others unaffected. This feature shapes each step of our identification argument. Specifically, by interpreting the wage distribution as a mixture—the mixture step of our identification argument—any variable that affects the mixture components also affects the mixture weights, and vice versa. As a result, we cannot apply well-known identification strategies for mixture models that rely on excluded variables (see Henry et al., 2014; Compiani and Kitamura, 2016; Jochmans et al., 2017). Instead, we rely on the *generalised finite mixture* representation for identification, as mentioned above.

In the second quantile step of our identification arguments—where we address selection on $\epsilon_{n,t}$ to recover the deterministic wage $\varphi(\cdot)$—the same lack of exclusion restrictions reappears: every state variable affects wages at all jobs and job choices. Classical identification arguments for the static Roy model leverage excluded regressors with rich support that shift wages only in one job, so that for some workers, their choice of job is independent of their unobserved characteristics. Since such regressors are unavailable here, we cannot follow that route. Instead, we adapt this "at-infinity" logic by exploiting the rich support of wages. Intuitively, at extreme wage quantiles, the idiosyncratic productivity shocks $\epsilon_{n,t}$ dominate the deterministic component of wages $\varphi(\cdot)$ in determining a worker's job assignment. Namely, for very high values of a job-specific productivity shock, the corresponding job becomes by far a worker's best assignment. Hence, conditioning on working in a job is essentially equivalent to conditioning on a tail event of the shock in that job. By mapping tail probabilities to quantiles and comparing groups of workers at suitably matched high quantiles (after a simple selection re-indexing), the contribution of $\epsilon_{n,t}$ to wages cancels out, allowing us to recover the deterministic wage $\varphi(\cdot)$ as desired.

As for the rest of this overview, Section 3.1 summarises the identification literature on the static

Roy model and how we adapt it in the second step of our argument to address selection on $\epsilon_{n,t}$. Building on this summary, Section 3.2 presents our identification steps in greater detail. Section 4 contains all formal arguments—readers primarily interested in our empirical application may skip it.

3.1 A Review of the Identification of the Roy Model

As discussed by French and Taber (2011), identification for the Roy model can be achieved either semiparametrically or even nonparametrically. For an intuitive understanding of the common approach and the challenges associated with adapting it to our setting, consider, *throughout Section 3.1*, a *simplified static* version of the wage equation in (8), namely,

$$w_n = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\} w_n(d) = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n) + \epsilon_n(d)\big], \tag{9}$$

obtained by removing the dependence on the second-best firm $D'_{n,t}$—so all wage components previously indexed by d' are now indexed by d only—as well as the dependence on the efficiency type e_n and the time subscript t. Here, $\mathcal{D} \coloneqq \{0, 1\}$ denotes two job alternatives, and the compensating differential $\Psi(\cdot)$ does not arise in this static version of the model. For the purposes of this section, the state vector $s_{n,t}$ is replaced by covariates X_n, which are assumed to be *observed* by the econometrician (whereas in our setting, the law of motion for $s_{n,t}$ will be identified first via the wage-mixture step, as previewed above). The vector of shocks $\epsilon_n \coloneqq (\epsilon_n(0), \epsilon_n(1))$ is independent of X_n. As is well known, identifying the deterministic wage components $y(1, X_n)$ and $y(0, X_n)$ in equation (9) is difficult due to selection of D_n based on ϵ_n. To see why, observe that

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

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

The Case with Exclusion Restrictions. One way to address selection on ϵ_n is to rewrite (9) as

$$w_n = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n) + \lambda(d, X_n) + u_n(d)\big], \tag{10}$$

where $u_n(d) \coloneqq \epsilon_n(d) - \lambda(d, X_n)$ and hence, by construction, $\mathbb{E}(u_n(d) \mid D_n = d, X_n) = \mathbb{E}(\epsilon_n(d))$, which is tipically normalized to zero. Then, if X_n can be split into two components, $X_n^\dagger$ and X_n^*, such that $y(d, X_n)$ depends only on $X_n^\dagger$ and $\lambda(d, X_n)$ depends only on X_n^*—often referred to as *exclusion*

restrictions—it becomes possible to identify $y(d, X_n^\dagger)$, provided that certain additional assumptions on $y(\cdot)$ and $\lambda(\cdot)$ hold (Ahn and Powell, 1993; Newey, 2009; Das et al., 2003)

Another way to address selection on ϵ_n consists of relying on worker-job-specific covariates with a sufficiently rich support that influence the wage in one job only, representing another type of exclusion restriction (Heckman and Honoré, 1990). In particular, suppose we can express (9) as

$$w_n = \sum_{d \in \{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n(d)) + \epsilon_n(d)\big], \tag{11}$$

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

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

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

$$\lim_{x_0 \to -\infty} \Pr(D_n = 1 \mid X_n = (x_0, x_1)) = 1, \tag{12}$$

and by the law of total probability,

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

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

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

and $y(1, x_1)$ is identified from knowledge of $\lim_{x_0 \to -\infty} \mathbb{E}(w_n \mid D_n = 1, X_n = (x_0, x_1))$—hence, the phrase *identification at infinity* (Chamberlain, 1986; Heckman, 1990). In summary, condition (12) eliminates the impact of selection from the first moment of the wage distribution of a group of individuals with extreme values of $X_n(0)$. For this group, the expected potential wage in job

1 conditional on choosing job 1, which is *observed* from the data, is equal to the unconditional expected potential wage in job 1, which is generally *unobserved* from the data.

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

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

The Case without Exclusion Restrictions: Our Approach. Without covariates that serve as exclusion restrictions, we adapt an "at-infinity" argument that exploits the rich support of wages—rather

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

[10] *Firm*-specific covariates *fixed* at the worker level—for instance, firm size or revenues—are often available in datasets, yet do not provide enough variation for identification in the Roy model.

than excluded covariates—to address the problem of selection on ϵ_n. Specifically, far out in the upper tail—at extremely high wages—selection into a job (say, job 1) is easy to account for: conditional on receiving a very high wage in job 1, the probability of actually being observed in job 1 converges to a constant. In practical terms, selection merely rescales the extreme right tail of the observed wage distribution, so that selection just shifts the tail up or down but does *not* change how fast it thins out as wages grow. Under the assumption that the wage distribution is well-behaved in that far-right region—namely, continuous and strictly increasing, a regularity feature of many distributions (including both light-tailed and heavy-tailed)—tail probabilities and quantiles are one-to-one related. This property lets us express extreme quantiles of *observed wages* as extreme quantiles of *potential* job-1 *wages*, evaluated at a slightly adjusted quantile that corrects for the selection scaling. We then compare any group x to a reference group $\bar{x}$ whose job-1 intercept $y(1, \bar{x})$ is normalised to zero. We choose sufficiently high quantiles for each group so that, after the selection correction, both groups are effectively evaluated at the same quantile of their potential wage distributions. With this alignment, the shock component loads identically across groups and cancels when we difference the two extreme observed quantiles, without imposing any further restrictions. What remains is precisely the structural component of the wage for group x relative to the reference, $y(1, x) - y(1, \bar{x})$. Because $y(1, \bar{x})$ is normalised to zero, that difference equals the target parameter $y(1, x)$, which is therefore identified (up to a location normalisation).

We now formalize this result for the simplified static wage equation (9) in Proposition 3. We focus on job 1, but a symmetric argument applies to job 0. In Section 4, we present the analogue of Proposition 3 for our general class of dynamic models.

Proposition 3 (Deterministic Wage Component)**.** *Assume:*

(i) (Exogeneity.) $\epsilon_n(1)$ *is independent of* X_n*.*

(ii) (Supports.) For each realisation x *of* X_n*,*

$$\omega(x) \coloneqq \sup\{u : \Pr(w_n(1) \leq u \mid X_n = x) < 1\} = +\infty,$$
$$\omega_{\text{obs}}(x) \coloneqq \sup\{u : \Pr(w_n \leq u \mid D_n = 1, X_n = x) < 1\} = +\infty,$$
$$0 < \Pr(D_n = 1 \mid X_n = x) \leq 1.$$

(iii) (Tail Limit.) There exists an (unknown) constant $q_1 \in (0, 1]$ *such that for each realisation* x *of* X_n*,*

$$\lim_{w \to +\infty} \Pr\big(D_n = 1 \mid X_n = x,\ w_n(1) > w\big) = q_1.$$

(iv) (Tail Regularity.) For each realisation x of X_n, there exist (unknown) thresholds $w_x < +\infty$ and $w_x^{\text{obs}} < +\infty$ such that the cumulative distribution functions $F_{w_n(1)|X_n=x}$ and $F_{w_n|D_n=1,X_n=x}$ are continuous and strictly increasing on $(w_x, +\infty)$ and $(w_x^{\text{obs}}, +\infty)$, respectively.

(v) (Normalization.) There exists a known realisation $\bar{x}$ of X_n with $y(1, \bar{x}) = 0$.

For each realisation x of X_n, define

$$c(1,x) \;\coloneqq\; \frac{q_1}{\Pr(D_n = 1 \mid X_n = x)} \;\in (0, \infty).$$

Let $\{\tau_{\bar{x}}^{(k)}\}_{k\geq 1} \subset (0,1)$ be any sequence with $\tau_{\bar{x}}^{(k)} \to 1$ as $k \to +\infty$. Define

$$1 - \tau_x^{(k)} \;\coloneqq\; \frac{c(1,x)}{c(1,\bar{x})}\left(1 - \tau_{\bar{x}}^{(k)}\right).$$

Then,

$$\lim_{k\to+\infty}\left[\, Q_{w_n \,|\, D_n=1, X_n=x}\big(\tau_x^{(k)}\big) \;-\; Q_{w_n \,|\, D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) \,\right] \;=\; y(1,x). \tag{13}$$

Hence, $y(1,x)$ is identified (up to the location normalization at $\bar{x}$).

The proof of this result proceeds in three steps. First, under Assumptions (ii)–(iii), the right tail of the *selected* wage distribution—*observed* in the data—is asymptotically proportional to the right tail of the *potential* wage distribution—*unobserved* in the data. Namely, for each realisation x of X_n,

$$\Pr(w_n > w \mid D_n = 1, X_n = x) \;\sim\; c(1,x)\,\Pr\big(w_n(1) > w \mid X_n = x\big) \qquad (w \to +\infty), \tag{14}$$

with $c(1,x) = q_1/\Pr(D_n = 1 \mid X_n = x) \in (0,\infty)$. Intuitively, selection only multiplies the far-right tail of the potential wage distribution by the constant $c(1,x)$.

Second, Assumption (iv) lets us invert (14) on the tail. The result is the following relation between the selected quantile and the potential quantile:

$$Q_{w_n \,|\, D_n=1, X_n=x}(\tau) = Q_{w_n(1) \,|\, X_n=x}\Big(1 - \tfrac{1-\tau}{c(1,x)} + o_x(1-\tau)\Big), \qquad \tau \to 1, \tag{15}$$

where the remainder satisfies $o_x(1-\tau)/(1-\tau) \to 0$ as $\tau \to 1$. By Assumption (i) and the decomposition $w_n(1) = y(1,x) + \epsilon_n(1)$, $Q_{w_n(1) \,|\, X_n=x}(u) = y(1,x) + Q_{\epsilon_n(1)}(u)$, so (15) becomes

$$Q_{w_n \,|\, D_n=1, X_n=x}(\tau) = y(1,x) + Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau}{c(1,x)} + o_x(1-\tau)\Big), \qquad \tau \to 1. \tag{16}$$

Third, we apply (16) twice: first at (x, τ_x) and then at $(\bar{x}, \tau_{\bar{x}})$, where the levels $\tau_x, \tau_{\bar{x}} \to 1$ are

chosen so that the inner indices match,

$$1 - \frac{1-\tau_x}{c(1,x)} \;=\; 1 - \frac{1-\tau_{\bar{x}}}{c(1,\bar{x})}. \tag{17}$$

In particular, one convenient choice that guarantees (17) is $1-\tau_x = \frac{c(1,x)}{c(1,\bar{x})}(1-\tau_{\bar{x}})$. Using (17) and the normalization $y(1,\bar{x}) = 0$ from Assumption (v), we obtain

$$\begin{aligned} Q_{w_n \mid D_n=1, X_n=x}(\tau_x) &= y(1,x) \;+\; Q_{\epsilon_n(1)}\Big(u + o_x(1-\tau_x)\Big) \qquad \tau_x \to 1, \\ Q_{w_n \mid D_n=1, X_n=\bar{x}}(\tau_{\bar{x}}) &= 0 \;+\; Q_{\epsilon_n(1)}\Big(u + o_{\bar{x}}(1-\tau_{\bar{x}})\Big) \qquad \tau_{\bar{x}} \to 1. \end{aligned}$$

Subtracting the second display from the first yields

$$\begin{aligned} Q_{w_n \mid D_n=1, X_n=x}(\tau_x) \;-\; Q_{w_n \mid D_n=1, X_n=\bar{x}}(\tau_{\bar{x}}) = \; & y(1,x) \;+\; Q_{\epsilon_n(1)}\Big(u + o_x(1-\tau_x)\Big) \\ & -\; Q_{\epsilon_n(1)}\Big(u + o_{\bar{x}}(1-\tau_{\bar{x}})\Big) \qquad \tau_x, \tau_{\bar{x}} \to 1. \end{aligned} \tag{18}$$

Since $o_x(1-\tau_x),\ o_{\bar{x}}(1-\tau_{\bar{x}}) \to 0$ and $Q_{\epsilon_n(1)}$ is continuous near 1 by Assumption (iv), the difference of the two error–quantile terms in (18) vanishes as $\tau_x, \tau_{\bar{x}} \to 1$. Hence the left–hand side converges to $y(1,x)$. In summary, the proof hinges on two ideas: (a) selection preserves the *rate* of tail decay up to a constant, and (b) by working with quantiles and carefully *reindexing* the tail probability, we can subtract out the shock and recover the deterministic component $y(1,x)$.

To clarify Assumptions (i)–(v) in Proposition 3: Assumption (i) is the standard exogeneity condition in Roy models. Assumption (ii) requires that both the potential wages $w_n(1) \mid X_n = x$ and the observed, selected wages $w_n \mid (D_n = 1, X_n = x)$ have unbounded right support. This requirement is not essential: a bounded-right-endpoint analogue tracks convergence to the finite right endpoint—rather than to $+\infty$—with only minor modifications. In particular, exactly one of the following cases obtains: (a) $\omega(x) = \omega_{\text{obs}}(x) = +\infty$; (b) $\omega(x) = \omega_{\text{obs}}(x) < +\infty$; (c) $\omega_{\text{obs}}(x) < \omega(x) \leq +\infty$.[11] Case (a) is the setting covered by Proposition 3. Under (b), Proposition 3 and its proof go through with minimal edits—replace limits as $w \to +\infty$ with limits as $w \to \omega(x)$. Under (c)—where the right endpoint of the observed, selected wage distribution can differ from (and be finite relative to) that of the potential wage distribution, so selection affects not only the distributional shape but also the support of observed wages—the identification result retains the spirit of Proposition 3, but extra care is needed in taking limits because the two endpoints differ. In Appendix B.2, we treat case (c) and further show that, when finite, the right and left endpoints of the potential wages $w_n(1) \mid X_n = x$ and shock $\epsilon_n(1)$ can be nonparametrically identified. Assumption (ii) also requires that, for every

[11] We ignore $\omega_{\text{obs}}(x) > \omega(x)$ as $\text{supp}(w_n | D_n = 1, X_n = x) \subseteq \text{supp}(w_n(1) | X_n = x)$ implies $\omega_{\text{obs}}(x) \leq \omega(x)$.

realisation x of X_n, the probability of choosing job 1 is strictly positive, $\Pr(D_n = 1 \mid X_n = x) > 0$. This is only for expositional simplicity: the framework also allows $\Pr(D_n = 1 \mid X_n = x) = 0$ for some x, in which case $y(1, x)$ is not identified at those x.

Assumption (iii) imposes a common tail-selection limit $q_1 \in (0, 1]$, independent of x, for $\Pr(D_n = 1 \mid X_n = x,\ w_n(1) > w)$ as $w \to +\infty$. Existence and positivity of this limit imply that the right tail of the selected wages, $\Pr(w_n > w \mid D_n = 1, X_n = x)$, and the right tail of the potential wages, $\Pr(w_n(1) > w \mid X_n = x)$, are asymptotically proportional as $w \to +\infty$ (equation (14)). Invariance of the limit across x—that is, when the potential wage for job 1 is very large, the effect of X_n on the probability of selecting job 1 becomes negligible—ensures that the indices $\{\tau_x^{(k)}\}_{k\geq 1}$ in the identification claim (20) can be computed from the data *without* knowing q_1. Specifically,

$$1 - \tau_x^{(k)} := \frac{c(1,x)}{c(1,\bar{x})}\left(1 - \tau_{\bar{x}}^{(k)}\right) = \frac{\Pr(D_n = 1 \mid X_n = \bar{x})}{\Pr(D_n = 1 \mid X_n = x)}\left(1 - \tau_{\bar{x}}^{(k)}\right).$$

For a micro-foundation of Assumption (iii), see Lemma 1 in Appendix B.1, which reproduces Corollary 4.1 in D'Haultfoeuille and Maurel (2013).

Assumption (iv) is a tail-regularity condition: continuity and strict monotonicity of the relevant CDFs on far-right intervals ensure a one-to-one mapping between tail probabilities and quantiles, which justifies the quantile reindexing step of the proof (equation (15)). This condition is satisfied by many parametric families, including both thin-tailed and fat-tailed distributions. Finally, Assumption (v) is a location normalisation: as in standard Roy models, wages are identified only up to an additive constant. Fixing $y(1, \bar{x}) = 0$ pins down the wage level in our setting. Alternatively, the error term can be normalised to have zero unconditional mean or median (French and Taber, 2011).[12]

To complete the argument, we now consider the unconditional joint distribution of the vector of shocks $\epsilon_n := (\epsilon_n(1), \epsilon_n(0))$. A well-known negative result by Tsiatis (1975) shows that, in competing-risks models without covariates, the joint distribution of the latent risks is not identified. By contrast, Heckman and Honoré (1989) establish that under sufficiently rich covariate variation—specifically, with at least as many continuous covariates as there are causes of failure among other conditions—the joint distribution can be identified nonparametrically. Translated to the Roy setting studied here, this implies that without at least as many continuous covariates as there are jobs, one cannot nonparametrically identify the joint distribution of ϵ_n. Standard matched employer–employee

[12] See also D'Haultfoeuille and Maurel (2013), who identify the deterministic wage component in a static Roy model without exclusion restrictions by exploiting the extreme tails of the shock distribution. Our Assumption (iii) corresponds to their Assumption 3. Whereas they work under a thin-tailed assumption (their Assumption 2), we extend the argument to allow wage (and log-wage) distributions with fat tails, which is important for plausibly modeling the U.S. wage distribution.

data, for instance, the LEHD for the United States, typically record worker attributes in coarse, discrete categories such as education and intervals or occupation codes. Moreover, in our dynamic framework, a state variable that can be treated as approximately continuous may exist, the belief $P_{n,t}$, but, as discussed earlier, it is a *single* probability distribution over a worker's general ability. Consequently, the requisite continuous variation *per alternative* is absent, rendering the Heckman and Honoré (1989) strategy infeasible in our setting. In view of this, we proceed by focusing on the marginal distributions of $\epsilon_n(1)$ and $\epsilon_n(0)$. To recover their joint distribution, we either add an explicit independence assumption, impose a parametric copula, or work with Fréchet–Höffding bounds for partial identification.[13]

Regarding the marginal shock distributions, we have seen above that, under the assumptions of Proposition 3, in the far-right tail the survival function of the observed (selected) wages is asymptotically proportional to the survival function of the corresponding potential wages (equation (14)). Intuitively, selection stops "tilting" the tail and only rescales it by a constant. Because this rescaling cancels when we take *ratios* of tail probabilities (or differences of high quantiles), the proposition lets us nonparametrically recover the *shape* of each shock's extreme right tail: how fast tail probabilities decay, how heavy the tail is, high-quantile growth rates, and extreme support points when finite. What this does *not* deliver is the full marginal distribution: the proposition yields asymptotic tail information but leaves the interior unrestricted. If, however, one specifies a parametric family for $\epsilon_n(1)$ and $\epsilon_n(0)$, the same tail-ratio argument produces a finite system whose solution identifies the parameter vectors governing the two marginals.[14]

Corollary 1 (Identification of the Shock Distribution)**.** *Let Assumptions (i) to (v) of Proposition 3 hold for each $d \in \{0,1\}$ so that $y(d,x)$ is identified for each $d \in \{0,1\}$ and realisation x of X_n.*

(a) (Marginal Identification.) Assume $\epsilon_n(1)$ belongs to a known parametric family indexed by the $p_1 \times 1$ vector or parameters $\mu_1 \in M_1 \subseteq \mathbb{R}^{p_1}$. Fix any realisation x of X_n and choose p_1+1

[13] Assuming independence between the shocks $\epsilon_n(1)$ and $\epsilon_n(0)$ in the static Roy model (9) can be restrictive, because these shocks are the sole source of unobserved heterogeneity and may, in principle, embody substantial correlation across potential wages. By contrast, in our broader class of dynamic models this assumption is less consequential: structural correlation across potential wages is captured by latent state variables—e_n and $P_{n,t}$—so the productivity shocks can be treated as residual errors.

[14] Unlike in dynamic discrete choice models, where the parameters governing the distribution of idiosyncratic shocks are typically not point identified, here we can obtain point identification of these parameters by exploiting the additional information provided by the wage distribution.

distinct large thresholds $0 < w_0 < w_1 < \cdots < w_{p_1}$. *Define the function*

$$\Phi_{1,x} : M_1 \to \mathbb{R}^{p_1}, \;\; \Phi_{1,x}(\mu_1) \;:=\; \left(\frac{S_{\epsilon_n(1)}(w_1 - y(1,x); \mu_1)}{S_{\epsilon_n(1)}(w_0 - y(1,x); \mu_1)}, \ldots, \frac{S_{\epsilon_n(1)}(w_{p_1} - y(1,x); \mu_1)}{S_{\epsilon_n(1)}(w_0 - y(1,x); \mu_1)} \right),$$

where $S_{\epsilon_n(1)}$ *denotes the survival function of* $\epsilon_n(1)$. *If* $\Phi_{1,x}$ *is injective, then the parameter* μ_1 *is identified. An analogous statement holds for* $\epsilon_n(0)$.

(b) (Joint Identification.) Let F_{ϵ_n} *denote the joint CDF of* ϵ_n, *and* $F_{\epsilon_n(1)}(\cdot; \mu_1)$, $F_{\epsilon_n(0)}(\cdot; \mu_0)$ *the identified marginal CDFs from part (a).*

- *Independence. If* $\epsilon_n(1)$ *and* $\epsilon_n(0)$ *are mutually independent, then*

$$F_{\epsilon_n}(v_1, v_0) \;=\; F_{\epsilon_n(1)}(v_1; \mu_1)\, F_{\epsilon_n(0)}(v_0; \mu_0) \quad \forall\, (v_1, v_0) \in \mathbb{R}^2.$$

- *Parametric copula. If a copula* C_μ *is specified so that*

$$F_{\epsilon_n}(v_1, v_0) \;=\; C_\mu\big(F_{\epsilon_n(1)}(v_1; \mu_1),\, F_{\epsilon_n(0)}(v_0; \mu_0)\big) \quad \forall\, (v_1, v_0) \in \mathbb{R}^2,$$

  and the copula parameter μ *is known, then* F_{ϵ_n} *is identified.*

- *No dependence restrictions (partial identification). Absent further restrictions on the dependence between* $\epsilon_n(1)$ *and* $\epsilon_n(0)$, *the joint CDF is partially identified by the sharp Fréchet–Höffding bounds:*

$$\max\Big\{F_{\epsilon_n(1)}(v_1; \mu_1) + F_{\epsilon_n(0)}(v_0; \mu_0) - 1,\; 0\Big\} \;\le\; F_{\epsilon_n}(v_1, v_0) \;\le\; \min\Big\{F_{\epsilon_n(1)}(v_1; \mu_1),\; F_{\epsilon_n(0)}(v_0; \mu_0)\Big\} \quad \forall\, (v_1, v_0) \in \mathbb{R}^2.$$

In Appendix D, which contains the proof of Corollary 1, we provide examples of common parametric families that satisfy the injectivity condition in (a), including both thin-tailed and heavy-tailed distributions.

We conclude by noting that the quantile approach in Proposition 3 extends to wage specifications in which the shock $\epsilon_n(1)$ is multiplied by a scale function $\sigma(1, X_n)$; see Proposition 13 in Appendix B.4. This covers, for example, equilibrium wage equations arising in search models, where conditional heteroskedasticity is an inherent feature (Bagger et al., 2014); see Appendix C for the extension of our identification arguments to search models. More broadly, our quantile approach does not rely on the exact mechanism that generates job choices $D_{n,t}$ in the class of models we study

and can accommodate arbitrary dependence between those choices and the unobserved shocks $\epsilon_{n,t}$. It is therefore quite general and applies to any class of models that produces a wage equation with a structure resembling equation (9).[15][16]

3.2 Our Identification Approach

Here we re-sketch our identification strategy, organised by the classes of primitives of interest and building on the review in Section 3.1. The full, formal arguments appear in Section 4.

Information Technology, Deterministic Wage Component, and Shock Distribution. Return to the general wage equation (8). To adapt Proposition 3 to our setting for identifying the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$, we must first know the distribution of $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$. The sampling process does not reveal this distribution directly because $P_{n,t}$ and e_n are unobserved. Consequently, we must identify it. We proceed in three steps. First, we express the distribution of $w_{n,t}$ conditional on $(H_{n,1}, D_n^t)$—which is observed under Assumption 1—as a mixture over worker n's efficiency e_n and signal history $a_n^{t-1} \coloneqq (a_{n,1}, \ldots, a_{n,t-1})$. Using existing results on the identification of mixture models—in particular, the assumption that the wage distribution admits a *generalised finite mixture* representation (a finite mixture whose components are (potentially continuous) Gaussian mixtures)—we identify the mixture weights and components of the wage mixture (Proposition 4).

Second, by concatenating the mixture weights across periods, we identify the signal distribution conditional on the latent ability θ_n and the prior belief function. In turn, we recover the posterior belief in each period by recursively computing it from (3) (Proposition 5).

Third, given the identification of the learning process and the fact that, in the model, the vector of state variables $s_{n,t}$ is a deterministic function of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$, it follows that we identify the distribution of $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$ (Proposition 9).

Having recovered the distribution of $w_{n,t}$ given $(D_{n,t}, s_{n,t})$, we adapt the quantile argument of Proposition 3 to identify the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ (Proposition 10). Lastly, once this deterministic wage component is recovered, we identify the unconditional distribu-

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

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

tion of the vector of productivity shocks $\epsilon_{n,t}$ by adapting Corollary 1 (Corollary 4).

Law of Motion of the State and CCPs. The wage mixture weights not only help us identify the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ but are also key to identifying other objects of interest, such as the law of motion of the state variables, $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$, and the CCPs, $\Pr(D_{n,t} \mid s_{n,t})$ (Propositions 7 and 8). Intuitively, recall that the mixture weights essentially determine the distribution of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ in each period, which in turn governs the state variables $s_{n,t}$ and subsequently the occupation choices. Thus, by appropriately combining these weights across periods, it becomes natural to recover the law of motion of $s_{n,t}$ and the CCPs. Notably, in contrast to the typical method of recovering CCPs from agents' discrete choices in dynamic models, here the CCPs are identified from the continuous part of the data, that is, the wage distribution.

