HOOVER INSTITUTION
ECONOMICS WORKING PAPERS

# Careers in Firms: the Role of Learning and Human Capital

Elena Pastorino*

Economics Working Paper 19118

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010

February 14, 2022

Workers experience wage increases not only when they move across firms but also when they progress through a firm's internal hierarchy of jobs.Workers who are promoted to higher-level jobs within a firm, typically after good performance, tend to enjoy further promotions and wage increases. Workers who are not promoted tend to instead experience negative changes in real wages, either because of the eroding effect of inflation on nominal wages or because of outright wage cuts. These features of careers are commonly interpreted as resulting from firms and workers *learning* about workers' ability based on their performance and workers *acquiring human capital* through experience. To date, however, little is known about the relative importance of these two sources of wage growth and dispersion. Using administrative data on one firm first analyzed by Baker, Gibbs, and Holmström (1994a,b), I estimate a structural model in which firms and workers acquire information about workers' ability and workers accumulate human capital when employed. The model fits well the rich patterns of job and wage mobility in the data. I find evidence for a key mechanism through which learning about ability shapes the dynamics of wages. Learning not only directly affects wages through the impact of current beliefs about ability on current wages but also indirectly affects them through promotions, by improving over time the sorting of workers to jobs according to their ability. Through this indirect effect, which is absent from existing empirical models of learning, I estimate that learning about ability is a crucial determinant of the growth and dispersion of individual wages.

Keywords: Careers, Job Assignment, Uncertainty, Learning, Experimentation, Human Capital, Selection, Wage Growth, Wage Dispersion
JEL Classification: D22, D83, J24, J31, J44, J62



* I am grateful to the editor and four anonymous referees for very helpful comments. I am indebted to Petra Todd and Ken Wolpin for their advice, and to Moshe Buchinsky, John Geweke, Mike Keane, and Fabian Lange for their suggestions. Jasper Bagger, John Cochrane, Zvi Eckstein, Eric French, Limor Golan, Cristina Gualdani, Lance Lochner, Bob Miller, Theo Papageorgiou, Fabien Postel-Vinay, Rob Sauer, Isaac Sorkin, Yasutora Watanabe, and especially Pete Klenow, Eddie Lazear, Jeremy Lise, Luigi Pistaferri, and Mike Waldman, have offered many valuable comments. I thank George Baker for kindly sharing the data and Bengt Holmström for his support of the project. Sam Bailey, Ross Batzer, and Zhen Huo provided excellent research assistance and Joan Gieseke invaluable editorial help.

A major question in labor economics is how wages are determined and evolve over time. Wage growth is known to be closely related to job mobility (Topel and Ward (1992) and Keane and Wolpin (1997)) not just *across* firms, as workers pursue better job opportunities through turnover (Buchinsky et al. (2010) and Bagger et al. (2014)), but also *within* firms, as workers advance through a firm's internal hierarchy of jobs (Lazear (1992, 2009), Gibbons and Waldman (1999a,b, 2006), and Lazear and Shaw (2007)). For instance, the wages of promoted workers are 28 percent higher after five years, whereas those of unpromoted workers grow only by 7 percent over five years, owing to wage increases at promotion and higher wages at higher job levels (Baker and Holmström (1995); see also Waldman (2013) and Frederiksen et al. (2017)). Workers who are promoted or experience wage increases once are more likely to receive future promotions and wage increases. Thus, understanding the determinants of job mobility within firms is central to explaining wage growth.

Yet, the dynamics of wages within firms cannot be described solely by the progression of workers through a firm's job hierarchy. For example, a large fraction of the changes in real wages that workers experience yearly in a firm are negative, which leads both to a sizable variability in individual wages over time and to a significant dispersion in wages at any level of the job hierarchy (Waldman (2013)). Uncovering the sources of this variability is key to interpreting the forces that shape wage profiles, the nature of labor income risk, and, ultimately, the origins of inequality.

A common interpretation of these features of careers is that job and wage mobility arise from a gradual process of firms and workers *learning* about workers' ability, which is uncertain when workers enter the labor market, and of workers *acquiring human capital* with experience in the market (Rubinstein and Weiss (2006)). Intuitively, as workers discover their talents and accumulate new skills through their success and failure on the job, they eventually settle in the jobs and firms that best match, and most reward, their productivity. Indeed, learning and human capital acquisition have long been recognized as key sources of the dynamics of jobs and wages. To date, though, their relative importance is debated (Neal and Rosen (2000), Rubinstein and Weiss (2006), and Neal (2017)). One reason is the empirical challenge of the required measurement exercise. When jobs differ in the output they generate and in the opportunities they offer for information and human capital acquisition, job and wage mobility are the result of a complex dynamic selection process, as workers match with the jobs and firms that best fit their accumulating information and skills and will determine their future information and human capital. Here, I show that this process, ignored by the empirical literature so far, is the critical channel through which learning affects wages.

The goal of this paper is to assess the importance for careers of learning about ability and human capital acquisition within a model that accounts for detailed patterns of job and wage mobility in firms. Specifically, I estimate how workers sort into firms and jobs and the extent to which the resulting assignment and wage process is governed by the acquisition of new information about workers' ability and of new skills by workers. I use rich panel data on the wages, jobs, and performance of all managers (supervisory workers) of a U.S. firm first analyzed by Baker, Gibbs, and Holmström (1994a,b)—henceforth, BGH. Based on these data, I infer how the organization of production within a firm and, in turn, the allocation of workers to jobs affect the acquisition of information about workers' ability and of workers' human capital and, conversely, how these two investment processes influence the dynamics of jobs and wages in a firm.[1]

Three main findings emerge. First, in contrast to the literature (Gibbons et al. (2005), Lluis (2005), and Hunnes (2012)), I find the contribution of learning to wages to be sizable: the additional wage growth and dispersion due to it amount to one-quarter of the cumulative wage growth and dispersion over the first seven years at the firm—the rest is due to human capital acquisition. This novel result stems from accounting for both the *direct* and *indirect* effects of learning on wages. I estimate that the direct effect, which captures the impact of current beliefs about ability on current wages in a job, is small as consistent with the literature, which has focused solely on it. But by estimating the process for beliefs, human capital, jobs, and wages, I can also measure the indirect effect of learning on wages resulting from its impact on the dynamics of promotions. I find that this effect is responsible for almost the entire impact of learning on wages.

This indirect effect operates through the endogenous process by which managers are progressively *selected* to higher levels of the job hierarchy whenever higher ability and acquired skills are more valuable at higher-level jobs. Intuitively, as information about managers' ability and their human capital accumulate and are revealed by performance, managers who perform well advance to the jobs that are most suited to their ability and skills. Since wages at higher-level jobs are on average much higher and more dispersed than those at lower-level jobs, this sorting process is key to the growth and dispersion of wages with tenure. The combined patterns of promotions, wages, and job performance in my data suggest that this indirect effect is indeed present: wage growth and dispersion primarily occur as higher performance leads to promotions to jobs at which managers are paid higher and more variable wages. This process is closely related to the sorting mechanisms in Heckman and Singer (1984) and Cameron and Heckman (1998, 2001), who emphasize the role of dynamic selection on observed and unobserved characteristics for careers and educational attainment.[2]

The second main finding is that the differential informativeness of jobs, which is absent from models of careers, is a crucial determinant of wage profiles. To see why, note that according to my estimates, the firm's lowest job level, at which managers are hired, is the most informative about ability, but wages at this level are the lowest. The informational benefit of this job level implies that the firm prefers to employ managers at this level even when beliefs about ability imply they would produce more at higher levels—this force *depresses* wages in *early* tenures relative to the case of equally informative jobs. But learning about ability is faster at this entry level, so when managers are promoted to higher levels, their priors and wages are on average higher than in the case of equally informative jobs—this force *increases* wages in *later* tenures. Hence, the different informativeness of jobs explains why wage returns may be small in early tenures; accounts for the markedly convex relationship between wages and job levels, or tenure, in firms (Waldman (2013)); and rationalizes the wage increases paid even to experienced workers at promotion—a known puzzle for learning models (Gibbons and Waldman (2006))—as compensating differentials for switching from more to less informative jobs.[3]

[1] The BGH data are well known in the literature because they include information on performance and many of their features, which have been replicated by other studies, are now considered stylized facts about careers. See Waldman (2013) and Frederiksen et al. (2017).

[2] The idea that selection is a key force shaping careers builds on the pioneering insight of Heckman (1979) and work by Heckman (1981, 1990), Heckman and Honoré (1990), Heckman and Sedlacek (1985, 1990), and Sauer (1998). Search frictions and performance incentives are also sources of wage growth and dispersion. Like most of the literature on the careers of non-executives, my paper focuses on learning and human capital. See Gayle et al. (2015) for a model of promotions, turnover, and incentives that accounts for the relationship between executive pay and firm size.

[3] In this case, wage premia at promotion compensate workers for the *discrete* loss in information—namely, in the option value of learning about

The third main finding concerns the measurement of learning. As I have argued, the firm acquires information about ability by first assigning managers to jobs that are especially informative, although they contribute little to output. Since these jobs pay lower wages, the benefits of learning delay promotions to higher-paying positions. The speed of a promotion, then, does not necessarily reflect the amount of learning that occurs before it, as often conjectured. In particular, it may take time for learning to have a positive effect on wages through promotions. As a result, to measure the impact of learning on wages, it is critical, as I show, to capture its *cumulative* effect on job assignment.[4] However, popular instrumental-variable methods (Gibbons et al. (2005)) cannot be used for this purpose, as they abstract from the estimation of the dynamics of beliefs and job assignment. When jobs differ in informativeness, I prove that these methods cannot be applied to estimate the direct effect of learning either. The reason is that this approach relies on using variables such as past jobs as instruments for the assigned job in the wage equation. Intuitively, when jobs are equally informative, the additional performance information on which a new assignment is based is random—namely, independent of job assignment and previous information, as beliefs are martingales. But when jobs vary in informativeness, any information acquired is endogenous to the choice of job, so past jobs or wages do not provide suitable instruments for current jobs.

Formally, I consider a labor market in which a worker's ability is symmetrically unobserved to all and correlated across jobs and firms, as in Gibbons and Waldman (1999b, 2006)—hereafter, GW. Production in firms is organized in *jobs* that differ in the output they generate, the information they provide about ability, and the human capital workers acquire when assigned to them. In particular, differently from influential learning models of the labor market (Jovanovic (1979), Miller (1984), and Flinn (1986)), in my model, ability is transferable across jobs and firms, so job transitions occur after good and bad performance. Also, the speed of learning about ability differs across jobs. For instance, performance at entry-level jobs, which usually entail simple tasks of limited value to a firm, may provide a less noisy signal about ability than performance at higher-level jobs, which normally involve more complex tasks of greater value to a firm.[5] Employed workers stochastically accumulate human capital, which depends on their ability and experience at jobs and firms. Firms compete à la Bertrand in segmented, skill-specific submarkets for workers in wages and jobs, possibly facing repeated trade-offs between output today and information and workers' human capital tomorrow. As output technologies can differ across firms, this competition implies that wages in general differ from expected output.

Central to the identification of the model, which I formally establish, is that the BGH data inform not only about managers' jobs and wages but also about the yearly evaluation of each manager's performance, which provides direct evidence on the signals at each job that the firm and its managers use to learn about ability.[6] I estimate the model by maximum likelihood using eight years of observations on managers entering the firm between 1970 and 1979, imposing all the model restrictions.[7] The estimates shed light on several characteristics of the process of information acquisition at

ability—resulting from their new jobs and so can be large, even for very small changes in uncertainty about ability.

[4] As this cumulative effect arises from the sorting of managers to jobs, which is also affected by human capital, the impact of learning could be small even once its indirect effect on promotions is taken into account. I find instead that through this effect, learning has a large impact on wages.

[5] See Prescott and Visscher (1980), Holmström and Tirole (1989), and Jovanovic and Nyarko (1997) on this trade-off, and Harris and Holmström (1982) and MacDonald (1982) for related models of symmetric learning about ability.

[6] Without performance information, identification holds under suitable invertibility conditions. See Section 2.1 in the Supplementary Appendix.

[7] The model captures well the rich patterns of jobs and wages described, as well as finer moments of the data, including the tenure profiles of

the firm. First, uncertainty about ability at entry is substantial, and the job level at which managers are first assigned is the most informative but the least productive. Second, by comparing the estimated wage growth with that in the counterfactual scenario in which learning is absent—beliefs about ability are not updated—I find that learning contributes to more than 25 percent of wage growth and dispersion over the first seven years at the firm. In the absence of learning, wages would increase much more slowly, owing to the lower speed of promotions, and be significantly less dispersed. Third, I estimate learning to be a gradual process: uncertainty about ability declines slowly over time, thereby significantly reducing the pace of transitions to higher levels relative to the counterfactual case in which ability is learned in a single period. Both this persistent uncertainty and the differential informativeness of jobs are responsible for a substantial compression of wage growth in early tenures, and together they account for the curvature of wages in job levels and tenure.[8]

As for human capital, I estimate that the human capital acquired at the firm accounts for most of the growth and dispersion in wages, but like learning, it affects wages primarily through its impact on the dynamics of job assignment. It does so in two ways. First, as managers acquire skills that make them more productive at higher levels, the firm promotes them to jobs to which they are progressively better suited and at which they are correspondingly paid higher wages. Second, although stochastic, human capital acquisition increases productivity on average and so makes demotions less likely for any decrease in beliefs about ability after low performance.

Overall, many features of careers in my data are similar to those documented in the literature, in terms of the patterns of job assignment, the size of wage increases at promotion, the curvature of wages in job levels and tenure, and the magnitude of wage growth on the job (Topel (1991), Belzil and Bognanno (2008), Buchinsky et al. (2010), Waldman (2013), Song et al. (2019), and Frederiksen et al. (2017)). Along these dimensions, the firm I study is comparable to those studied in other work. By focusing on detailed data on one such firm, I can examine aspects of careers that are difficult to measure in more representative worker or matched employer-employee data—for instance, the relationship between jobs, wages, and performance, and the characteristics of a firm's internal organization as reflected in the job hierarchy—but are important to gauging the determinants of job and wage mobility and distinguishing among them. Since my analysis is limited to one firm, however, results must be subject to further verification to establish their generality. The availability of data on more firms would also make it possible to relax the assumptions I have imposed on market structure and the wage-setting mechanism and to dispense with the maintained restrictions of symmetry in the informational and human capital process across firms. The linkage of the firm I study to the rest of the market crucially relies on them. With multi-firm data, additional sources of productivity differences and systematic uncertainty across firms, including about their demand conditions, could be incorporated. My estimates provide just a peek into one firm's wage policy that hopefully offers a first step towards a reassessment of the role of learning for wages.

In the literature, learning about worker productivity and job mobility in the labor market have been investigated

job-to-job transitions at the firm and of separations and the distributions of performance and wages at the main job levels in each tenure.