Output (and Human-Capital) Technology and Compensating Differential. Beyond identifying the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot)+\Psi(\cdot)$, we further separate within $\varphi(\cdot)$ the output (and human-capital) technology $y(\cdot)$ from the compensating differential $\Psi(\cdot)$. Specifically, the market-wide job allocation can be represented as a pseudo-planner (single-agent) dynamic discrete decision problem. Therefore, given the CCPs and the distribution of productivity shocks, $y(\cdot)$ is identified by standard arguments for dynamic discrete choice models (for instance, Magnac and Thesmar, 2002; see Proposition 11). With $y(\cdot)$ in hand, $\Psi(\cdot)$ follows residually from $\varphi(\cdot)$ (Corollary 5, Part I). Similarly, once $y(\cdot)$ is known, we can net out the deterministic labor-input component $\ell(\cdot)$, thereby completing the recovery of the output (human-capital) technology (Corollary 5, Part II).

4 Formal Identification Argument

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

4.1 Relevant Market

Our overview in Section 3 has been silent on the role played by the second-best firm $D'_{n,t}$, which appears in the wage equation and remains unobserved under Assumption 1. This was intentional: we wanted the reader to focus on other, more pressing identification challenges that arise in our framework. With the standard data assumed under Assumption 1 and in the absence of further restrictions on the model, $D'_{n,t}$ is *not* identified, as is well understood in the literature. This paper does not provide new results on that front. In this section, we therefore introduce a standard assumption that allows us to sidestep this non-identification problem and proceed with the analysis.

Assumption 2 (Relevant Market)**.**(i) In each period t, conditional on the worker's state $s_{n,t}$, the set of firms making offers—worker n's "relevant market" or "choice set"—depends only on $s_{n,t}$ and not

on $\epsilon_{n,t}$, and has size 2. We denote this relevant market by $\mathcal{D}_t(s_{n,t}) \subseteq \mathcal{D}$, with $|\mathcal{D}_t(s_{n,t})| = 2$. (ii) The correspondence $s_{n,t} \mapsto \mathcal{D}_t(s_{n,t})$ is known to the econometrician. $\diamond$

Assumption 2(i) requires that, in each period t, the set of firms offering a job to worker n (worker n's "relevant market" or "choice set") is a function of $s_{n,t}$ only and contains exactly two firms. Limiting offers to a small number of firms is realistic in many labor markets. Restricting this set to depend only on $s_{n,t}$ and to contain exactly two offers is technically helpful: conditional on the first-best firm $D_{n,t}$ and the state $s_{n,t}$, the second-best firm $D'_{n,t}$ entering the wage equation (8) is already implicitly conditioned on and need not be modelled as an additional stochastic index. Therefore, the distribution of $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$ is fully determined by the distribution of $\epsilon_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$, which is used in the arguments below.

Assumption 2(ii) requires that the correspondence from $s_{n,t}$ to worker n's choice set $\mathcal{D}_t(s_{n,t})$ be known to the econometrician. Combined with Assumption 2(i), this implies that, conditional on the first-best firm $D_{n,t}$ and the state $s_{n,t}$, the second-best firm $D'_{n,t}$ is not only already implicitly conditioned on but also known to the researcher. While Assumption 2(ii) is not required to identify the information technology, law of motion of the state, and CCPs, we exploit it to identify components of the wage equation that are indexed by both the first- and second-best firm, namely $\varphi(\cdot)$, $y(\cdot)$, and $\Psi(\cdot)$, in order to avoid any labelling indeterminacy. We will be explicit about when and how each part of Assumption 2 is used.

This two-firm, known choice-set assumption is standard in the empirical labor literature and preserves the familiar incumbent-poacher structure of search models.

4.2 Wage Mixture

In this section, we represent the cross-sectional wage distribution at time t, conditional on worker n's observed initial human capital $H_{n,1}$ and occupational history $D_n^t \coloneqq (D_{n,1}, \ldots, D_{n,t})$, as a mixture over latent classes indexed by the efficiency type e_n and by the history of noisy performance signals $a_n^{t-1} \coloneqq (a_{n,1}, \ldots, a_{n,t-1})$ about θ_n. Specifically, by the law of total probability, the conditional distribution of wages $w_{n,t}$ can be expressed as the following mixture:

$$\begin{aligned}\Pr\big(w_{n,t} \leq w \mid H_{n,1},\ D_n^t\big) = \sum_{(e,a^{t-1}) \in \mathcal{E} \times \mathcal{A}^{t-1}} & \Pr\big(w_{n,t} \leq w \mid H_{n,1},\ D_n^t,\ e_n = e,\ a_n^{t-1} = a^{t-1}\big) \\ & \times \Pr\big(e_n = e,\ a_n^{t-1} = a^{t-1} \mid H_{n,1},\ D_n^t\big),\end{aligned} \tag{19}$$

where the sets $\mathcal{E}$, $\mathcal{A}$, and $\mathcal{A}^{t-1}$ denote the (unconditional) finite supports of e_n, $a_{n,t}$, and a_n^{t-1}, respectively; e and a^{t-1} denote generic realisations of e_n and a_n^{t-1}, respectively, with $a^{t-1} \coloneqq (a_1, \ldots, a_{t-1})$;

$\Pr\big(w_{n,t} \leq w \mid H_{n,1}, D_n^t, e_n = e, a_n^{t-1} = a^{t-1}\big)$ corresponds to a mixture component; and $\Pr\big(e_n = e, a_n^{t-1} = a^{t-1} \mid H_{n,1}, D_n^t\big)$ corresponds to the associated mixture weight. In what follows, we show identification of these mixture components and weights.

Before presenting the assumptions and results, we introduce some useful notation. Because of Assumption 2(i), not all realisations of the observables $(H_{n,1}, D_n^t)$ in $\mathcal{H} \times \mathcal{D}^t$ need occur with strictly positive probability. We henceforth denote by $\mathcal{H}_t^{\text{eff}} \subseteq \mathcal{H} \times \mathcal{D}^t$ the set of realisations (h, d^t) of $(H_{n,1}, D_n^t)$ such that $\Pr(H_{n,1} = h, D_n^t = d^t) > 0$, with $d^t \coloneqq (d_1, \ldots, d_t)$. Similarly, conditional on $(H_{n,1}, D^t)$, not all realisations of the unobservables (e_n, a_n^{t-1}) in $\mathcal{E} \times \mathcal{A}^{t-1}$ need occur with strictly positive probability. We henceforth denote by $\mathcal{L}_{h,d^t}^{\text{eff}} \subseteq \mathcal{E} \times \mathcal{A}^{t-1}$ the set of realisations (e, a^{t-1}) of (e_n, a_n^{t-1}) such that $\Pr\big(e_n = e,\ a_n^{t-1} = a^{t-1} \mid H_{n,1} = h,\ D_n^t = d^t\big) > 0$. Assumption 3 sets out the conditions used to identify the mixture components and weights in equation (19).

Assumption 3 (Generalised Finite Mixture)**.**For each $t \geq 1$ and $(h, d^{t-1}, d) \in \mathcal{H}_t^{\text{eff}}$, assume:

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

$$f_{d'_t,e}(r \mid H_{n,1} = h, D_n^t = d^t, e_n = e, a_n^{t-1} = a^{t-1}) = \int_{(\mu,\sigma^2) \in \mathcal{G}_{h,d^t,e,a^{t-1}}} \mathcal{N}(r; \mu, \sigma^2) d\pi(\mu, \sigma^2; h, d^t, e, a^{t-1}),$$

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

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

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

(iv) *(Continuity and Measurability.)* For each $(e, a^{t-1}) \in \mathcal{E} \times \mathcal{A}^{t-1}$, the map

$$(\mu, \sigma^2) \mapsto \mathcal{N}(r; \mu, \sigma^2)$$

is continuous on $\mathcal{G}_{h,d^t,e,a^{t-1}}$ for every $r \in \mathbb{R}$, and the map

$$(r, \mu, \sigma^2) \mapsto \mathcal{N}(r; \mu, \sigma^2)$$

is Borel-measurable on $\mathbb{R} \times \mathbb{R} \times (0, \infty)$.

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

$\diamond$

Proposition 4 formalises the identification result under Assumption 3.

Proposition 4 (Wage Mixture)**.** *Let Assumptions 1, 2(i), and 3 hold. Then, for each* $1 \leq t \leq T$ *and* $(h, d^t) \in \mathcal{H}^{\textit{eff}}_t$*:*

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

(ii) The probability $\Pr\big(w_{n,t} \leq w \mid H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1}\big)$ *is identified for each* $(e, a^{t-1}) \in \mathcal{L}^{\textit{eff}}_{h,d^t}$ *and* $w \in \mathbb{R}$*.*

(iii) The set $\mathcal{L}^{\textit{eff}}_{h,d^t} \subseteq \mathcal{E} \times \mathcal{A}^{t-1}$ *is identified.*

Assumption 3(i) imposes that, conditional on $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1})$, the productivity shock of the second-best firm d'_t at time t, $\epsilon_{n,t}(d'_t, e)$, is distributed as a mixture of a (possibly uncountable) family of Gaussian distributions. Combined with Assumptions 2(i) and 3(ii), this implies that the wage mixture in equation (19) is a *finite mixture whose components are (possibly uncountable) Gaussian mixtures*.[17] To see why, recall our wage equation (8) and that, in the model, the state vector $s_{n,t} \coloneqq (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ is a deterministic function of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$. Hence, by conditioning each mixture component in (19) on $(H_{n,1}, D_n^t, e_n, a_n^{t-1})$, we implicitly condition

[17] We remark that the wage mixture in equation (19) is *not* assumed to be a finite mixture of Gaussians.

on the state $s_{n,t}$ entering (8) as well. In turn, under Assumption 2(i), this also entails implicitly conditioning on the second-best firm $D'_{n,t}$ entering (8). Therefore, each mixture component $\Pr\big(w_{n,t} \leq w \mid H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1}\big)$ in (19) is fully determined by the distribution of the productivity shock of the second-best firm $D'_{n,t} = d'_t \in \mathcal{D}$ at time t, $\epsilon_{n,t}(d'_t, e)$, conditional on $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1})$. By Assumption 3(i), this distribution is a mixture of a (possibly uncountable) family of Gaussian distributions, and since e_n and a_n^{t-1} can take only finitely many values under Assumption 3(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.

We call a distribution admitting a finite mixture representation whose components are (possibly uncountable) Gaussian mixtures a *generalised finite mixture*. Such two-layer mixture models are known to approximate any distribution arbitrarily well (Nguyen and McLachlan, 2019). This class is therefore well suited to model general distributions that need not follow a standard parametric form. This generality is particularly important in our setting, where, as explained above, each mixture component in (19) is fully determined by the distribution of the productivity shock of the second-best firm, $\epsilon_{n,t}(d'_t, e)$, conditional on $(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1})$ and, therefore, is "contaminated" by workers selecting jobs based on the unobserved vector of shocks $\epsilon_{n,t}$. We should *not* expect this conditional distribution to coincide with the unconditional distribution of $\epsilon_{n,t}(d'_t, e)$ (that is, the family need *not* be closed under conditioning), nor to have a "standard" parametric form such as Normal or Gumbel.[18] Instead, this conditional distribution is endogenously determined by how workers and firms make decisions within the model. Therefore, we must rely on assumptions that allow for *very flexible* distributions, as Assumption 3(i) does.

Assumption 3(ii) posits that efficiency e_n and signal a_n^{t-1} have finite unconditional supports, $\mathcal{E}$ and $\mathcal{A}^{t-1}$, with known cardinalities. Proposition 4(iii) identifies the conditional support of these random objects, namely the subset $\mathcal{L}^{\text{eff}}_{h,d^t} \subseteq \mathcal{E} \times \mathcal{A}^{t-1}$. In Appendix A, we discuss how to relax this assumption and allow e_n and $a_{n,t}$ to be continuous multidimensional random vectors.

Assumptions 3(iii) and (iv) are regularity conditions requiring that all Gaussian means and variances (μ, σ^2) that can arise lie in a bounded rectangle, with variances bounded away from 0 and ∞, and that all Gaussian densities involved vary continuously with (μ, σ^2), with the kernels measurable

[18] By "closure under conditioning" we mean that, if a random vector ϵ follows a given parametric family F_θ, then for any selection event A defined in terms of ϵ, the conditional distribution $\epsilon \mid A$ still belongs to the same family, i.e. $\epsilon \mid A \sim F_{\theta'}$ for some θ'. This is a very strong requirement and is satisfied only by a few special families, such as i.i.d. type-I extreme value (Gumbel) shocks in multinomial logit models.

in all their arguments.

Lastly, Assumption 3(v) is a standard separation condition in the identification of mixture models and requires that the mixing distributions $\{\pi(\cdot; h, d^t, e, a^{t-1}) : (e, a^{t-1}) \in \mathcal{L}^{\text{eff}}_{h,d^t}\}$ place all their mass on pairwise disjoint sets $G_{h,d^t,e,a^{t-1}} \subseteq \mathcal{G}_{h,d^t,e,a^{t-1}}$ in the (μ, σ^2)-space. Otherwise, they could not be separately distinguished. Importantly, it does not require the densities $\{f_{d'_t,e}(r \mid H_{n,1} = h, D^t_n = d^t, e_n = e, a^{t-1}_n = a^{t-1}) : (e, a^{t-1}) \in \mathcal{L}^{\text{eff}}_{h,d^t}\}$, and so the wage mixture components, to have disjoint supports, and instead allows them to overlap arbitrarily.

We establish Proposition 4 as a straightforward application of Bruni and Koch (1985). Specifically, the wage mixture (19) corresponds to the mixture model discussed in Section 4.c of Bruni and Koch (1985) and is shown to be identified under Assumptions 1, 2(i), and 3.

As an alternative to the approach set out by Assumption 3 and Proposition 4, we have examined the applicability of two identification strategies in the mixture-model literature. The first approach uses exclusion restrictions, that is, variables that enter either the mixture weights or the mixture components, but not both (Henry et al., 2014; Compiani and Kitamura, 2016; Jochmans et al., 2017). The second approach considers the joint distribution of the entire vector of wages $(w_{n,1}, \ldots, w_{n,T})$, rather than focusing on the cross-sectional wage distribution at each time t as in (19), and relies on assumptions that simplify the temporal dependence of wages, such as conditional independence or Markovianity, together with a constant number of latent classes over time (Hall and Zhou, 2003; Allman et al., 2009; Kasahara and Shimotsu, 2009; Bonhomme et al., 2016a,b). Neither approach is suitable for our framework. In the class of models we consider, exclusion restrictions do not arise: any variable that affects the conditional distribution of efficiency types and signals also affects the conditional distribution of wages, and vice versa. Moreover, wage observations are neither conditionally independent over time nor Markovian, and the number of latent classes over which we mix increases with t because of the growing dimension of the vector of performance signals a^{t-1}_n. More broadly, human capital accumulation and learning imply that there is no sufficiently time-invariant structure for the wage time series to be useful for identification. Given these considerations, we view the approach set out by Assumption 3 and Proposition 4 as the best compromise between generality and the key features of class of models we study.

Despite the extreme flexibility of the class of mixtures embraced by Assumption 3, Proposition 4 can also be extended to fully nonparametric mixture families. In particular, Aragam et al. (2020) propose a criterion known as the "clusterability" condition, which is sufficient for identification. Intuitively, this condition essentially requires that the mixture components are "sufficiently distinct,"

as quantified by an appropriate distance measure—a notion that can already be found in Teicher (1961, 1963)'s earlier discussion of mixture identifiability. Not only is this condition met in the setting described by Bruni and Koch (1985), but it applies to a wide range of other mixtures.

Finite mixture models can only be identified up to the labeling of their components because the likelihood is invariant to permutation of the components. In our setting, we can resolve the labeling indeterminacy by examining the moments of the mixture components, for example, by using their variances to order them with respect to e_n and their means to order them with respect to a_n^{t-1}.

Lastly, as an immediate implication of Proposition 4, 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 2 (Signal Distribution)**.** *Assume:*

(i) Assumption 1 holds.

(ii) The wage mixture weights in (19) *are identified at times* t *and* $t+1$*, with* $t \in \{1, \dots, T-1\}$*. See Proposition 4 for sufficient conditions.*

Then, the conditional signal distribution

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

is identified for each $(a^t, h, d^t, e) \in \mathcal{A}^t \times \mathcal{H} \times \mathcal{D}^t \times \mathcal{E}$ *such that* $\Pr\big(H_{n,1} = h,\ D_n^t = d^t\big) > 0$ *and* $\Pr\big(e_n = e \mid H_{n,1} = h,\ D_n^t = d^t\big) > 0$*, where* $a^t \coloneqq (a_1, \dots, a_t)$ *and* $d^t \coloneqq (d_1, \dots, d_t)$*.*

Corollary 3 (Signal Distribution and Job Retention)**.** *Assume:*

(i) Assumption 1 holds.

(ii) The wage mixture weights in (19) *are identified at times* $t+2$ *and* $t+3$*, with* $t \in \{1, \dots, T-3\}$*. See Proposition 4 for sufficient conditions.*

Then, the conditional distribution of three consecutive signals at job d*,*

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

is identified for each $(a_t, a_{t+1}, a_{t+2}, h, d, e) \in \mathcal{A}^3 \times \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ *such that* $\Pr\big(H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d\big) > 0$*, and* $\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d\big) > 0$*.*

4.3 Information Technology

In this section, we use the wage mixture weights in (19), identified by Proposition 4, to recover each firm's information technology, that is, the prior and posterior beliefs generated under any sequences of jobs and signals. To this end, we first introduce an assumption that disciplines the distribution of signals conditional on the latent ability θ_n.

Assumption 4 (Signal Distribution Conditional on Ability)**.** (i) $\mathcal{A} \coloneqq \{\bar{a}, \underline{a}\}$ and $\Theta \coloneqq \{\bar{\theta}, \underline{\theta}\}$.

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

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

(iii) The distribution of $a_{n,t}$ conditional on $(H_{n,1}, D_{n,t}, e_n, \theta_n)$ is time-invariant, with

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

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

for each $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$, and $\alpha(h,d,e) > \beta(h,d,e)$.

$\diamond$

Proposition 5 formalises the identification result under Assumption 4.

Proposition 5 (Information Technology)**.** *Suppose that:*

(i) Assumption 4 hold.

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

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

is identified for each $(a_t, a_{t+1}, a_{t+2}) \in \mathcal{A}^3$, and the conditional distribution of the initial signal at job d,

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

is identified for each $a \in \mathcal{A}$. Condition (ii) is required to hold for each $(h, d, e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$, possibly with t varying across (h, d, e). See Corollaries 2 and 3 for sufficient conditions.

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 $(h, d, e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$.

Assumption 4(i) imposes that the signal $a_{n,t}$ and the latent ability θ_n have finite supports. 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 A, we discuss how this assumption can be relaxed to allow for continuous and multidimensional $a_{n,t}$ and θ_n. Assumption 4(ii) imposes that signals are independent over time conditional on the history of jobs and θ_n, while Assumption 4(iii) requires that the distribution of $a_{n,t}$ conditional on the chosen job and θ_n is time-invariant and described by the parameters $\alpha(h, d, e)$ and $\beta(h, d, e)$. This is a standard requirement for identifying the information technology: if that distribution varied over time, we could not recover belief dynamics solely from observing workers switching jobs. Assumption 4(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 4, Proposition 5 also builds on Corollaries 2 and 3. In particular, condition (ii) of Proposition 5 requires the identification of the conditional distribution of three consecutive signals at job d to be identified, for which sufficient conditions are provided by Corollary 2, and of the conditional distribution of the initial signal at job d, for which sufficient conditions are provided by Corollary 3. These sufficient conditions essentially amount to the identification of certain wage mixture weights in (19). See the remark at the end of the proof of Proposition 5 in Appendix D for a clarification of these sufficient conditions.

The proof of Proposition 5 is straightforward and makes clear where each restriction is used. Under Assumption 4, we can represent the *identified* (under condition (ii) of Proposition 5) distribution $\Pr\big(a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d,\ e_n = e\big)$ 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 therefore be identified using the results in Blischke (1964, 1978) for binomial mixtures. This explains why condition (ii) of Proposition 5 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 4, we can represent the *identified* (under condition (ii) of Proposi-

tion 5) distribution $\Pr(a_{n,1} = a \mid H_{n,1} = h,\ D_{n,1} = d,\ e_n = e)$ 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 $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e)$ 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 $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$.

4.4 Law of Motion of the State and Conditional Choice Probabilities

In this section, we use the wage mixture weights and components in (19), identified by Proposition 4, together with the information technology identified by Proposition 5, 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 some intuition afterwards.

Proposition 6 (Unconditional Distribution of the State)**.** *Let $t \in \{1, \ldots, T\}$. Suppose that:*

(i) Assumption 1 holds.

(ii) The wage mixture weights in (19) *are identified at time t. See Proposition 4 for sufficient conditions.*

(iii) $\alpha(h, d, e)$, $\beta(h, d, e)$, the prior belief $\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h,\ e_n = e)$, and the set of realizations of the posterior belief $P_{n,t}$ are identified for each $(h, d, e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$. See Proposition 5 for sufficient conditions.

Then,

(i) The map from realizations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realizations of $s_{n,t} \coloneqq (H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ is identified. Denote this map by g_t, that is,

$$(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1}) \mapsto s_{n,t} = g_t(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1}),$$

and denote by $\mathcal{I}_t$ the image of g_t.

(ii) The unconditional distribution of $s_{n,t}$ on $\mathcal{I}_t$ is identified. We denote by $\mathcal{S}_t \subseteq \mathcal{I}_t$ the set of all $s \in \mathcal{I}_t$ such that $\Pr(s_{n,t} = s) > 0$, and hereafter refer to $\mathcal{S}_t$ as the support of $s_{n,t}$.

The proof of Proposition 6 is straightforward. Given that the set $\mathcal{E}$ is known under Assumption 3(ii), $H_{n,1}$ is observed, $\kappa_{n,t}$ is a known function of D_n^{t-1}, and $P_{n,t}$ is identified by Proposition 5, we identify the map g_t from realisations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realisations of $s_{n,t} \coloneqq$

$(H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$. The joint distribution of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ is identified from Proposition 4(i) at time t. Given knowledge of this distribution and the map g_t from realisations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realisations of $s_{n,t}$, we identify the unconditional distribution of $s_{n,t}$ on $\mathcal{S}_t$.

Proposition 7 (Law of Motion of the State)**.** *Suppose that:*

(i) Assumption 1 holds.

(ii) For $t \in \{2, \ldots, T\}$, the wage mixture weights in (19) *are identified at times $t-1$ and t, together with the state maps, g_{t-1} and g_t, and supports, $\mathcal{S}_{t-1}$ and $\mathcal{S}_t$. See Propositions 4 and 6 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 $s \in \mathcal{S}_t$, $d \in \mathcal{D}$, and $\tilde{s} \in \mathcal{S}_{t-1}$ such that $\Pr(D_{n,t-1} = d \mid s_{n,t-1} = \tilde{s}) > 0$.

Proposition 8 (Conditional Choice Probabilities)**.** *Suppose that:*

(i) Assumption 1 holds.

(ii) For $t \in \{1, \ldots, T\}$, the wage mixture weights in (19) *are identified at times t, together with the state map, g_t, and support, $\mathcal{S}_t$. See Propositions 4 and 6 for sufficient conditions.*

Then, the conditional choice probability,

$$\Pr(D_{n,t} = d \mid s_{n,t} = s),$$

is identified for each $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$.

Proposition 7 follows directly from combining the wage mixture weights at times $t-1$ and t, while Proposition 8 relies on Bayes' rule and the wage mixture weights at time t. Proposition 8 is particularly interesting because, in contrast to the typical approach of deriving CCPs from agents' discrete choices in dynamic models, here the CCPs are identified from the continuous part of the data, namely the wage distribution.

4.5 Deterministic Wage Component and Productivity Shocks

In this section, we identify the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ and the distribution of the productivity-shock vector $\epsilon_{n,t}$ by extending Proposition 3 and Corollary 1 from the simple static Roy model to our general class of dynamic models.

Since the deterministic wage component $\varphi(\cdot)$ is indexed by both the first- and second-best firms, $D_{n,t}$ and $D'_{n,t}$, the arguments in this section exploit Assumption 2(ii)—which has not been used so far—in order to identify $\varphi(\cdot)$ *without any labelling indeterminacy* with respect to the second-best firm's identity. In particular, as preliminary ingredients, we recover the distribution of wages $w_{n,t}$ conditional on $(D_{n,t}, D'_{n,t}, s_{n,t})$ and the distribution of $D_{n,t}$ conditional on $(D'_{n,t}, s_{n,t})$.

Proposition 9 (Conditional Wage Distribution and Choice Probabilities)**.** *Suppose that:*

(i) Assumptions 1 and 2 holds.

(ii) For $t \in \{1, \dots, T\}$, the wage mixture weights and components in (19) *at time t are identified, together with the state map, g_t, and support, $\mathcal{S}_t$. See Propositions 4 and 6 for sufficient conditions.*

Then, the conditional wage distribution

$$\Pr(w_{n,t} \leq w \mid D_{n,t} = d,\ D'_{n,t} = d',\ s_{n,t} = s),$$

is identified for each $w \in \mathbb{R}$, $s \in \mathcal{S}_t$, and $(d, d') \in \mathcal{D}^2$ such that $\Pr(D_{n,t} = d,\ D'_{n,t} = d' \mid s_{n,t} = s) > 0$*. The conditional choice probability*

$$\Pr(D_{n,t} = d \mid D'_{n,t} = d',\ s_{n,t} = s),$$

is identified for each $s \in \mathcal{S}_t$ and $(d, d') \in \mathcal{D}^2$ such that $\Pr(D'_{n,t} = d' \mid s_{n,t} = s) > 0$*.*

Intuitively, the distribution of $w_{n,t}$ conditional on $(H_{n,1}, D_n^t, e_n, a_n^{t-1})$ is identified from the wage mixture components in (19) at time t. Using this distribution, together with the map g_t from realisations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realisations of $s_{n,t}$, we identify the distribution of $w_{n,t}$ conditional on $(D_{n,t}, s_{n,t})$. Furthermore, under Assumption 2(i), conditioning on $(D_{n,t}, s_{n,t})$ also implicitly conditions on the second-best firm $D'_{n,t}$, which enters the wage equation (8). Under Assumption 2(ii), we know which firm is $D'_{n,t}$. Therefore, we identify the distribution of $w_{n,t}$ conditional

on $(D_{n,t}, D'_{n,t}, s_{n,t})$. Similarly, the distribution of $D_{n,t}$ conditional on $(D'_{n,t}, s_{n,t})$ is identified by Proposition 8, using the wage mixture components in (19) at time t, and Assumption 2.

Based on Proposition 9, we can now extend Proposition 3 and Corollary 1 to identify the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ and the distribution of the productivity-shock vector $\epsilon_{n,t}$. We begin by introducing notation to formalize the results. Given $e \in \mathcal{E}$, recall that $s_{n,t}(e)$ denotes the vector $s_{n,t}$ evaluated at $e_n = e \in \mathcal{E}$ and $\epsilon_{n,t}(e) \coloneqq (\epsilon_{n,t}(d, e) : d \in \mathcal{D})$. Let $\mathcal{S}_t(e) \subseteq \mathcal{S}_t$ denote the support of $s_{n,t}(e)$, identified by Proposition 6. Given $(d, d', e) \in \mathcal{D}^2 \times \mathcal{E}$, let $\mathcal{S}_t(d, d', e) \subseteq \mathcal{S}_t(e)$ be the set of realizations s of $s_{n,t}(e)$ such that $\Pr\big(D_{n,t} = d,\, D'_{n,t} = d' \,\big|\, s_{n,t}(e) = s\big) > 0$, identified by Proposition 9. Given this notation, we can adapt Assumptions (i) to (v) of Proposition 3 to our setting.

Assumption 5 (Exogeneity)**.**Let $t \in \{1, \ldots, T\}$ and $e \in \mathcal{E}$. $\epsilon_{n,t}(e)$ is independent of $s_{n,t}(e)$. ◇

Assumption 6 (Supports)**.**Let $t \in \{1, \ldots, T\}$, $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$. Then,

$$\inf\{w : \Pr(w_{n,t}(d, d', e) \leq w \mid s_{n,t}(e) = s) > 0\} = -\infty,$$
$$\inf\{w : \Pr(w_{n,t}(d, d', e) \leq w \mid D_{n,t} = d, D'_{n,t} = d', s_{n,t}(e) = s) > 0\} = -\infty.$$

◇

Assumption 7 (Tail Limit)**.**Let $t \in \{1, \ldots, T\}$, $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$. There exists an (unknown) constant $q_{t,d,d',e} \in (0, 1]$ such that

$$\lim_{w \to -\infty} \Pr\big(D_{n,t} = d, D'_{n,t} = d' \mid s_{n,t}(e) = s,\ w_{n,t}(d, d', e) < w\big) = q_{t,d,d',e}.$$

◇

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

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

Proposition 10 (Deterministic Wage)**.** *Let $t \in \{1, \ldots, T\}$. Assume:*

(i) The set $\mathcal{S}_t(d, d', e)$ is identified for each $(d, d') \in \mathcal{D}^2$ and $e \in \mathcal{E}$. The conditional probability $\Pr\big(w_{n,t} \leq w \,\big|\, D_{n,t} = d,\ D'_{n,t} = d',\ s_{n,t}(e) = s\big)$ is identified for each $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$. See Proposition 9 for sufficient conditions.

(ii) Assumptions 5 to 9 hold.

For $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$, define

$$c(t, d, d', s) \ := \ \frac{q_{t,d,d',e}}{\Pr(D_{n,t} = d, D'_{n,t} = d' \mid s_{n,t}(e) = s)} \ \in (0, \infty).$$

Let $\{\tau_{\bar{s}}^{(k)}\}_{k \geq 1} \subset (0, 1)$ be any sequence with $\tau_{\bar{s}}^{(k)} \to 1$ as $k \to +\infty$. Define

$$1 - \tau_s^{(k)} \ := \ \frac{c(t, d, d', s)}{c(t, d, d', \bar{s})} \big(1 - \tau_{\bar{s}}^{(k)}\big).$$

Then,

$$\lim_{k \to +\infty} \Big[Q_{w_{n,t} \,|\, D_{n,t}=d, D'_{n,t}=d', s_{n,t}(e)=s}\big(\tau_s^{(k)}\big) \ - \ Q_{w_{n,t} \,|\, D_{n,t}=d, D'_{n,t}=d', s_{n,t}(e)=\bar{s}}\big(\tau_{\bar{s}}^{(k)}\big) \Big] \ = \ \varphi(d, d', s). \tag{20}$$

Hence, $\varphi(d, d', s)$ is identified for each $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$.

Assumptions 5–9 mirror Assumptions (i) to (v) in Proposition 3, which are stated for the simple static Roy model; we refer the reader to Section 3.1, where Proposition 3 is introduced, for a discussion of the role of each assumption, including a sufficient condition for Assumption 7 that restricts dependence among the shocks (see Lemma 1 in Appendix B.1).

There is one minor difference worth highlighting. In Assumptions 6–8, we focus on the *left* extreme tails of the wage distributions, unlike the original construction of Proposition 3, which considers the right tails. The proof works symmetrically for left tails. Focusing on the left tails ensures that the quantity $q_{t,d,d',e}$ remains strictly positive. By contrast, if we let $w \to +\infty$ (rather than $w \to -\infty$), then $\Pr\big(D_{n,t} = d, D'_{n,t} = d' \mid s_{n,t}(e) = s,\ w_{n,t}(d, d', e) > w\big)$ would go to zero due to the equilibrium pricing mechanism in our model, which mirrors a second–price auction. Recall that, in the class of models we study, the equilibrium wage for job d equals the expected output at the second–best firm d', $y(d', s_{n,t}(e)) + \epsilon_{n,t}(d', e)$, plus a compensating differential $\Psi(d, d', s_{n,t}(e))$. Letting the wage of job d go to $+\infty$ while holding $s_{n,t}(e)$ fixed would effectively push the second–best firm's productivity shock $\epsilon_{n,t}(d', e)$—and hence the expected output and wage that firm d' could offer—to $+\infty$. This would alter the equilibrium ranking of firms, making the former first-best firm d no longer the best choice for worker n and driving $q_{t,d,d',e}$ to zero. Focusing on left tails avoids this issue.

Moreover, note that assuming unbounded left support for both the observed selected wages and the potential wages (Assumption 6) is not essential. For example, it is often the case that the observed selected wages are bounded *away from zero* in the data. A bounded–left-endpoint analogue proceeds by tracking convergence to the finite endpoint rather than to $-\infty$, with only minor adjustments. See the discussion following Proposition 3 and Appendix B.2 for details on the bounded case. In Appendix B.2, we further show that, when finite, the right and left endpoints of the potential wages and shocks can be nonparametrically identified.

Also note that Assumption 9 imposes a location normalisation at one state for each *pair* of first- and second-best firms, whereas Assumption (v) in Proposition 3 imposes a normalisation at one state per firm. As discussed in connection with Proposition 3, wages in Roy-type models are identified only up to an additive constant, so location normalisations are needed to pin down levels. Given that potential wages in our class of models are indexed by both the first- and second-best firms, while in the static Roy model they are indexed only by the first-best firm, it is natural that more location normalisations are required here.

Lastly, nonparametric identification of the joint distribution of the shock vector $\epsilon_{n,t}(e)$ is not feasible for the same reason as in the simple static Roy model of Section 3.1: we lack at least as many continuous state variables as there are jobs (Tsiatis, 1975; Heckman and Honoré, 1989). In view of this, we focus on recovering the marginal distributions of each $\epsilon_{n,t}(d, e)$ and show that, if $\epsilon_{n,t}(d, e)$ belongs to a parametric family, its governing parameters are identified—by extending Corollary 1 to our general class of dynamic models. To identify the joint distribution of $\epsilon_{n,t}(e)$, we either add an explicit independence assumption, impose a parametric copula, or work with Fréchet–Höffding bounds for partial identification.[19][20]

Corollary 4 (Identification of the Shock Distribution)**.** *For each $t \geq 1$, $d \in \mathcal{D}$, and $e \in \mathcal{E}$, let $S_{d,e}$ denote the marginal survival function of $\epsilon_{n,t}(d, e)$. Let F_e denote the joint CDF of $\epsilon_{n,t}(e)$ and $F_{d,e}$ the marginal CDF of $\epsilon_{n,t}(d, e)$. Assume:*

[19] As noted in Footnote 13, assuming independence between the wage shocks in the static Roy model can be restrictive, because these shocks are the sole source of unobserved heterogeneity and may, in principle, embody substantial correlation across potential wages. By contrast, in our broader class of dynamic models this assumption is less consequential: structural correlation across potential wages is captured by latent state variables—e_n and $P_{n,t}$—so the productivity shocks can be treated as residual errors.

[20] Restricting $\epsilon_{n,t}(d, e)$ to a parametric family does not render the generalised finite-mixture approach we use in Proposition 4 superfluously general. Even if the *unconditional* distribution of $\epsilon_{n,t}(d', e)$ is parametrically specified—as prescribed by Corollary 4—the *conditional* distribution of $\epsilon_{n,t}(d', e)$ given $D_{n,t}$—which determines the wage-mixture components in Equation (19)—typically does not belong to the same parametric family. The only common case exhibiting "closure under conditioning" is i.i.d. Type-I extreme value (Gumbel). However, the i.i.d. Gumbel specification is well known to be ill-suited for dynamic discrete-choice models, as it implies Independence of Irrelevant Alternatives (IIA) and leads to unrealistic substitution patterns.

(i) The set $\mathcal{S}_t(d, d', e)$ is identified for each $(d, d') \in \mathcal{D}^2$ and $e \in \mathcal{E}$. The conditional probability $\Pr\big(w_{n,t} \leq w \,\big|\, D_{n,t} = d,\ D'_{n,t} = d',\ s_{n,t}(e) = s\big)$ is identified for each $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$. See Proposition 9 for sufficient conditions.

(ii) Assumptions 5 to 9 hold, implying that $\varphi(d, d', s)$ is identified for each $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, $s \in \mathcal{S}_t(d, d', e)$, and $t \geq 1$ (Proposition 10).

(iii) For each $d \in \mathcal{D}$ and $e \in \mathcal{E}$, $\epsilon_{n,t}(d, e)$ belongs to a known parametric family indexed by the $p_{d,e} \times 1$ vector of parameters $\mu_{d,e} \in M_{d,e} \subseteq \mathbb{R}^{p_{d,e}}$

Fix $d \in \mathcal{D}$ and $e \in \mathcal{E}$. Consider some $d' \in \mathcal{D} \setminus \{d\}$, $s \in \mathcal{S}_t(d, d', e)$, and $t \geq 1$. Choose $p_{d,e}+1$ distinct large thresholds $0 < w_0 < w_1 < \cdots < w_{p_{d,e}}$. Define the function $\Phi_{d',s} : M_{d,e} \to \mathbb{R}^{p_{d,e}}$ as

$$\Phi_{d',s}(\mu_{d,e}) \ := \ \left(\frac{S_{d,e}(w_1 - \varphi(d, d', s); \mu_{d,e})}{S_{d,e}(w_0 - \varphi(d, d', s); \mu_{d,e})}, \ldots, \frac{S_{d,e}(w_{p_{d,e}} - \varphi(d, d', s); \mu_{d,e})}{S_{d,e}(w_0 - \varphi(d, d', s); \mu_{d,e})} \right).$$

If $\Phi_{d',s}$ is injective, then the parameter $\mu_{d,e}$ is identified. Moreover, if the shocks $\{\epsilon_{n,t}(d, e)\}_{d \in \mathcal{D}}$ are mutually independent across $d \in \mathcal{D}$, then the joint distribution of $\epsilon_{n,t}(e)$ is identified as the product of the identified marginals. Alternatively, if a copula C_{μ_e} is specified so that

$$F_e(v_1, \ldots, v_{|\mathcal{D}|}) \ = \ C_{\mu_e}\big(F_{1,e}(v_1; \mu_{1,e}), \ldots, F_{|\mathcal{D}|,e}(v_{|\mathcal{D}|}; \mu_{|\mathcal{D}|,e})\big) \quad \forall (v_1, \ldots, v_{|\mathcal{D}|}) \in \mathbb{R}^{|\mathcal{D}|},$$

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

$$\max\Big\{ \sum_{d \in \mathcal{D}} F_{d,e}(v_d; \mu_{d,e}) - (|\mathcal{D}| - 1),\ 0 \Big\} \ \leq \ F_e(v_1, \ldots, v_{|\mathcal{D}|}) \ \leq \ \min_{d \in \mathcal{D}} F_{d,e}(v_d; \mu_{d,e}).$$

4.6 Output, Human-Capital Technology, and Compensating Differential

Once the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ is identified, the remaining objects to recover are the output (human-capital) technology $y(\cdot)$ and its deterministic labour-input component $\ell(\cdot)$, together with the compensating differential $\Psi(\cdot)$. We first show how to identify $y(\cdot)$ using standard arguments for dynamic discrete choice models. Given $y(\cdot)$, the components $\ell(\cdot)$ and $\Psi(\cdot)$ can then be obtained residually from $y(\cdot)$ and $\varphi(\cdot)$, respectively.

We start with the simplest case, in which the model's equilibrium is efficient. Indeed, under Assumption 2(i), the equilibrium can be shown to be efficient. In an efficient equilibrium, job choices

maximise the expected present discounted value of output. Hence, a worker's choice of firm solves a planning problem: a social planner assigns a job to each worker in each period. In other words, the market-wide equilibrium allocation (matching workers to firms) reduces to a single-agent dynamic decision problem. It follows that standard identification arguments for dynamic discrete choice models (for instance, Magnac and Thesmar, 2002) can be applied to identify $y(\cdot)$ from observed job choices.

To elaborate, let $Y\big(s_{n,t}(e), \epsilon_{n,t}(e)\big)$ denote the expected present discounted value of output produced by worker n with efficiency $e_n = e \in \mathcal{E}$ (equivalently, the expected present discounted social surplus) at state $\big(s_{n,t}(e), \epsilon_{n,t}(e)\big)$. Then

$$Y\big(s_{n,t}(e), \epsilon_{n,t}(e)\big) = \max_{d \in \mathcal{D}_t(s_{n,t}(e))} \Big\{ y\big(d, s_{n,t}(e)\big) + \epsilon_{n,t}(d,e) \\ + \ \delta\left[1 - \eta(\kappa_{n,t}, d)\right] \mathbb{E}\big[Y\left(s_{n,t+1}(e), \epsilon_{n,t+1}(e)\right) \ \big| \ s_{n,t}(e),\ d\big] \Big\}.$$

By Propositions 7 and 8, the law of motion of $s_{n,t}(e)$, as well as the CCPs, are identified. The joint distribution F_e of the shock vector $\epsilon_{n,t}(e)$ is identified by Corollary 4. The exogenous separation rate $\eta(\kappa_{n,t}, d)$ is nonparametrically identified by the fraction of employed workers at firm d with given $\kappa_{n,t}$ who exit at the end of the period. Therefore, $y\big(d, s_{n,t}(e)\big)$ is identified following Magnac and Thesmar (2002) under standard normalisations in dynamic discrete choice models.

We now state the formal normalisation conditions and the identification result. Given $(d, e) \in \mathcal{D} \times \mathcal{E}$, let $\mathcal{S}_t(d,e) \subseteq \mathcal{S}_t(e)$ be the set of realizations s of $s_{n,t}(e)$ such that $\Pr\big(D_{n,t} = d \,\big|\, s_{n,t}(e) = s\big) > 0$, identified by Proposition 8.

Assumption 10 (Normalisation)**.**For each $e \in \mathcal{E}$, $d \in \mathcal{D}$, and $s \in \bigcup_t \mathcal{S}_t(d,e)$, there exists $\tilde{d} \in \mathcal{D}$ such that $s \in \bigcup_t \mathcal{S}_t(\tilde{d}, e)$ and $y(\tilde{d}, s) = 0$. $\diamond$

Assumption 10 normalises $y(\cdot)$ to zero at one firm for each state. More precisely, to identify $y(d,s)$ for some $d \in \mathcal{D}$ and $s \in \bigcup_t \mathcal{S}_t(e)$, the worker must be able—at state realisation s—to choose between d and at least one other firm $\tilde{d}$ with strictly positive probability, and $y(\tilde{d}, s)$ is set to zero.

Proposition 11 (Output (Human-Capital) Technology)**.** *Let $t \in \{1, \ldots, T\}$. Suppose that:*

(i) The law of motion of state, $\Pr(s_{n,t} \mid D_{n,t-1}, s_{n,t-1})$, and the CCPs, $\Pr(D_{n,t} \mid s_{n,t})$, are identified for each $t \in \{1, \ldots, T\}$. See Propositions 7 and 8 for sufficient conditions.

(ii) The joint distribution F_e of the shock vector $\epsilon_{n,t}(e)$ is identified for each $e \in \mathcal{E}$. See Corollary 4 for sufficient conditions.

(iii) The discount factor δ is known.

(iv) The separation rates $\{\eta(\kappa_{n,t}, d)\}_{d\in\mathcal{D}}$ are identified (immediate consequence of Assumption 1).

(v) Assumption 10 holds.

Then, the output (human-capital) technology $y(d, s)$ is identified for each $d \in \mathcal{D}$, $e \in \mathcal{E}$, $s \in \mathcal{S}_t(d, e)$.

When firms consist of multiple jobs—for example, as in our empirical application where two firms make wage offers each period and each firm operates multiple jobs—the equilibrium can be inefficient. Even so, the identification of $y(\cdot)$ proceeds analogously to Proposition 11. Specifically, the main difference in the multi-job case is that the market-wide equilibrium allocation problem does not solve the planning problem but instead solves the pseudo-planning problem of maximizing the match surplus for each firm $d \in \mathcal{D}$. In this scenario, the one-period surplus when firm d does not employ the worker equals the deterministic component of the wage paid by the employing firm. Since the latter is identified, standard dynamic discrete choice arguments applied to each pseudo-planning problem can once again be used to establish the identification of $y(\cdot)$. With $y(\cdot)$ known, then $\ell(\cdot)$ and $\Psi(\cdot)$ can then be obtained residually from $y(\cdot)$ and $\varphi(\cdot)$, respectively.

Corollary 5 (Compensating Differential and Deterministic Labor-Input Component)**.** *Let $t \in \{1, \ldots, T\}$. Suppose that:*

(i) The deterministic wage component $\varphi(d, d', s)$ is identified for each $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$. See Proposition 10 for sufficient conditions.

(ii) The output (human-capital) technology $y(d, s)$ is identified for each $d \in \mathcal{D}$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, e)$. See Proposition 11 for sufficient conditions.

Then, the compensating differential $\Psi(d, d', s)$ is identified for each $(d, d') \in \mathcal{D}^2$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, d', e)$. Moreover, under the additional normalisation $\mathbb{E}(a_{n,t}(d, e) \mid s_{n,t}(e)) = 0$ for each $d \in \mathcal{D}$, $e \in \mathcal{E}$, and $s \in \mathcal{S}_t(d, e)$, the deterministic labour-input component $\ell(h, \kappa; d, e)$ is identified for each $(h, \kappa) \in \mathcal{H} \times \mathcal{K}_t$.

4.7 Discussion: Longitudinal vs. Cross-Sectional Dimension

In this section, we provide a high-level overview of our identification strategy, focusing on how it leverages both the longitudinal and cross-sectional dimensions of the data. First, we use the longitudinal dimension to identify the information technology, the law of motion of the state, and the CCPs. Indeed, these primitives are identified by concatenating the wage mixture weights across periods.

Next, we identify the deterministic wage component $\varphi(\cdot) \coloneqq y(\cdot) + \Psi(\cdot)$ by adapting Proposition 3's approach in each period, thereby drawing on the cross-sectional dimension. Because $\varphi(\cdot)$ is left nonparametrically specified and depends on $s_{n,t}$—whose support may vary across periods—$\varphi(\cdot)$ is effectively a time-varying function and must therefore be identified in each period, making the longitudinal dimension less helpful here. Finally, we once again leverage the dynamic dimension of the model to identify the output (and human-capital) technology $y(\cdot)$, and in turn, the compensating differential $\Psi(\cdot)$ and deterministic labour-input component $\ell(\cdot)$.

A strength of our identification approach is its limited reliance on workers' mobility across jobs over time. Nonetheless, some heterogeneous variation in job choices—akin to job mobility—facilitates identification of the output (human-capital) technology $y(\cdot)$ and the compensating differential $\Psi(\cdot)$. Regarding $y(\cdot)$, recall the standard normalizations in Assumption 10 (as in the dynamic discrete choice literature), which fix the value of $y(\cdot)$ at one firm for each state. These deliver nontrivial identification only if, at the same state, workers can choose employment at *at least two* different firms with strictly positive probability. Regarding $\Psi(\cdot)$, note that for a given state realization $s \in \mathcal{S}_t(e)$, the compensating differential of firm d relative to firm d', $\Psi(d, d', s)$, is obtained by subtracting the expected output $y(d', s)$—identified from observing d' as the first-best firm at state s—from the deterministic wage component $\varphi(d, d', s)$—identified from observing d' as the second-best firm at state s. Therefore, to identify $\Psi(d, d', s)$, it must occur with strictly positive probability (at state s) that firm d' is both first-best and second-best across observations.

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

4.8 Estimation

In this section, we describe the procedure used to estimate the wage-equation parameters in our empirical application. The procedure mirrors the structure of the identification arguments (a mixture step followed by a quantile step), while introducing some parametric assumptions and other simplifications to preserve tractability.

In a preliminary step, we estimate the learning process by constructing performance measures from existing earnings data available in the LEHD dataset. While this data does not provide direct

performance information, we can infer it from changes in observed earnings. Specifically, we focus on individual-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 (quarters $t-3$ to $t-1$). We then define an observation of high performance pay as the dollar value of earnings in quarter t if those earnings are more than 50% higher than this calculated lagged average earnings. Our identification procedure is more general and robust, as it does not rely on the availability of these performance measures, whose construction typically involves additional assumptions.

Furthermore, we treat each worker's second-best firm, $D'_{n,t}$, as known in every period. In particular, in the empirical application, we construct $D'_{n,t}$ by considering the worker's labor markets defined by industry and geographical location for classes of observationally equivalent workers (defined by gender and education).

Our observables therefore consist of

$$(w_{n,t}, D_{n,t}, D'_{n,t}, H_{n,1}, \kappa_{n,t}, P_{n,t}).$$

Next, we assume that, for each second-best firm $D'_{n,t} = d' \in \mathcal{D}$ and efficiency type $e_n = e$, the productivity shock $\epsilon_{n,t}(d', e)$ is normally distributed conditional on $(D_{n,t}, D'_{n,t}, H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$, with mean and variance allowed to depend flexibly on $(D_{n,t}, D'_{n,t}, H_{n,1}, \kappa_{n,t}, P_{n,t}, e_n)$ to account for the potential selection of $D_{n,t}$ and $D'_{n,t}$ based on $\epsilon_{n,t}$. Therefore, the conditional cross-sectional wage distribution at time t is a finite mixture of Normal distributions:

$$\begin{aligned}
&\Pr(w_{n,t} \leq w \mid D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, \kappa_{n,t} = \kappa, P_{n,t} = p) \\
&\quad = \sum_{e \in \mathcal{E}} \Pr(e_n = e \mid D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, \kappa_{n,t} = \kappa, P_{n,t} = p) \\
&\qquad\qquad \times \Pr(w_{n,t} \leq w \mid D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, \kappa_{n,t} = \kappa, P_{n,t} = p, e_n = e),
\end{aligned}$$

where each mixture component is normally distributed:

$$\begin{aligned}
w_{n,t} \mid D_{n,t} = d,\ D'_{n,t} = d',\ H_{n,1} = h,\ \kappa_{n,t} = \kappa, \\
P_{n,t} = p,\ e_n = e \sim \mathcal{N}\Big(&\varphi(d, d', h, \kappa, p, e) + \mu(d, d', h, \kappa, p, e), \\
&\sigma^2(d, d', h, \kappa, p, e)\Big).
\end{aligned}$$

with $\mu(d, d', h, \kappa, p, e)$ and $\sigma^2(d, d', h, \kappa, p, e)$ denoting the unknown conditional mean and variance of $\epsilon_{n,t}(d', e)$, respectively.

For each mixture component, we parameterized the deterministic wage, $\varphi(d, d', h, \kappa, p, e) :=$

$y(d', h, \kappa, p, e) + \Psi(d, d', h, \kappa, p, e)$, with the finite-dimensional vector of parameters $\beta_e(e, d, d') \coloneqq (\beta_y(e, d, d'), \beta_\Psi(e, d, d'))$. In particular, for $y(\cdot; \beta_y(e, d, d'))$, we assume a specification that depends linearly on worker experience and the beliefs $P_{n,t}$ (see equation 23). For $\Psi(\cdot; \beta_\psi(e, d, d'))$, we assume a flexible quartic polynomial on workers' experience, κ, and beliefs, p, interacted with initial human capital, h.

For each latent class $e_n = e \in \mathcal{E}$, we estimate $\beta_e(e, d, d')$ using the extremal quantile regression approach of D'Haultfoeuille et al. (2018), implemented in the `eqregsel` Stata command (D'Haultfoeuille et al., 2020). This procedure estimates selection-corrected linear quantile regression coefficients at extreme quantiles by exploiting the behavior of the conditional wage distribution in the upper tail, in the spirit of our identification arguments. Under the regularity conditions in D'Haultfoeuille et al. (2018), the resulting estimator is consistent and asymptotically normal, and we compute standard errors using the bootstrap procedure provided by `eqregsel`. This inner extremal quantile regression is nested inside an outer maximum-likelihood step, where we fit a finite mixture of Normal distributions with $|\mathcal{E}|$ components and estimate the mixture weights $\Pr(e_n = e \mid D_{n,t} = d, D'_{n,t} = d', H_{n,1} = h, \kappa_{n,t} = \kappa, P_{n,t} = p)$, for example using the `fmm` routine in Stata.

5 The Impact of Sorting on Earnings Inequality

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

Building on the theoretical insights from the class of models we study, we argue that the AKM estimates of the correlation between firm and worker effects may be understated because two key forces are omitted. First, the *compensating differential* can dampen the direct impact of worker and firm characteristics on wages, because it compensates a worker for the forgone future wage returns from the human capital and information that could have been acquired by accepting competing firms' offers. Second, *endogenous matching frictions*, namely a worker's acquisition of human

capital and the gradual resolution of uncertainty about ability, may prevent high-type workers from immediately joining the most productive firms. For instance, workers might persistently choose less-productive firms that offer valuable training or learning opportunities, which challenges the common presumption that workers always sort into the most productive match given their observed and *fixed* unobserved productive characteristics. To empirically validate these conjectures, we provide both simulation-based evidence and empirical evidence about them.

5.1 An Illustrative Simulation Exercise

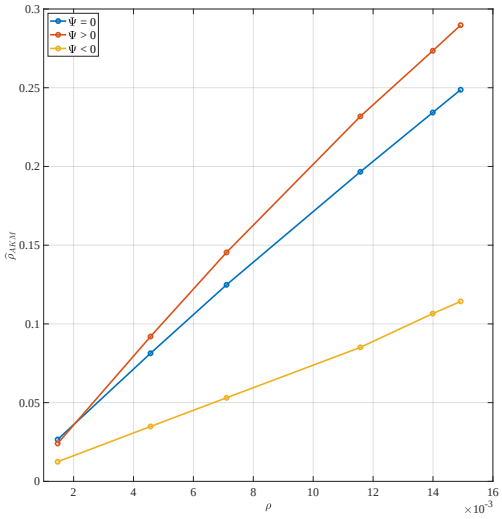
In our first application, we simulate an economy based on a data-generating process (DGP) that captures the main features of our class of models, while introducing a few simplifications to facilitate direct comparison with the AKM framework. Specifically, we remove the wage equation's dependence on the second-best firm (thus eliminating the need to impose Assumption 2) and assume away selection on $\epsilon_{n,t}$, since neither is present in AKM. Under these simplifications, workers' wages follow equation (8), are parameterized as:

$$\begin{aligned} w_{n,t} \;=\; & \sum_{d\in\mathcal{D}}\sum_{e\in\mathcal{E}} \mathbb{1}\{D_{n,t}=d,\, e_n{=}e\} \times \Big[e + \beta_0(d) + \beta_1(d,e)\, H_{n,1} + \beta_2(d,e)\, \kappa_{n,t} + \beta_3(d,e)\, P_{n,t} \\ & + \Psi\big(H_{n,1},\, \kappa_{n,t},\, P_{n,t};\, \psi(d,e)\big) + \epsilon_{n,t}(d,e)\Big], \end{aligned} \tag{21}$$

where the output technology $y(\cdot)$ consists of an AKM-style sum of worker and firm effects, $e + \beta_0(d)$, plus first-order terms in $H_{n,1}$, $\kappa_{n,t}$, and $P_{n,t}$ governed by the parameters $\beta_1(d,e)$, $\beta_2(d,e)$, and $\beta_3(d,e)$. The compensating differential $\Psi(\cdot)$ is approximated by a truncated Taylor expansion, which includes higher-order and interaction (cross) terms in $H_{n,1}, \kappa_{n,t}$, and $P_{n,t}$, governed by the parameters $\psi(d,e)$. $H_{n,1}$ consists of gender and education, while $\kappa_{n,t}$ incorporates age.