[8] Relatedly, Lochner et al. (2018) identify the role of changes in the returns to unobserved skills and in the variance of unobserved skills and of transitory non-skill shocks for the increase in U.S. residual wage inequality from the 1980s onward. See also Lochner and Shin (2014) on the importance of unobserved skills for the evolution of log earning residuals. Ghosh (2007) proposes a model of learning and human capital acquisition in which worker turnover across firms arises from disutility shocks to continuing employment with a same firm.

by Berkovec and Stern (1991), Farber and Gibbons (1996), Altonji and Pierret (2001), Antonovics and Golan (2012), and Sanders (2016).[9] Pries and Rogerson (2005) consider a model with learning and matching frictions to account for differences in turnover between the U.S. and Europe. Moscarini (2005) integrates a job-matching model à la Jovanovic (1979) into an equilibrium model with search frictions and finds that idiosyncratic productivity risk dampens learning and worker sorting, thus compressing wage inequality. Nagypál (2007) proposes a model of match-specific informational capital that integrates two learning processes—namely, learning about match productivity and about the best use of a technology in a match—and finds support for a gradual version of the former from the response of worker turnover to first-order Markov price shocks.[10] Papageorgiou (2014) analyzes a search model of occupational mobility with learning in which ability is correlated across occupations, documenting the importance of comparative advantage for occupational sorting. All this work, though, abstracts from within-firm job and wage mobility. Papageorgiou (2018) enriches Papageorgiou (2014) with occupational mobility in firms to account for the relationship between firm size and wages.

As for learning in firms, Garicano and Van Zandt (2013) investigate the role of organizations for acquiring information. Chiappori et al. (1999) provide evidence of learning and downward wage rigidity for executives of a French state-owned firm. Gibbons et al. (2005) study sectoral and inter-industry wage differentials based on the framework of GW. As noted, this strand of work on careers in firms is based on models of job assignment with general rather than firm-, match-, or occupation-specific ability, unlike nearly all of the literature in the previous paragraph. Only when ability is correlated across firms, jobs, or occupations, do workers turn over across them not only after bad performance, as in Jovanovic (1979), Miller (1984), and Flinn (1986), but also after good performance, as observed in the data. Building on Gibbons et al. (2005), Lluis (2005) and Hunnes (2012) assess the importance of comparative advantage and learning for worker mobility in firms as well as across occupations. Using information on wages and performance to estimate a perfectly competitive labor-market model without job assignment, Kahn and Lange (2014) document that learning and stochastic productivity changes are important determinants of the variance of wages in the BGH data. My results, based on a framework that extends GW to asymmetric firms and differentially informative jobs, imply that the selection of managers to jobs through promotions is central to the growth and dispersion of wages in the BGH firm.

As for the rest of the paper, Section 1 examines the data, Section 2 describes the model, Section 3 establishes identification and discusses model specification, Section 4 presents the model estimates, Section 5 contains the counterfactual exercises, and Section 6 concludes. See Appendixes A to C and the Supplementary Appendix (S.A.) for omitted details.

# 1 Data

The data, first analyzed by BGH, consist of the personnel records of all management employees of a medium-sized U.S. firm in a service industry observed between 1969 and 1988. The firm's job hierarchy consists of eight levels, Levels 1

[9] Analogously to Farber and Gibbons (1996), Altonji and Pierret (2001) show how the importance of learning can be inferred from the impact on wages of individual attributes that are not observed by employers, such as AFQT test scores, but, because of learning, should be progressively reflected in wages. This is in contrast to early signals about ability, such as schooling, whose importance for wages should correspondingly decline.

[10] The progressively greater selectivity implied by the first learning process leads matches of low and high tenure to be vulnerable to termination after adverse price shocks, owing to the countervailing effects of a higher match quality and a lower option value of learning on match-specific capital as tenure increases. By contrast, under the second learning process, matches of the shortest tenure are the most vulnerable to termination.

to 8 (chairman-CEO), but is clearly divided into two distinct parts: Levels 1 to 4, which contain nearly all observations (97.6 percent), and Levels 5 to 8, which consist of top management positions.[11] Given the small number of observations at Levels 3 and higher, I use BGH's original Levels 1 and 2 and aggregate Levels 3 to 8 into a single level, henceforth referred to as Level 3. The number of managers varies from 400 to 1,078 at Level 1, from 375 to 1,254 at Level 2, and from 496 to 3,199 at Level 3 across years, which suggests that capacity constraints on managerial employment at any level, or strong forms of fixed complementarity among managers, may not be too relevant. Ratings of managers' performance are available for about two-thirds of the original records and range from 1, the highest, to 5, the lowest. Ratings of 3 to 5, though, represent a very small fraction of all ratings: 36,750 of the 45,673 observations are 1 or 2. Hence, I combine ratings of 2 to 5 into a single measure of "0," obtaining a binary classification of *high* and *low* performance. I focus on 1,426 managers who enter managerial positions between 1970 and 1979 at Level 1 and have at least 16 years of education, experience no change in the recorded number of years of education during their first 10 years at the firm, and have no level information missing. No such manager is missing age or education information or is older than 45 years at entry. Because of the high separation rate from the firm each year (Table 1), I restrict attention to managers' first 8 years at the firm.[12] For completeness, I also estimated the model on a larger sample that includes entrants at the original Levels 1 to 4, which displays very similar features to those of the sample of entrants at Level 1. The estimates of the key parameters based on it are almost indistinguishable from those reported below (see the S.A.).

**Job Assignments and Separations.** Table 1 displays two main features. First, separation rates are high in all tenures. By the seventh year, over half the managers hired at Level 1 have separated from the firm and are no longer employed by it. Second, the percentage of managers assigned to Level 1 rapidly decreases with tenure, whereas that of managers assigned to Levels 2 and 3 first increases and then decreases with tenure. Table A.1 converts these distributions into level-specific hazard rates of separation, retention at the same level, and promotion to the next level by tenure. Note that whereas separation hazards at each level are approximately constant, promotion hazards initially increase and then decrease with tenure—as is common, promotions are by one level and demotions (almost) never occur. Interpreting the data through the lens of a learning and stochastic human capital acquisition model rationalizes all these features of the data. When higher ability and acquired skills are more valuable at higher-level jobs, managers advance through a firm's hierarchy, and possibly turn over across firms, as information and human capital accumulate. Such a process can also account for the nonmonotone tenure profile of promotion rates in two ways, which the model will incorporate. First, as

[11] To see that levels are not just pay grades, note that if levels were determined only by pay, then the firm would promote managers once their pay reached a certain threshold. So, promoted managers would always belong to the top percentiles of the wage distribution at the previous level, which is not the case in the BGH data; see BGH (1994a, Table VII). I use information on managers' year of entry, age, education, job level, salary, and performance rating but not on bonus pay, as it is unavailable before 1981 and otherwise missing for no fewer than 45.8 percent of managers in each tenure with higher percentages, up to 100, in early tenures; see Ekinci et al. (2019) for an analysis of bonus pay in the BGH data.

[12] Like BGH, I consider ratings as year-end variables. Since managers with a performance of 2 display assignment and wage profiles similar to those of managers with a performance of 3, 4, or 5, performance worse than 2 does not seem to convey information different from that conveyed by a performance of 2, as is consistent with the evidence that supervisors are averse to negative evaluations. So, intermediate ones are effectively equivalent to poor ones; see Section 3.1. In the BGH data, the average age of managers is 39 years with a standard deviation of 10 years, and their average number of years of education is 15 with a standard deviation of 2 years, from a minimum of 12 (high school) to a maximum of 23 (Ph.D.). Age and education display little variation across entrant cohorts. BGH report that the share of minorities and women at the firm increased over time; my copy of the data does not include information on race or gender. I consider entrants at Level 1 between 1970 and 1979 to be able to compare my results with those of BGH, as most of their analysis also concerns this group, and to avoid an excessive right censoring of the careers of later entrants. I exclude entrants in 1969, since it is unclear from the data in which year managers observed in 1969 entered.

the best managers, according to their perceived ability and human capital, are promoted out of a level, the remaining ones necessarily face worse promotion prospects. Second, if acquired human capital makes a manager on average more productive at a given level than at any other, then the probability of promotion out of that level eventually decreases.

**Performance Ratings.** Table 2 shows two patterns for ratings. First, the percentages of high ratings at Levels 1 and 2 decrease with tenure. Second, high ratings are more likely at Level 2 than at Level 1 in any tenure.[13] As shown in Table A.14 in the S.A., the percentage of high ratings is also higher among promoted than among unpromoted managers and, for promoted managers, higher among those promoted early in their tenure in a level relative to those promoted later. All these features of the data are consistent with the idea that ability and acquired human capital affect performance and are more valuable at higher levels. If managers receiving high ratings earlier in their tenure in a level are characterized, on average, by higher priors about their ability and higher human capital, then they are naturally the first assigned to higher levels.This selection process leads managers with greater ability and human capital to progress through the job hierarchy and so explains the higher performance of managers at higher levels compared with that of managers at lower levels.As a result, performance is lower for unpromoted than for promoted managers, and it worsens with tenure in a level.

**Wages.** Table 3 displays the distribution of (real in 1988 U.S. dollars) wages at the firm by level and tenure, and Table A.2 reports statistics on the distribution of wage changes by tenure. Three features emerge. First, wages are on average higher at higher levels, and the spread of the distribution of wages tends to increase with the level—see the note to Table 3. Since, as discussed, managers at higher levels tend to exhibit higher performance ratings, managers at higher levels on average perform better and are paid more. These characteristics of the data are consistent with a sorting process whereby managers with higher perceived ability and acquired human capital are promoted over time to higher levels and receive higher wages. As beliefs about managers' ability and their realized human capital increasingly differ over time, wages become more dispersed. Second, as is apparent from Table A.2, negative wage changes are quite frequent—over 20 percent in each tenure—as is compatible with a learning and stochastic human capital process. Intuitively, if low performance lowers beliefs about ability and implies lower levels of realized human capital, then it reduces the value of a manager's contribution to output and thus wages. Third, although wages tend to increase with tenure, their growth is nonmonotone, which confirms that wages are not governed simply by managers' progression through the job hierarchy.

**Case for Integrated Model.** As is common in firm-level data, in the BGH data, wages increase with tenure, job level, and performance; promotions and wage increases occur after good performance and are correlated over time; demotions (almost) never occur, despite the fact that wage decreases are frequent after low performance; and wage dispersion at each level is substantial. (See BGH on the serial correlation of promotions and wage increases and on the relationship between wages and performance.) I interpret these patterns of the data as resulting from a stochastic process of information and human capital acquisition, whereby managers accumulate information about their ability in addition to new skills when employed and progressively advance to the jobs (and firms) at which higher ability and human capital are most valuable.

[13] As ratings at Level 3 are missing for no fewer than 35 percent of managers assigned to Levels 3 and higher, with much higher proportions of missing values in lower tenures, I do not use information on ratings at Level 3 in estimation and so omit statistics on them. See footnote 27.

Crucially, ability must be correlated across jobs for managers to be promoted after *good* performance: were ability independent across jobs, job transitions would occur only after *bad* performance—as in typical bandit problems.

To see the importance of combining learning and stochastic human capital acquisition to rationalize all these aspects of careers and appreciate their separate roles, observe that learning and job assignment can account not just for mobility across jobs and firms but, unlike random productivity shocks, also for the observed serial correlation in wage increases, promotions, and performance (GW and Waldman (2013)). Learning further leads to wage increases and promotions after high performance and to wage decreases and demotions after low performance. As for acquired human capital, note that, on average, it gives rise to promotions and wage increases by augmenting managers' productivity and so accounts for the low frequency of demotions relative to promotions. Then, these two mechanisms can in principle be distinguished as follows. For a given frequency of wage increases and promotions, the higher the serial correlation in wage increases and promotions, the larger the role of unobserved ability—as in the runs test for the non-randomness of data. In addition, the higher the frequency of wage decreases, the greater the scope for learning. Yet, for a given frequency of wage decreases, the lower the frequency of demotions relative to that of promotions, the larger the role of human capital.[14]

The patterns of job assignments, performance ratings, and wages at the firm are also consistent with the notion that learning affects wages not only *directly*, as managers with higher performance ratings, who are thus perceived to be of higher ability, are paid more on average, but also *indirectly*. In particular, higher performance ratings, and so higher beliefs about ability, are associated with a greater chance of promotions, which lead to higher and more dispersed wages—a force the model will capture. That is, job assignment importantly mediates the impact of learning on wages.

## 2 A Model of Careers

I consider a labor market for an occupation (managers) in discrete time indexed by $t \geq 1$ (experience), in which two firms, $f = A, C$, compete for workers and use workers' effective labor or *human capital* at their jobs as the only input to production. Firms and workers discount the future by the factor $\delta \in (0, 1)$. Workers differ in their ability and human capital, which evolves stochastically over time according to a process that depends on a worker's ability, employing firm, and assigned job. Since ability is unobserved to all, firms and workers learn about it based on realized output.

**Firms and Workers.** Firms produce a homogeneous good sold in a perfectly competitive market at a price of one. I will occasionally refer to firm $A$ as *my firm*, the firm in my data, and firm $C$ as $A$'s *competitor*. Production in each firm $f$ is organized in a set $K^f$ of *jobs*, which can be interpreted as groupings of productive tasks possibly different across firms; for instance, tasks are known to be grouped by function, such as accounting or marketing, in certain firms and by skill requirements in others. As in GW, a worker is characterized by a level of ability $\theta \in \{\alpha, \beta\}$, unobserved to all, including the worker, and human capital, described next. Each $\theta$ corresponds to a vector of probabilities of success at the jobs of each firm, $\alpha = \{\alpha_{fk}\}$ and $\beta = \{\beta_{fk}\}$ with $\alpha_{fk} \geq \beta_{fk}$, as detailed below. I refer to $\alpha$ as *high ability* and $\beta$ as *low ability*.[15]

[14] In particular, since human capital, on average, increases productivity, acquired human capital offsets the adverse impact of low performance on beliefs about ability, so that managers are retained at a level, rather than demoted to a lower level, even after repeated low performance.

[15] By allowing technologies to differ across firms, this setup nests the case of perfect competition when firms have the same technology, a differentiated duopoly when firms have different technologies, or even a virtual monopoly when, say, $A$'s technology is uniformly more productive

**Human Capital and Output.** A worker in $t$ is endowed with human capital acquired before entry in the market, $H_1$, with *general* experience at the firms in the market, $H_t^G=\{H_{fkt}^G\}$, and with *specific* experience at their jobs, $H_t^S=\{H_{fkt}^S\}$. The production function of general $j=G$ and job-specific $j=S$ human capital at job $k$ of firm $f$ is

$$h_{fkt}^j=\ln H_{fkt}^j=a_{fkt}^j+g_k^j(i_{k_1}^j,\dots,i_{k_{t-1}}^j)+\varepsilon_{fkt}^j, \tag{1}$$

with $h_{fk1}^j=a_{fk1}^j+\varepsilon_{fk1}^j$. The term $a_{fkt}^j$ is stochastic and its distribution depends on a worker's ability, $i_{k_\tau}^j$ is the investment in period $\tau\leq t-1$, and $\varepsilon_{fkt}^j$ is a zero-mean i.i.d. human capital or *productivity shock*, realized at the start of $t$, capturing any temporary unexpected change in human capital unrelated to ability. Whereas general human capital accumulates with time spent in the market, job-specific human capital is acquired with time spent at a job: $i_{k_\tau}^G$ equals the constant $\bar{i}$ if a worker is employed at any job of either firm in $\tau$ and zero otherwise, whereas $i_{k_\tau}^S$ equals the constant $\bar{i}_{k_\tau}$ if a worker is employed at job $k_\tau$ of either firm in $\tau$ and zero otherwise. All investments, $a_{fkt}^j$, and $\varepsilon_{fkt}^j$ are observed by all.