We calibrate the wage parameters and other simulation features to match key earnings moments from PSID, a representative survey of U.S. households with panel information from 1968 to 2022. These moments include wage growth, earnings life-cycle patterns (both first and higher-order), and cross-sectional earnings inequality. We also include as targets the AKM-type moments from Song et al. (2019), which derive from SSA. This calibration ensures that our simulated economy reflects both the broader U.S. earnings distribution and the key features highlighted by the AKM framework; see Appendix E for details. As mentioned, the literature centered on the AKM framework measures the impact of sorting on earnings inequality based on firm and worker (linear) complementarities in the output technology $y(\cdot)$. Correspondingly, this literature focuses on the fraction of the total wage variance attributable to the covariance between worker and firm effects or, by the notation of the

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



wage equation in (21), $\rho \coloneqq \text{Cov}(e_n, \beta_0(D_{n,t}))/\text{Var}(w_{n,t})$.

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

$$w_{n,t} = \sum_{d \in \mathcal{D}} \sum_{e \in \mathcal{E}} \mathbb{1}\{D_{n,t} = d, e_n = e\} \Big[e + \beta_0(d) + \beta_1 H_{n,1} + \beta_2 \kappa_{n,t} + \beta_3 P_{n,t} + \epsilon_{n,t}(d,e) \Big], \tag{22}$$

where the compensating differential $\Psi(\cdot)$ is omitted and the parameters β_1, β_2, and β_3 are assumed to be invariant across (d, e). Our findings suggest that when $\Psi(\cdot)$ is negative—implying that workers match with firms offering human capital and information gains with future returns higher than competitors—$\hat{\rho}_{\text{AKM}}$ underestimates ρ because the omitted $\Psi(\cdot)$ attenuates the firm and worker complementarities in output technology $y(\cdot)$. Conversely, when $\Psi(\cdot)$ is positive—so that workers match with firms offering human capital and information gains with future returns lower than competitors—$\hat{\rho}_{\text{AKM}}$ overestimates ρ because the omitted $\Psi(\cdot)$ enhances the firm and worker complementarities in output technology $y(\cdot)$. Figure 1 illustrates these patterns. On the horizontal axis, we plot the increasing values of ρ used to generate our data, while on the vertical axis we report the corresponding AKM estimates. The blue line corresponds to the case $\Psi(\cdot) = 0$ in the true DGP—though it does not perfectly coincide with the 45-degree line because, even though $\Psi(\cdot) = 0$ in the true DGP, the parameters $\beta_1(d, e)$, $\beta_2(d, e)$, and $\beta_3(d, e)$ still vary across (d, e), whereas AKM incorrectly treats them as invariant across (d, e). The red line shows results when workers $\Psi(\cdot) > 0$ (leading to an upward bias), and the yellow line shows results when $\Psi(\cdot) < 0$ (leading to a downward bias).[21]

[21] We adjust standard AKM estimates for low-mobility bias, following the methodology of Bonhomme et al. (2023).

5.2 Monte Carlo Simulation

The previous results ignored two key features of our model, both of which we now bring back. In particular, we next simulate the economy of our model, with only a few simplifications that allow for a simpler solution of the model. Specifically, we assume that the economy comprises only two firm types (d), whereas workers can be of many types (e). Production takes a linear form that depends on a type-dependent intercept and on the beliefs about a worker's ability, $P_{n,t}$. We further assume that within the firm, there are two possible jobs, denoted by $k \in \{H, L\}$, with high and low skill requirements, but that entail identical opportunities for information and human capital acquisition, and that workers have a comparative advantage depending on their own observable skills—namely, high-skill workers produce more, on average, than low-skilled workers in the high-skilled job whereas low-skill workers produce more, on average, than high-skilled workers in the low-skilled job.

Simulation. As in our previous exercise, we simulate data from our model parameterized so that it matches standard cross-sectional moments of the earnings distribution in the U.S., for instance, life-cycle profiles of average and standard deviation of (log) earnings, standard deviation, skewness, and kurtosis of earnings growth, and measures of top-earnings concentration. In simulating the model, we solve it by backward induction. That is, starting from a given final period T, we assume that the continuation value for every worker is equal to zero after period T. Exploiting the efficiency of our equilibrium, we determine optimal equilibrium allocations by solving for the planner's problem of choosing for a worker, given each beginning-of-period state, the optimal firm and job. We then use equation (7) to determine a worker' wage at each time and state. Conveniently, since there are only two types of firms in this economy, we can easily identify the second-best firm for each worker.

Given the solution of the planner's problem and the equilibrium wage in period T, we can proceed one period backward and solve the problem in period $T-1$ fully knowing the continuation match surplus value function for each worker. In any such period, a worker's assignment to a particular firm and job depends on the present value of the worker's output, but wages now also contain the compensating differential term, which we calculate given the worker's match surplus continuation value with the current first- and second-best firm from next period on. We proceed in this fashion for 30 periods, so our model captures most of the standard working life cycle.

In our simulation, worker's output is denoted by $Y\left(s_{n,t}, \epsilon_{L,n,t}, \epsilon_{H,n,t}\right)$ with $s_{n,t} = (H_{1,t}, \kappa_{n,t}, P_{n,t}, e)$, that is, output depends on $H_{n,t}$, that captures the initial human capital of an individual (e.g. education), $\kappa_{n,t}$ represents experience, and $P_{n,t}$ captures the beliefs of the worker's type. Output is also affected by $(\epsilon_{L,n,t}, \epsilon_{H,n,t})$ which are idiosyncratic job-specific productivity shocks for low and high

skill job types. Then, for a worker type e, the planner's problem is

$$\begin{aligned} Y\left(s_{n,t},\epsilon_{L,n,t},\epsilon_{H,n,t}\right) &= \max_{d\in D_e}\left\{\tilde{y}\left(d,s_{n,t}\right)+\delta(1-\eta(\kappa_{n,t},d))\mathbb{E}\left[Y\left(s_{n,t+1},\epsilon_{L,n,t+1},\epsilon_{H,n,t+1}|s_{n,t},d\right)\right]\right\} \\ \tilde{y}\left(d,s_{n,t}\right) &= \max\left\{y_L\left(d,s_{n,t}\right)+\epsilon_{L,n,t},y_H\left(d,s_{n,t}\right)+\epsilon_{H,n,t}\right\} \\ y\left(d,s_{n,t}\right) &= \begin{cases} y_L\left(d,s_{n,t}\right) & \text{if } y_L\left(d,s_{n,t}\right)+\epsilon_{L,n,t}\geq y_H\left(d,s_{n,t}\right)+\epsilon_{H,n,t} \\ y_H\left(d,s_{n,t}\right) & \text{otherwise,} \end{cases} \end{aligned}$$

where

$$P_{n,t+1} = \begin{cases} \frac{\alpha_{h,d,e}P_{n,t}}{\alpha_{h,d,e}P_{n,t}+\beta_{h,d,e}(1-P_{n,t})} & \text{if } a_{n,t}\left(d,e\right)\leq\overline{a}\left(\theta\right) \\ \frac{\left(1-\alpha_{h,d,e}\right)P_{n,t}}{\left(1-\alpha_{h,d,e}\right)P_{n,t}+\left(1-\beta_{h,d,e}\right)(1-P_{n,t})} & \text{otherwise} \end{cases},$$

$$y_L(d,s_{n,t}) = \zeta_{0L}(d,\cdot)+\zeta_{1L}(d,\cdot)P_{n,t} \text{ and } y_H(d,s_{n,t})=\zeta_{0H}(d,\cdot)+\zeta_{1H}(d,\cdot)P_{n,t}, \quad (23)$$

with $\zeta_{0L}(d,\cdot)=\zeta_{0L}(d,H_{n,t},\kappa_{n,t},e)>\zeta_{0H}(d,\cdot)=\zeta_{0H}(d,H_{n,t},\kappa_{n,t},e)$ and $\zeta_{1L}(d,\cdot)=\zeta_{1L}(d,H_{n,t},$-$\kappa_{n,t},e)\leq\zeta_{1H}(d,\cdot)=\zeta_{1H}(d,H_{n,t},\kappa_{n,t},e)$ to capture the idea that low-ability workers have a comparative advantage at the low-skill job whereas high-ability workers have a comparative advantage at the high-skill job. We can then express wages and compensating differential as

$$w_{n,t}\left(d,d',e\right)=\begin{cases} y_L\left(d',s_{n,t}\right)+\epsilon_{L,n,t}+\Psi(d,d',s_{n,t}) & \text{if } y_L\left(d',s_{n,t}\right)+\epsilon'_{L,n,t}\geq y_H\left(d',s_{n,t}\right)+\epsilon'_{H,n,t} \\ y_H\left(d',s_{n,t}\right)+\epsilon_{H,n,t}+\Psi(d,d',s_{n,t}) & \text{otherwise} \end{cases},$$

where $\Psi\left(d,d',s_{n,t}\right)$ is the difference between $\delta[1-\eta\left(\kappa_{n,t},d'\right)]\mathbb{E}\left[Y\left(s_{n,t+1},\epsilon_{L,n,t+1},\epsilon_{H,n,t+1}|s_{n,t},d'\right)\right]$ and $\delta\left[1-\eta\left(\kappa_{n,t},d\right)\right]\mathbb{E}\left[Y\left(s_{n,t+1},\epsilon_{L,n,t+1},\epsilon_{H,n,t+1}|s_{n,t},d\right)\right]$.

Estimation. The goal of our Monte Carlo exercise is to show that, despite its complexity, our model can be easily estimated using relatively standard empirical methods that combine quantile regression and mixture model estimation. We start by simulating an economy comprised of two firm types and a large number of workers characterized by their unobserved type, their skill, and initial human capital, h. We also assume that workers and firms share initial beliefs about worker types, p, which are updated following Bayes' rule using output realization, $y(d,s_{n,t})$, as a noisy signal of workers' true ability. Our entire panel is consistent on 1 million workers simulated between ages 25 and 55 (30 periods).

Next, we aim to recover the key parameters of the production function (the $\zeta's$). For numerical expediency, we approximate the compensating differential through a flexible-enough polynomial that well captures the shape of the compensating differential. In particular, we estimate the mixture

model over the distribution of (unobserved) worker types e, where wages are assumed to follow

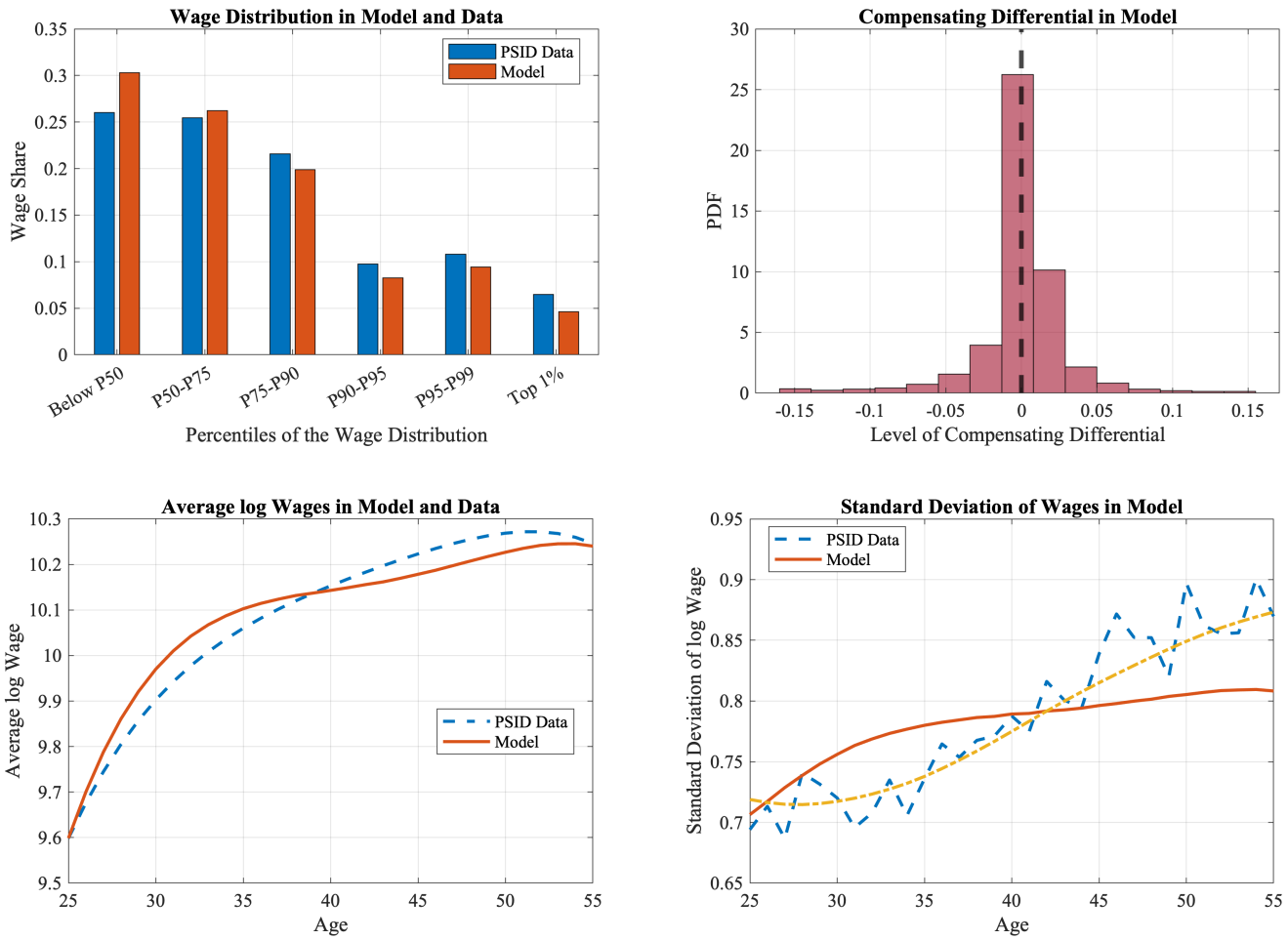
$$w_{n,t}(d,d',k) = \underbrace{\beta_0\left(d',k\right)+\beta_1\left(d',k\right)\mathbb{I}_H+\beta_3\left(d',k\right)\kappa_{n,t}}_{\zeta_{0k}(d',\cdot)}+\underbrace{\beta_4\left(d',k\right)}_{\zeta_{1k}(d',\cdot)}P_{n,t}+\Psi(d,d',s_{n,t})+\epsilon_{n,t}, \quad (24)$$

for a given worker type in a given firm type d and job type $k \in \{L,H\}$, which are treated as observables. Here, the $\mathbb{I}_H$ is a dummy that captures the initial human capital of the individual—for instance, educational attainment. The first two components aim to capture the worker's output, whereas the third term is the compensating differential, $\Psi(d,d',s_{n,t})$. We approximate Ψ via a flexible polynomial form on initial human capital, experience, and beliefs—which turns out to provide a very accurate approximation to the true compensating differential. Notice that if there was no selection in our model, we could estimate this mixture model via maximum likelihood to obtain estimates of the distribution of the underlying worker type, allowing for a flexible polynomial to capture $\Psi(d,d',s_{n,t})$—for instance, estimated using the `mfp` command in Stata. In our model, however, the presence of selection on observables and unobservables precludes using OLS for such estimation. To address this, we apply the extremal quantile selection procedure proposed by D'Haultfoeuille et al. (2018) to estimate semiparametric selection models. We modify their procedure in two ways. First, we incorporate a polynomial fitting step for Ψ within the quantile regression step, which we use to estimate (24). Second, we impose an alternative normalization to be able to recover the full intercept of expected output ($\beta_0(d',k)$), in the wage equation, by normalizing instead the relevant quantile of the (productivity) shock in the wage equation.

Figure 2 shows the results for our calibrated economy. The top left panel shows the distribution of wages, $w_{n,t}(d,d',k)$ generated by our model, which resembles the distribution of labor earnings in the U.S..The top bottom right of Figure 2 shows the distribution of the compensating differential, $\Psi(d,d',s_{n,t})$, which is typically negative, indicating that for a large number of workers in our simulated economy, the compensating differential reduces wages relative to output as they trade human capital accumulation and learning opportunities for higher current wages. Our simulation also matches an increase in average labor income of about 60% for workers between 25 and 55 years old, as observed in U.S. PSID data (bottom-left panel). The bottom right panel shows the standard deviation of labor earnings, which increases by 11 percentage points, which is about 2/3 of the increase observed in PSID over the same period.

Given the simulated economy, the next step is to validate our estimation procedure by estimating the key parameters of 24 to retrieve the coefficients underlying our simulation. The results of

Figure 2: Wages Distribution and Compensating Differential in Simulation



Note: Distribution of worker wages (top left), compensating differential (top right), mean of labor earnings over the life cycle (bottom left), and standard deviation of labor earnings over the life cycle (bottom right). Results based on a simulation of 10 million workers for 44 periods. Top left panel shows annual labor earnings data from PSID for workers between 22 and 65 who are employed head of households. Bars show the share of labor earnings accounted for by different percentiles of the labor earnings distribution.

this exercise are shown in Table 1, which reports the estimates for our simulation with two worker types and two firm types at a few selected quantiles of the income distribution. Here, we focus our discussion on workers with the lowest level of ability (lowest value of $H_{n,t}$); a similar one applies to the other type. Our key result is that our estimation recovers the key parameters of the production function—the linear component shown in column Value—and the fit improves as we move to lower ranks of the wage distribution, which is consistent with the intuition of D'Haultfoeuille et al. (2018): lower quantiles better capture the selection of the job at the first- best vs. the second-best firm. In fact, our procedure is the most accurate when we focus on the bottom 0.1% of the wage distribution, which is close to the optimal quantile obtained using D'Haultfoeuille et al. (2018).

Table 1: Comparison of Parameters and Model Estimates

Coefficient	Value	OLS	Quantile Regressions at				
			0.1%	0.5%	1%	5%	10%
Worker Type 1							
β_0	0.49	0.53	0.49	0.49	0.49	0.49	0.49
$\beta_{H_{n,t}}$	0.17	0.19	0.15	0.16	0.16	0.17	0.18
$\beta_{\kappa_{n,t}}$	12.30	10.02	12.31	12.31	12.31	12.29	11.89
$\beta_{P_{n,t}}$	0.11	3.29	0.12	0.08	0.07	0.06	0.35
Worker Type 2							
β_0	0.21	0.25	0.21	0.20	0.21	0.21	0.21
$\beta_{H_{n,t}}$	0.20	0.14	0.10	0.10	0.10	0.10	0.11
$\beta_{\kappa_{n,t}}$	9.61	6.15	9.56	9.40	9.45	9.15	8.76
$\beta_{P_{n,t}}$	0.14	3.18	0.14	0.25	0.23	0.49	0.85

Note: results from model simulation. Column "Value" indicates the calibrated values of the production function. The rest of the columns are the estimated values of these parameters from a mixture estimation. The underlying model is a fractional polynomial estimation that includes a linear term for $H_{n,t}$, $\kappa_{n,t}$ and $P_{n,t}$ and a flexible polynomial of these variables to capture the compensating differential, $\mathbf{\Psi_{n,t}}$. The estimation is done using a quantile regression at different quantiles of the wage distribution.

5.3 Empirical Application

Having validated our estimation procedure, we now move to estimate the wage equation (21) using U.S. employer-employee match data, namely LEHD data. This rich dataset provides quarterly labor earnings for all workers across 21 states—including California, Florida, and Pennsylvania—from the mid-1990s to 2022. We directly observe each worker's current firm, wage, gender, education, and age. Performance measures—in the model's notation, signal $a_{n,t}$—are not directly observed, and we built a procedure to infer them from workers' variable pay. The idea is that the quantiles of the variable pay distribution identify performance measures to the extent that variable pay is monotonic in performance. Based on these extracted performance measures, we are able to estimate $P_{n,t}$ for each worker n and period t and so treat $P_{n,t}$ as a "covariate" in the subsequent wage estimation step. Note that this construction of $P_{n,t}$ is not necessary for our more general identification framework, where we show how to identify the distribution of $P_{n,t}$ from the wage mixture. Additional details on the procedure used to construct $P_{n,t}$ are provided in Appendix F.

To estimate the wage equation (21), we assume that $\epsilon_{n,t}(d, e)$ is normally distributed conditional on $e_n = e$. Consequently, if there was no selection on $\epsilon_{n,t}$, the distribution of $w_{n,t}$ conditional on $(D_{n,t}, H_{n,1}, \kappa_{n,t}, P_{n,t})$ would be a finite (because $\mathcal{E}$ is finite) mixture of Normal distributions. In that case, we could simply use the Stata command `fmm` to estimate the wage parameters, since it combines the estimation of Normal mixtures with OLS. As in our simulation exercise, if there were no worker efficiency types e_n, we could rely on the Stata package `eqregsel` to run extremal quantile regressions that account for selection on $\epsilon_{n,t}$ (D'Haultfoeuille et al., 2018, 2020). To accommodate both aspects simultaneously, we follow our estimation results and use our implementation of

`eqregsel` developed by D'Haultfoeuille and Maurel (2013) to allow for mixture estimation and polynomial approximation of the compensating differential.

Our empirical results (currently awaiting approval from the Census Bureau before disclosure) corroborate the findings from our simulations. In particular,the AKM estimates of ρ based on the wage equation (22) fall below our own estimates of this parameter. This is because the estimated $\Psi(\cdot)$ is, on average, negative, suggesting that workers tend to match primarily with firms offering human capital and informational gains associated with high future wage returns. This finding indicates that our model provides a potential avenue for the resolution of the puzzle of low sorting.

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

6 Conclusion

In this paper, we examine the empirical content of a large class of dynamic matching models of the labor market with ex-ante heterogeneous firms and workers, symmetric uncertainty and learning about workers' ability, and firm monopsony power. We allow workers' ability and human capital, acquired before and after entry in the labor market, to be general across firms and jobs to varying degrees. Such a class nests and extends known models that have been used to study worker turnover across firms, occupational choice, wage differentials across jobs, firms, and occupations, and wage inequality across workers and over the life cycle.

We provide a novel argument to establish that these models are identified under intuitive conditions, solely from data on job choices and wages. In particular, we do not rely on any additional information that could facilitate the identification of the learning process, such as proxies for beliefs or direct measures of signals about ability. Moreover, we do not impose any restrictions on endogenous variables or on the dynamics of states, choices, and outcomes. Instead, our argument

rests on conditions that allow for arbitrary patterns of selection based on endogenously time-varying unobservables, are easy to verify, impose minimal data requirements, and yield a simple constructive estimator of the primitives of interest, as shown in our empirical application.

Using this framework, we revisit an outstanding puzzle regarding the role of labor market sorting for wage inequality. We demonstrate that ignoring the dynamics of the matching process between firms and workers due to human capital acquisition and learning about ability—and the resulting compensating differentials in wages when firms differ in the human capital and information opportunities they offer—can lead to a systematically underestimating the importance of sorting for wages.

References

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

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

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

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

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.

ARELLANO, M. AND S. BONHOMME (2017): "Quantile Selection Models With an Application to Understanding Changes in Wage Inequality," *Econometrica*, 85, 1–28.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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*.

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.

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

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

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

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

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

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

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.

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.

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.

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.

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

A Extensions of Identification Argument

We discuss here extensions of our identification framework.

Support $\mathcal{E}$ of Efficiency e_n. To identify the wage mixture (19), Assumption 3(ii) imposes that $\mathcal{E}$ is finite. If e_n is continuous (and potentially multidimensional), then the wage mixture (19) becomes a *continuous* mixture of potentially continuous Gaussian mixtures, making identification more challenging. This impasse can be easily resolved by assuming away selection of $D_{n,t}$ based on $\epsilon_{n,t}$. In that case, Assumption 3(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 (19) is a continuous mixture of Normals, whose identification is established by Bruni and Koch (1985)'s Theorem 1.

Support $\mathcal{A}$ of Signal $a_{n,t}$. As for $\mathcal{E}$, Proposition 4 can also be adapted to cases where $a_{n,t}$ is continuous (and potentially multidimensional), provided that there is no selection of $D_{n,t}$ based on $\epsilon_{n,t}$. To identify the learning process in Proposition 5, Assumption 4(i) imposes that $\mathcal{A}$ has a cardinality of two. As explained in Section 4.4, 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 $a_{n,t}$, provided that the signal distribution remains an identifiable mixture. For instance, if $a_{n,t}$ is distributed as a continuous and multivariate Gaussian mixture conditional on θ_n, then the signal distribution would then be a finite mixture of continuous and multivariate Gaussian mixtures (finite because Θ is finite), which remains identifiable according to Bruni and Koch (1985), as discussed in their Section 4.9.

Support Θ of Ability $\theta_{n,t}$. To identify the learning process in Proposition 5, Assumption 4(i) requires that Θ has cardinality two. This restriction allows us to model the signal distribution as a binomial mixture over the unobserved ability θ_n with *two* components. The binomial aspect arises because $\mathcal{A}$ has cardinality two, and the two components of this binomial mixture correspond to the cardinality of Θ. This mixture is identifiable, as shown by Blischke (1964) and Blischke (1978), provided that the number of periods where workers are observed at each given job d is at least $2r-1=3$, where $r=|\Theta|=2$ represents the number of mixture components (see Appendix D for more details). Keeping $\mathcal{A}$ 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 $a_{n,t}$ follows a multivariate Normal distribution conditional on θ_n, then the signal distribution is a continuous mixture of multivariate Normals, identified by Bruni and Koch (1985)'s Theorem 1.

B Additional Results

B.1 Micro-Fundation of Assumption (iii) of Proposition 3

Lemma 1 shows that if the productivity shocks are "sufficiently independent," then Assumption (iii) of Proposition 3 holds.

Lemma 1 (Moderate Dependence)**.** *Let Assumptions (i)–(ii) of Proposition 3 hold. For some* $q_1 \in (0, 1]$*, let*

$$\lim_{u \to +\infty} \Pr\big(\epsilon_n(0) \leq \epsilon_n(1) + a \,\big|\, \epsilon_n(1) \geq u\big) = q_1 \qquad \textit{for all } a \in \mathbb{R}. \tag{25}$$

Then, Assumption (iii) of Proposition 3 holds:

$$\lim_{w \to +\infty} \Pr\big(D_n = 1 \mid X_n = x,\ w_n(1) \geq w\big) = q_1 \quad \textit{for every } x.$$

Moreover, if $\epsilon_n(0)$ *and* $\epsilon_n(1)$ *are independent, then* $q_1 = 1$*.*

Proof. Fix a realisation x of X_n and $w \in \mathbb{R}$. In the static Roy model,

$$\begin{aligned}\Pr(D_n = 1 \mid X_n = x,\ w_n(1) \geq w) &= \Pr\big(y(1,x) + \epsilon_n(1) \geq y(0,x) + \epsilon_n(0) \,\big|\, X_n = x,\ y(1,x) + \epsilon_n(1) \geq w\big) \\ &= \Pr\big(\epsilon_n(0) \leq \epsilon_n(1) + y(1,x) - y(0,x) \,\big|\, \epsilon_n(1) \geq w - y(1,x)\big),\end{aligned} \tag{26}$$

where the last equality uses Assumption (i) of Proposition 3.[22]

Set $u \coloneqq w - y(1,x)$, so $w \to +\infty$ iff $u \to +\infty$. Applying (25) with $a = y(1,x) - y(0,x)$ gives

$$\lim_{w \to +\infty} \Pr\big(\epsilon_n(0) \leq \epsilon_n(1) + a(x) \,\big|\, \epsilon_n(1) \geq w - y(1,x)\big) = q_1. \tag{28}$$

[22] In our dynamic generalised equilibrium Roy model, we can likewise obtain an equation analogous to (26). In fact, under Assumption 2(i), the equilibrium is efficient. In an efficient equilibrium, job choices maximise the expected present discounted value of output. Therefore, for any $(d, d', e, s) \in \mathcal{D}^2 \times \mathcal{E} \times \mathcal{S}_t(d, d', e)$ and $w \in \mathbb{R}$,

$$\begin{aligned}&\Pr\big(D_{n,t} = d \,\big|\, s_{n,t}(e) = s,\ w_{n,t}(d, d', e) \leq w\big) \\ &= \Pr\big(Y(d,s) + \epsilon_{n,t}(d,e) \ \geq\ Y(d',s) + \epsilon_{n,t}(d',e) \,\big|\, s_{n,t}(e) = s,\ \varphi(d,d',s) + \epsilon_{n,t}(d',e) \leq w\big) \\ &= \Pr\big(\epsilon_{n,t}(d,e) \ \geq\ \epsilon_{n,t}(d',e) + Y(d',s) - Y(d,s) \,\big|\, \epsilon_{n,t}(d',e) \leq w - \varphi(d,d',s)\big),\end{aligned} \tag{27}$$

where $Y(d,s) + \epsilon_{n,t}(d,e)$ is the expected present discounted value of output for firm d in state s after productivity shocks have realised; the last equality uses Assumption 5.