When a worker is employed at job $k$ of firm $f$ in $t$, $a_{fkt}^j$ equals $\overline{a}_{fkt}^j$ with probability $\pi_{fk}(\theta)$ and $\underline{a}_{fkt}^j$ otherwise, where $\pi_{fk}(\theta)$ is $\alpha_{fk}$ for a high-ability worker and $\beta_{fk}$ for a low-ability worker. Thus, ability influences the evolution of human capital by affecting the probabilities of the realizations of $a_{fkt}^G$ and $a_{fkt}^S$, which are assumed to be Bernoulli distributed and perfectly correlated for consistency with the essentially binary and one-dimensional measure of performance in the data. (If the support of $a_{fkt}^j$ depended on $\theta$, ability could be inferred in one period.) As $\alpha_{fk}\geq\beta_{fk}$, the human capital of a high-ability worker is on average higher than that of a low-ability worker. Denote by $\overline{h}_{fkt}^j$ and $\underline{h}_{fkt}^j$ realized human capital after $\overline{a}_{fkt}^j$ ("success") and $\underline{a}_{fkt}^j$ ("failure") occur, which I refer to as *high* and *low performance*, respectively.[16]

Normalizing each worker's labor supply to one, a worker's (log) output at job $k$ of firm $f$ in $t$ is given by

$$y_{fkt}=h_{fkt}^1+h_{fkt}^G+h_{fkt}^S, \tag{2}$$

with $h_{fkt}^1$ function of $h_1=\ln H_1$. Denote output at the end of $t$ by $\overline{y}_{fkt}$ and $\underline{y}_{fkt}$, respectively, if $(\overline{h}_{fkt}^G,\overline{h}_{fkt}^S)$ and $(\underline{h}_{fkt}^G,\underline{h}_{fkt}^S)$ are realized. At the start of $t$, let human capital be summarized by $\kappa_t=(h_1,\{i_{k_\tau}^j\}_{\tau=1}^{t-1})$, $E(y_{fkt}|\theta,\kappa_t)$ be the expected output at job $k$ of firm $f$ conditional on $(\theta,\kappa_t)$, and $p_t$ be the prior that a worker is of high ability. By (1) and (2), *expected output* conditional on $(p_t,\kappa_t)$, which is the average of $E(y_{fkt}|\theta,\kappa_t)$ across the two possible values of $\theta$, is

$$y_f(p_t,\kappa_t,k)\;=\;\underbrace{\underline{y}_{fkt}^e+\beta_{fk}(\overline{y}_{fkt}-\underline{y}_{fkt})}_{E(y_{fkt}|\beta,\kappa_t)}\;+\;\underbrace{(\alpha_{fk}-\beta_{fk})(\overline{y}_{fkt}-\underline{y}_{fkt})}_{E(y_{fkt}|\alpha,\kappa_t)-E(y_{fkt}|\beta,\kappa_t)}p_t\;=\;d_{fk}(\kappa_t)+e_{fk}(\kappa_t)p_t, \tag{3}$$

with $\underline{y}_{fkt}^e$ defined as $\underline{y}_{fkt}-\varepsilon_{fkt}^G-\varepsilon_{fkt}^S$. An $\alpha$-worker has an absolute advantage at job $k$, since $E(y_{fkt}|\alpha,\kappa_t)\geq E(y_{fkt}|\beta,\kappa_t)$ if $\alpha_{fk}\geq\beta_{fk}$, and a comparative advantage at $k'$ over $k$ if $E(y_{fk't}|\alpha,\kappa_t)/E(y_{fkt}|\alpha,\kappa_t)\geq E(y_{fk't}|\beta,\kappa_t)/E(y_{fkt}|\beta,\kappa_t)$.

**Information and Beliefs.** Firms and workers share the initial prior belief $p_1$ that a worker is of high ability—in estimation, I allow for different initial priors across workers. Since firms and workers learn about ability from performance,

than $C$'s. I assume that the value of not working in the market considered is low enough that a worker is employed each period.

[16]This process generalizes to a setting with multi-job firms and learning the process in Bagger et al. (2014), $h_t=a+g(t)+\varepsilon_t$, where $a$ is known worker productivity, $g(t)$ is a deterministic trend for acquired human capital, and $\varepsilon_t$ is a zero-mean shock. In light of the data, I abstract from differences in $g_k^j(\cdot)$ across firms. See Appendix B.1 for common laws of motion of human capital that can be expressed in cumulative form as (1).

which is observed by all, learning is symmetric: firms and workers share the same information, and thus the same prior, in any $t$.[17] Each period after production occurs, firms and workers update their beliefs about a worker's ability according to Bayes's rule. Given $p_t$, this leads to two possible values of $p_{t+1}$ after high ($H$) and low ($L$) performance,

$$P_{fHk}(p_t) = \frac{\alpha_{fk} p_t}{\alpha_{fk} p_t + \beta_{fk}(1-p_t)} \quad \text{or} \quad P_{fLk}(p_t) = \frac{(1-\alpha_{fk}) p_t}{(1-\alpha_{fk}) p_t + (1-\beta_{fk})(1-p_t)}, \tag{4}$$

respectively. Since the probability of high performance varies across jobs, the *informativeness* of jobs and so the *speed of learning* about ability differ across them, ranging from the case of no learning ($\alpha_{fk} = \beta_{fk}$), when the posterior always equals the prior, to that of complete learning ($\alpha_{fk} = 1$ and $\beta_{fk} = 0$), when the posterior belief that a worker is of high ability is one or zero, depending on the worker's actual ability, after just one period of employment. Note that jobs with the same expected output can be differentially informative about ability: as $\alpha_{fk}$ and $\beta_{fk}$ change, expected output in (3) can be kept constant by adjusting $\overline{y}_{fkt}$ and $\underline{y}_{fkt}$. In particular, low output variability jobs (with $\overline{y}_{fkt}$ close to $\underline{y}_{fkt}$) can be more or less informative than high output variability ones (with $\overline{y}_{fkt}$ much larger than $\underline{y}_{fkt}$).[18]

**Separations.** As information and human capital accumulate, workers' evolving absolute and comparative advantages across jobs lead naturally to *endogenous* separations between workers and firms. I also account for *exogenous* separations unrelated to ability or human capital by allowing a worker to leave the market at the end of each period $t$ with probability $1-\eta_{fk}(\kappa_t)$, which depends on the worker's employing firm $f$, assigned job $k$, and human capital as captured by $\kappa_t$.

**Timing.** At the start of $t$, productivity shocks are realized. Next, firms simultaneously submit their wage and job offers to workers for the period.[19] Then, each worker decides which offer to accept. Lastly, the offered wage is paid; performance, human capital, and output are realized; beliefs about ability are updated; and separation shocks occur. As firms commit to the period offers they make, the timing of wage payments in a period is immaterial. I refer to $y_f(p_t, \kappa_t, k) + \varepsilon_{fkt}$ with $\varepsilon_{fkt} = \varepsilon^G_{fkt} + \varepsilon^S_{fkt}$ as a worker's *conditional expected output* before performance is realized. Without loss, I focus on the competition between the two firms for one worker. Denote by $(w_t, k_t) = \{w_{ft}, k_{ft}\}$ the vector of each firm $f$'s wage and job offer and by $l_t = \{l_{ft}\}$ the vector of the worker's decisions to accept ($l_{ft} = 1$) or reject ($l_{ft} = 0$) each offer.

**Equilibrium.** I restrict attention to robust Markov perfect equilibria (Bergemann and Välimäki (1996)). The state firms face when making their wage and job offers is $(s_t, \varepsilon_t)$, where $s_t = (p_t, \kappa_t)$, $p_t$ is the prior that the worker is of high ability, $\kappa_t = (h_1, \{i^j_{k_\tau}\}_{\tau=1}^{t-1})$ summarizes human capital, and $\varepsilon_t = \{\varepsilon_{fkt}\}$ collects all realized productivity shocks. The state the worker faces when choosing among offers consists of $(s_t, \varepsilon_t)$ and firms' wage and job offers, $(w_t, k_t)$. An equilibrium consists of offer strategies $w_{ft} = w_f(s_t, \varepsilon_t)$ and $k_{ft} = k_f(s_t, \varepsilon_t)$ for each firm $f$, an acceptance strategy $l_t = l(s_t, \varepsilon_t, w_t, k_t)$ for the worker with typical element $l_{ft} = l_f(s_t, \varepsilon_t, w_t, k_t)$ for each $f$, and belief-updating rules $P_{fHk}(p_t)$

[17] This setup captures the notion that in professional labor markets, resumes, references, recommendation letters, and other information gathered by human resource departments can accurately convey performance at previous jobs (Oyer and Shaefer (2011)). Maintaining that performance or human capital is commonly observed is less restrictive than it may seem: firms other than a worker's current employer can infer a worker's performance in the previous period from the worker's wage in the current period. Then, observing wages is analogous to observing performance.

[18] By Blackwell's informativeness criterion, job $k'$ is *more informative* than $k$ if the posterior after performance is observed at $k'$ second-order stochastically dominates that at $k$. Relative to the perfectly competitive setup of GW, where human capital, $h_{fkt} = \theta f(x_t) + \varepsilon_{fkt}$, depends on ability $\theta$, experience, $x_t$, and learning noise $\varepsilon_{fkt}$, in my setup the speed of learning differs across jobs, expected human capital and output vary across firms and jobs, human capital is acquired with experience at firms and jobs, and productivity shocks are present in addition to learning noise.

[19] Productivity shocks stochastically affect the value of a job to a firm across periods and workers so that the variability of job offers over time and among workers is not due just to the process of learning about ability and of human capital acquisition.

and $P_{fLk}(p_t)$ for each $f$ and $k$ such that in each period *i*) the worker maximizes the (expected present discounted) value of wages; *ii*) both firms maximize the (expected present discounted) value of profits; *iii*) the non-employing firm is indifferent between employing and not employing the worker at the job that maximizes its (expected present discounted) value of profits; and *iv*) beliefs are updated as in (4). Given the firms' strategies, the worker's strategy satisfies

$$W(s_t, \varepsilon_t, w_t, k_t) = \max_{\{l_f\}} \sum_f l_f \left\{ w_{ft} + \delta\eta_{fk_{ft}}(\kappa_t) \int_{\varepsilon_{t+1}} EW(s_{t+1}, \varepsilon_{t+1}, w_{t+1}, k_{t+1}|s_t, k_{ft})dG \right\}, \tag{5}$$

where $(w_{t+1}, k_{t+1})$ is the future set of offers, $G$ is the c.d.f. of future productivity shocks, $\varepsilon_{t+1}$, the expectation $EW(\cdot)$ is over performance at the accepted job, $k_{ft}$, and $(s_{t+1}, \varepsilon_{t+1}) = (P_{fHk_{ft}}(p_t), \kappa_{t+1}, \varepsilon_{t+1})$ or $(s_{t+1}, \varepsilon_{t+1}) = (P_{fLk_{ft}}(p_t),$-$\kappa_{t+1}, \varepsilon_{t+1})$. In evaluating an offer, the worker weighs the offered wage, $w_{ft}$, against the prospect of acquiring information and human capital at job $k_{ft}$. Given the worker's and the competitor's strategies, firm $f$'s strategy satisfies

$$\Pi^f(s_t, \varepsilon_t) = \max_{w,k} \Big( l_{ft} \Big\{ y_f(s_t, k) + \varepsilon_{fkt} - w + \delta\eta_{fk}(\kappa_t) \int_{\varepsilon_{t+1}} E\Pi^f(\cdot|s_t, k)dG \Big\} + l_{f't}\delta\eta_{f'k_{f't}}(\kappa_t) \int_{\varepsilon_{t+1}} E\Pi^f(\cdot|s_t, k_{f't})dG \Big), \tag{6}$$

$f' \neq f$, where $l_{f't} = l_{f'}(s_t, \varepsilon_t, w_t, k_t)$ is the worker's acceptance decision about the offer of firm $f'$ and $k_{f't}$ is the job offered by firm $f'$ in $t$. Note that each firm takes into account the option value of not employing the worker in a period and attracting the worker in some future period, which arises from the information revealed by performance and the human capital acquired by the worker at the competing firm. In particular, by (5) and (6), workers and firms *experiment* in that they contemplate sacrificing current wages and output to acquire more information about ability. Unlike in a standard multi-armed bandit problem, in which a decision maker repeatedly chooses among alternatives (arms) with uncertain *independent* rewards thus learning about their distribution, here the returns from firms' jobs (the arms) depend on the worker's unobserved ability, which is *common* across them. By *iii*), I require a Markov perfect equilibrium to be *robust* in that if firm $f'$ employs the worker, then the offer by the losing firm $f$, $f \neq f'$, must make it indifferent between not employing (left side of (7)) and employing (right side of (7)) the worker; that is, $\Pi^f(s_t, \varepsilon_t|f') = \Pi^f(s_t, \varepsilon_t|f)$, or

$$\delta\eta_{f'k_{f't}}(\kappa_t) \int_{\varepsilon_{t+1}} E\Pi^f(\cdot|s_t, k_{f't})dG = \max_{w,k} \left\{ y_f(s_t, k) + \varepsilon_{fkt} - w + \delta\eta_{fk}(\kappa_t) \int_{\varepsilon_{t+1}} E\Pi^f(\cdot|s_t, k)dG \right\}. \tag{7}$$

This refinement implies that wages and so equilibrium are unique. Intuitively, it rules out offers by $f$ with a present value of wages higher than firm $f$'s present value of output, net of the continuation value from not employing the worker.[20]

**Employment, Job Assignment, and Wages.** With ability correlated across jobs and firms, each firm's job offer policy and the worker's acceptance policy solve a *dependent* bandit problem, and so they do not admit a simple index form, as in independent bandit problems. Employment and job assignment can be nonetheless characterized as the solution to two pseudo-planning problems, which are natural analogues of the stopping problem that leads to an optimal (Gittins) index policy for independent bandits. Formally, denote by $V^f(s_t, \varepsilon_t)$ the maximal value of firm $f$'s match surplus at state $(s_t, \varepsilon_t)$ or *match surplus value*, which is defined as the sum of the worker's value function $W(s_t, \varepsilon_t)$, expressed using

[20] Offers with a present value of wages smaller than this net present value of output are already ruled out by equilibrium, since the losing firm must obtain a higher value of profits by not employing than by employing the worker; that is, the left side of (7) must weakly exceed the right side of it in any Markov perfect equilibrium. This notion of robust equilibrium generalizes that of cautious Markov perfect equilibrium introduced by Bergemann and Välimäki (1996) to a setting of competition among firms in jobs as well as wages.

the equilibrium dependence of wage and job offers on $(s_t, \varepsilon_t)$, or *value of wages* and firm $f$'s value function $\Pi^f(s_t, \varepsilon_t)$ or *value of profits*. Let $V^f(s_t, \varepsilon_t|f')$, $W(s_t, \varepsilon_t|f')$, and $\Pi^f(s_t, \varepsilon_t|f')$ be, respectively, firm $f$'s match surplus value, the worker's value of wages, and firm $f$'s value of profits conditional on firm $f'$ employing the worker.

I establish two properties. First, given each firm's choice of job, the employing firm at each state is the one generating the largest sum of values to all, $S(s_t, \varepsilon_t) = \Pi^A(s_t, \varepsilon_t) + W(s_t, \varepsilon_t) + \Pi^C(s_t, \varepsilon_t)$; that is, it is determined by a planning problem in which the planner's choice of job is restricted to the two jobs that maximize each firm's value of profits. Thus, conditional on a firm's choice of job, a worker's choice of firm is efficient. Second, each firm's choice of job at each state maximizes the value of its *output*, $\overline{V}^f(s_t, \varepsilon_t)$, defined in Proposition 1, regardless of whether the firm employs the worker, as if it were the only firm in the market. This result is due to the worker's indifference between the two firms' offers and the non-employing firm's indifference between employing and not employing the worker in equilibrium.[21]

**Proposition 1.** *The employing firm is determined in equilibrium by the policy that solves*

$$S(s_t, \varepsilon_t) = \max_f \left\{ y_f(s_t, k_{ft}) + \varepsilon_{fk_{ft}t} + \delta\eta_{fk_{ft}}(\kappa_t) \int_{\varepsilon_{t+1}} ES(s_{t+1}, \varepsilon_{t+1}|s_t, k_{ft}) dG \right\}, \tag{8}$$

*where* $k_{ft} = k_f(s_t, \varepsilon_t)$ *solves* $\overline{V}^f(s_t, \varepsilon_t) = \max_{k \in K^f} \left\{ y_f(s_t, k) + \varepsilon_{fkt} + \delta\eta_{fk}(\kappa_t) \int_{\varepsilon_{t+1}} E\overline{V}^f(s_{t+1}, \varepsilon_{t+1}|s_t, k) dG \right\}$.

As for equilibrium wages, the logic generalizes the one familiar from a static model of Bertrand competition. In such a model, the wage the employing firm pays is sufficiently high that a competitor cannot match it and obtain positive profits. For example, suppose for simplicity that the expected output of firms $A$ and $C$ when they employ the worker is $y_{At}$ and $y_{Ct}$, respectively. If $y_{At} > y_{Ct}$, then the worker is employed by $A$ at wage $w_{At} = w_{Ct} = y_{Ct}$, whereas if $y_{At} < y_{Ct}$, then the worker is employed by $C$ at wage $w_{At} = w_{Ct} = y_{At}$. Thus, a worker is typically paid less than output, unless the two firms share the same technology, in which case $w_{At} = w_{Ct} = y_{At} = y_{Ct}$, as under perfect competition.

Similarly, in the dynamic case, the employing firm must pay a wage high enough that a competitor cannot match it and obtain a positive value of profits—net of the continuation value from not employing the worker. The wage rule, though, is more general: a worker can be paid *more* or less than (conditional) expected output. For instance, $A$'s value of future profits can be high enough to justify paying a wage higher than expected output in a period if $A$ can sufficiently improve its future employment and assignment decisions based on the information or human capital acquired by employing the worker. When productivity shocks are Gumbel distributed (maximum) and any job $k$ entails the same opportunity for information acquisition ($\alpha_{Ak} = \alpha_{Ck}$ and $\beta_{Ak} = \beta_{Ck}$) and risk of exogenous separation ($\eta_{Ak}(\kappa_t) = \eta_{Ck}(\kappa_t)$) across firms, as I assume in Section 3, wages take a particularly simple form. See Section 2.2 in the S.A. for the general case.

**Proposition 2.** *Let productivity shocks be Gumbel distributed with mean* $0$ *and variance* $\pi^2/6$. *The equilibrium wage is*

$$w_f(s_t, \varepsilon_t) = y_{f'}(s_t, k_{ft}) - \ln \Pr(k_{f't} = k_{ft}|f_t = f', s_t) + \varepsilon_{f't} \tag{9}$$

*when the worker is employed by firm* $f$, $f' \neq f$, *if* $\alpha_{Ak} = \alpha_{Ck}$, $\beta_{Ak} = \beta_{Ck}$, *and* $\eta_{Ak}(\kappa_t) = \eta_{Ck}(\kappa_t)$ *at each job* $k$.

By Proposition 2, the wage paid by, say, firm $A$ can be conveniently expressed as the sum of the worker's expected output at $C$ at the job chosen by $A$, $y_C(s_t, k_{At})$, the negative log probability that $C$ chooses the same job as $A$ (the event

[21] These two efficiency properties do not imply overall efficiency, as neither firm internalizes the impact of its choices on its competitor's profits.

$\{k_{Ct}=k_{At}\}$) conditional on employing the worker (the event $\{f_t=C\}$), and the productivity shock to $C$'s match surplus value, $\varepsilon_{Ct}$.[22] The second term of (9) is a premium that accounts for $C$'s job offer typically differing from $A$'s. Intuitively, Bertrand competition implies that $C$ bids its entire match surplus value from employing the worker at its offered job—net of the continuation value from not employing—which I do not observe. But this value can be expressed in terms of $C$'s value had $C$ chosen the same job as $A$, which I do observe, up to an adjustment term for the possibility that $A$'s preferred job may not be $C$'s optimal choice. By standard Gumbel properties, this term is given by $\ln \Pr(k_{Ct}=k_{At}|\cdot)$. By the symmetry in the state law of motion across firms, the remaining terms reduce to $y_C(s_t, k_{At})$ and $\varepsilon_{Ct}$.[23]

# 3 Empirical Analysis

In this section, I discuss the assumptions maintained in estimation, the identification of the model, and the estimated specification. As I only observe workers, henceforth managers, at $A$, I denote by $t=1$ the year of entry in $A$ from now on.

## 3.1 Preliminaries

Recall that $\kappa_t = (h_1, \{i^j_{k_\tau}\}_{\tau=1}^{t-1})$. I maintain that the discount factor, $\delta$, is known, that productivity shocks are Gumbel distributed (maximum) with mean zero and variance $\pi^2/6$, and that firms $A$ and $C$ have three jobs or job *levels* each with common probabilities of high performance by manager ability, $\{\alpha_{fk}, \beta_{fk}\}$, and rates of exogenous separation, $\{1-\eta_{fkt}\}$ with $\eta_{fkt}=\eta_{fk}(\kappa_t)$. I refer to these probabilities as $\alpha_k$, $\beta_k$, and $\eta_{kt}$. See Section 2.2 in the S.A. for the identification of the model when the information and human capital processes and the risk of exogenous separation differ across firms, provided that data from multiple firms are available.

**Unobserved Heterogeneity.** As managers may differ along dimensions known to managers and firms—say, through job interviews—but unobserved to the econometrician, in the spirit of Heckman (1981) I assume that each manager is of skill type $i=1,\dots,I$, which affects the initial priors about ability, output, and wages. Based on changes in likelihood values and the Akaike information criterion, I set $I=4$. Let $p_{i1}$ be the initial prior that a manager of skill type $i$ is of high ability, $s_{it}=(p_{it},\kappa_t,i)$ be a manager's state at the beginning of $t$, and $q_i=\Pr(i|L_{A1}=1)$ be the probability of skill type $i$, where $L_{ft}$ is a manager's level at firm $f$ in $t$, which corresponds to job $k_{ft}=k_f(s_{it},\varepsilon_t)$ in the model. I allow firm $A$ to face different competitors for managers of different skill types so that firm $C$ varies across $i$. Thus, I let $a^j_{Ckt}$ depend on $i$, and denote its two possible values for each $j$, $k$, $t$, and $i$ by $\overline{a}^j_{Ckt}(i)$ and $\underline{a}^j_{Ckt}(i)$. Expected output in (3) then becomes

$$y_f(s_{it},k)=\underline{y}^e_{fkt}(i)+\beta_k[\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)]+(\alpha_k-\beta_k)[\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)]p_{it}=d_{fk}(\kappa_t,i)+e_{fk}(\kappa_t,i)p_{it}. \quad (10)$$

**Performance Ratings.** Like BGH, I interpret recorded ratings as noisy measures of performance and allow their error to be biased, as in Keane and Wolpin (1997) and Keane and Sauer (2009).[24] Formally, denote by $E_0(k,t)=\Pr(R^o_{At}=$

[22] The term $\varepsilon_{f't}$ is the random component of the maximum surplus $C$ can generate by employing the worker. See the proof of the proposition.

[23] Note that if the two firms had the same technologies and a worker experienced the same productivity shocks across them, then the market would be perfectly competitive and firms would choose the same jobs. In this case, the paid wage would equal a worker's conditional expected output as in the static case, $w_A(s_t,\varepsilon_t)=y_A(s_t,k_{At})+\varepsilon_{Ak_{At}t}$, where $y_A(s_t,k_{At})+\varepsilon_{Ak_{At}t}$ plays the role of $y_{At}$ in the above example.

[24] It is well known that ratings measure performance imperfectly and that the associated error may be systematic. For instance, supervisors tend to assign uniform ratings to employees regardless of their true performance, potentially leading to repeated misreporting of actual performance. See Baker et al. (1988), Murphy (1992), Prendergast (1999), and, for a study that uses direct measures of worker output, Lazear et al. (2015).

$1|R_{At}=0, L_{At}=k, t)$ the probability of a recorded *high* rating, $R^o_{At}=1$, when *low* performance is realized, $R_{At}=0$, at Level $k$ in $t$ so that $R^o_{At}$ is the performance rating observed by the econometrician, whereas $R_{At}$ is the true performance observed by firms and managers. Similarly, denote by $E_1(k,t)=\Pr(R^o_{At}=0|R_{At}=1, L_{At}=k,t)$ the probability of a recorded *low* rating when *high* performance is realized at Level $k$ in $t$. I specify error rates as simple logistic functions,

$$E_0(k,t)=\frac{e^{\rho_0+\rho_2(k)[t\times 1\{k=1\}+(t-1)\times 1\{k=2\}+(t-2)\times 1\{k=3\}]}}{1+e^{\rho_0+\rho_2(k)[t\times 1\{k=1\}+(t-1)\times 1\{k=2\}+(t-2)\times 1\{k=3\}]}} \text{ and } E_1(k,t)=\frac{1-E_0(k,t)}{1-E_0(k,t)+E_0(k,t)e^{\rho_1}}, \tag{11}$$

with $\rho_1>0$ so that the probability of a high rating increases with that of high performance, $\Pr(R^o_{At}=1|R_{At}=1, L_{At}=k,t)>\Pr(R^o_{At}=1|R_{At}=0, L_{At}=k,t)$. I allow error rates to vary across levels and tenures, since performance appraisal is often job-specific and may be more thorough at certain stages of a manager's career at the firm. As no manager is observed at Level 2 before $t=2$ or Level 3 before $t=3$, I let error rates at these levels depend on $t-1$ and $t-2$.

**Wages.** I allow for measurement error in wages, $\varepsilon^m_{it}$, that can differ across skill types and is zero-mean Gumbel distributed (minimum) with standard deviation $\sigma_{Aik}$. Let $\epsilon_{Ait}=\lambda^\varepsilon_{ik}\varepsilon_{Ct}+\varepsilon^m_{it}$, with $\lambda^\varepsilon_{ik}=\sqrt{6}\sigma_{Aik}/\pi$. For a manager of skill type $i$,

$$w_{Ait}=Ew_{Ait}(k)+\epsilon_{Ait}=y_C(s_{it},k)-\ln\Pr(L_{Ct}=k|f_t=C,s_{it})+\epsilon_{Ait} \tag{12}$$

is the recorded wage at Level $k$ of firm $A$ at state $(s_{it},\varepsilon_t)$ by (9), where $\epsilon_{Ait}$ is a zero-mean logistically distributed shock with standard deviation $\sigma_{Aik}$, since it is the difference between two Gumbel-distributed (maximum) random variables.

## 3.2 Identification

Identifying a partial-equilibrium model of the labor market from data on one firm is a known challenge. The approach I follow combines information on wages, job assignments, and performance, and exploits two features of the model: *i*) the symmetry in the opportunities for information acquisition across firms; and *ii*) the relationship between wages and firm productivity implied by Bertrand competition. By this symmetry, the law of motion of beliefs about ability can be recovered from just the distribution of performance ratings at $A$. Given a manager's prior, human capital, and skill type, Bertrand competition implies that a manager's expected wage at $A$ depends only on $C$'s productivity and job offer.

A first key observation is that by (12), the level distribution of wages at $A$ in each tenure, conditional on managers' histories of assignments and performance ratings at $A$, is an (identifiable) finite mixture of the logistic distributions of wages of managers of each *skill type* with any possible history of *true performance*, given their histories of assignments and performance ratings at $A$. The identified weights of these mixtures pin down the distribution of initial priors and, over time, both the distribution of classification error in performance ratings and the joint probabilities of managers' histories of assignments, true performance, and performance ratings at $A$ by skill type. As repeated ratings information at $A$ identifies the distribution of true performance, from these joint probabilities, conditional assignment probabilities at $A$ can be recovered. These assignment probabilities, in turn, identify expected output at $A$ at each state by standard discrete choice arguments, and so the human capital and output process, up to a level normalization.

A second key observation is that mean wages at $A$ at any job level are the sum of $C$'s *expected output* at the same level and the *probability* that $C$ chooses it by (12). But assignment probabilities at $C$ are determined by $C$'s expected match surplus value and so ultimately by $C$'s expected output. Thus, it is easy to establish not only that $C$'s expected match surplus value can be recovered from mean wages at $A$ but also that differences in mean wages at $A$ across levels map into differences in $C$'s expected output across levels. Intuitively, then, it is possible to show that mean wages at $A$ identify $C$'s expected output by job level at each state, up to two level normalizations—relative to standard discrete choice arguments, an additional normalization is required, since assignments at $C$ are unobserved. In estimation, I dispense with these two normalizations by setting $\{y_f(s_{it}, 2)\}$ so that the probability of assignment to Level 2 (the "reference" level) is equal across firms and by estimating $\{y_C(s_{it}, k)\}$ relative to $\{y_A(s_{it}, k)\}$; see Section 3.3. Alternatively, identification can be achieved by imposing exclusionary restrictions on the dependence of $y_C(s_{it}, k)$ on $s_{it}$ across levels.