By (26), (28) is precisely

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

which is Assumption (iii) of Proposition 3.

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

$$\begin{aligned} \Pr(\epsilon_n(0) \leq \epsilon_n(1) + a \mid \epsilon_n(1) \geq u) &= \mathbb{E}\Big[\Pr(\epsilon_n(0) \leq \epsilon_n(1) + a \mid \epsilon_n(1)) \,\Big|\, \epsilon_n(1) \geq u \Big] \\ &= \mathbb{E}\Big[F_{\epsilon_n(0)}\big(\epsilon_n(1) + a\big) \,\Big|\, \epsilon_n(1) \geq u \Big], \end{aligned} \tag{29}$$

where the second line uses independence of $\epsilon_n(0)$ and $\epsilon_n(1)$. Since $F_{\epsilon_n(0)}$ is nondecreasing,

$$F_{\epsilon_n(0)}(u + a) \;\leq\; F_{\epsilon_n(0)}\big(\epsilon_n(1) + a\big) \;\leq\; 1 \quad \text{on the event } \{\epsilon_n(1) \geq u\}.$$

Taking conditional expectations yields the bounds

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

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

$$\lim_{u\to+\infty} \Pr(\epsilon_n(0) \leq \epsilon_n(1) + a \mid \epsilon_n(1) \geq u) = 1,$$

so $q_1 = 1$. □

Remark. Suppose that $\epsilon_n(1)$ and $\epsilon_n(0)$ are jointly normal—or lognormal. If $\text{cov}(\epsilon_n(1), \epsilon_n(0)) < \text{Var}(\epsilon_n(1))$—or if $\text{cov}(\log(\epsilon_n(1)), \log(\epsilon_n(0))) < \text{Var}(\log(\epsilon_n(1)))$—then (25) holds with $q_1 = 1$. 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.

B.2 Proposition 3 with Bounded Support

Proposition 12 establishes identification of the deterministic wage components in the case where the potential wages $w_n(1) \mid X_n = x$ and the observed, selected wages $w_n \mid (D_n = 1, X_n = x)$ have different right endpoints. The identification result retains the spirit of Proposition 3, but extra care is needed in taking limits because the two endpoints differ. We further show that, when finite, the right and left endpoints of the potential wages $w_n(1) \mid X_n = x$ are identified (Corollary 6).

Proposition 12 (Deterministic Wage Component with Bounded Supports)**.** *Assume:*

(i) (Exogeneity.) $\epsilon_n(1)$ *is independent of* X_n*.*

(ii) (Supports.) For each realisation x *of* X_n*,*

$$\omega(x) \coloneqq \sup\{u : \Pr(w_n(1) \le u \mid X_n = x) < 1\} \le +\infty,$$
$$\omega_{\text{obs}}(x) \coloneqq \sup\{u : \Pr(w_n \le u \mid D_n = 1, X_n = x) < 1\} < \omega(x),$$
$$0 < \Pr(D_n = 1 \mid X_n = x) \le 1,$$

with $\omega(x)$ *and* $\omega_{\text{obs}}(x)$ *potentially unknown.*

(iii) (Relative Tail Decay.) For each realisation x *of* X_n*, define*

$$r_x(u) \;\coloneqq\; \Pr\big(D_n = 1 \mid X_n = x,\; w_n(1) > Q_{w_n(1)|X_n=x}(u)\big), \qquad u \in (0,1).$$

There exists an (unknown) constant $q_1 \in (0, +\infty)$ *such that, for every* x *and a fixed reference* $\bar{x}$*,*

$$\lim_{u \to 1} \frac{r_x(u)}{r_{\bar{x}}(u)} = q_1.$$

(iv) (Tail Regularity.) For each realisation x *of* X_n*, there exist (unknown) thresholds* $w_x < +\infty$ *and* $w_x^{\text{obs}} < \omega_{\text{obs}}(x)$ *such that* $F_{w_n(1)|X_n=x}$ *and* $F_{w_n|D_n=1,X_n=x}$ *are* continuous and strictly increasing *on* $(w_x, +\infty)$ *and* $(w_x^{\text{obs}}, \omega_{\text{obs}}(x))$*, respectively. Moreover,* $F_{w_n|D_n=1,X_n=x}$ *is* continuous at the endpoint*:*

$$\lim_{w \to \omega_{\text{obs}}(x)} F_{w_n|D_n=1,X_n=x}(w) = 1.$$

(v) (Normalization.) There exists a known realisation $\bar{x}$ *of* X_n *with* $y(1, \bar{x}) = 0$*.*

For each realisation x *of* X_n*, let* $\{\tau_{\bar{x}}^{(k)}\}_{k \ge 1} \subset (0,1)$ *be any sequence with* $\tau_{\bar{x}}^{(k)} \to 1$ *as* $k \to +\infty$*. Define*

$$1 - \tau_x^{(k)} \;=\; \frac{\Pr(D_n = 1 \mid X_n = \bar{x})}{\Pr(D_n = 1 \mid X_n = x)} \big(1 - \tau_{\bar{x}}^{(k)}\big). \tag{30}$$

Then,

$$\lim_{k \to +\infty} \Big[\, Q_{w_n \,|\, D_n=1, X_n=x}\big(\tau_x^{(k)}\big) \;-\; Q_{w_n \,|\, D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) \,\Big] \;=\; y(1,x). \tag{31}$$

Hence, $y(1,x)$ *is identified (up to the normalization at* $\bar{x}$*).*

Proof. To facilitate reading, we divide the proof into steps and box the key equations in each step.

Step 1 (Bayes rule). Fix a realisation x of X_n. For any real w, Bayes' rule gives

$$
\begin{aligned}
S_{w_n|D_n=1,X_n=x}(w) &= \frac{\Pr\big(w_n(1) > w \mid X_n = x\big)\ \Pr\big(D_n = 1 \mid X_n = x,\ w_n(1) > w\big)}{\Pr(D_n = 1 \mid X_n = x)} \\
&= S_{w_n(1)|X_n=x}(w)\ \frac{\Pr\big(D_n = 1 \mid X_n = x,\ w_n(1) > w\big)}{\Pr(D_n = 1 \mid X_n = x)}.
\end{aligned}
\tag{32}
$$

Define

$$
r_x(u) \ :=\ \Pr\big(D_n = 1 \mid X_n = x,\ w_n(1) > Q_{w_n(1)|X_n=x}(u)\big), \qquad u \in (0,1).
$$

Evaluating (32) at $w = Q_{w_n(1)|X_n=x}(u)$ yields

$$
\boxed{S_{w_n|D_n=1,X_n=x}\big(Q_{w_n(1)|X_n=x}(u)\big) \ =\ (1-u)\,\frac{r_x(u)}{\Pr(D_n = 1 \mid X_n = x)}, \qquad u \in (0,1).}
\tag{33}
$$

Step 2 (Behaviour of the composed survival near the observed endpoint). By Assumption (iv), there exist thresholds $w_x < \omega(x)$ and $w_x^{\mathrm{obs}} < \omega_{\mathrm{obs}}(x)$ such that $F_{w_n(1)|X_n=x}$ is continuous and strictly increasing on $(w_x, \omega(x))$, and $F_{w_n|D_n=1,X_n=x}$ is continuous and strictly increasing on $(w_x^{\mathrm{obs}}, \omega_{\mathrm{obs}}(x))$. Define

$$
u_x^* := F_{w_n(1)|X_n=x}(w_x), \qquad \tau_x^* := F_{w_n|D_n=1,X_n=x}(w_x^{\mathrm{obs}}),
$$

so $Q_{w_n(1)|X_n=x} : (u_x^*, 1) \to (w_x, \omega(x))$ and $Q_{w_n|D_n=1,X_n=x} : (\tau_x^*, 1) \to (w_x^{\mathrm{obs}}, \omega_{\mathrm{obs}}(x))$ are strictly increasing. Because $\omega_{\mathrm{obs}}(x) < \omega(x)$, set

$$
\bar{u}_x := \sup\big\{u \in (u_x^*, 1) :\ Q_{w_n(1)|X_n=x}(u) < \omega_{\mathrm{obs}}(x)\big\} \in (u_x^*, 1).
$$

Then, $Q_{w_n(1)|X_n=x}(u) \to \omega_{\mathrm{obs}}(x)$ as $u \to \bar{u}_x$. Since $Q_{w_n(1)|X_n=x}(u)$ is increasing and $\omega_{\mathrm{obs}}(x)$ is finite, there exists $\tilde{u}_x \in (u_x^*, \bar{u}_x)$ such that $Q_{w_n(1)|X_n=x}(u) \in (w_x^{\mathrm{obs}}, \omega_{\mathrm{obs}}(x))$ for all $u \in (\tilde{u}_x, \bar{u}_x)$. On that interval the map

$$
u \ \longmapsto\ S_{w_n|D_n=1,X_n=x}\big(Q_{w_n(1)|X_n=x}(u)\big)
$$

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 $(\tilde{u}_x, \bar{u}_x)$. By the endpoint continuity in Assumption (iv),

$$
\lim_{w \to \omega_{\mathrm{obs}}(x)} F_{w_n|D_n=1,X_n=x}(w) = 1,
$$

and therefore

$$
\boxed{\lim_{u \to \bar{u}_x} S_{w_n|D_n=1,X_n=x}\big(Q_{w_n(1)|X_n=x}(u)\big) = 0.}
\tag{34}
$$

Step 3 (Exact tail matching). By the continuity and strict decrease of $u \mapsto S_{w_n|D_n=1,X_n=x}(Q_{w_n(1)|X_n=x}(u))$ on $(\tilde{u}_x, \bar{u}_x)$ and its limit 0 as $u \to \bar{u}_x$, there exists $\tilde{\tau}_x \in (\tau_x^*, 1)$ such that for every $\tau \in (\tilde{\tau}_x, 1)$ there is a unique $u_x(\tau) \in (\tilde{u}_x, \bar{u}_x)$ solving

$$S_{w_n|D_n=1,X_n=x}\big(Q_{w_n(1)|X_n=x}(u_x(\tau))\big) = 1 - \tau.$$

Combining this with $S_{w_n|D_n=1,X_n=x}(Q_{w_n|D_n=1,X_n=x}(\tau)) = 1 - \tau$ for all $\tau \in (\tau_x^*, 1)$ and the strict decrease of $w \mapsto S_{w_n|D_n=1,X_n=x}(w)$ on $(w_x^{\text{obs}}, \omega_{\text{obs}}(x))$ yields

$$\boxed{Q_{w_n|D_n=1,X_n=x}(\tau) = Q_{w_n(1)|X_n=x}(u_x(\tau)) \qquad \text{for all } \tau \in (\tilde{\tau}_x, 1).} \tag{35}$$

Moreover,

$$\boxed{\lim_{\tau \to 1} u_x(\tau) = \bar{u}_x.} \tag{36}$$

Step 4 (Cross-x τ-alignment and the product identity). Fix any sequence $\{\tau_{\bar{x}}^{(k)}\}_{k \geq 1} \subset (0, 1)$ with $\tau_{\bar{x}}^{(k)} \to 1$. For each x, define

$$\boxed{1 - \tau_x^{(k)} \;=\; \frac{\Pr(D_n = 1 \mid X_n = \bar{x})}{\Pr(D_n = 1 \mid X_n = x)} \big(1 - \tau_{\bar{x}}^{(k)}\big).} \tag{37}$$

Let $u_x^{(k)} := u_x(\tau_x^{(k)})$ and $u_{\bar{x}}^{(k)} := u_{\bar{x}}(\tau_{\bar{x}}^{(k)})$. Using (33) at $u = u_x^{(k)}$ and $u = u_{\bar{x}}^{(k)}$,

$$1 - \tau_x^{(k)} \;=\; (1 - u_x^{(k)}) \, \frac{r_x(u_x^{(k)})}{\Pr(D_n = 1 \mid X_n = x)}, \qquad 1 - \tau_{\bar{x}}^{(k)} \;=\; (1 - u_{\bar{x}}^{(k)}) \, \frac{r_{\bar{x}}(u_{\bar{x}}^{(k)})}{\Pr(D_n = 1 \mid X_n = \bar{x})}.$$

Divide the two equalities and use (37) to obtain

$$\boxed{\frac{(1 - u_x^{(k)}) \, r_x(u_x^{(k)})}{(1 - u_{\bar{x}}^{(k)}) \, r_{\bar{x}}(u_{\bar{x}}^{(k)})} \;=\; 1.} \tag{38}$$

Step 5 (Aligning tail probabilities across covariates). Under Assumption (iii),

$$\lim_{k \to +\infty} \frac{r_x(u_x^{(k)})}{r_{\bar{x}}(u_{\bar{x}}^{(k)})} = q_1 \in (0, +\infty).$$

By (36), $u_{\bar{x}}^{(k)} \to \bar{u}_{\bar{x}}$ and $u_x^{(k)} \to \bar{u}_x$. Since $\bar{u}_{\bar{x}}, \bar{u}_x < 1$ and $r_{\bar{x}}(\cdot), r_x(\cdot)$ are continuous near those limits (by Assumption (iv)), (38) implies

$$\lim_{k \to +\infty} \frac{1 - u_x^{(k)}}{1 - u_{\bar{x}}^{(k)}} = 1,$$

and therefore

$$\boxed{\lim_{k\to+\infty}\left(u_x^{(k)} - u_{\bar{x}}^{(k)}\right) = 0.} \tag{39}$$

Step 6 (Identification by differencing). Assumption (i) implies

$$Q_{w_n(1)|X_n=x}(u) \;=\; y(1,x) \;+\; Q_{\epsilon_n(1)}(u) \qquad \text{for all } u \in (0,1).$$

Apply (35) at $\tau = \tau_x^{(k)}$ and at $\tau = \tau_{\bar{x}}^{(k)}$:

$$\begin{aligned} Q_{w_n|D_n=1,X_n=x}\big(\tau_x^{(k)}\big) &= y(1,x) + Q_{\epsilon_n(1)}(u_x^{(k)}), \\ Q_{w_n|D_n=1,X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) &= y(1,\bar{x}) + Q_{\epsilon_n(1)}(u_{\bar{x}}^{(k)}). \end{aligned} \tag{40}$$

By (39) and continuity of $Q_{\epsilon_n(1)}$ near the upper tail,

$$\lim_{k\to+\infty}\Big(Q_{\epsilon_n(1)}(u_x^{(k)}) - Q_{\epsilon_n(1)}(u_{\bar{x}}^{(k)})\Big) = 0.$$

Subtracting the equations in (40) and using $y(1,\bar{x}) = 0$ (Assumption (v)) yields the identification result:

$$\boxed{\lim_{k\to+\infty}\Big[\,Q_{w_n\,|\,D_n=1,X_n=x}\big(\tau_x^{(k)}\big) \;-\; Q_{w_n\,|\,D_n=1,X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big)\,\Big] \;=\; y(1,x).}$$

□

Remark. The only substantive difference between Proposition 3 and Proposition 12—apart from the support restrictions in Assumption (ii)—is that Assumption (iii) in the unbounded case is replaced, in the bounded case, by a *relative tail decay* condition. For reference, Assumption (iii) of Proposition 3 posits that there exists an (unknown) constant $q_1 \in (0,1]$ such that, for each realisation x of X_n,

$$\lim_{w\to\omega(x)} \Pr\big(D_n = 1 \mid X_n = x,\; w_n(1) > w\big) = q_1. \tag{41}$$

The requirement (41) is too strong—and in fact necessarily violated—under a strict support gap $\omega_{\text{obs}}(x) < \omega(x)$ (Assumption (ii) of Proposition 12). To see this, Bayes' rule (Step 1 of the proof) implies, for any realisation x of X_n and any w,

$$S_{w_n|D_n=1,X_n=x}(w) = \; S_{w_n(1)|X_n=x}(w)\,\frac{\Pr(D_n = 1 \mid X_n = x,\; w_n(1) > w)}{\Pr(D_n = 1 \mid X_n = x)}.$$

For any $w \in (\omega_{\text{obs}}(x), \omega(x))$ we have $S_{w_n|D_n=1,X_n=x}(w) = 0$ while $S_{w_n(1)|X_n=x}(w) > 0$ and

$\Pr(D_n = 1 \mid X_n = x) > 0$, hence

$$\Pr(D_n = 1 \mid X_n = x,\ w_n(1) > w) = 0 \quad \text{for all } w \in (\omega_{\text{obs}}(x), \omega(x)).$$

Therefore the tail selection probability collapses to zero as $w \to \omega(x)$, forcing $q_1 = 0$ in (41). A positive limit could arise only in the case $\omega_{\text{obs}}(x) = \omega(x)$, which is excluded by Assumption (ii). This is why we adopt a *relative* tail condition in place of (41): it governs the *rate* at which tail probabilities vanish across x (via ratios) rather than imposing a common nonzero limit that cannot hold under a support gap.

B.3 Identification of Support Endpoints

Corollary 6 shows that, when finite, the right and left endpoints of the potential wages $w_n(1) \mid X_n = x$ are nonparametrically identified. Intuitively, for each x, the endpoints of the observed, selected wage distribution, $\underline{\omega}_{\text{obs}}(x)$ and $\omega_{\text{obs}}(x)$, are read directly from extreme quantiles: very low quantiles approach the lower endpoint and very high quantiles approach the upper endpoint. Because the deterministic part of wages $y(1, x)$ is already known by Proposition 12, we can "shift" these observed endpoints to learn about the latent endpoints of both the shock $\epsilon_n(1)$ and the potential wage $w_n(1) = y(1, x) + \epsilon_n(1)$. Selection trims the extremes, so the observed support sits inside the latent one: $\underline{\omega}_{\text{obs}}(x) \geq \underline{\omega}(x)$ and $\omega_{\text{obs}}(x) \leq \omega(x)$, with $\omega(x) = y(1, x) + \omega_\epsilon$ and $\underline{\omega}(x) = y(1, x) + \underline{\omega}_\epsilon$. Taking the best (tightest) such shifts across x gives bounds:

$$\sup_x \{\omega_{\text{obs}}(x) - y(1, x)\} \leq \omega_\epsilon, \qquad \underline{\omega}_\epsilon \leq \inf_x \{\underline{\omega}_{\text{obs}}(x) - y(1, x)\},$$

and adding back $y(1, x)$ yields corresponding tighest bounds for $\omega(x)$ and $\underline{\omega}(x)$. Moreover, if there exists a covariate value $x^\star$ where selection does *not* truncate the top ($\omega_{\text{obs}}(x^\star) = \omega(x^\star)$), the upper latent endpoint is revealed by the extreme quantile at $x^\star$:

$$\omega_\epsilon = \lim_{\tau \to 1} \Big\{ Q_{w_n \mid D_n = 1, X_n = x^\star}(\tau) - y(1, x^\star) \Big\},$$

and then $\omega(x) = y(1, x) + \omega_\epsilon$ for every x. A symmetric argument applies to the lower endpoint if selection does not truncate the bottom at some $x^\dagger$.

Corollary 6 (Identification of finite right and left endpoints of $\epsilon_n(1)$ and $w_n(1)$)**.** *Maintain the assumptions of Proposition 12, implying that $y(1, x)$ is identified for each realisation x of X_n. In addition, assume finite and distinct endpoints, with two-sided tail regularity: for each realisation x*

of X_n,

$$\underline{\omega}(x) \coloneqq \inf\{u : \Pr(w_n(1) \leq u \mid X_n = x) > 0\} > -\infty,$$
$$\omega(x) \coloneqq \sup\{u : \Pr(w_n(1) \leq u \mid X_n = x) < 1\} < +\infty,$$
$$\underline{\omega}_{\text{obs}}(x) \coloneqq \inf\{u : \Pr(w_n \leq u \mid D_n = 1, X_n = x) > 0\} > \underline{\omega}(x) > -\infty,$$
$$\omega_{\text{obs}}(x) \coloneqq \sup\{u : \Pr(w_n \leq u \mid D_n = 1, X_n = x) < 1\} < \omega(x) < +\infty,$$

with $F_{w_n(1)|X_n=x}$ continuous and strictly increasing on $(\underline{\omega}(x),\, w_x) \cup (w'_x,\, \omega(x))$ for some $w_x < w'_x$, and $F_{w_n|D_n=1,X_n=x}$ continuous and strictly increasing on $(\underline{\omega}_{\text{obs}}(x),\, w_x^{\text{obs}}) \cup ((w_x^{\text{obs}})',\, \omega_{\text{obs}}(x))$ for some $w_x^{\text{obs}} < (w_x^{\text{obs}})'$, as well as continuous at both endpoints:

$$\lim_{w \to \underline{\omega}_{\text{obs}}(x)} F_{w_n|D_n=1,X_n=x}(w) = 0, \qquad \lim_{w \to \omega_{\text{obs}}(x)} F_{w_n|D_n=1,X_n=x}(w) = 1.$$

Define the shock $\epsilon_n(1)$ (finite) endpoints as:

$$\underline{\omega}_\epsilon \coloneqq \inf\{u \in \mathbb{R} : F_{\epsilon_n(1)}(u) > 0\} > -\infty, \qquad \omega_\epsilon \coloneqq \sup\{u \in \mathbb{R} : F_{\epsilon_n(1)}(u) < 1\} < +\infty.$$

Then:

(a) (Observed wage endpoints are identified.) For every realisation x of X_n, $\underline{\omega}_{\text{obs}}(x)$ and $\omega_{\text{obs}}(x)$ are identified:

$$\underline{\omega}_{\text{obs}}(x) = \lim_{\tau \to 0} Q_{w_n|D_n=1,X_n=x}(\tau), \qquad \omega_{\text{obs}}(x) = \lim_{\tau \to 1} Q_{w_n|D_n=1,X_n=x}(\tau).$$

(b) (Sharp bounds for latent endpoints.) For every realisation x of X_n, a lower (resp. upper bound) bound for ω_ϵ (resp. $\underline{\omega}_\epsilon$) and a lower (resp. upper bound) bound for $\omega(x)$ (resp. $\underline{\omega}(x)$) are identified:

$$L_\epsilon \coloneqq \sup_{x'} \left\{\omega_{\text{obs}}(x') - y(1,x')\right\} \;\leq\; \omega_\epsilon, \qquad U_\epsilon \coloneqq \inf_{x'} \left\{\underline{\omega}_{\text{obs}}(x') - y(1,x')\right\} \;\geq\; \underline{\omega}_\epsilon,$$
$$\underline{\omega}(x) \;\leq\; \min\Big\{\, \underline{\omega}_{\text{obs}}(x),\;\; y(1,x) + U_\epsilon \,\Big\}, \qquad \omega(x) \;\geq\; \max\Big\{\, \omega_{\text{obs}}(x),\;\; y(1,x) + L_\epsilon \,\Big\}.$$

Moreover, these bounds are sharp under the stated assumptions.

(c) (Right endpoint point identification under no top truncation at some $x^\star$.) If there exists a known realisation $x^\star$ of X_n with $\omega_{\text{obs}}(x^\star) = \omega(x^\star)$ (i.e. the finite right endpoint of the selected observed wages equals the finite right endpoint of the potential wages; in other words, selection does not affect the rightmost support of wages at x^*), then, for every realisation x of X_n,

ω_ϵ and $\omega(x)$ are identified:

$$\omega_\epsilon = \lim_{\tau \to 1} \Big[Q_{w_n | D_n = 1, X_n = x^\star}(\tau) - y(1, x^\star) \Big],$$
$$\omega(x) = y(1, x) + \lim_{\tau \to 1} \Big[Q_{w_n | D_n = 1, X_n = x^\star}(\tau) - y(1, x^\star) \Big].$$

(d) (Left endpoint point identification under no bottom truncation at some $x^\dagger$.) If there exists a known realisation $x^\dagger$ of X_n with $\underline{\omega}_{\text{obs}}(x^\dagger) = \underline{\omega}(x^\dagger)$ (i.e. the finite left 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 $x^\dagger$), then, for every realisation x of X_n, $\underline{\omega}_\epsilon$ and $\underline{\omega}(x)$ are identified:

$$\underline{\omega}_\epsilon = \lim_{\tau \to 0} \Big[Q_{w_n | D_n = 1, X_n = x^\dagger}(\tau) - y(1, x^\dagger) \Big],$$
$$\underline{\omega}(x) = y(1, x) + \lim_{\tau \to 0} \Big[Q_{w_n | D_n = 1, X_n = x^\dagger}(\tau) - y(1, x^\dagger) \Big].$$

Proof. We present the proof for right endpoints; the argument for left endpoints is symmetric.
(a) Fix any realisation x of X_n. By Assumption (iv) of Proposition 12, $F_{w_n | D_n = 1, X_n = x}$ is continuous and strictly increasing on $(w_x^{\text{obs}}, \omega_{\text{obs}}(x))$ and continuous at the endpoint. Therefore, its upper quantiles converge to the endpoint, yielding (a).
(b) Fix any realisation x of X_n. Assumption (i) of Proposition 12 implies

$$\omega(x) \;=\; y(1, x) + \omega_\epsilon. \tag{42}$$

By Assumption (ii) of Proposition 12, $\omega_{\text{obs}}(x) < \omega(x)$ for each x, so

$$\omega_{\text{obs}}(x) - y(1, x) \;<\; \omega(x) - y(1, x) \;=\; \omega_\epsilon.$$

Taking the supremum over x yields a lower bound for ω_ϵ. Adding $y(1, x)$ gives a lower bound for $\omega(x)$. These bounds are the best possible (sharp) without further restrictions.
(c) Fix any realisation x of X_n. If there exists a known realisation $x^\star$ of X_n with $\omega_{\text{obs}}(x^\star) = \omega(x^\star)$, then by (a), $\omega(x^\star)$ is identified:

$$\omega(x^\star) = \lim_{\tau \to 1} Q_{w_n | D_n = 1, X_n = x^\star}(\tau).$$

Using (42) written for x^* and recalling that $y(1, x^\star)$ is identified by Proposition 12 gives $\omega_\epsilon = \omega(x^\star) - y(1, x^\star)$. We plug this into (42) and complete the proof. □

B.4 Proposition 3 with Location and Scale Parameters

Propositions 3 and 12 extend to wage specifications in which the shock $\epsilon_n(1)$ is multiplied by a scale parameter $\sigma(1, X_n) > 0$:

$$w_n = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\} w_n(d) = \sum_{d\in\{0,1\}} \mathbb{1}\{D_n = d\}\big[y(d, X_n) + \sigma(d, X_n)\epsilon_n(d)\big]. \tag{43}$$

Proposition 13 (Identification of $y(1,\cdot)$ and $\sigma(1,\cdot)$)**.** *Assume:*

(i) (Exogeneity.) $\epsilon_n(1)$ *is independent of* X_n*.*

(ii) (Supports.)[23] *For each realisation* x *of* X_n*,*

$$\omega(x) \coloneqq \sup\{u : \Pr(w_n(1) \le u \mid X_n = x) < 1\} = +\infty,$$
$$\omega_{\text{obs}}(x) \coloneqq \sup\{u : \Pr(w_n \le u \mid D_n = 1, X_n = x) < 1\} = +\infty,$$
$$0 < \Pr(D_n = 1 \mid X_n = x) \le 1.$$

(iii) (Tail Limit.) There exists an (unknown) constant $q_1 \in (0, 1]$ *such that for every realisation* x *of* X_n*,*

$$\lim_{w\to+\infty} \Pr\big(D_n = 1 \mid X_n = x,\ w_n(1) > w\big) = q_1.$$

(iv) (Tail Regularity.) For each realisation x *of* X_n*, there exist (unknown) thresholds* $w_x < +\infty$ *and* $w_x^{\text{obs}} < +\infty$ *such that the cumulative distribution functions* $F_{w_n(1)|X_n=x}$ *and* $F_{w_n|D_n=1,X_n=x}$ *are continuous and strictly increasing on* $(w_x, +\infty)$ *and* $(w_x^{\text{obs}}, +\infty)$*, respectively.*

(v) (Normalization.) There exists a known realisation $\bar{x}$ *of* X_n *with* $y(1, \bar{x}) = 0$ *and* $\sigma(1, \bar{x}) = 1$*.*

For each realisation x *of* X_n*, define*

$$c(1, x) \ \coloneqq\ \frac{q_1}{\Pr(D_n = 1 \mid X_n = x)} \ \in (0, \infty).$$

Fix the following sequences

$$\tau_{\bar{x}}^{(k)} \coloneqq 1 - 2^{-k}, \qquad \tilde{\tau}_{\bar{x}}^{(k)} \coloneqq 1 - 3^{-k}, \qquad k = 1, 2, \ldots$$

and, for any x*, define the* x*-specific re-indexed tails*

$$1 - \tau_x^{(k)} \coloneqq \frac{c(1, x)}{c(1, \bar{x})}\big(1 - \tau_{\bar{x}}^{(k)}\big), \qquad 1 - \tilde{\tau}_x^{(k)} \coloneqq \frac{c(1, x)}{c(1, \bar{x})}\big(1 - \tilde{\tau}_{\bar{x}}^{(k)}\big).$$

[23] We focus on the case of unbounded supports. The bounded-support case follows analogously, with the technical modifications highlighted in Appendix B.2.