**Summary.** The argument features four steps. In Step 1, I show that classification error rates, $\{E_0(k,t), E_1(k,t)\}$, and the distribution of performance, $\{\alpha_k, \beta_k\}$, at each level are identified from wages and performance ratings at $A$. In Step 2, I prove that the initial priors about ability, $\{p_{i1}\}$, are also identified from ratings. Since, as I show, the probability masses of the initial priors, $\{q_i\}$, are identified from the weights of the identified mixture distribution of wages in the first year at the firm, these two steps identify the learning process. In Step 3, I establish that expected output at $A$, up to that at one level, and the exogenous separation rates $\{1-\eta_{kt}\}$ are identified from conditional assignment probabilities at $A$, which can be recovered from the weights of the identified mixture distributions of wages in each tenure. In Step 4, I show that $C$'s "ex ante" match surplus value, which governs wages at $A$, and, up to two level normalizations, $C$'s expected output by level are identified at each state from wages at $A$. The wage means $\{Ew_{Ait}(k)\}$ and standard deviations $\{\sigma_{Aik}\}$ are identified by Lemma 1 below. As it will become apparent in Section 3.3, the expressions for $\{y_f(s_{it}, k)\}$ provide (linear) moment conditions from which the human capital and output parameters of interest can be easily recovered.

**Argument.** By (12), the density of wages of managers with initial human capital $h_1$ in the first year at $A$ is a finite mixture of the logistic densities of wages of each manager skill type $i$, $f(w_{A1}|L_{A1}=1, h_1) = \sum_i q_i f(w_{A1}|L_{A1}=1, h_1, i)$, where $q_i = \Pr(i|L_{A1}=1)$ is the typical weight, independent of $h_1$, and $f(w_{A1}|L_{A1}=1, h_1, i)$ is the typical component density with standard deviation $\sigma_{Ai1}$. Since such a mixture is identified up to its labeling with respect to $i$ (Al-Hussaini and Ahmad (1981, Proposition 1) and Shi et al. (2014, Theorem 1)), so are its weights and components. In the second year, wages depend on the updated beliefs about managers' ability based on their true performance. Thus, the density of wages of managers with initial human capital $h_1$ and first-year rating $R^o_{A1}$ at level $L_{A2}$ is a finite mixture of the logistic densities of wages of each manager skill type $i$ with each possible first-year performance $R_{A1}$, $f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, h_1) = \sum_{R_{A1}, i} \Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1) f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, R_{A1}, h_1, i)$. The weights and component densities of this mixture are then identified up to their labeling with respect to $i$ and $R_{A1}$. The products of the probabilities $\{\Pr(L_{A1}=1, R^o_{A1}, L_{A2}, h_1)\}$, known from the data, and the identified weights $\{\Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)\}$ pin down the probabilities of managers' histories by skill type, $\{\Pr(L_{A1}=1, R^o_{A1}, R_{A1}, L_{A2}, h_1, i)\}$. A similar logic applies to $t>2$ and to wages conditional only on managers' initial human capital and histories of assignments at the firm.

**Lemma 1.** *Let* $\mathcal{H}_{Lt} = (L_{A1}=1, \ldots, L_{At-1})$, $\mathcal{H}_{LRt} = (L_{A1}=1, R^o_{A1}, \ldots, L_{At-1}, R^o_{At-1})$, *and* $\mathcal{H}_t \in \{\mathcal{H}_{Lt}, \mathcal{H}_{LRt}\}$. *Order* $\{\sigma_{Aik}\}$ *by* $i$ *for each* $k$. *The density of wages at Level* $L_{At}$ *is an identified finite mixture up to its labeling by* $\{R_{A\tau}\}_{\tau=1}^{t-1}$,

$$f(w_{At}|\mathcal{H}_t,L_{At},h_1)=\sum_{R_{A1},\ldots,R_{At-1},i}\Pr(R_{A1},\ldots,R_{At-1},i|\mathcal{H}_t,L_{At},h_1)f(w_{At}|\mathcal{H}_t,L_{At},R_{A1},\ldots,R_{At-1},h_1,i). \quad (13)$$

Thus, up to labeling, the component weights, means, and standard deviations of these mixtures can be recovered. As shown in the proof of the lemma, ordering $\{\sigma_{Aik}\}$ by $i$ at each level is sufficient to order mixture components by $i$. In the proof of Proposition 5, I show that $\rho_1>0$ in (11) resolves any label indeterminacy with respect to true performance.

**Step 1: Distributions of Classification Error and Performance.** Omitting the dependence on $L_{A1}$ and $h_1$ for simplicity, the probability $\Pr(R^o_{A1},R_{A1},L_{A2},i)$, as just noted, is identified from the mixture density of wages at Level $L_{A2}$ in the second year at the firm conditional on $R^o_{A1}$, $f(w_{A2}|R^o_{A1},L_{A2})$. By Lemma 1, $f(w_{A2}|L_{A2})$ is also an identified mixture density with weights $\{\Pr(R_{A1},i|L_{A2})\}$. Then, multiplying each such weight by the corresponding known probability $\Pr(L_{A2})$ yields that $\Pr(R_{A1},L_{A2},i)$ is identified. In turn, the ratio of $\Pr(R^o_{A1},R_{A1},L_{A2},i)$ to $\Pr(R_{A1},L_{A2},i)$ pins down $\Pr(R^o_{A1}|R_{A1},i)$, which is independent of $i$.[25] By a similar logic, $\Pr(R^o_{A2}|R_{A2})$ is identified as well. By (11), from $\Pr(R^o_{A1}|R_{A1})$ and $\Pr(R^o_{A2}|R_{A2})$, the error rates $E_0(1,t)$ and $E_1(1,t)$, $t=1,2$, and so the parameters $(\rho_0,\rho_1,\rho_2(1))$, are recovered. An analogous argument holds for $\{\rho_2(k)\}_{k\geq 2}$. Thus, the distribution of classification error is identified.

Note that the distribution of performance ratings of managers retained at Level $k$ is a binomial mixture of the distribution of true performance of managers of high and low ability with parameters $(\alpha_k,\beta_k)$, contaminated by classification error. Once the error distribution is identified, this mixture is identified as follows. Consider managers at Level 1 up to $t=3$. With the error rates $E_0(1,t)$ and $E_1(1,t)$ recovered, the probability of high performance in $t\leq 3$ at Level 1,

$$\Pr(R_{A1}=\ldots=R_{At}=1|L_{A1}=L_{A2}=L_{A3}=1,h_1)=[\textstyle\sum_i q_{i3}(h_1)p_{i1}]\alpha_1^t+[1-\sum_i q_{i3}(h_1)p_{i1}]\beta_1^t, \quad (14)$$

is identified from that of a high rating in $t\leq 3$ at Level 1, $\Pr(R^o_{A1}=\ldots=R^o_{At}=1|L_{A1}=L_{A2}=L_{A3}=1,h_1)$, where $q_{i3}(h_1)$ denotes $\Pr(i|L_{A1}=L_{A2}=L_{A3}=1,h_1)$. Condition (14) evaluated in $t=1,2,3$ yields a system of three equations in three unknowns, $\alpha_1$, $\beta_1$, and the average prior $\sum_i q_{i3}(h_1)p_{i1}$, with a unique solution (if $\alpha_1>\beta_1$). Similarly, the performance ratings of managers employed at Levels 2 and 3 for three years identify $\{\alpha_k,\beta_k\}$ at these levels.[26]

**Proposition 3.** *The classification error parameters $(\rho_0,\rho_1,\rho_2(1))$ are identified from the distribution of wages in $t=2,3$ of managers at Level 1 of A in $t=1,2$ with either a high or low performance rating in $t=1,2$. The parameter $\rho_2(k)$ for $k\in\{2,3\}$ is identified from the distribution of wages in $t=k+1$ of managers at Level $k$ of A in $t=k$ with either a high or low performance rating in $t=k$. The probabilities of high performance $\{\alpha_k,\beta_k\}$ at Level $k\geq 1$ are identified from three years of consecutive observations on the distribution of performance ratings at the corresponding level.*

**Step 2: Initial Priors.** As just discussed, $\alpha_1$, $\beta_1$, and the average prior $\sum_i q_{i3}(h_1)p_{i1}$ for each $h_1$ are recovered from performance ratings at Level 1 in the first three years. By the analogue of (14), the average prior $\sum_i q_{it}(h_1)p_{i1}$ is similarly identified in $t\geq 4$ with $q_{it}(h_1)=\Pr(i|L_{A1}=\ldots=L_{At}=1,h_1)$ recovered for each $h_1$ by Lemma 1. Thus, knowledge of $\sum_i q_{it}(h_1)p_{i1}$ and $\{q_{it}(h_1)\}$ from $t=3$ to $t=6$ provides a system of four linearly independent equations,

[25] With $f(w_{A2}|R^o_{A1},L_{A2},R_{A1},i)$ independent of $R^o_{A1}$, $\{f(w_{A2}|L_{A2},R_{A1},i)\}$ can be recovered from $f(w_{A2}|R^o_{A1},L_{A2})$ or $f(w_{A2}|L_{A2})$, as these mixtures have identical components through which each weight $\Pr(R_{A1},i|R^o_{A1},L_{A2})$ can be paired with the corresponding one $\Pr(R_{A1},i|L_{A2})$.

[26] It is important that wage and rating distributions be identifiable mixtures of the distributions of wages of managers of each skill type and of true performance of managers of each ability, respectively. I specify them as logistic and binomial, but they would be identified for many alternative specifications. See Hunter et al. (2007) for semiparametric symmetric finite mixtures and Lindsay (1995) for general mixing distributions.

$q_{1t}(h_1)p_{11}+\ldots+q_{4t}(h_1)p_{41}=\sum_i q_{it}(h_1)p_{i1}$ for each such $t$, in the four unknowns $\{p_{i1}\}$, which are then identified.[27]

**Proposition 4.** *If* $\{\Pr(i|L_{A1}=\ldots=L_{At}=1,h_1)\}$, $3\leq t\leq 6$, $(\alpha_1,\beta_1)$, *and the classification error parameters at Level 1 are identified, then* $\{p_{i1}\}$ *are identified from the distribution of performance ratings at Level 1 of A from* $t=3$ *to* $t=6$.

**Step 3: Expected Output and Exogenous Separations.** I first argue that conditional assignment probabilities at $A$ in each tenure are identified for each type. Omit the dependence on $L_{A1}$ and $h_1$ and recall that $\{\Pr(R_{A1},L_{A2},i)\}$ are identified by Step 1 from the wage density $f(w_{A2}|L_{A2})$, $\{\alpha_k,\beta_k,p_{i1}\}$ are identified by Steps 1 and 2 so that the probability $\Pr(R_{A1}|i)=\alpha_1 p_{i1}+\beta_1(1-p_{i1})$ is identified for each $R_{A1}$ and $i$, and $\Pr(i)=q_i$ is identified for each $i$ by Lemma 1. Thus, so is $\Pr(L_{A2}|R_{A1},i)$, as it is the ratio of $\Pr(R_{A1},L_{A2},i)$ to $\Pr(R_{A1}=1|i)\Pr(i)$. This logic extends to any $t$.

As for expected output at $A$, abstract first from exogenous separations. By the properties of Gumbel distributions, the expected value of $\overline{V}^A(s_{it},\varepsilon_{it})$ in Proposition 1 conditional on $s_{it}$ can be expressed as $\ln\sum_k \exp\{\overline{v}^A(s_{it},k)\}$, where

$$\overline{v}^A(s_{it},k)=y_A(s_{it},k)+\delta E[\overline{v}^A(s_{it+1},k)-\ln\Pr(L_{At+1}=k|f_{t+1}=A,s_{it+1})|s_{it},k] \tag{15}$$

and $\Pr(L_{At}=k|f_t=A,s_{it})=\exp\{\overline{v}^A(s_{it},k)\}/\sum_{k'}\exp\{\overline{v}^A(s_{it},k')\}$ is the identified conditional assignment probability to Level $k$. For fixed probabilities, (15) is a functional equation defining a contraction mapping and so admits a unique solution. With the state law of motion and conditional assignment probabilities recovered, if $y_A(s_{it},2)$ is known, then $\overline{v}^A(s_{it},2)$ is identified. At Levels 1 and 3, in turn, $\overline{v}^A(s_{it},k)$ is identified as $\overline{v}^A(s_{it},2)+\ln\Pr(L_{At}=k|f_t=A,s_{it})-\ln\Pr(L_{At}=2|f_t=A,s_{it})$, and so is $y_A(s_{it},k)$ as the difference between $\overline{v}^A(s_{it},k)$ and $\delta E[\cdot]$ in (15). With exogenous separations, defining $\tilde{y}(s_{it})$ as the identified value $Ew_{Ait}(k)+\ln\Pr(L_{At}=k|f_t=A,s_{it})$, a similar argument applies.[28]

**Proposition 5.** *Let* $\{y_A(s_{it},2)\}$ *be known, and* $y_A(s_{it},k)=y_C(s_{it},k)=\tilde{y}(s_{it})$, $k\geq 1$, $t\geq 2$, *at one* $s_{it}$. *Then,* $\{y_A(s_{it},k)\}$, $k=1,3$, *at all other* $s_{it}$, *and* $\{\eta_{kt}\}$ *are identified from assignments and wages at A if* $\{p_{i1},q_i,\alpha_k,\beta_k\}$ *are identified.*

**Step 4: Wages.** By using (12), it is easy to express wage means, which are identified by Lemma 1, as

$$Ew_{Ait}(k)=\ln\textstyle\sum_{k'}\exp\{v^C(s_{it},k')\}-\delta\eta_{kt}E(\ln\sum_{k'}\exp\{v^C(s_{it+1},k')\}|s_{it},k),$$

which defines a functional equation that pins down $C$'s expected match surplus value, $\ln\sum_{k'}\exp\{v^C(s_{it},k')\}$. I now show how $C$'s expected output $y_C(s_{it},k)=d_{Ck}(\kappa_t,i)+e_{Ck}(\kappa_t,i)p_{it}$ can also be recovered. By (12) and the fact that conditional assignment probabilities sum to one, $y_C(s_{it},1)$ can be expressed as a known function $m_1(s_{it})$ of mean wages, $y_C(s_{it},2)$, and $y_C(s_{it},3)-y_C(s_{it},1)$.[29] If $y_C(s_{it},2)$ and $y_C(s_{it},3)-y_C(s_{it},1)$ are known at two values of $p_{it}$ for each $\kappa_t$ and $i$, then $d_{C1}(\kappa_t,i)+e_{C1}(\kappa_t,i)p_{it}=m_1(s_{it})$ evaluated at these two values of $p_{it}$ for managers with the same values of $\kappa_t$ and $i$ provides a system of two linearly independent equations in $d_{C1}(\kappa_t,i)$ and $e_{C1}(\kappa_t,i)$, which are then recovered. This logic applies to all $\kappa_t$ and $i$. Once $y_C(s_{it},1)$ and $y_C(s_{it},2)$ are known, by (12) the conditional probabilities of assignment to Levels 1-2 are identified from $\{Ew_{Ait}(k)\}$, and residually, so is the conditional probability of assignment

[27] Rating information at Levels 2 or 3 would alternatively identify $\{p_{i1}\}$. With $\{p_{i1}\}$ recovered, the proof of Proposition 4, as shown, can be adapted to establish $\{\alpha_k,\beta_k\}_{k\geq 2}$ are identified. See the S.A. (Section 2.1) for an argument à la Hu and Shum (2012) that does not require ratings.

[28] The probabilities of employment at $A$ and $C$ are equal if $y_A(s_{it},k)$ and $y_C(s_{it},k)$ equal the identified value $\tilde{y}(s_{it})$ at one $s_{it}$ for each $k$. In this case, the survival probability $\eta_{k_{t-1}t-1}$ is identified from the probability that a manager at Level $k_{t-1}$ of $A$ in $t-1$ is retained at $A$ at such $s_{it}$.

[29] This function is $m_1(s_{it})=\ln[e^{Ew_{Ait}(2)}-e^{y_C(s_{it},2)}]-\ln[e^{Ew_{Ait}(2)-Ew_{Ait}(1)}+e^{Ew_{Ait}(2)-Ew_{Ait}(3)+y_C(s_{it},3)-y_C(s_{it},1)}]$. Alternatively, knowledge of $y_C(s_{it},2)$ and the existence of a component of $s_{it}$ affecting, say, $y_C(s_{it},1)$ but not $y_C(s_{it},3)$ would allow to recover $y_C(s_{it},1)$ and $y_C(s_{it},3)$ at each state. So, by imposing simple exclusionary restrictions, one level normalization would suffice to pin down $C$'s expected output.

to Level 3. This probability and $Ew_{Ait}(3)$ identify $y_C(s_{it},3)$. A similar argument holds if $y_C(s_{it},1)+y_C(s_{it},3)$ is known instead of $y_C(s_{it},3)-y_C(s_{it},1)$ or, in general, if two sums of, or differences in, $C$'s expected output are known.[30]

**Proposition 6.** *The expected match surplus value of $C$ is identified at each state and tenure from wages at $A$. Suppose one of the following is known at two values of $p_{it}$ for each $\kappa_t$ and $i$: a) $y_C(s_{it},k)$ and either $y_C(s_{it},k')+y_C(s_{it},k'')$ (or $y_C(s_{it},k')-y_C(s_{it},k'')$) for some $k$, $k'$, and $k''$ with $k'\neq k''$; or b) $y_C(s_{it},k)+y_C(s_{it},k')$ (or $y_C(s_{it},k)-y_C(s_{it},k')$) for some $k\neq k'$ and $y_C(s_{it},k'')+y_C(s_{it},k''')$ (or $y_C(s_{it},k'')-y_C(s_{it},k''')$) for some $k''\neq k'''$ with either $k''$ or $k'''$ different from $k$ and $k'$. Then, $\{y_C(s_{it},k)\}$ are identified from wages at $A$ if $\{p_{i1},\alpha_k,\beta_k\}$ are identified.*

### 3.3 Empirical Specification

In this section, I present the empirical specification of the model. See Appendix B for omitted details.[31]

**Human Capital and Output.** I specify $g_k^G(i_{k_1}^G,\ldots,i_{k_{t-1}}^G)=g_k^G(t-1)$ in (1) so that, as commonly assumed, only the sum of past investments matters for the accumulation of general human capital and $g_k^S(i_{k_1}^S,\ldots,i_{k_{t-1}}^S)=g_k^S(i_{k_{t-1}}^S)$ in (1) so that only the last-period investment is relevant for the acquisition of job-specific human capital. This latter process is flexible enough to capture the benefit of training before a promotion when $g_k^S(i_{k_{t-1}}^S)>0$ in $t$ for $k>k_{t-1}$, as is consistent with much of the evidence that training often occurs right before a promotion or is effectively a prerequisite for it (Cobb-Clark and Dunlop (1999)); the potential loss of job-specific human capital upon promotion when $g_k^S(i_{k_{t-1}}^S)=0$ in $t$ for $k>k_{t-1}$; or any constraints that may induce a cost of reallocating managers across levels when $g_k^S(i_{k_{t-1}}^S)<0$ in $t$ for $k\neq k_{t-1}$.[32] Then, $s_{it}$ reduces to $(p_{it},h_1,t-1,k_{t-1},i)$, and the human capital functions in (1) become

$$h_{fkt}^G=a_{fkt}^G(i)+g_k^G(t-1)+\varepsilon_{fkt}^G \quad\text{and}\quad h_{fkt}^S=a_{fkt}^S(i)+g_k^S(i_{k_{t-1}}^S)+\varepsilon_{fkt}^S, \tag{16}$$

where $a_{fkt}^j(i)=a_{fkt}^j+a_{fk}^j(i)$, $a_{fkt}^j\in\{\overline{a}_{fkt}^j,\underline{a}_{fkt}^j\}$, and $a_{fk}^j(i)\in\{\overline{a}_{fk}^j(i),\underline{a}_{fk}^j(i)\}$. As discussed, $a_{Ak}^j(i)$ is zero.

Let $h_{fkt}^1=b_k(h_1)$ common across firms and let $h_1=(e_1,x_1,y_1)$ denote education at entry, $e_1$, experience (age) at entry, $x_1$, and year of entry, $y_1$, in firm $A$. As shown in Appendix B.2, expected output in (10) can thus be expressed as

$$y_f(s_{it},k)=\underbrace{\underline{y}_{fkt}^e(i)+\beta_k[\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)]}_{b_k(h_1)+d_{fkt}(k_{t-1},i)}+\underbrace{(\alpha_k-\beta_k)[\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)]}_{e_{fkt}(i)}p_{it}, \tag{17}$$

with $\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)=\sum_j(\overline{a}_{fkt}^j-\underline{a}_{fkt}^j)$. I specify $b_k(h_1)=b_{0k}+b_k^x x_1+b_k^{xx}x_1^2+b_k^e e_1+\sum_m b_{ym}1\{y_1=m\}$, with $m$ ranging from 0 (1970) to 9 (1979). This formulation accounts for different degrees of transferability of human capital across job levels that can be empirically detected. Namely, define human capital acquired at *firm $f$* to be *task-general* if the tenure profiles of the parameters $\{d_{fkt}(k_{t-1},i),e_{fkt}(i)\}$ are positively correlated across levels (Gibbons and Waldman (2006)) and *task-specific* otherwise. This notion captures the idea that managers may become more productive at all levels with

[30] Intuitively, knowing, say, $y_C(s_{it},1)+y_C(s_{it},2)$ and $y_C(s_{it},1)-y_C(s_{it},3)$ or $y_C(s_{it},1)+y_C(s_{it},2)$ and $y_C(s_{it},2)+y_C(s_{it},3)$ is equivalent.

[31] As the sample covers only the first eight years of managers at firm $A$, I assume that $\{\eta_{kt}\}$ are constant from $t=8$ on, in that $\eta_{kt}=\eta_{k7}$, $t\geq 8$.

[32] These forms for $\{g_k^j(\cdot)\}$, as discussed, can be derived from standard laws of motion; the identification results can be easily extended to more general forms for them. Baker et al. (1994b) argue that training takes place at the firm, since managers are rewarded for good performance primarily with salary increases rather than bonuses, supporting the notion that these raises reward permanent rather than transitory increases in productivity.

experience at a firm. Define, instead, human capital acquired at *Level $k$ of firm $f$* to be *task-general* if it is at least as productive at Level $k' \neq k$ as it is at Level $k$, $d_{fk't}(k,i) \geq d_{fkt}(k,i)$, and *task-specific* otherwise. This notion captures the idea that experience at a job level ($k$) may increase productivity at other levels ($k'$) even more. As in Cameron and Heckman (2001), to conserve on parameters, for transitions with relatively few observations, I do not estimate any of the associated parameters and set them equal to their reference value (zero). For transitions with a similar number of observations, I maintain their parameters are equal.[33]

**Wages.** By (12), wages at firm $A$ depend on the probability of job assignments at firm $C$, which are not observed in my data. It can be shown, though, that assignment probabilities at $C$ can be expressed as simple functions of the identified assignment probabilities at $A$ under the assumption of free entry of firms in the labor market, no-recall of previously employed managers (satisfied in my data), and the normalization of $\{y_f(s_{it},2)\}$, as is consistent with Propositions 5 and 6, so that $\Pr(L_{Ct}=2|f_t=C,s_{it})=\Pr(L_{At}=2|f_t=A,s_{it})$ at each $s_{it}$. In this case, $Ew_{Ait}(k)$ reduces to

$$Ew_{Ait}(k) = \underbrace{y_A(s_{it},k)+y_C(s_{it},2)-y_A(s_{it},2)}_{\omega(i,h_1,k)+\omega_{4kt}\times(t-1)+(\omega_{5i}+\omega_{5kt})p_{it}} - \ln\Pr(L_{At}=k|f_t=A,s_{it}), \tag{18}$$

where $\omega(i,h_1,k)=\omega_{0ik}+\omega_{1k}x_1+\omega_{2k}x_1^2+\omega_{3k}e_1+\sum_m \omega_{ym}1\{y_1=m\}$. See Lemma 2 in Appendix B.3 and equation (68) in Appendix B.7 for the simple mapping between the parameters in (18) and those of human capital and output at the two firms. The parameters in (18) are easy to interpret: $\omega_{1k}$, $\omega_{2k}$, and $\omega_{3k}$ measure the impact of human capital at entry in the firm; $\omega_{ym}$ allows for aggregate conditions at the time of hiring, which BGH find important, to persistently affect the value of a manager; and $\omega_{4kt}\times(t-1)$ and $\omega_{5i}+\omega_{5kt}$ measure, respectively, the effects of the deterministic component of the human capital acquired at the firm and of ability.[34] As per Lemma 1, I maintain that $\sigma_{Aik} \geq \sigma_{A(i+1)k}$.

# 4 Estimation Results

This section discusses the estimates of the model parameters obtained by full-information, full-solution maximum likelihood, which are all significant at the 1 percent level. In estimation, I set the discount factor to 0.95. As Tables 1 to 3 and A.1 show, the model well captures the patterns of managers' separations, their transitions across levels, and the level distributions of performance ratings and wages at the firm. I discuss model fit and additional results in Appendix B.[35]

**Uncertainty and Learning.** Two findings emerge. First, the initial prior distribution in Table 4A, with $p_{i1}$ estimated as $\exp(\phi_{i1})/[1+\exp(\phi_{i1})]$, implies a large degree of uncertainty about managers' ability and dispersion in information at entry in the firm. That is, the average initial prior probability that a newly hired manager is of high ability is 0.473, but initial priors range from 0.338 to 0.607 across skill types, with a standard deviation of 0.102 (see the table note).

[33] I normalize $d_{A3t}(L1)$, $d_{A1t}(L2)$, $d_{A1t}(L3)$, and $d_{A2t}(L3)$, where $Lk$ denotes the previous-period level, $k_{t-1}$, to zero, since multi-level promotions (first normalization) and demotions (last three normalizations) are rarely observed. I also impose a form of symmetry in that $d_{Ak't}(k) = -d_{Akt}(k')$, $k' > k$, where $d_{Akt}(k')$ can be interpreted as a demotion cost that I normalize to zero. The parameters left to estimate are then $\eta_{1t}$ and those of $b_k(h_1)$ in $t \geq 1$, $(d_{A1t}(L1), e_{A1t}, e_{A2t}, \eta_{2t})$ in $t \geq 2$, $(d_{A2t}(L2), e_{A3t}, \eta_{3t})$ in $t \geq 3$, and $d_{A3t}(L3)$ in $t \geq 4$.

[34] As in the previous footnote, of the parameters $\{d_{Ckt}(k_{t-1},i)\}$, only $\{d_{Ckt}(k,i)\}$ are estimated and captured by $\omega_{4kt}\times(t-1)$ in (18). Since $\{\omega_{5kt}\}$ did not vary much with $k$ or $t$ over the relevant range of values, I only estimated them as part of the robustness exercise in Appendix B.11.

[35] Section 5 in the S.A. reports estimates from a larger sample that includes managers entering the firm at levels higher than Level 1. Although I estimate a richer specification on this larger sample, the estimates of the key parameters across the two samples are very close.

Second, the estimates of $\{\alpha_k, \beta_k\}$ in Table 4B imply that learning about ability is gradual, as $\{\alpha_k\}$ and $\{\beta_k\}$ are far from 0 and 1, and suggest that Level 1 is the most informative level, thus playing an important role by allowing the firm to gather information about managers.[36] Specifically, by Blackwell's informativeness criterion, the ordering of job levels by informativeness from lowest to highest is 2, 3, and 1—note that the differences $\alpha_2 - \alpha_1$ and $\beta_2 - \beta_1$ are significant at the 1 percent level. Interestingly, the ordering of job levels from lowest to highest by the probability of high performance of a manager of high or low ability is the opposite: 1, 3, and 2. As a result, when deciding on a manager's assignment, the firm faces a clear trade-off between a manager's output and the information it can acquire about ability. For instance, although employment at Level 1 generates the most information, a manager of either ability has the lowest probability of high output at this level. The greater informativeness of Level 1 mostly accounts for the premium that managers receive upon promotion to Level 2—a mean percentage raise relative to unpromoted workers—consistent with the estimates in BGH (1994a, Table VI). Such a premium compensates managers for switching to a less informative job level—that is, for reaching higher levels at lower priors and so receiving lower wages than they would have received if they had been assigned to Level 1 for a longer period. See Appendix B.8 for the estimates of classification error in performance.

**Human Capital and Output.** The estimates of the output parameters $\{d_{Akt}(k_{t-1}), e_{Akt}\}$, expressed in thousands with $k_{t-1} = Lk$, are reported in Table A.4; all unreported parameters are set to zero, the value to which $d_{A2t}(\cdot)$ is normalized.[37] Three features emerge. First, both the parameters $\{e_{A2t}\}$ and $\{e_{A3t}\}$ start at low values, increase to peak at intermediate tenures ($t = 2$ and $t = 4$ for $\{e_{A2t}\}$ and $t = 4$ for $\{e_{A3t}\}$), and essentially flatten out thereafter. This positive correlation in the tenure profiles of $\{e_{A2t}\}$ and $\{e_{A3t}\}$ suggests that the human capital acquired at firm $A$ is task-general between Levels 2 and 3 by the definition in Section 3.3. In particular, since $e_{Akt}$, which equals $E(y_{Akt}|\alpha, \kappa_t) - E(y_{Akt}|\beta, \kappa_t)$ by (3), captures the output return to high ability, this estimated return tends to increase at Levels 2 and 3 with tenure.