Then,

$$\sigma(1,x) = \lim_{k\to+\infty} \frac{Q_{w_n|D_n=1,X_n=x}(\tau_x^{(k)}) - Q_{w_n|D_n=1,X_n=x}(\tilde{\tau}_x^{(k)})}{Q_{w_n|D_n=1,X_n=\bar{x}}(\tau_{\bar{x}}^{(k)}) - Q_{w_n|D_n=1,X_n=\bar{x}}(\tilde{\tau}_{\bar{x}}^{(k)})},$$

and

$$y(1,x) = \lim_{k\to+\infty} \Big[Q_{w_n|D_n=1,X_n=x}(\tau_x^{(k)}) - \sigma(1,x)\, Q_{w_n|D_n=1,X_n=\bar{x}}(\tau_{\bar{x}}^{(k)}) \Big].$$

Hence $y(1,x)$ and $\sigma(1,x)$ are identified (under the location and scale normalizations at $\bar{x}$).

Proof. Fix a realisation x of X_n. For any threshold w, Bayes' rule gives

$$\Pr(w_n > w \mid D_n = 1, X_n = x) = \frac{\Pr(w_n(1) > w \mid X_n = x)\ \Pr(D_n = 1 \mid X_n = x,\ w_n(1) > w)}{\Pr(D_n = 1 \mid X_n = x)}.$$

Letting $w \to +\infty$ and using (iii),

$$\Pr(w_n > w \mid D_n = 1, X_n = x)\ \sim\ c(1,x)\ \Pr\big(w_n(1) > w \mid X_n = x\big), \qquad (w \to +\infty), \tag{44}$$

where $c(1,x) \coloneqq \frac{q_1}{\Pr(D_n=1|X_n=x)} \in (0,\infty)$ and "$\sim$" denotes that the ratio of the two sides converges to 1.

Write $S_{1,x}(w) \coloneqq S_{w_n(1)|X_n=x}(w)$ and $S_x(w) \coloneqq S_{w_n|D_n=1,X_n=x}(w)$. Then, (44) reads as

$$S_x(w)\ \sim\ c(1,x)\, S_{1,x}(w) \qquad (w \to +\infty). \tag{45}$$

By (ii), both right endpoints are $+\infty$; by (iv), the upper-tail CDFs $F_{w_n(1)|X_n=x}$ and $F_{w_n|D_n=1,X_n=x}$ 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 2,

$$Q_{w_n\,|\,D_n=1,X_n=x}(\tau)\ =\ Q_{w_n(1)\,|\,X_n=x}\Big(1 - \frac{1-\tau}{c(1,x)}\ +\ o_x(1-\tau)\Big) \qquad (\tau \to 1), \tag{46}$$

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

From $w_n(1) = y(1,x) + \sigma(1,x)\epsilon_n(1)$ and Assumption (i), for all $u \in (0,1)$,

$$Q_{w_n(1)\,|\,X_n=x}(u)\ =\ y(1,x) + \sigma(1,x) Q_{\epsilon_n(1)}(u).$$

Plugging into (46) gives

$$Q_{w_n\,|\,D_n=1,X_n=x}(\tau)\ =\ y(1,x)\ +\ \sigma(1,x) Q_{\epsilon_n(1)}\Big(1 - \frac{1-\tau}{c(1,x)}\ +\ o_x(1-\tau)\Big) \qquad (\tau \to 1). \tag{47}$$

Scale. Evaluate (47) at $\tau = \tau_x^{(k)}$ and $\tau = \tilde{\tau}_x^{(k)}$:

$$\begin{aligned} Q_{w_n \mid D_n=1, X_n=x}\big(\tau_x^{(k)}\big) &= y(1,x) + \sigma(1,x) Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_x^{(k)}}{c(1,x)} + o_x(1-\tau_x^{(k)})\Big), \\ Q_{w_n \mid D_n=1, X_n=x}\big(\tilde{\tau}_x^{(k)}\big) &= y(1,x) + \sigma(1,x) Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_x^{(k)}}{c(1,x)} + o_x(1-\tilde{\tau}_x^{(k)})\Big), \end{aligned} \qquad (k \to +\infty). \tag{48}$$

Take the difference between the two equations in (48):

$$\begin{aligned} \Delta_x^{(k)} &:= Q_{w_n \mid D_n=1, X_n=x}(\tau_x^{(k)}) - Q_{w_n \mid D_n=1, X_n=x}(\tilde{\tau}_x^{(k)}) \\ &= \sigma(1,x)\Big[Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_x^{(k)}}{c(1,x)} + o_x(1-\tau_x^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_x^{(k)}}{c(1,x)} + o_x(1-\tilde{\tau}_x^{(k)})\Big)\Big] \; (k \to +\infty). \end{aligned} \tag{49}$$

Repeat analogous steps for $\tau = \tau_{\bar{x}}^{(k)}$ and $\tau = \tilde{\tau}_{\bar{x}}^{(k)}$ and use the normalisations in (v):

$$\begin{aligned} \Delta_{\bar{x}}^{(k)} &:= Q_{w_n \mid D_n=1, X_n=\bar{x}}(\tau_{\bar{x}}^{(k)}) - Q_{w_n \mid D_n=1, X_n=\bar{x}}(\tilde{\tau}_{\bar{x}}^{(k)}) \\ &= \Big[Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tilde{\tau}_{\bar{x}}^{(k)})\Big)\Big] \qquad (k \to +\infty). \end{aligned} \tag{50}$$

By the definition of $\tau_x^{(k)}$ and $\tilde{\tau}_x^{(k)}$,

$$1 - \frac{1-\tau_x^{(k)}}{c(1,x)} \;=\; 1 - \frac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})}, \qquad 1 - \frac{1-\tilde{\tau}_x^{(k)}}{c(1,x)} \;=\; 1 - \frac{1-\tilde{\tau}_{\bar{x}}^{(k)}}{c(1,\bar{x})}.$$

Therefore, (49) and (50) can be written as

$$\begin{aligned} \Delta_x^{(k)} &= \sigma(1,x)\Big[Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_x(1-\tau_x^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_x(1-\tilde{\tau}_x^{(k)})\Big)\Big], \\ \Delta_{\bar{x}}^{(k)} &= \Big[Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tilde{\tau}_{\bar{x}}^{(k)})\Big)\Big], \end{aligned} \quad (k \to +\infty). \tag{51}$$

Take the ratio between the two equations in (51). (iv) implies that $Q_{\epsilon_n(1)}$ is continuous and strictly increasing near 1. Since $\tau_{\bar{x}}^{(k)} = 1 - 2^{-k}$ and $\tilde{\tau}_{\bar{x}}^{(k)} = 1 - 3^{-k}$ are distinct for all k, the denominator of the ratio is nonzero for all large k. By continuity and $o_x(1-\tau_x^{(k)}), o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)}) \to 0$,

$$\lim_{k \to +\infty} \frac{Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_x(1-\tau_x^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_x(1-\tilde{\tau}_x^{(k)})\Big)}{Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tilde{\tau}_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tilde{\tau}_{\bar{x}}^{(k)})\Big)}.$$

Therefore,

$$\sigma(1,x) = \lim_{k \to +\infty} \frac{Q_{w_n \mid D_n=1, X_n=x}(\tau_x^{(k)}) - Q_{w_n \mid D_n=1, X_n=x}(\tilde{\tau}_x^{(k)})}{Q_{w_n \mid D_n=1, X_n=\bar{x}}(\tau_{\bar{x}}^{(k)}) - Q_{w_n \mid D_n=1, X_n=\bar{x}}(\tilde{\tau}_{\bar{x}}^{(k)})}.$$

Location. Evaluate (47) at $\tau = \tau_x^{(k)}$ and, with $x = \bar{x}$, at $\tau = \tau_{\bar{x}}^{(k)}$:

$$\begin{aligned} Q_{w_n \mid D_n=1, X_n=x}\big(\tau_x^{(k)}\big) &= y(1,x) + \sigma(1,x) Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_x^{(k)}}{c(1,x)} + o_x(1-\tau_x^{(k)})\Big), \\ Q_{w_n \mid D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) &= Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big), \end{aligned} \qquad (k \to +\infty), \tag{52}$$

where we use the normalisations in (v). By the definition of $\tau_x^{(k)}$,

$$1 - \frac{1-\tau_x^{(k)}}{c(1,x)} \;=\; 1 - \frac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})}.$$

Therefore, (52) can be written as

$$\begin{aligned} Q_{w_n \mid D_n=1, X_n=x}\big(\tau_x^{(k)}\big) &= y(1,x) + \sigma(1,x) Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_x^{(k)}}{c(1,\bar{x})} + o_x(1-\tau_x^{(k)})\Big), \\ Q_{w_n \mid D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) &= Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big), \end{aligned} \qquad (k \to +\infty). \tag{53}$$

Subtracting the two equations in (53):

$$\begin{aligned} &Q_{w_n \mid D_n=1, X_n=x}\big(\tau_x^{(k)}\big) - \sigma(1,x) Q_{w_n \mid D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) \\ &= y(1,x) + \sigma(1,x)\Big[Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_x(1-\tau_x^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(1 - \tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big]\ (k \to +\infty). \end{aligned}$$

Also note that $o_x(1-\tau_x^{(k)}) \to 0$ and $o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)}) \to 0$ as $k \to +\infty$. Therefore, by continuity of $Q_{\epsilon_n(1)}$ near 1 under Assumption (iv),

$$Q_{\epsilon_n(1)}\Big(u + o_x(1-\tau_x^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(u + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big) = o(1), \qquad u \coloneqq 1 - \frac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} \qquad (k \to +\infty).$$

Therefore,

$$\lim_{k \to +\infty} \Big[\, Q_{w_n \mid D_n=1, X_n=x}\big(\tau_x^{(k)}\big) \;-\; \sigma(1,x) Q_{w_n \mid D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) \,\Big] \;=\; y(1,x)$$

as desired. □

C Extension to Search Models

The quantile approach in Proposition 13 can be used to identify key parameters in equilibrium wage equations with inherent conditional heteroskedasticity, as in standard search models. For example, consider a specification in the spirit of Bagger et al. (2014), where the potential wage of worker n at

time t in firm $d \in \mathcal{D}$ is

$$w_{n,t}(d) = \omega\, \gamma_d^{\alpha_d}\, H_{n,t}^{1-\alpha_d}\, \epsilon_{n,t}(d) \;+\; (1-\omega)(1-\delta)\, U_d\big(H_{n,t}\big), \tag{54}$$

with $0 < \omega < 1$ the worker's bargaining weight, $\gamma_d > 0$ firm d's productivity, $\alpha_d \in [0,1)$ the elasticity of wages with respect to γ_d, and $\delta \in (0,1)$ the discount factor. Here $H_{n,t}$ denotes human capital at time t with support $\mathcal{H}_t$, and

$$U_d\big(H_{n,t}\big) \coloneqq z \;+\; \delta\, \mathbb{E}_{\epsilon_{n,t} \sim F_{\epsilon_{n,t}}} \left[\, f\big(S\big(H_{n,t}, \epsilon_{n,t};\, \omega, \alpha_d, \gamma_d, \delta\big)\big)\right],$$

is the flow value of non-market time, where z is the flow value of unemployment; $S(\cdot; \omega, \alpha_d, \gamma_d, \delta)$ is the match–surplus function; $f(\cdot)$ is a functional of the surplus; and the expectation is taken with respect to the shock vector $\epsilon_{n,t}$ 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 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.

Case 1: $H_{n,t}$ is Observed (or Unobserved with Known Distribution and Support). For simplicity, we focus on the case of potential and observed selected wages with unbounded supports. The bounded-support case follows analogously—together with the possibility of identifying the extreme support endpoints—with the technical modifications highlighted in Appendix B.2.

Proposition 14, Corollary 7, and Corollary 8 stated below follow immediately from Proposition 13 and Corollary 1. Specifically, for Proposition 14: replace X_n with $H_{n,t}$ and define

$$y(d, H_{n,t}) \coloneqq (1-\omega)(1-\delta)\, U_d(H_{n,t}), \qquad \sigma(d, H_{n,t}) \coloneqq \omega\, \gamma_d^{\alpha_d}\, H_{n,t}^{1-\alpha_d};$$

then Proposition 13 identifies $y(d, H_{n,t})$ and $\sigma(d, H_{n,t})$. For Corollary 7: 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 by Corollary 1. For Corollary 8: once $y(d, H_{n,t})$, $\sigma(d, H_{n,t})$, and $F_{\epsilon_{n,t}}$ are identified, then the parameters α_d, γ_d, and z follow directly.

Proposition 14 (Identification of $y(d, H_{n,t})$ and $\sigma(d, H_{n,t})$)**.** *For each firm $d \in \mathcal{D}$ and period $t \geq 1$, assume:*

(i) (Exogeneity.) $\epsilon_{n,t}(d)$ is independent of $H_{n,t}$.

(ii) (Supports.) For each $h \in \mathcal{H}_t$,

$$\sup\{u : \Pr(w_{n,t}(d) \le u \mid H_{n,t} = h) < 1\} = +\infty,$$
$$\sup\{u : \Pr(w_{n,t} \le u \mid D_{n,t} = d, H_{n,t} = h) < 1\} = +\infty,$$
$$0 < \Pr(D_{n,t} = d \mid H_{n,t} = h) \le 1.$$

(iii) (Tail Limit.) There exists a constant $q_{t,d} \in (0, 1]$ such that for every $h \in \mathcal{H}_t$,

$$\lim_{w \to +\infty} \Pr\big(D_{n,t} = d \mid H_{n,t} = h,\ w_{n,t}(d) > w\big) = q_{t,d}.$$

(iv) (Tail Regularity.) For each $h \in \mathcal{H}_t$, there exist thresholds $w_{h,t,d} < +\infty$ and $w^{\text{obs}}_{h,t,d} < +\infty$ 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 $(w_{h,t,d}, +\infty)$ and $(w^{\text{obs}}_{h,t,d}, +\infty)$, respectively.

(v) (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 $t \ge 1$.

Corollary 7 (Identification of $F_{\epsilon_{n,t}}$)**.** *Let $F_{\epsilon_{n,t}}$ denote the joint CDF of $\epsilon_{n,t}$ and $F_{\epsilon_{n,t}(d)}$ the marginal CDF of $\epsilon_{n,t}(d)$. Let $S_{\epsilon_{n,t}(d)}$ denote the survival function of $\epsilon_{n,t}(d)$. Maintain Assumptions (i) to (v) of Proposition 14, implying that $y(d, h)$ and $\sigma(d, h)$ are identified for each $d \in \mathcal{D}$, $h \in \mathcal{H}_t$, and $t \ge 1$. For each period $t \ge 1$:*[24]

(a) (Marginal Identification.) For each firm $d \in \mathcal{D}$, assume $\epsilon_{n,t}(d)$ belongs to a known parametric family indexed by the $p_{t,d} \times 1$ vector or parameters $\mu_{t,d} \in M_{t,d} \subseteq \mathbb{R}^{p_{t,d}}$. Fix any $h \in \mathcal{H}_t$ and choose $p_{t,d}+1$ distinct large thresholds $0 < w_0 < w_1 < \cdots < w_{p_{t,d}}$. Define the function

$$\Phi_{t,d,h} : M_{t,d} \to \mathbb{R}^{p_{t,d}},\ \Phi_{t,d,h}(\mu_{t,d}) := \left(\frac{S_{\epsilon_{n,t}(d)}\left(\frac{w_1 - y(d,h)}{\sigma(d,h)}; \mu_{t,d}\right)}{S_{\epsilon_{n,t}(d)}\left(\frac{w_0 - y(d,h)}{\sigma(d,h)}; \mu_{t,d}\right)}, \ \ldots, \ \frac{S_{\epsilon_{n,t}(d)}\left(\frac{w_{p_{t,d}} - y(d,h)}{\sigma(d,h)}; \mu_{t,d}\right)}{S_{\epsilon_{n,t}(d)}\left(\frac{w_0 - y(d,h)}{\sigma(d,h)}; \mu_{t,d}\right)} \right).$$

If $\Phi_{t,d,h}$ is injective, then 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.

[24] Note that our identification result in Corollary 7 is established period by period; accordingly, the shocks $\epsilon_{n,t}$ are allowed to be correlated across periods.

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 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:

$$\max\Big\{\sum_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d;\mu_{t,d}) - (|\mathcal{D}|-1),\;0\Big\} \;\le\; F_{\epsilon_{n,t}}(v_1,\ldots,v_{|\mathcal{D}|}) \;\le\; \min_{d\in\mathcal{D}} F_{\epsilon_{n,t}(d)}(v_d;\mu_{t,d})$$
$$\forall\,(v_1,\ldots,v_{|\mathcal{D}|})\in\mathbb{R}^{|\mathcal{D}|}.$$

Corollary 8 (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 realisation h of $H_{n,t}$, and for some period $t\ge 1$ (see Proposition 14 for sufficient conditions). Assume also that the joint distribution of the shock vector, $F_{\epsilon_{n,t}}$, is identified for the same period $t\ge 1$ (see Corollary 7 for sufficient conditions). Then the parameters α_d, γ_d, and z are identified for each $d\in\mathcal{D}$.*

Proof. *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 (v) of Proposition 14, because the ratio $\sigma(d,h)/\sigma(d,h')$ is identified without any such normalization, as shown in the proof of Proposition 13. Moreover, to identify α_d we do not need to know the distribution of $\epsilon_{n,t}$, $F_{\epsilon_{n,t}}$.

Step 2: Identification of γ_d from $\sigma(d,H_{n,t})$. For any $h\in\mathcal{H}_t$, we have

$$\sigma(d,h) = \omega\,\gamma_d^{\alpha_d}\,h^{\,1-\alpha_d}.$$

Solving for γ_d yields

$$\gamma_d \;=\; \big(\sigma(d,h)\omega^{-1}\,h^{\,\alpha_d-1}\big)^{1/\alpha_d}.$$

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 (v) of Proposition 14. Moreover, as with α_d, to identify γ_d we do not need to know the distribution of $\epsilon_{n,t}$, $F_{\epsilon_{n,t}}$.

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

$$z \;=\; \frac{y(d,h)}{(1-\delta)(1-\omega)} \;-\; \delta\,\mathbb{E}_{\epsilon_{n,t}\sim F_{\epsilon_{n,t}}}\big[\,f\big(S\big(h,\epsilon_{n,t};\,\omega,\alpha_d,\gamma_d,\delta\big)\big)\big]\,.$$

Note that 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 14. Moreover, unlike α_d and γ_d, to identify z we need to know the distribution of $\epsilon_{n,t}$, $F_{\epsilon_{n,t}}$. □

Case 1: Alternative Proof. Rather than relying on Proposition 14, we can use a quantile-based approach that skips the nonparametric identification of $y(d,H_{n,t})$ and $\sigma(d,H_{n,t})$ as intermediate steps and instead leverages directly the parametric structure of the wage equation in (54). We show how this approach works to identify α_d and γ_d in Proposition 15. We provide a more detailed comparison between the two approaches at the end of Proposition 15 and its proof. Define

$$y(d,H_{n,t}) := (1-\omega)(1-\delta)\,U_d(H_{n,t}), \qquad M_{n,t}(d) := \omega\,\gamma_d^{\alpha_d} H_{n,t}^{1-\alpha_d}\,\epsilon_{n,t}(d).$$

Proposition 15 (Identification of α_d and γ_d)**.** *For each firm $d \in \mathcal{D}$ and some period $t \geq 1$, 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) (Quantile Reminder Negligible Relative to Human Capital.) For each $p \in (0,1)$, define the conditional quantile reminder

$$R_{t,d}(p) := Q_{\log w_{n,t}|D_{n,t}=d}(p) - \Big\{\log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\,Q_{\log H_{n,t}|D_{n,t}=d}(p)\Big\}. \quad (55)$$

Then, the contribution of this reminder to the upper observed, selected wages $w_{n,t} \mid D_{n,t} = d$

grows strictly more slowly than the contribution of human capital $H_{n,t}$. That is,

$$\lim_{p\to 1} \frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} = 0. \tag{56}$$

(iii) (Normalisation.) The upper tail of the remainder $R_{t,d}(p)$ has a known finite limit. That is,

$$\lim_{p\to 1} R_{t,d}(p) \;=\; L_{t,d}, \tag{57}$$

and $L_{t,d}$ is known.

Assume in addition 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. *Step 1: Identification of α_d.* Fix a firm $d \in \mathcal{D}$. Using the structure of the wage equation,

$$\log w_{n,t}(d) = \log M_{n,t}(d) + \log\left(1 + \tfrac{y(d,H_{n,t})}{M_{n,t}(d)}\right). \tag{58}$$

Using the definition of $M_{n,t}(d)$,

$$\log M_{n,t}(d) = \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\log H_{n,t} + \log\epsilon_{n,t}(d). \tag{59}$$

Therefore, substituting (58) in (59),

$$\log w_{n,t}(d) = \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\log H_{n,t} + \log\epsilon_{n,t}(d) + \log\left(1 + \tfrac{y(d,H_{n,t})}{M_{n,t}(d)}\right). \tag{60}$$

Now condition on $D_{n,t} = d$ and apply the conditional quantile operator $Q_{\cdot|D_{n,t}=d}(p)$ to both sides of (60). Using only that adding a constant shifts quantiles, we get

$$Q_{\log w_{n,t}|D_{n,t}=d}(p) = \log\omega + \alpha_d \log\gamma_d + Q_{(1-\alpha_d)\log H_{n,t}+\log\epsilon_{n,t}(d)+\log(1+\frac{y(d,H_{n,t})}{M_{n,t}(d)})\,|D_{n,t}=d}(p). \tag{61}$$

Define the conditional quantile remainder: $R_{t,d}(p)$

$$R_{t,d}(p) := Q_{\log w_{n,t}|D_{n,t}=d}(p) - \Big[\log\omega + \alpha_d\log\gamma_d + (1-\alpha_d)\,Q_{\log H_{n,t}|D_{n,t}=d}(p)\Big]. \tag{62}$$

Plug-in (61) into (62):

$$\begin{aligned} R_{t,d}(p) = {} & \log\omega + \alpha_d\log\gamma_d + Q_{(1-\alpha_d)\log H_{n,t}+\log\epsilon_{n,t}(d)+\log(1+\frac{y(d,H_{n,t})}{M_{n,t}(d)})\,|D_{n,t}=d}(p) \\ & - \Big[\log\omega + \alpha_d\log\gamma_d + (1-\alpha_d)\,Q_{\log H_{n,t}|D_{n,t}=d}(p)\Big]. \end{aligned}$$

Thus,

$$R_{t,d}(p) = Q_{(1-\alpha_d)\log H_{n,t}+\log\epsilon_{n,t}(d)+\log(1+\frac{y(d,H_{n,t})}{M_{n,t}(d)})\,|D_{n,t}=d}(p) - (1-\alpha_d)\,Q_{\log H_{n,t}|D_{n,t}=d}(p),$$

and

$$Q_{\log w_{n,t}|D_{n,t}=d}(p) = \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\,Q_{\log H_{n,t}|D_{n,t}=d}(p) + R_{t,d}(p). \tag{63}$$

Fix any $\bar{p} \in (0,1)$ and define

$$\Delta_W(p) := Q_{\log w_{n,t}|D_{n,t}=d}(p) - Q_{\log w_{n,t}|D_{n,t}=d}(\bar{p}),$$

$$\Delta_H(p) := Q_{\log H_{n,t}|D_{n,t}=d}(p) - Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p}),$$

and

$$\Delta_R(p) := R_{t,d}(p) - R_{t,d}(\bar{p}).$$

Subtracting (63) evaluated at p and at $\bar{p}$ yields, for all $p \in (0,1)$,

$$\Delta_W(p) = (1-\alpha_d)\,\Delta_H(p) + \Delta_R(p). \tag{64}$$

By Assumption (i),

$$\lim_{p\to 1} Q_{\log H_{n,t}|D_{n,t}=d}(p) = +\infty,$$

and hence

$$\lim_{p\to 1} \Delta_H(p) = +\infty. \tag{65}$$

Assumption (ii) states that

$$\lim_{p\to 1} \frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} = 0. \tag{66}$$

By combining (65) and (66), we can show that

$$\lim_{p\to 1} \frac{\Delta_R(p)}{\Delta_H(p)} = 0. \tag{67}$$

Indeed, write

$$\frac{\Delta_R(p)}{\Delta_H(p)} = \frac{R_{t,d}(p) - R_{t,d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p) - Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})} = \frac{\frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} - \frac{R_{t,d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)}}{1 - \frac{Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)}}. \tag{68}$$

As $p \to 1$, Assumptions (i) and (ii) imply

$$\frac{R_{t,d}(p)}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} \to 0, \qquad \frac{R_{t,d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} \to 0, \qquad \frac{Q_{\log H_{n,t}|D_{n,t}=d}(\bar{p})}{Q_{\log H_{n,t}|D_{n,t}=d}(p)} \to 0,$$

so the numerator of (68) converges to $0 - 0 = 0$ and the denominator to $1 - 0 = 1$, which yields (67).

Now divide both sides of (64) by $\Delta_H(p)$:

$$\frac{\Delta_W(p)}{\Delta_H(p)} = (1 - \alpha_d) + \frac{\Delta_R(p)}{\Delta_H(p)}. \tag{69}$$

Taking limits as $p \to 1$ on both sides of (69) and using (67), we obtain

$$\lim_{p\to 1} \frac{\Delta_W(p)}{\Delta_H(p)} = \lim_{p\to 1} \left\{ (1 - \alpha_d) + \frac{\Delta_R(p)}{\Delta_H(p)} \right\} = 1 - \alpha_d.$$

This identifies $1 - \alpha_d$ and hence α_d. Note that we do not use the normalisation in Assumption (iv) to identify α_d. Assumption (iv) will be used below to identify γ_d.