Second, human capital acquired at Level 3 is task-specific and so implies that demotions to lower levels are costly for the firm, which helps explain why demotions are rare. To see why, recall from Section 3.3 that the more task-general the human capital accumulated at Level $k$ is, the larger the difference $d_{Ak't}(k) - d_{Akt}(k)$, $k' \neq k$, is. Given that $d_{A3t}(L3)$ is estimated to be large and positive in $4 \leq t \leq 7$ and, as explained in footnote 33, $d_{A1t}(L3)$ and $d_{A2t}(L3)$ equal zero, managers at Level 3 in $t-1$ and $t$ in intermediate to high tenures produce more in $t$ than they would if they were demoted in $t$ from Level 3 to Level 1, as $d_{A3t}(L3) > d_{A1t}(L3)$ and, for $t > 3$, $e_{A3t} > e_{A1t}$, or to Level 2, as $d_{A3t}(L3) > d_{A2t}(L3)$; that is, human capital acquired at Level 3 is highly task-specific. Moreover, as argued in footnote 33, $d_{A3t}(L1)$ and $d_{A3t}(L2)$ equal zero. Thus, managers at Level 3 in $t-1$ and $t$, $4 \leq t \leq 7$, produce more in $t$ than they would if they were promoted in $t$ to Level 3 from Level 1, as $d_{A3t}(L3) > d_{A3t}(L1)$, or from Level 2, as $d_{A3t}(L3) > d_{A3t}(L2)$.

[36] To measure the speed of learning, I compute the number of consecutive years of high performance that is necessary to infer with a 90 percent chance that a manager is of high ability. This speed varies across levels: starting at the average prior, it takes 20 years to reach this level of certainty at Level 1, whereas this process takes 23 years at Levels 2 or 3. That this speed also varies across skill types is apparent from the fact that, for example, this level of certainty for managers of the fourth skill type is reached just after 15, 18, and 17 years, respectively, at Levels 1, 2, and 3. This finding that learning is gradual is consistent with those in Nagypál (2007). From her Figure 7, the convergence of beliefs, reflected in the tenure profiles of match quality and output, occurs past the tenth year of tenure. This estimated graduality is also in line with the large degree of uncertainty about ability at entry in the firm that I estimate, even for managers with several years of labor market experience.

[37] See below the estimates of $(\omega_{1k}, \omega_{2k}, \omega_{3k}, \{\omega_{ym}\})$ for the estimates of $\{b_k(h_1)\}$. Note that $d_{A11}(\cdot)$ and $e_{A11}$ are not estimated and set, respectively, to 1,000 and zero, since all managers in the sample are first observed employed at Level 1.

Third, relative to low-ability workers, high-ability workers have not only an absolute advantage at all job levels, as the constraint $e_{Akt} > 0$, which I did not impose in estimation, is satisfied, but also a comparative advantage at Level 2 over 1, as $e_{A2t} > e_{A1t}$, and, in intermediate to high tenures, at Level 3 over 1, as $e_{A3t} > e_{A1t}$.[38]

These estimates imply that expected output is higher at Levels 2 and 3 than at Level 1 ($d_{Akt}(k)$, and in nearly all tenures, $e_{Akt}$, $k = 2, 3$, are higher than at Level 1); tends to increase at Levels 2 and 3 over time; and tends to be more sensitive to ability at higher levels ($e_{Akt} \geq e_{A1t}$) in mid to high tenures. Thus, higher ability and acquired human capital tend to make managers more productive at higher levels, which contribute more to output than lower levels.

**Wages.** The estimates in Tables 5A to 5C are similar to those in the literature in terms of the returns to education and age, the size of wage increases at promotion, the curvature of wages in job levels and tenure, and the implied magnitude of wage growth on the job. Along these dimensions, then, the firm I study is comparable to those in other work.

The estimates reveal key features of the wage process at the firm. By Table 5B, human capital acquired before entry in the firm, as captured by age and education at entry, has a significant impact on wages. For example, at Levels 1 and 2, the effect of an additional year of age, evaluated at the average age at entry of 29.71 years, is a 1.0 percent increase in annual (log) wages, since $\omega_1 + 2\omega_2(29.71) = 0.0102$, where $\omega_{1k} = \omega_1$, $\omega_{2k} = \omega_2$, and $\omega_{3k} = \omega_3$ are common across Levels 1 and 2; see (18) and Appendix B.7 for details. The effect of an additional year of schooling, $\omega_3$, is a 2.2 percent increase in wages. At Level 3, the corresponding figures are 0.4 percent ($\omega_{13} + 2\omega_{23}(29.71) = 0.0041$) and 2.1 percent ($\omega_{33} = 0.021$). These estimated effects are comparable to those in the literature over the same period. For instance, Belzil and Bognanno (2008, Table 1) estimate coefficients of 0.0127 and 0.0494 for the impact of age and education on (log) wages from a large multi-firm sample of U.S. executives between 1981 and 1988. Perhaps not surprisingly, since I focus on highly educated workers, my estimate of the marginal effect of education on wages is smaller than theirs.

The estimates of $\{\omega_{0ik}\}$ in Table 5A imply that promotions lead to sizable permanent wage increases, which are higher at higher levels: a promotion from Level 1 to 2 entails an increase in annual wages between \$590 (fourth skill type) and \$1,188 (first skill type), whereas a promotion from Level 2 to 3 entails an increase between \$4,638 (fourth skill type) and \$6,667 (second skill type).[39] Overall, average wages are convex in job levels, and so in tenure, since higher levels are reached only over time—see the first three entries in the first column of Table 6 for the model and the note to Table 3 for the data. This result mirrors the higher average output of managers with higher ability and human capital at higher levels. All these findings are in line with the literature on internal labor markets (Gibbons and Waldman (1999a,b), Belzil and Bognanno (2008), and Waldman (2013)).

As for the human capital acquired at the firm, the coefficient on tenure in Table 5B is quite small, which implies that the deterministic component of the human capital acquired at the firm, in contrast to that acquired before entry in the firm, has a limited *direct* effect on wages—acquired human capital still has an important *indirect* effect on wages through

[38] Comparative advantage at Level $k'$ over $k$ can be expressed as $e_{fk't}(i)/e_{fkt}(i) \geq [b_{k'}(h_1) + d_{fk't}(k_{t-1}, i)]/[b_k(h_1) + d_{fkt}(k_{t-1}, i)]$ by the definition after (3) and (17). As $b_k(h_1) + d_{Akt}(k_{t-1})$ does not increase fast with the job level, a sufficient condition is effectively $e_{Ak't} > e_{Akt}$.

[39] These increases are calculated by differencing across $k$ the estimates of $\{\omega_{0ik}\}$ once converted from logs to levels; say, $\$590 = e^{\omega_{042}} - e^{\omega_{041}}$. The estimated year-of-entry effects in Table A.6 are consistent with the 1973–1982 recessions, which depressed the wages of entrants in the firm in those years. I set $\omega_{ym} = 0$ for $0 \leq m \leq 3$, the reference years, and $\omega_{y4} = \omega_{y5}$ in accordance with the severity of those recession years.

its impact on promotions, as illustrated in Section 5.[40] In terms of ability, since the coefficients $\{\omega_{5i}\}$ on the prior in Table 5C are positive and significant, average wages increase with the prior so high-ability managers are indeed valued more than low-ability managers—recall that $\omega_{5kt}$ in (18) is estimated as part of the robustness exercise in Appendix B.11.[41]

**Decomposing Wage Growth.** The model implies an increase of 19.4 percent in average wages over the first seven years at the firm (18.5 percent in the unbalanced panel that includes separating managers) for an average yearly growth rate of 3.2 percent; see Table 6. Such growth is consistent with that in Topel (1991, Table 2) and the lower end of the range estimated by Buchinsky et al. (2010, Figure 2) from the Panel Study of Income Dynamics. It is also consistent with the estimates in Song et al. (2019) from U.S. social security data.[42] Note that by (18), the average wage of a manager of skill type $i$ at Level $k$ can be decomposed into four terms: *i)* $\omega(i, h_1, k)$, which captures the effect of unobserved skills, $i$, human capital at entry, $h_1$, and job assignment, $k$; *ii)* $\omega_{4kt} \times (t-1)$, which captures the *direct* effect of the deterministic component of the human capital acquired at the firm; *iii)* $\omega_{5i} p_{it}$, which captures the *direct* effect of learning about the systematic stochastic component of human capital, $\{a^j_{fkt}\}$, by (1); and *iv)* the (log) conditional assignment probability to Level $k$.[43] Over these first seven years, the term $\omega(i, h_1, k)$ accounts for more than 98 percent of the increase in wages. It would then seem that learning accounts for only a trivial percentage of wage growth, as is consistent with the literature.

But this inference is incorrect: the term $\omega(i, h_1, k)$ depends on the assigned job, $k$, which changes as information and human capital accumulate. In simple notation that suppresses $\{\varepsilon^j_{fkt}\}$, measuring the *average cumulative effect* of learning on wages requires accounting for how a change in $p_{it}$ affects the current job $k_t = k(p_{it}, \kappa_t)$ and wage $w_t = w(p_{it}, \kappa_t, k_t)$, which determine beliefs $p_{it+1} = p(p_{it}, k_t)$, (expected) human capital $\kappa_{t+1} = \kappa(p_{it}, \kappa_t, k_t)$, and the assigned job $k_{t+1} = k(p_{it+1}, \kappa_{t+1})$ next period. In turn, beliefs, human capital, and the assigned job next period influence the next-period wage $w_{t+1} = w(p_{it+1}, \kappa_{t+1}, k_{t+1})$ as well as beliefs, human capital, and the assigned job in the subsequent period, and so on. Since wages greatly differ across levels through $\omega(i, h_1, k)$, the belief process can have a large effect on wage growth through its *indirect* effect on the path of promotions, although its direct effect through $\omega_{5i}$ is small. As Section 5 shows, learning leads managers progressively revealed to be of high ability to advance to higher levels, at which they are paid higher wages. This previously neglected effect accounts for almost all of the impact of learning on wage growth.

**Instrumental-Variable Approach: A Discussion.** The result that the direct or *marginal contemporaneous effect* of learning on wages, captured here by $\omega_{5i} + \omega_{5kt}$, is small is consistent with the findings of Gibbons et al. (2005), Lluis (2005), and Hunnes (2012), who focus only on it and find it to be small or insignificant.[44] As I show next, their approach, though, relies critically on the assumption that jobs are equally informative about ability. Without it, their methods

[40] I specify $\omega_{41t} = \omega_{412} I(t < 4) + \omega_{414} I(t \geq 4)$ with $\omega_{414} = -\omega_{412}$ to conserve on parameters, given the much smaller fraction of managers assigned to Level 1 after the third year of tenure. Since the parameters $\{\omega_{4kt}\}$ proved to be negligible at Levels 2 and 3 and impossible to estimate with any precision, they are omitted from the baseline specification but estimated as part of the robustness exercise presented in Appendix B.11.

[41] The standard deviation of the wage shock $\sigma_{Aik}$ does not increase with the job level, $k$. Yet, the model implies that the standard deviation of wages is higher at higher levels, as in the data: \$6,936 at Level 1, \$7,077 at Level 2, and \$8,046 at Level 3; see the first column of Table 6 for the model and the note to Table 3 for the data counterparts. Thus, the predicted increase in wage dispersion with the job level is generated by the endogenous mechanisms of the model of learning, human capital acquisition, and job assignment, rather than by idiosyncratic unexplained factors.

[42] Song et al. (2019) estimate wage growth both for a year comparable to my sample, 1982, and for a more recent year. They find that in 1982, the median wages of employees with two to four years of tenure and with five or more years are, respectively, \$29,763 and \$45,010. Since my sample consists of managers, the corresponding median wages, \$42,493 and \$49,544, are naturally higher but are broadly consistent with theirs.

[43] The estimate of $\omega_{5kt}$, part of the robustness exercise in Appendix B.11, is relatively small. The role of the last term of (18) is also negligible.

[44] Replicating the analysis of Lluis (2005) and Hunnes (2012) on my data, I also found the impact of learning on wages to be insignificant.

cannot be applied. To see why, consider Lluis (2005), which is based on GW and is the paper closest in spirit to mine, in which workers can be of high or low ability with initial beliefs about it and performance normally distributed. Using the "rank dummy" $D_{ikt}$ for worker $i$'s rank (job, occupation, or sector) $k$ in $t$, worker $i$'s wage in rank $k$ in $t$ equals expected output up to measurement error, $w_{it}=\sum_{k=1}^{K} D_{ikt}[d_k+X_{it}\beta_k+c_k\theta^e_{it}f(x_{it})]+\mu_{it}$, where $X_{it}$ captures observable characteristics, $\theta^e_{it}$ is the mean of the posterior beliefs about ability, $f(\cdot)$ is the human capital function, $x_{it}$ is labor market experience, and $\mu_{it}$ is the error with $E(\mu_{it}|X_{i1},\ldots,X_{iT},D_{ik1},\ldots,D_{ikT},\theta_i)=0$. Firms are assumed to be identical but an analogous expression for $w_{it}$ is obtained if, for instance, expected output at each job differed across firms by a constant or all parameters differed across firms by a multiplicative factor common across jobs. Note that since the assigned rank depends on $\theta^e_{it}$, $\theta^e_{it}$ is correlated with $\{D_{ikt}\}$. But wages in $t$ and $t-1$ yield expressions for $\theta^e_{it}$ and $\theta^e_{it-1}$ that, once substituted into the martingale condition $\theta^e_{it}=\theta^e_{it-1}+u_{it}$, where $u_{it}$ is an i.i.d. shock, give

$$\frac{w_{it}}{\sum_k D_{ikt}c_k f(x_{it})}=\frac{\sum_k D_{ikt}(d_k+X_{it}\beta_k)}{\sum_k D_{ikt}c_k f(x_{it})}+\frac{w_{it-1}-\sum_k D_{ikt-1}(d_k+X_{it-1}\beta_k)}{\sum_k D_{ikt-1}c_k f(x_{it-1})}+\underbrace{u_{it}+\frac{\mu_{it}}{\sum_k D_{ikt}c_k f(x_{it})}-\frac{\mu_{it-1}}{\sum_k D_{ikt-1}c_k f(x_{it-1})}}_{e_{it}}. \tag{19}$$

In the complete information case ($\theta^e_{it}=\theta^e_{it-1}=\theta_i$), an expression similar to (19) can be obtained, just without the term $u_{it}$. Note that (19) is a random-coefficient panel-data model with unobserved individual time-varying effects, in which $D_{ikt}$ is endogenous, as it depends on $\theta^e_{it}$ and so is correlated with $e_{it}$ through $u_{it}$, and $w_{it-1}$ is correlated with $\mu_{it-1}$. Yet, (19) can be estimated by nonlinear instrumental-variable methods using interactions between $D_{ikt-1}$ and $D_{ikt-2}$ as instruments for $D_{ikt}$—and analogous interactions as instruments for $w_{it-1}$.[45] Lluis (2005, p. 751) finds that relative to the complete information case, the parameters of (19) are imprecisely estimated and their estimates are difficult to reconcile with economic intuition when the wage equation is estimated assuming that learning is present.