Step 2: Identification of γ_d. Rearranging (63) gives

$$R_{t,d}(p) = Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1 - \alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) - \big(\log \omega + \alpha_d \log \gamma_d\big).$$

Taking limits as $p \to 1$ on both sides and using the tail normalisation (57), we obtain

$$L_{t,d} = \lim_{p\to 1} R_{t,d}(p) = \lim_{p\to 1} \Big\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1 - \alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \Big\} - \big(\log \omega + \alpha_d \log \gamma_d\big).$$

Hence,

$$\log \omega + \alpha_d \log \gamma_d = \lim_{p\to 1} \Big\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1 - \alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \Big\} - L_{t,d}.$$

Solving for $\alpha_d \log \gamma_d$ yields

$$\alpha_d \log \gamma_d = \lim_{p\to 1} \Big\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1 - \alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \Big\} - L_{t,d} - \log \omega,$$

so that

$$\gamma_d = \exp\left(\frac{1}{\alpha_d} \left[\lim_{p\to 1} \Big\{ Q_{\log w_{n,t}|D_{n,t}=d}(p) - (1 - \alpha_d)\, Q_{\log H_{n,t}|D_{n,t}=d}(p) \Big\} - L_{t,d} - \log \omega \right] \right).$$

The right-hand side 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 constant $L_{t,d}$, the known bargaining parameter ω, and the already identified α_d. Thus, γ_d is identified. □

Remark. We now compare the identification approach of Proposition 14, Corollary 7, and Corollary 8 (hereafter, the "*first approach*") with the identification approach of Proposition 15 (hereafter, the "*second approach*") for recovering α_d and γ_d. To recap, 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 $\sigma(d, H_{n,t}) \coloneqq \omega\,\gamma_d^{\alpha_d} H_{n,t}^{1-\alpha_d}$, α_d is then identified from *ratios* of $\sigma(d,h)$ at different values of h, which do not depend on any normalisation for $\sigma(d,\cdot)$. The parameter γ_d is identified from the *level* of $\sigma(d,h)$ at some h and therefore requires a normalisation, for example $\sigma(d,\bar{h}) = 1$ for some $\bar{h}$.

The *second approach* does not pass through the intermediate identification of $\sigma(d, H_{n,t})$, but instead works directly with the log wage equation

$$\log w_{n,t}(d) = \log\omega + \alpha_d \log\gamma_d + (1-\alpha_d)\log H_{n,t} + \text{error},$$

and studies conditional upper quantiles of $\log w_{n,t}$ and $\log H_{n,t}$ given $D_{n,t} = d$. Under Assumptions (i)–(ii), we obtain 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 differences in p and ratios of the form $\Delta_W(p)/\Delta_H(p)$ identify the *slope* $1-\alpha_d$ without any normalisation. The parameter γ_d is then recovered from an *intercept*–type tail normalisation on 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 $Q_{\log H_{n,t}|D_{n,t}}(p)$ in the upper tail: the slope $1-\alpha_d$ is identified from the limiting ratio of quantile differences, while the intercept $\log\omega + \alpha_d\log\gamma_d$ is pinned down by a normalisation on the tail behaviour of the composite error.

Thus, both approaches are fundamentally based on *upper–tail identification*. The *first approach* looks at the upper tail of $w_{n,t}$ conditional on $(H_{n,t}, D_{n,t})$, while the *second approach* looks at the upper tail of $\log w_{n,t}$ and $\log H_{n,t}$ conditional on $D_{n,t}$. Moreover, in both approaches, α_d is identified via a slope argument, while γ_d requires a normalisation condition.

In the *first approach*, the key restriction is a *tail limit condition* (Assumption (iii) of Proposition 14), which requires that, conditional on human capital $H_{n,t}$, the probability of working at firm

d given very high *potential* wages $w_{n,t}(d)$ converges to a firm–specific constant $q_{t,d}$ as $w \to +\infty$. This stabilisation of selection in the upper tail is what allows the tail behaviour of the *potential* wage distribution to be recovered from the observed, selected wages. In the *second approach*, the key restrictions is a *dominance condition* (Assumptions (ii) of Proposition 15) on the quantile reminder which produces an asymptotically linear relation between $Q_{\log w_{n,t}|D_{n,t}=d}(p)$ and $Q_{\log H_{n,t}|D_{n,t}=d}(p)$ as $p \to 1$, from which the slope and intercept can be identified.

Case 2: $H_{n,t}$ is Unobserved with Unknown Distribution and Support. We proceed in two steps. First, in Proposition 16, to account for the fact that $\mathcal{H}_t$ is unknown, we work in the human-capital *rank space* by mapping $H_{n,t}$ to its quantile (percentile) index via its CDF:

$$U_{n,t} \coloneqq F_{H_{n,t}}(H_{n,t}).$$

We then identify the rank-mapped primitives

$$y^\circ(d, U_{n,t}) \;\coloneqq\; y\big(d, F^{-1}_{H_{n,t}}(U_{n,t})\big), \quad \sigma^\circ(d, U_{n,t}) \;\coloneqq\; \sigma\big(d, F^{-1}_{H_{n,t}}(U_{n,t})\big),$$

defined on the support of $U_{n,t}$, rather than on the support of $H_{n,t}$. Second, in Proposition 10, assuming 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, we identify α_d, γ_d, and z.

Proposition 16 (Identification of $y^\circ(d, U_{n,t})$ and $\sigma^\circ(d, U_{n,t})$)**.** *Given $d \in \mathcal{D}$ and $t \geq 1$, let $\mathcal{U}_{t,d} \subseteq (0,1)$ be the set of realisations 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 $t \geq 1$, assume:*

(i) (Exogeneity.) $\epsilon_{n,t}(d)$ is independent of $U_{n,t}$.

(ii) (Supports.) For each $u \in \mathcal{U}_{t,d}$,

$$\sup\{w : \Pr(w_{n,t}(d) \leq w \mid U_{n,t} = u) < 1\} = +\infty,$$
$$\sup\{w : \Pr(w_{n,t} \leq w \mid D_{n,t} = d, U_{n,t} = u) < 1\} = +\infty.$$

(iii) (Tail Limit.) There exists an (unknown) constant $q_{t,d} \in (0,1]$ such that for every $u \in \mathcal{U}_{t,d}$,

$$\lim_{w \to +\infty} \Pr\big(D_{n,t} = d \mid U_{n,t} = u,\; w_{n,t}(d) > w\big) = q_{t,d}.$$

(iv) (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.*

(v) (Normalisation.) 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 13 after a change of conditioning variable from the latent value $H_{n,t}$ to its rank $U_{n,t} \coloneqq F_{H_{n,t}}(H_{n,t})$. 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 $y(d, h)$ and $\sigma(d, h)$ at the unknown levels h. □

Corollary 9 (Identification of $F_{\epsilon_{n,t}}$)**.** *Let* $F_{\epsilon_{n,t}}$ *denote the joint CDF of* $\epsilon_{n,t}$ *and* $F_{\epsilon_{n,t}(d)}$ *the marginal CDF of* $\epsilon_{n,t}(d)$*. Let* $S_{\epsilon_{n,t}(d)}$ *denote the survival function of* $\epsilon_{n,t}(d)$*. Maintain Assumptions (i) to (v) of Proposition 16, implying 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* $t \geq 1$*. For each period* $t \geq 1$*:*

(a) (Marginal Identification.) For each firm $d \in \mathcal{D}$*, assume* $\epsilon_{n,t}(d)$ *belongs to a known parametric family indexed by the* $p_{t,d} \times 1$ *vector or parameters* $\mu_{t,d} \in M_{t,d} \subseteq \mathbb{R}^{p_{t,d}}$*. Fix any* $u \in \mathcal{U}_{t,d}$ *and choose* $p_{t,d}+1$ *distinct large thresholds* $0 < w_0 < w_1 < \cdots < w_{p_{t,d}}$*. Define the function*

$$\Phi_{t,d,u} \colon M_{t,d} \to \mathbb{R}^{p_{t,d}},\ \Phi_{t,d,u}(\mu_{t,d}) \coloneqq \left(\frac{S_{\epsilon_{n,t}(d)}\left(\frac{w_1 - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)}{S_{\epsilon_{n,t}(d)}\left(\frac{w_0 - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)}, \ \ldots, \ \frac{S_{\epsilon_{n,t}(d)}\left(\frac{w_{p_{t,d}} - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)}{S_{\epsilon_{n,t}(d)}\left(\frac{w_0 - y^\circ(d,u)}{\sigma^\circ(d,u)}; \mu_{t,d}\right)} \right).$$

If $\Phi_{t,d,u}$ *is injective, then 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 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:

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

Corollary 10 (Identification of α_d, γ_d, and z)**.** *For each firm $d \in \mathcal{D}$ and for some period $t \ge 1$, assume that:*

(i) $y^\circ(d,u)$ and $\sigma^\circ(d,u)$ are identified for each $u \in \mathcal{U}_{t,d}$ (see Proposition 16 for sufficient conditions).

(ii) The distribution $F_{\epsilon_{n,t}}$ of $\epsilon_{n,t}$ is identified (see Corollary 9 for sufficient conditions).

(iii) There exist two distinct ranks $u_a \neq u_b$ in $\mathcal{U}_{t,d}$ such that the corresponding human-capital levels $h_a \coloneqq F^{-1}_{H_{n,t}}(u_a)$ and $h_b \coloneqq F^{-1}_{H_{n,t}}(u_b)$ are known to the researcher.

Then α_d, γ_d, and z are identified for each $d \in \mathcal{D}$.

Proof. *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 the two ranks $u_a \neq u_b$ and their corresponding levels $h_a \coloneqq F^{-1}_{H_{n,t}}(u_a)$ and $h_b \coloneqq F^{-1}_{H_{n,t}}(u_b)$. 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 yields

$$\alpha_d \ =\ 1 \ -\ \frac{\log\big(\sigma^\circ(d,u_a)/\sigma^\circ(d,u_b)\big)}{\log(h_a/h_b)}.$$

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: Identify γ_d from $\sigma^\circ(d,U_{n,t})$. Using any anchored pair (u_*,h_*) with $* \in \{a,b\}$,

$$\sigma^\circ(d,u_*) = \omega\,\gamma_d^{\alpha_d} h_*^{1-\alpha_d} \quad\Longrightarrow\quad \gamma_d = \left(\frac{\sigma^\circ(d,u_*)}{\omega\,h_*^{1-\alpha_d}}\right)^{1/\alpha_d},$$

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 and (α_d,γ_d) are now known, while $F_{\epsilon_{n,t}}$ is known by assumption,

z is identified as

$$z = \frac{y^{\circ}(d, u_*)}{(1-\delta)(1-\omega)} - \delta\, \mathbb{E}_{\epsilon_{n,t} \sim F_{\epsilon_{n,t}}}\Big[f\big(S(h_*, \epsilon_{n,t}; \omega, \alpha_d, \gamma_d, \delta)\big)\Big].$$

□

D Omitted Proofs

Proof of Proposition 3. Fix a realisation x of X_n. For any threshold w, Bayes' rule gives

$$\Pr(w_n > w \mid D_n = 1, X_n = x) = \frac{\Pr(w_n(1) > w \mid X_n = x)\ \Pr(D_n = 1 \mid X_n = x,\ w_n(1) > w)}{\Pr(D_n = 1 \mid X_n = x)}.$$

Letting $w \to +\infty$ and using (iii),

$$\Pr(w_n > w \mid D_n = 1, X_n = x)\ \sim\ c(1,x)\ \Pr\big(w_n(1) > w \mid X_n = x\big), \qquad (w \to +\infty), \tag{70}$$

where $c(1,x) \coloneqq \frac{q_1}{\Pr(D_n=1 \mid X_n = x)} \in (0, \infty)$ and "$\sim$" denotes that the ratio of the two sides converges to 1.

Write $S_{1,x}(w) \coloneqq S_{w_n(1) \mid X_n = x}(w)$ and $S_x(w) \coloneqq S_{w_n \mid D_n = 1, X_n = x}(w)$. Then, (70) reads as

$$S_x(w)\ \sim\ c(1,x)\, S_{1,x}(w) \qquad (w \to +\infty). \tag{71}$$

By (ii), both right endpoints are $+\infty$; by (iv), the upper-tail CDFs $F_{w_n(1) \mid X_n = x}$ and $F_{w_n \mid D_n = 1, X_n = x}$ 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 2,

$$Q_{w_n \mid D_n = 1, X_n = x}(\tau)\ =\ Q_{w_n(1) \mid X_n = x}\Big(1 - \frac{1-\tau}{c(1,x)}\ +\ o_x(1-\tau)\Big) \qquad (\tau \to 1), \tag{72}$$

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

From $w_n(1) = y(1,x) + \epsilon_n(1)$ and Assumption (i), for all $u \in (0,1)$,

$$Q_{w_n(1) \mid X_n = x}(u)\ =\ y(1,x) + Q_{\epsilon_n(1)}(u).$$

Plugging into (72) gives

$$Q_{w_n \mid D_n = 1, X_n = x}(\tau)\ =\ y(1,x)\ +\ Q_{\epsilon_n(1)}\Big(1 - \frac{1-\tau}{c(1,x)}\ +\ o_x(1-\tau)\Big) \qquad (\tau \to 1). \tag{73}$$

Let $\{\tau_{\bar{x}}^{(k)}\}_{k\geq 1} \subset (0,1)$ with $\tau_{\bar{x}}^{(k)} \to 1$. Define

$$1-\tau_x^{(k)} \;:=\; \frac{c(1,x)}{c(1,\bar{x})}\,\big(1-\tau_{\bar{x}}^{(k)}\big). \tag{74}$$

Since $c(1,x), c(1,\bar{x}) \in (0,\infty)$, we have $\tau_x^{(k)} \in (0,1)$ for all large k and $\tau_x^{(k)} \to 1$ as $k \to +\infty$. Note also that by (74), $1-\tau_x^{(k)} = (c(1,x)/c(1,\bar{x}))(1-\tau_{\bar{x}}^{(k)})$, so $1-\tau_x^{(k)}$ and $1-\tau_{\bar{x}}^{(k)}$ are of the same order. Evaluate (73) at $\tau = \tau_x^{(k)}$ and, with $x = \bar{x}$, at $\tau = \tau_{\bar{x}}^{(k)}$:

$$\begin{aligned} Q_{w_n \,|\, D_n=1, X_n=x}\big(\tau_x^{(k)}\big) &= y(1,x) + Q_{\epsilon_n(1)}\Big(1-\tfrac{1-\tau_x^{(k)}}{c(1,x)} + o_x(1-\tau_x^{(k)})\Big), \\ Q_{w_n \,|\, D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) &= y(1,\bar{x}) + Q_{\epsilon_n(1)}\Big(1-\tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big), \end{aligned} \qquad (k \to +\infty). \tag{75}$$

By construction (74),

$$1-\frac{1-\tau_x^{(k)}}{c(1,x)} \;=\; 1-\frac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})}.$$

Therefore, (75) can be written as

$$\begin{aligned} Q_{w_n \,|\, D_n=1, X_n=x}\big(\tau_x^{(k)}\big) &= y(1,x) + Q_{\epsilon_n(1)}\Big(1-\tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_x(1-\tau_x^{(k)})\Big), \\ Q_{w_n \,|\, D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big) &= y(1,\bar{x}) + Q_{\epsilon_n(1)}\Big(1-\tfrac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} + o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big), \end{aligned} \qquad (k \to +\infty). \tag{76}$$

Also note that $o_x(1-\tau_x^{(k)}) \to 0$ and $o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)}) \to 0$ as $k \to +\infty$. Therefore, by continuity of $Q_{\epsilon_n(1)}$ near 1 under Assumption (iv),

$$Q_{\epsilon_n(1)}\Big(u+o_x(1-\tau_x^{(k)})\Big) - Q_{\epsilon_n(1)}\Big(u+o_{\bar{x}}(1-\tau_{\bar{x}}^{(k)})\Big) = o(1), \qquad u := 1-\frac{1-\tau_{\bar{x}}^{(k)}}{c(1,\bar{x})} \qquad (k \to +\infty).$$

Subtracting the two equations in (76) and using the normalisation $y(1,\bar{x}) = 0$ from Assumption (v):

$$\lim_{k\to+\infty}\Big[\,Q_{w_n \,|\, D_n=1, X_n=x}\big(\tau_x^{(k)}\big) \;-\; Q_{w_n \,|\, D_n=1, X_n=\bar{x}}\big(\tau_{\bar{x}}^{(k)}\big)\,\Big] \;=\; y(1,x).$$

This proves the claim. □

*

Lemma 2 (Survival-to-quantile index inversion)**.** *Under (ii) and (iv), for fixed realisation x of X_n and some $c(1,x) \in (0,\infty)$,*

$$S_{w_n|D_n=1,X_n=x}(w) \;\sim\; c(1,x)\, S_{w_n(1)|X_n=x}(w) \qquad (w \to +\infty), \tag{77}$$

if and only if

$$Q_{w_n|D_n=1,X_n=x}(\tau) \;=\; Q_{w_n(1)|X_n=x}\Big(1-\frac{1-\tau}{c(1,x)}+o_x(1-\tau)\Big) \qquad (\tau\to 1), \tag{78}$$

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

Proof. (⇒) Assume

$$S_x(w)\sim c(1,x)\,S_{1,x}(w) \qquad (w\to+\infty). \tag{79}$$

Fix $\tau\to 1$ and define

$$w_\tau \coloneqq Q_{w_n|D_n=1,X_n=x}(\tau). \tag{80}$$

By (ii), $\omega_{\text{obs}}(x)=+\infty$, so $w_\tau\to+\infty$ as $\tau\to 1$. For τ close enough to 1, w_τ lies in the tail region where (iv) applies; thus, by continuity on the tail and (80),

$$F_{w_n|D_n=1,X_n=x}(w_\tau)=\tau,$$

which is equivalent to

$$S_x(w_\tau)=1-\tau. \tag{81}$$

From (79) evaluated at $w=w_\tau$ and (81), we get

$$S_{1,x}(w_\tau)=\frac{1-\tau}{c(1,x)}+o_x(1-\tau) \qquad (\tau\to 1), \tag{82}$$

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

Define

$$u_\tau \coloneqq F_{w_n(1)|X_n=x}(w_\tau)=1-S_{1,x}(w_\tau). \tag{83}$$

By (82)–(83),

$$u_\tau=1-\frac{1-\tau}{c(1,x)}+o_x(1-\tau) \qquad (\tau\to 1). \tag{84}$$

Since $u_\tau\to 1$, for τ close enough to 1 we have u_τ in the tail index range where $F_{w_n(1)|X_n=x}$ is invertible; combining this with (83),

$$w_\tau=\big(F_{w_n(1)|X_n=x}\big)^{-1}(u_\tau)=Q_{w_n(1)|X_n=x}(u_\tau) \qquad (\tau\to 1). \tag{85}$$

Substituting (84) into (85) yields

$$w_\tau = Q_{w_n(1)|X_n=x}\Big(1 - \frac{1-\tau}{c(1,x)} + o_x(1-\tau)\Big) \qquad (\tau \to 1). \tag{86}$$

Combining (86) with (80) gives (78).

(⇐) Conversely, assume

$$Q_{w_n|D_n=1,X_n=x}(\tau) = Q_{w_n(1)|X_n=x}\Big(1 - \frac{1-\tau}{c(1,x)} + o_x(1-\tau)\Big) \qquad (\tau \to 1). \tag{87}$$

Fix $\tau \to 1$ and set

$$u_\tau \coloneqq 1 - \frac{1-\tau}{c(1,x)} + o_x(1-\tau). \tag{88}$$

Then, (87) becomes

$$Q_{w_n|D_n=1,X_n=x}(\tau) = Q_{w_n(1)|X_n=x}(u_\tau) \qquad (\tau \to 1). \tag{89}$$

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

$$F_{w_n(1)|X_n=x}\big(Q_{w_n(1)|X_n=x}(u_\tau)\big) \;=\; u_\tau. \tag{90}$$

Applying $F_{w_n(1)|X_n=x}$ to both sides of (89) and using (88)-(90) gives

$$F_{w_n(1)|X_n=x}\big(Q_{w_n|D_n=1,X_n=x}(\tau)\big) = 1 - \frac{1-\tau}{c(1,x)} + o_x(1-\tau) \qquad (\tau \to 1).$$

Equivalently, in survival notation,

$$S_{1,x}\big(Q_{w_n|D_n=1,X_n=x}(\tau)\big) = \frac{1-\tau}{c(1,x)} + o_x(1-\tau) \qquad (\tau \to 1). \tag{91}$$

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

$$S_x\big(Q_{w_n|D_n=1,X_n=x}(\tau)\big) \;=\; 1-\tau. \tag{92}$$

Define

$$w_\tau \;\coloneqq\; Q_{w_n|D_n=1,X_n=x}(\tau). \tag{93}$$

Then, by tail continuity under (iv),

$$S_x(w_\tau) = 1-\tau. \tag{94}$$

From (92), (93), and (94), with $r_x(\tau) \coloneqq o_x(1-\tau)/(1-\tau) \to 0$,

$$\frac{S_x(w_\tau)}{S_{1,x}(w_\tau)} = \frac{1-\tau}{\frac{1-\tau}{c(1,x)}\big(1 + c(1,x) r_x(\tau)\big)} = c(1,x)\,\frac{1}{1 + c(1,x) r_x(\tau)} = c(1,x)\{1 + o(1)\},$$

whence

$$S_x(w_\tau) \;\sim\; c(1,x)\, S_{1,x}(w_\tau) \qquad (\tau \to 1).$$

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

$$S_x(w) \;\sim\; c(1,x)\, S_{1,x}(w) \qquad (w \to +\infty),$$

which is (77). □

Proof of Corollary 1, Part (a). Fix a realisation x of X_n. By Bayes' rule, for any $w \in \mathbb{R}$,

$$S_{w_n | D_n=1, X_n=x}(w) = S_{w_n(1) | X_n=x}(w)\, \frac{\Pr\big(D_n = 1 \mid X_n = x,\; w_n(1) > w\big)}{\Pr(D_n = 1 \mid X_n = x)}.$$

Using Assumption (iii) of Proposition 3,

$$S_{w_n | D_n=1, X_n=x}(w) \;\sim\; c(1,x)\, S_{w_n(1) | X_n=x}(w) \qquad (w \to +\infty), \tag{95}$$

where $c(1,x) \coloneqq q_1/\Pr(D_n = 1 \mid X_n = x) \in (0, \infty)$. Take two thresholds $w_1, w_2 > 0$ and let $\min\{w_1, w_2\} \to +\infty$. Dividing (95) at $w = w_1$ and $w = w_2$ gives

$$\lim_{\min\{w_1,w_2\} \to +\infty} \frac{S_{w_n | D_n=1, X_n=x}(w_1)}{S_{w_n | D_n=1, X_n=x}(w_2)} \;= \frac{S_{w_n(1) | X_n=x}(w_1)}{S_{w_n(1) | X_n=x}(w_2)}. \tag{96}$$

By Assumption (i) of Proposition 3, $w_n(1) = y(1,x) + \epsilon_n(1)$. Hence,

$$S_{w_n(1) | X_n=x}(w) = \Pr\big(\epsilon_n(1) > w - y(1,x)\big) = S_{\epsilon_n(1)}\big(w - y(1,x)\big). \tag{97}$$

Substituting (97) into (96) yields

$$\lim_{\min\{w_1,w_2\} \to +\infty} \frac{S_{w_n | D_n=1, X_n=x}(w_1)}{S_{w_n | D_n=1, X_n=x}(w_2)} \;=\; \frac{S_{\epsilon_n(1)}\big(w_1 - y(1,x)\big)}{S_{\epsilon_n(1)}\big(w_2 - y(1,x)\big)}. \tag{98}$$

Now choose $p_1 + 1$ distinct large thresholds $0 < w_0 < w_1 < \cdots < w_{p_1}$ and form the p_1 ratios

$$R_j(x) \;\coloneqq\; \frac{S_{w_n | D_n=1, X_n=x}(w_j)}{S_{w_n | D_n=1, X_n=x}(w_0)}, \qquad j = 1, \ldots, p_1.$$

Applying (98) with $(w_1, w_2) = (w_j, w_0)$ for each j and letting all thresholds be large gives

$$R_j(x) \;\longrightarrow\; \frac{S_{\epsilon_n(1)}\big(w_j - y(1,x);\mu_1\big)}{S_{\epsilon_n(1)}\big(w_0 - y(1,x);\mu_1\big)}, \qquad j = 1, \ldots, p_1,$$

so the observed vector $\big(R_1(x), \ldots, R_p(x)\big)$ converges to $\Phi_{1,x}(\mu_1)$ as defined in the statement. If $\Phi_{1,x}$ is injective, this limit uniquely determines μ_1, establishing identification.

□

Examples of Corollary 1. Fix a realisation x of X_n. Let $0 < w_0 < w_1 < w_2$ be large thresholds and define

$$R_j(x) \;\coloneqq\; \frac{S_{w_n|D_n=1,X_n=x}(w_j)}{S_{w_n|D_n=1,X_n=x}(w_0)}, \qquad j = 1, 2.$$

We now investigate whether the map from the shock parameters to the 2-vector $\big(R_1(x), R_2(x)\big)$ is injective for three common parametric families.

Normal. $\epsilon_n(1) \sim \mathcal{N}(\mu, \sigma^2)$, parameters (μ, σ). Let

$$z_j = \frac{w_j - y(1,x) - \mu}{\sigma}, \qquad z_0 = \frac{w_0 - y(1,x) - \mu}{\sigma}.$$

Then,

$$R_j(x) = \frac{1 - \Phi(z_j)}{1 - \Phi(z_0)} = \frac{1 - \Phi\big(z_0 + \Delta_j/\sigma\big)}{1 - \Phi(z_0)}, \qquad \Delta_j \coloneqq w_j - w_0 > 0.$$

For fixed σ, $R_j(x)$ is strictly decreasing in z_0; for fixed z_0, $R_j(x)$ is strictly decreasing in $1/\sigma$ because $\Delta_j > 0$ and $1 - \Phi$ is strictly decreasing. Hence,

$$(\mu, \sigma) \;\longmapsto\; \big(R_1(x), R_2(x)\big)$$

is injective.

Lognormal. $\epsilon_n(1) \sim \mathrm{LN}(m, \sigma^2)$, parameters (m, σ). We have

$$R_j(x) = \frac{1 - \Phi\Big(\frac{\log(w_j - y(1,x)) - m}{\sigma}\Big)}{1 - \Phi\Big(\frac{\log(w_0 - y(1,x)) - m}{\sigma}\Big)} = \frac{1 - \Phi\big(z_0 + \Delta_j^{\log}/\sigma\big)}{1 - \Phi(z_0)},$$

with

$$z_0 = \frac{\log(w_0 - y(1,x)) - m}{\sigma}, \qquad \Delta_j^{\log} = \log\Big(\frac{w_j - y(1,x)}{w_0 - y(1,x)}\Big) > 0.$$

The same monotonicity logic as in the Normal case (now in $\Delta_j^{\log}$) implies injectivity of

$$(m, \sigma) \;\longmapsto\; \big(R_1(x), R_2(x)\big).$$

Shifted Pareto. $\epsilon_n(1) \sim \mu + \mathrm{Par}(\alpha, t_{\min})$, parameters (μ, α) (the scale $t_{\min} > 0$ cancels). For $t > \mu + t_{\min}$,

$$S_{\epsilon_n(1)}(t; \mu, \alpha, t_{\min}) = \Big(\frac{t-\mu}{t_{\min}}\Big)^{-\alpha} \quad \Rightarrow \quad R_j(x) = \left(\frac{w_j - y(1,x) - \mu}{w_0 - y(1,x) - \mu}\right)^{-\alpha}.$$

Taking logs,

$$\log R_j(x) \;=\; -\alpha \,\log\Big(\frac{w_j - y(1,x) - \mu}{w_0 - y(1,x) - \mu}\Big), \qquad j = 1, 2.$$

With two distinct j's these give two equations in (μ, α), each strictly monotone in μ on the admissible region $w_\ell - y(1,x) - \mu > 0$, with common slope $-\alpha$. Hence

$$(\mu, \alpha) \;\longmapsto\; \big(R_1(x), R_2(x)\big)$$

is injective; $t_{\min}$ drops out of the ratios and is not identified.

Proof of Corollary 2. We show how to identify the conditional signal distribution

$$\Pr\big(a_n^t = a^t \mid H_{n,1} = h,\; D_n^t = d^t, e_n = e\big), \tag{99}$$

for each $1 \le t \le T-1$ and $(a^t, h, d^t, e) \in \mathcal{A}^t \times \mathcal{H} \times \mathcal{D}^t \times \mathcal{E}$, where $a^t := (a_1, \ldots, a_t)$ and $d^t := (d_1, \ldots, d_t)$, such that $\Pr\big(H_{n,1} = h,\; D_n^t = d^t\big) > 0$ and $\Pr\big(e_n = e \mid H_{n,1} = h,\; D_n^t = d^t\big) > 0$.

By Proposition 4(i) at time $t+1$,

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

is identified for each $(e, a^t) \in \mathcal{E} \times \mathcal{A}^t$ and $(h, d^{t+1}) \in \mathcal{H} \times \mathcal{D}^{t+1}$ such that $\Pr(H_{n,1} = h,\; D_n^{t+1} = d^{t+1}) > 0$, where $d^{t+1} := (d_1, d_2, \ldots, d_t, d_{t+1}) = (d^t, d_{t+1})$. Using the law of total probability,

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

Therefore

$$\Pr\big(e_n = e,\; a_n^t = a^t \mid H_{n,1} = h,\; D_n^t = d^t\big), \tag{100}$$

is identified.

By Proposition 4(i) at time t,

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

is identified for each $(e, a^{t-1}) \in \mathcal{E} \times \mathcal{A}^{t-1}$ and $(h, d^t) \in \mathcal{H} \times \mathcal{D}^t$ such that $\Pr(H_{n,1} = h,\ D_n^t = d^t) > 0$, where $d^t \coloneqq (d_1, \ldots, d_t)$ and $a^{t-1} \coloneqq (a_1, \ldots, a_{t-1})$.

Therefore,

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_n^t = d^t\big), \tag{101}$$

is identified from

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

In turn, by combining (100) and (101), the conditional distribution in (99) is identified via the ratio

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

for any $e \in \mathcal{E}$ such that

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_n^t = d^t\big) > 0.$$

Proof of Corollary 3. We show how to identify the conditional signal distribution

$$\Pr\big(a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}, e_n = e\big), \tag{102}$$

for each $1 \leq t \leq T-3$ and $(a_t, a_{t+1}, a_{t+2}, h, d_t, d_{t+1}, d_{t+2}, e) \in \mathcal{A}^3 \times \mathcal{H} \times \mathcal{D}^3 \times \mathcal{E}$ such that $\Pr\big(H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big) > 0$ and $\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big) > 0$.

By Proposition 4(i) at time $t+3$,

$$\Pr\big(e_n = e,\ a_n^{t+2} = a^{t+2} \mid H_{n,1} = h,\ D_n^{t+3} = d^{t+3}\big),$$

is identified for each $(e, a^{t+2}) \in \mathcal{E} \times \mathcal{A}^{t+2}$ and $(h, d^{t+3}) \in \mathcal{H} \times \mathcal{D}^{t+3}$ such that $\Pr(H_{n,1} = h,\ D_n^{t+3} = d^{t+3}) > 0$, where $d^{t+3} \coloneqq (d_1, d_2, \ldots, d_{t-1}, d_t, d_{t+1}, d_{t+2}, d_{t+3}) = (d^{t-1}, d_t, d_{t+1}, d_{t+2}, d_{t+3})$ and

$a^{t+2} \coloneqq (a_1, a_2, \ldots, a_{t-1}, a_t, a_{t+1}, a_{t+2}) = (a^{t-1}, a_t, a_{t+1}, a_{t+2})$. Using the law of total probability,

$$\begin{aligned}
&\Pr\big(e_n = e,\ a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big) \\
&= \sum_{a^{t-1}} \sum_{d^{t-1},\, d_{t+3}} \Pr\big(e_n = e,\ a_n^{t+2} = (a^{t-1}, a_t, a_{t+1}, a_{t+2}) \mid H_{n,1} = h,\ D_n^{t+3} = (d^{t-1}, d_t, d_{t+1}, d_{t+2}, d_{t+3})\big) \\
&\qquad \times \Pr\big(D_n^{t-1} = d^{t-1},\ D_{n,t+3} = d_{t+3} \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big).
\end{aligned}$$

Therefore,

$$\Pr\big(e_n = e,\ a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big), \tag{103}$$

is identified.

By Proposition 4(i) at time $t+2$,

$$\Pr\big(e_n = e,\ a_n^{t+1} = a^{t+1} \mid H_{n,1} = h,\ D_n^{t+2} = d^{t+2}\big),$$

is identified for each $(e, a^{t+1}) \in \mathcal{E} \times \mathcal{A}^{t+1}$ and $(h, d^{t+2}) \in \mathcal{H} \times \mathcal{D}^{t+2}$ such that $\Pr(H_{n,1} = h,\ D_n^{t+2} = d^{t+2}) > 0$, where $d^{t+2} \coloneqq (d_1, d_2, \ldots, d_t, d_{t+1}, d_{t+2}) = (d^{t-1}, d_t, d_{t+1}, d_{t+2})$ and $a^{t+1} \coloneqq (a_1, a_2, \ldots, a_t, a_{t+1}) = (a^{t-1}, a_t, a_{t+1})$.

Using the law of total probability,

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

Therefore,

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big), \tag{104}$$

is identified.

In turn, by combining (103) and (104), the conditional distribution in (102) is identified via the ratio

$$\frac{\Pr\big(e_n = e,\ a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big)}{\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big)},$$

for any $e \in \mathcal{E}$ such that

$$\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d_t,\ D_{n,t+1} = d_{t+1},\ D_{n,t+2} = d_{t+2}\big) > 0.$$

Proof of Proposition 5. *Step 1: Identification of $\alpha(h,d,e)$ and $\beta(h,d,e)$.* In this step, we identify the conditional probabilities

$$\alpha(h,d,e) \qquad \text{and} \qquad \beta(h,d,e), \tag{105}$$

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

Proof. Fix $(h,d,e) \in \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ and $1 \leq t \leq T-3$ such that, for some $1 \leq t \leq T-3$, $\Pr\big(a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d, e_n = e\big)$ is identified for each $(a_t, a_{t+1}, a_{t+3}) \in \mathcal{A}^3$.

For any $(a_t, a_{t+1}, a_{t+2}) \in \mathcal{A}^3$, using the law of total probability and Assumption 4, we can write

$$\begin{aligned}
&\Pr\big(a_{n,t} = a_t,\ a_{n,t+1} = a_{t+1},\ a_{n,t+2} = a_{t+2} \mid H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d, e_n = e\big) \\
&\quad = \alpha(h,d,e)^{\sum_{\ell=0}^{2} \mathbb{1}\{a_{t+\ell} = \bar{a}\}} \big(1 - \alpha(h,d,e)\big)^{3 - \sum_{\ell=0}^{2} \mathbb{1}\{a_{t+\ell} = \bar{a}\}} q(h,d,e) \\
&\qquad + \beta(h,d,e)^{\sum_{\ell=0}^{2} \mathbb{1}\{a_{t+\ell} = \bar{a}\}} \big(1 - \beta(h,d,e)\big)^{3 - \sum_{\ell=0}^{2} \mathbb{1}\{a_{t+\ell} = \bar{a}\}} \big(1 - q(h,d,e)\big),
\end{aligned} \tag{106}$$

where

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

Equation (106) is a binomial mixture with two components and three trials. The left-hand side of (106) is identified by assumption. Following Blischke (1964, 1978), the weights and components of the binomial mixture in (106), $\{\alpha(h,d,e), \beta(h,d,e), q(h,d,e)\}$, 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 (106). 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 4(iii).

□

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

$$\Pr(\theta_n = \bar{\theta} \mid H_{n,1} = h, e_n = e), \tag{107}$$

for each $(h, e) \in \mathcal{H} \times \mathcal{E}$ such that, for some $d \in \mathcal{D}$, $\Pr\big(a_{n,1} = a \mid H_{n,1} = h,\ D_{n,1} = d,\ e_n = e\big)$ 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 $\Pr\big(a_{n,1} = a \mid H_{n,1} = h,\ D_{n,1} = d,\ e_n = e\big)$ is identified and $\alpha(h, d, e)$ and $\beta(h, d, e)$ are identified.

Using the law of total probability and Assumption 4(i) and (iii), we can write

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

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

$$p_1(h, e) = \begin{cases} \frac{\Pr(a_{n,1}=a \mid H_{n,1}=h,\ D_{n,1}=d,\ e_n=e) - \beta(h,d,e)}{\alpha(h,d,e) - \beta(h,d,e)} & \text{if } a = \bar{a}, \\ \frac{\Pr(a_{n,1}=a \mid H_{n,1}=h,\ D_{n,1}=d,\ e_n=e) - (1-\beta(h,d,e))}{\beta(h,d,e) - \alpha(h,d,e)} & \text{if } a = \underline{a}. \end{cases} \tag{108}$$

Therefore, $p_1(h, e)$ is identified if $\Pr(a_{n,1} = a \mid H_{n,1} = h,\ D_{n,1} = d,\ e_n = e)$ 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 4(iii).

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

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

is identified for each $(a_t, a_{t+1}, a_{t+2}) \in \mathcal{A}^3$ if:

(i) Assumption 1 holds.

(ii) The wage mixture weights in (19) are identified at times $t+2$ and $t+3$. See Proposition 4 for sufficient conditions.

(iii) $\Pr\big(H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d\big) > 0$ and $\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d,\ D_{n,t+1} = d,\ D_{n,t+2} = d\big) > 0$, where the first condition can be verified from the data and the second from the identification of $\mathcal{L}^{\text{eff}}_{h,d^{t+2}}$ under Proposition 4(iii) for each $d^{t+2} \in \mathcal{D}^{t+2}$.

By Corollary 2,

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

is identified for each $a \in \mathcal{A}$ if:

(i) Assumption 1 holds.

(ii) The wage mixture weights in (19) are identified at times 1 and 2. See Proposition 4 for sufficient conditions.

(iii) $\Pr\big(H_{n,1} = h,\ D_{n,1} = d\big) > 0$ and $\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,1} = d\big) > 0$, where the first condition can be verified from the data and the second from the identification of $\mathcal{L}^{\text{eff}}_{h,d}$ under Proposition 4(iii).

□

Proof of Proposition 7. Let $2 \leq t \leq T$, $s \coloneqq (h, \kappa, p, e) \in \mathcal{S}_t$, $d \in \mathcal{D}$, and $\tilde{s} \coloneqq (\tilde{h}, \tilde{\kappa}, \tilde{p}, \tilde{e}) \in \mathcal{S}_{t-1}$ such that $\Pr(D_{n,t} = d, s_{n,t-1} = \tilde{s}) > 0$. We have

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

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 $(h, e) = (\tilde{h}, \tilde{e})$,

$$\begin{aligned}
&\Pr(s_{n,t} = s \mid D_{n,t-1} = d, s_{n,t-1} = \tilde{s}) \\
&\quad = \Pr(\kappa_{n,t} = \kappa \mid D_{n,t-1} = d, \kappa_{n,t-1} = \tilde{\kappa}) \\
&\qquad \times \Pr(P_{n,t} = p \mid D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e).
\end{aligned} \tag{109}$$

In (109), $\Pr(\kappa_{n,t} = \kappa \mid D_{n,t-1} = d, \kappa_{n,t-1} = \tilde{\kappa})$ is known because $\kappa_{n,t}$ is a known function of D_n^{t-1} and $\kappa_{n,t-1}$ is a known function of D_n^{t-2}. From equation (3), p can take two values, $\{\bar{p}, \underline{p}\}$, depending on whether $a_{n,t-1} = \bar{a}$ or $a_{n,t-1} = \underline{a}$. Therefore, $\Pr(P_{n,t} = p \mid D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e)$ in (109) can be

$$\Pr(P_{n,t} = \bar{p} \mid D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) = \Pr(a_{n,t-1} = \bar{a} \mid D_{n,t-1} = d, H_{n,1} = h, e_n = e),$$

or

$$\Pr(P_{n,t} = \underline{p} \mid D_{n,t-1} = d, H_{n,1} = h, P_{n,t-1} = \tilde{p}, e_n = e) = \Pr(a_{n,t-1} = \underline{a} \mid D_{n,t-1} = d, H_{n,1} = h, e_n = e).$$

Moreover, $\Pr(a_{n,t-1} = a \mid D_{n,t-1} = d, H_{n,1} = h, e_n = e)$ is identified by Proposition 4(i) at times t and $t-1$; see Step (a) below. Therefore, $\Pr(s_{n,t} = s \mid D_{n,t-1} = d, s_{n,t-1} = \tilde{s})$ is identified.

Step (a): Useful to Identify the Law of Motion of the State. In this step, we identify the conditional signal distribution

$$\Pr\big(a_{n,t} = a \mid H_{n,1} = h,\ D_{n,t} = d, e_n = e\big), \tag{110}$$

for each $1 \leq t \leq T-1$ and $(a, h, d, e) \in \mathcal{A} \times \mathcal{H} \times \mathcal{D} \times \mathcal{E}$ such that $\Pr\big(H_{n,1} = h,\ D_{n,t} = d\big) > 0$ and $\Pr\big(e_n = e \mid H_{n,1} = h,\ D_{n,t} = d\big) > 0$.

Proof. By Proposition 4(i) at time $t+1$ and t,

$$\Pr(a_n^t = (a^{t-1}, a) \mid H_{n,1} = h, D_n^t = (d^{t-1}, d), e_n = e), \tag{111}$$

is identified for each $(e, a^{t-1}, a, h, d^{t-1}, d) \in \mathcal{E} \times \mathcal{A}^t \times \mathcal{H} \times \mathcal{D}^t$ such that $\Pr(H_{n,1} = h, D_n^t = (d^{t-1}, d)) > 0$ and $\Pr(e_n = e \mid H_{n,1} = h, D_n^t = (d^{t-1}, d)) > 0$, where $d^{t-1} \coloneqq (d_1, \ldots, d_{t-1})$ and $a^{t-1} \coloneqq (a_1, \ldots, a_{t-1})$. See Corollary 2.

Moreover,

$$\Pr(D_n^t = (d^{t-1}, d) \mid H_{n,1} = h, e_n = e) \tag{112}$$

is identified for each $(d^{t-1}, d, h, e) \in \mathcal{D}^t \times \mathcal{H} \times \mathcal{E}$ such that $\Pr(H_{n,1} = h \mid D_n^t = (d^{t-1}, d)) > 0$ and $\Pr(e_n = e \mid H_{n,1} = h, D_n^t = (d^{t-1}, d)) > 0$. This is because

$$\begin{aligned}
&\Pr(D_n^t = (d^{t-1}, d) \mid H_{n,1} = h, e_n = e) \\
&= \frac{\sum_{a^{t-1}} \Pr(e_n = e, a_n^{t-1} = a^{t-1} \mid H_{n,1} = h, D_n^t = (d^{t-1}, d)) \times \Pr(D_n^t = (d^{t-1}, d) \mid H_{n,1} = h)}{\sum_{d_1} \Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d_1) \times \Pr(D_{n,1} = d_1 \mid H_{n,1} = h)},
\end{aligned}$$

where:

- $\Pr(e_n = e, a_n^{t-1} = a^{t-1} \mid H_{n,1} = h, D_n^t = (d^{t-1}, d))$ is identified from Proposition 4(i) at time t for each $(e, a^{t-1}, h, d^{t-1}, d) \in \mathcal{E} \times \mathcal{A}^{t-1} \times \mathcal{H} \times \mathcal{D}^t$ such that $\Pr(H_{n,1} = h, D_n^t = (d^{t-1}, d)) > 0$.
- $\Pr(D_n^t = (d^{t-1}, d) \mid H_{n,1} = h)$ is known from the data for each $(d^{t-1}, d, h) \in \mathcal{D}^t \times \mathcal{H}$.
- $\Pr(D_{n,1} = d_1 \mid H_{n,1} = h)$ is known from the data for each $(d_1, h) \in \mathcal{D} \times \mathcal{H}$.
- $\Pr(e_n = e \mid H_{n,1} = h, D_{n,1} = d_1)$ is identified from Proposition 4(i) applied to the first period for each $(e, h, d_1) \in \mathcal{E} \times \mathcal{H} \times \mathcal{D}$ such that $\Pr(H_{n,1} = h, D_{n,1} = d_1) > 0$.

From (112),

$$\Pr(D_{n,t} = d \mid H_{n,1} = h, e_n = e) \tag{113}$$

is identified for each $(d, h, e) \in \mathcal{D} \times \mathcal{H} \times \mathcal{E}$ such that $\Pr(H_{n,1} = h \mid D_{n,t} = d) > 0$ and $\Pr(e_n = e \mid H_{n,1} = h, D_{n,t} = d) > 0$.

Using the law of total probability, for each $a \in \mathcal{A}$,

$$\begin{aligned}
&\Pr(a_{n,t} = a \mid D_{n,t} = d, H_{n,1} = h, e_n = e) \\
&= \sum_{d^{t-1}} \sum_{a^{t-1}} \Pr(a_n^t = (a^{t-1}, a) \mid H_{n,1} = h, D_n^t = (d^{t-1}, d), e_n = e) \times \Pr(D_n^t = (d^{t-1}, d) \mid H_{n,1} = h, e_n = e) \\
&\qquad \Big/ \Pr(D_{n,t} = d \mid H_{n,1} = h, e_n = e).
\end{aligned} \tag{114}$$

Therefore, the conditional probability in (110) is identified by (111) to (114). □

Proof of Proposition 8. Let $2 \leq t \leq T$. The conditional probability

$$\Pr(D_{n,t} = d \mid H_{n,1} = h, D_n^{t-1} = d^{t-1}, e_n = e, a_n^{t-1} = a^{t-1}), \tag{115}$$

is identified for each $(d, h, d^{t-1}, e, a^{t-1}) \in \mathcal{D} \times \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}) > 0$ and $\Pr(e_n = e, a_n^{t-1} = a^{t-1} \mid H_{n,1} = h, D_n^{t-1} = d^{t-1}) > 0$, where $d^{t-1} \coloneqq (d_1, \ldots, d_{t-1})$ and $a^{t-1} \coloneqq (a_1, \ldots, a_{t-1})$. This is because, by Bayes' rule,

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

where:

- $\Pr(e_n = e, a_n^{t-1} = a^{t-1} \mid H_{n,1} = h, D_n^t = (d^{t-1}, d))$ is identified by Proposition 4(i) at time t for each $(e, a^{t-1}, h, d^{t-1}, d) \in \mathcal{E} \times \mathcal{A}^{t-1} \times \mathcal{H} \times \mathcal{D}^t$ such that $\Pr(H_{n,1} = h, D_n^t = (d^{t-1}, d)) > 0$.
- $\Pr(D_{n,t} = d \mid H_{n,1} = h, D_n^{t-1} = d^{t-1})$ is known from the data for each $(d, h, d^{t-1}) \in \mathcal{D} \times \mathcal{H} \times \mathcal{D}^{t-1}$ such that $\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}) > 0$.
- $\Pr(e_n = e, a_n^{t-1} = a^{t-1} \mid H_{n,1} = h, D_n^{t-1} = d^{t-1})$ is identified, as shown in (100), from Proposition 4(i) at time t for each $(h, d^{t-1}, e, a^{t-1}) \in \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}) > 0$.

The joint distribution

$$\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}, e_n = e, a_n^{t-1} = a^{t-1}), \tag{116}$$

is identified from Proposition 4(i) at time t for each $(h, d^{t-1}, e, a^{t-1}) \in \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}) > 0$. Given (115), (116), and knowledge of the map g_t from realisations of $(H_{n,1}, D_n^{t-1}, e_n, a_n^{t-1})$ to realisations of $s_{n,t}$, 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 $a_{n,t-1}$ are not present in the derivations.

□

Proof of Proposition 9. Let $2 \leq t \leq T$. From Proposition 4(ii) at time t, we identify

$$\Pr\big(w_{n,t} \leq w \mid H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1}\big), \tag{117}$$

for each $(h, d^t, e, a^{t-1}) \in \mathcal{H} \times \mathcal{D}^t \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h,\ D_n^t = d^t) > 0$ and $\Pr(e_n = e,\ a_n^{t-1} = a^{t-1} \mid H_{n,1} = h,\ D_n^t = d^t) > 0$, where $d^t := (d_1, \ldots, d_t) = (d^{t-1}, d_t)$ and $a^{t-1} := (a_1, \ldots, a_{t-1})$.

From Proposition 4(i) at time t, we identify

$$\Pr\big(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1}\big), \tag{118}$$

for each $(h, d^t, e, a^{t-1}) \in \mathcal{H} \times \mathcal{D}^t \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h,\ D_n^t = d^t) > 0$.

From Proposition 4(i) at time t, we identify

$$\Pr\big(H_{n,1} = h,\ D_n^{t-1} = d^{t-1},\ e_n = e,\ a_n^{t-1} = a^{t-1}\big), \tag{119}$$

for each $(h, d^{t-1}, e, a^{t-1}) \in \mathcal{H} \times \mathcal{D}^{t-1} \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h,\ D_n^{t-1} = d^{t-1}) > 0$.

By taking the ratio between (118) and (119), we identify

$$\begin{aligned} &\Pr\big(D_{n,t} = d_t \mid H_{n,1} = h,\ D_n^{t-1} = d^{t-1},\ e_n = e,\ a_n^{t-1} = a^{t-1}\big) \\ &\qquad = \frac{\Pr(H_{n,1} = h,\ D_n^t = d^t,\ e_n = e,\ a_n^{t-1} = a^{t-1})}{\Pr(H_{n,1} = h,\ D_n^{t-1} = d^{t-1},\ e_n = e,\ a_n^{t-1} = a^{t-1})}, \end{aligned} \tag{120}$$

for each $(h, d^t, e, a^{t-1}) \in \mathcal{H} \times \mathcal{D}^t \times \mathcal{E} \times \mathcal{A}^{t-1}$ such that $\Pr(H_{n,1} = h,\ D_n^t = d^t) > 0$ and $\Pr(e_n = e,\ a_n^{t-1} = a^{t-1} \mid H_{n,1} = h,\ D_n^{t-1} = d^{t-1}) > 0$.

Let $d \in \mathcal{D}$ and $s \in \mathcal{S}_t$ such that $\Pr(D_{n,t} = d \mid s_{n,t} = s) > 0$. Using Bayes' rule, for each $w \in \mathbb{R}$,

we can write

$$
\begin{aligned}
&\Pr(w_{n,t} \leq w \mid D_{n,t} = d, s_{n,t} = s) \\
&= \sum_{\substack{(h,d^{t-1},e,a^{t-1}):\\ g(h,d^{t-1},e,a^{t-1})=s}} \Pr(w_{n,t} \leq w \mid H_{n,1} = h, D_n^{t-1} = d^{t-1}, D_{n,t} = d, e_n = e, a_n^{t-1} = a^{t-1}) \\
&\times \frac{\Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}, D_{n,t} = d, e_n = e, a_n^{t-1} = a^{t-1})}{\sum_{\substack{(h,d^{t-1},e,a^{t-1}):\\ g(h,d^{t-1},e,a^{t-1})=s}} \Pr(H_{n,1} = h, D_n^{t-1} = d^{t-1}, D_{n,t} = d, e_n = e, a_n^{t-1} = a^{t-1})}.
\end{aligned} \tag{121}
$$

All the components on the right-hand side of (121) are identified by (117) to (120). Therefore, $\Pr(w_{n,t} \leq w \mid D_{n,t} = d,\ s_{n,t} = s)$ is identified.

Lastly, we use Assumption 2 to identify $\Pr(w_{n,t} \leq w \mid D_{n,t} = d,\ D'_{n,t} = d',\ s_{n,t} = s)$ from $\Pr(w_{n,t} \leq w \mid D_{n,t} = d,\ s_{n,t} = s)$. Indeed, under Assumption 2(i), conditioning on $D_{n,t} = d$ and $s_{n,t} = s$ also implicitly conditions on the second-best firm $D'_{n,t}$ entering the wage equation (8). Moreover, by Assumption 2(ii), we know which firm is $D'_{n,t}$. Therefore, we identify $\Pr(w_{n,t} \leq w \mid D_{n,t} = d,\ D'_{n,t} = d',\ s_{n,t} = s)$.

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

The identification of $\Pr(D_{n,t} = d \mid D_{n,t} = d', s_{n,t} = s)$ follows directly from Proposition 8 and Assumption 2. □

E Details on Monte Carlo Simulation

E.1 Simulation Exercise

Here we describe the implementation of the exercise in Section 5.1. Data are generated by tracking 1,000,000 workers over 30 periods (from age 25 to 55). Upon entering the labor market, each worker is assigned an efficiency type e from K_1 possible values. To assign these values, the interval $[-2\sigma_1, 2\sigma_1]$ is divided into $K_1 - 1$ equal subintervals, with grid points $\{a_1, \ldots, a_{K_1}\}$ (with $\sigma_1 > 0$). For each worker, a number is drawn from the Normal distribution $\mathcal{N}(0, \sigma_1)$; if the number falls within the interval $[a_{j-1}, a_j]$, the worker is assigned efficiency type a_j; if the number falls below a_1, the worker is assigned efficiency type a_1; and if the number falls above a_{K_1}, the worker is assigned type a_{K_1}. Workers are also assigned gender, education level, and age. In addition, each worker is given an ability type θ_n from two possible levels: high ability $\bar{\theta}$ and low ability $\underline{\theta}$. The

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

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

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

The components of the wage equation (21) are specified as

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

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

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

with $\psi_{i,j}(d) \coloneqq \psi_{i,j}\exp(\beta_0(d))$. The moments from PSID that discipline simulation parameters are:

- the variance, skewness, and kurtosis of log-earnings;
- the variance of log-earnings at three ages:30, 50, and 65 years old;
- the variance of log earning within cells defined by age-gender-education groups. Age groups are 5-year groups for a total of 30 cells;

- the variance of two-years log-earnings growth;
- the growth of average log earnings between 30 and 50 years old;
- the growth of average log earnings between 50 and 65 years old;
- the college premium;
- the gender gap.

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

E.2 Monte Carlo Simulation

In this Section, we describe a few details of the simulation in Section 5.2. We also describe the data we use to calibrate the simulated economy.

The Data. To calibrate the economy to match key moments of the distribution of labor earnings in the US economy, we draw information from the Panel Study of Income Dynamics (PSID). The PSID is a multi-generational, household-level panel dataset that began in 1968 and tracks the same individuals and their lineal descendants across time. The survey provides annual observations between 1968 and 1996, and is biennial thereafter. The original study comprised about 5,000 households (families), and the study currently tracks over 9,000 households. The PSID provides detailed micro-data on various income sources (e.g., labor earnings, self-employment income), net worth/wealth, labor force participation, consumption/expenditures, and other factors. Additionally, the PSID collects information on the age, education level, occupation, and industry of the respondent (typically the head of the household).

From this dataset, we select observations of employed heads of household who are between 22 and 65 years old, have positive labor earnings in a particular period, and belong to the core PSID sample. Using this sample, we calculate the share of labor earnings within brackets of the income distribution, and the age profile of the mean and the standard deviation of labor earnings. To calculate the average labor income profile, we regress log labor earnings on a quartic polynomial in age and include year fixed effects. For the standard deviation, we calculate the cross-sectional standard deviation of log labor earnings.

F Details on Empirical Application

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

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

Third, we define three objects:

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

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