Now, when jobs are differentially informative, beliefs are updated differently depending on the assigned rank $D_{ikt}$, so the martingale condition becomes $\theta^e_{it}=\theta^e_{it-1}+\sum_k D_{ikt}u_{ikt}$ and the first component of the error term in (19) is $\sum_k D_{ikt}u_{ikt}$ rather than $u_{it}$. Past rank dummies are still valid instruments for $D_{ikt}$ in (19) if the jobs of *different* firms are equally informative.[46] However, when the speed of learning differs even by a constant amount across firms' jobs, the wage equation is no longer as in (19). In analogy to (9)—see the term $\ln\Pr(k_{Ct}=k_{At}|\cdot)$—an additional term for any difference in information across the jobs of different firms would be part of the error. Such a term depends on beliefs and so is correlated with past rank dummies. The same issue arises when the speed of learning varies across jobs, but not across firms, and firms differ in their rank productivity. In all these instances, the approach described is invalid.

# 5 The Role of Learning, Human Capital Acquisition, and Uncertainty

Here I assess counterfactually the role of learning, human capital acquisition, the differential informativeness of jobs, and persistent uncertainty about ability for the profiles of jobs and wages at the firm.

[45] To see why, observe that $D_{ikt-1}$ and $D_{ikt-2}$ are uncorrelated with both $u_{it}$—as they are functions of $u_{it-1}$ and $u_{it-2}$, which are independent of $u_{it}$ by assumption—and $\mu_{it}$, since $\mu_{it}$ is pure noise. I am grateful to Thomas Lemieux for helpful discussions.

[46] For instance, $D_{ikt-1}$ is uncorrelated with $e_{it}$ as $E[\sum_k D_{ikt}u_{ikt}|D_{ikt-1}]=0$ by the martingale condition, although $D_{ikt-1}$ is now dependent on the error term $e_{it}$, since $\sum_k D_{ikt}u_{ikt}$ is one of its components and $D_{ikt}$ and $D_{ikt-1}$ are correlated as they both depend on a worker's ability.

**Learning.** To evaluate the impact of learning on the dynamics of jobs and wages, I compare the implications of the baseline model with those of a *no learning* version of it, in which learning is absent as jobs are assumed to be uninformative about ability; namely, $\beta_k = \alpha_k$ in (4) for all $k$ so beliefs are not updated. Table 6 shows that learning generates faster wage growth and greater wage dispersion. Over the first seven years at the firm, the additional wage growth due to learning is a large portion of the wage growth without learning, 26 percent—that is, an increase from 15.4 to 19.4 relative to 15.4. The additional wage dispersion due to learning is 23 percent at Level 3—that is, an increase from \$6,534 to \$8,046 relative to \$6,534—with somewhat lower percentages at lower levels.[47] Job assignment is the key channel through which learning affects wages. Table A.7 displays the distribution of managers across levels by tenure in the baseline and no learning models. In the baseline model, by the third year of tenure, about 9 percent of managers have been promoted to Level 3, whereas in the no learning model, fewer than 1 percent have. By the fifth year, the fraction of managers at Level 3 is about 31 percent in the baseline model and about 16 percent in the no learning model. These results are consistent with the impact of learning on turnover documented by Nagypál (2007) and its effect on occupational mobility found by Papageorgiou (2014). Overall, learning implies more rapid promotions for managers progressively revealed to be of high ability and thus leads to higher wage growth and dispersion, since wages are higher and more variable at higher levels. So, the effect of learning on wages is sizable through its indirect impact on job assignment.

**Human Capital Acquisition.** Human capital has a sizable impact on wages. For instance, in the previous experiment, the progression of managers across levels is due only to the human capital they acquire at the firm. Acquired human capital improves managers' expected output, in particular at Levels 2 and 3, and so makes it profitable for the firm to eventually assign managers to these levels, at which they are paid higher and more dispersed wages. Over time, acquired human capital is also important to preventing demotions and thus reducing the wage loss that managers experience after low performance. These findings on the role of human capital are mirrored by the results of the experiment in which I simulate the model assuming that human capital acquisition does not take place at the firm. In this case, almost no promotion occurs, when promotions do occur, they are followed by demotions, virtually no wage growth accrues, and wages in any tenure are nearly as dispersed as at entry in the firm. Since accumulated human capital has a limited direct effect on wages, as implied by the small estimated wage coefficients on tenure $\{\omega_{4kt}\}$, these results imply that, much like learning, human capital affects wages primarily indirectly, through its impact on managers' job paths at the firm.

**Learning by Experimentation.** Here I focus on an exercise, referred to as *no experimentation*, in which all jobs are assumed to be equally informative about ability in that Levels 2 and 3 are made as informative as Level 1 by setting $\alpha_k$ and $\beta_k$, $k \geq 2$, equal to their estimated values at Level 1 in the belief-updating rules in (4), whereas all other parameters are unchanged. Results for the analogous experiments of equal informativeness as Levels 2 and 3 are reported in the S.A. (Tables A.11-A.13). Table A.8 shows that without experimentation, nearly all managers who do not separate from the firm are quickly promoted to Level 3. In particular, after the first two years of tenure, the proportion of managers assigned

[47] When $\beta_k = \alpha_k$, I simulate the model assuming that the initial prior is never updated but managers of each ability experience high and low performance with the same probabilities as in the baseline model. I follow the same approach for the remaining experiments that involve $\{\alpha_k, \beta_k\}$.

to Level 1 is very small, although learning takes place slowly over time. Experimentation proves to be quantitatively important not just for job mobility but also for wage growth. Indeed, fast promotions are accompanied by rapid wage growth, as Table 6 shows. Without experimentation, by the third year of tenure, wage growth is 17.6 percent, nearly twice as large as in the baseline model. By the fourth year, managers experience a higher wage growth (20.5 percent) than over the first seven years in the baseline model (19.4 percent). After that, wage growth is fairly flat. Overall, cumulative wage growth during the first seven years at the firm would be 20 percent higher—that is, an increase from 19.4 to 23.3 relative to 19.4. Wages at each level are only slightly less dispersed than in the baseline model. Thus, the differential informativeness of jobs is critical to accounting for the convexity of wages in job levels and tenure, as it leads to a much more gradual wage growth in early tenures than if jobs were equally informative.

**Persistent Uncertainty.** Since I estimate learning to be a gradual process, I conduct two experiments to assess the role of the persistence of uncertainty about ability. In the *fast learning at Level 1* case, jobs at Level 1 are made to be nearly perfectly informative about ability, with $\alpha_1 = 0.99$ and $\beta_1 = 0.01$, so that ability is (almost) fully learned after one period, whereas the other parameters are fixed at their baseline values (see Crawford and Shum (2005) for a similar approach). Table 6 shows the implications for wages and Table A.7 for job assignments. During the first seven years at the firm, faster learning at Level 1 leads wages to grow more than 60 percent, compared with approximately 20 percent in the baseline model. Indeed, if managers could sort to the best jobs given their ability and human capital after just one year at the firm, wage growth over the first two years (39.3 percent) would be more than twice as large as the cumulative wage growth over the first seven years in the baseline model (19.4). Wage dispersion at each level also increases: the standard deviation of wages at Level 3 is over five times larger than in the baseline model. Moreover, as Table A.7 illustrates, promotions to Level 3 occur more rapidly. By the third year of tenure, 20 percent of managers are already at Level 3. Hence, persistent uncertainty about ability, like learning through experimentation, substantially compresses wage growth especially in early years, thus helping account for the curvature of wages in job levels and tenure, and reduces wage dispersion. The case of *fast learning at Level 2* is similar; see Appendix B.12 for details.

## 6 Conclusion

This paper estimates a model of learning about ability, human capital acquisition with experience, and job assignment to account for rich patterns of careers in firms. Most of the literature has argued that the impact of learning on wages is small but has investigated only the *direct* effect of learning due to the impact of current beliefs about ability on current wages. By explicitly estimating the joint dynamics of beliefs, human capital, jobs, and wages, I can evaluate the total effect of learning on wages, which consists of both a direct and an *indirect* effect. Intuitively, learning also affects wages indirectly by influencing the dynamic selection process that leads managers who are progressively revealed to be of higher ability to be promoted to higher levels of a firm's job hierarchy. Since wages at higher job levels are higher and more variable, by stimulating promotions, learning critically contributes to wage growth and dispersion. I estimate that this indirect effect is large and accounts for almost all of the impact of learning on wages. However, as these results are based on data on one firm, they should be subject to further investigation to assess their validity. These findings nonetheless attest to the potential of learning and human capital models to account for salient features of careers in firms.

# References


**Al-Hussaini, E., and K. Ahmad** (1981): "On the Identifiability of Finite Mixtures of Distributions," *IEEE Transactions on Information Theory* 27(5), 664–668.

**Altonji, J., and C. Pierret** (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics* 116(1), 313–350.

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

**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(6), 1551–1596.

**Baker, G., M. Gibbs, and B. Holmström** (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics* 109(4), 881–919.

— (1994b): "The Wage Policy of a Firm," *Quarterly Journal of Economics* 109(4), 921–955.

**Baker, G., and B. Holmström** (1995): "Internal Labor Markets: Too Many Theories, Too Few Facts," *American Economic Review* 85(2), 255–259.

**Baker, G., M.C. Jensen, and K.J. Murphy** (1988): "Compensation and Incentives: Practice vs. Theory," *Journal of Finance* 48(3), 593–616.

**Belzil, C., and B. Bognanno** (2008): "Promotions, Demotions, Halo Effects, and the Earnings Dynamics of American Executives," *Journal of Labor Economics* 26(2), 287–310.

**Bergemann, D., and J. Välimäki** (1996): "Learning and Strategic Pricing," *Econometrica* 64(5), 1125–1149.

**Berkovec, J., and S. Stern** (1991): "Job Exit Behavior of Older Men," *Econometrica* 59(1), 189–210.

**Bernhardt, D.** (1995): "Strategic Promotion and Compensation," *Review of Economic Studies* 62(2), 315–339.

**Bernhardt, D., and D. Scoones** (1993): "Promotion, Turnover, and Preemptive Wage Offers," *American Economic Review* 83(4), 771–791.

**Buchinsky, M., D. Fougère, F. Kramarz, and R. Tchernis** (2010): "Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States," *Review of Economic Studies* 77(3), 972–1001.

**Cameron, S.V., and J.J. Heckman** (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy* 106(2), 262–333.

— (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy* 109(3), 455–499.

**Chiappori, P.-A., B. Salanié, and J. Valentin** (1999): "Early Starters versus Late Beginners," *Journal of Political Economy* 107(4), 731–760.

**Cobb-Clark, D.A., and Y. Dunlop** (1999): "The Role of Gender in Job Promotions," *Monthly Labor Review* 122(12), 32–38.

**Crawford, G.S., and M. Shum** (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica* 73(4), 1137–1173.

**Davis, S.** (1997): "Sorting, Learning, and Mobility When Jobs Have Scarcity Value: A Comment," *Carnegie-Rochester Conference Series on Public Policy* 46(1), 327–337.

**DeVaro, J., and M. Waldman** (2012): "The Signaling Role of Promotions: Further Theory and Empirical Evidence," *Journal of Labor Economics* 30(1), 91–147.

**Ekinci, E., A. Kauhanen, and M. Waldman** (2019): "Bonuses and Promotion Tournaments: Theory and Evidence," *Economic Journal* 129(622), 2342–2389.

**Farber, H.S., and R. Gibbons** (1996): "Learning and Wage Dynamics," *Quarterly Journal of Economics* 111(4), 1007–1047.

**Ferrall, C.** (1997): "Empirical Analysis of Occupational Hierarchies," *Journal of Human Resources* 32(1), 1–34.

**Ferrall, C., K.G. Salvanes, and E.Ø. Sørensen** (2009): "Wages and Seniority When Coworkers Matter: Estimating a Joint Production Economy Using Norwegian Administrative Data," Working Paper 1200, QED Working Paper 1200.

**Flinn, C.J.** (1986): "Wages and Job Mobility of Young Workers," *Journal of Political Economy* 94(3), 88–110.

**Frederiksen, A., B. Kriechel, and F. Lange** (2017): "Subjective Performance Evaluations and Careers," *Journal of Economic Behavior and Organization* 134, 408–429.

**Garicano, L., and T. Van Zandt** (2013): "Hierarchies and the Division of Labor," in *Handbook of Organizational Economics*, R. Gibbons and J. Roberts, Eds., Princeton, NJ: Princeton University Press, 604–654.

**Gayle, G.-L., L. Golan, and R.A. Miller** (2015): "Promotion, Turnover, and Compensation in the Executive Labor Market," *Econometrica* 83(6), 2293–2369.

**Ghosh, S.** (2007): "Job Mobility and Careers in Firms," *Labour Economics* 14(3), 603–621.

**Gibbons, R., L.F. Katz, T. Lemieux, and D. Parent** (2005): "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics* 23(4), 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(4), 1321–1358.

— (2006): "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," *Journal of Labor Economics* 24(1), 59–107.

**Gibbs, M.** (1995): "Incentive Compensation in a Corporate Hierarchy," *Journal of Accounting and Economics* 19(2–3), 247–277.

**Harris, M., and B. Holmström** (1982): "A Theory of Wage Dynamics," *Review of Economic Studies* 49(3), 315–333.

**Heckman, J.J.** (1979): "Sample Selection Bias as a Specification Error," *Econometrica* 47(1), 153–161.

**Heckman, J.J.** (1981): "The Incidental Parameters Problem and the Problem of Initial Condition in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data and Econometric Applications*, C.F. Manski and D.L. McFadden, Eds., Boston: MIT Press, 179–195.

**Heckman, J.J.** (1990): "Varieties of Selection Bias," *American Economic Review* 80(2), 313–318.

**Heckman, J.J., and B.E. Honoré** (1990): "The Empirical Content of the Roy Model," *Econometrica* 58(5), 1121–1149.

**Heckman, J.J., L. Lochner, and C. Taber** (1998): "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics* 1(1), 1–58.

**Heckman, J.J., and G.L. Sedlacek** (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy* 93(6), 1077–1125.

— (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics* 8(1), S329–S363.

**Heckman, J.J., and B. Singer** (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52(2), 271–320.

**Holmström, B., and J. Tirole** (1989): "The Theory of the Firm," in *Handbook of Industrial Organization*, Volume I, R. Schmalensee and R. Willig, Eds., Amsterdam: North-Holland, 61–133.

**Hu, Y., and M. Shum** (2012): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," *Journal of Econometrics* 171(1), 32–44.

**Hunnes, A.** (2012): "Testing the Role of Comparative Advantage and Learning in Wage and Promotion Dynamics," *International Journal of Manpower* 33(5), 556–582.

**Hunter, D.R., S. Wang, and T.P. Hettmansperger** (2007): "Inference for Mixtures of Symmetric Distributions," *Annals of Statistics* 35(1), 224–251.

**Jovanovic, B.** (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy* 87(5), 972–990.

**Jovanovic, B., and Y. Nyarko** (1997): "Stepping-Stone Mobility," *Carnegie-Rochester Conference Series on Public Policy* 46(1), 289–325.

**Kahn, L., and F. Lange** (2014): "Employer Learning, Productivity, and the Earnings Distribution: Evidence from Performance Measures," *Review of Ecomomic Studies* 81(4), 1575–1613.

**Keane, M.P., and R.M. Sauer** (2009): "Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior," *Econometrica* 77(3), 975–991.

**Keane, M.P., and K.I. Wolpin** (1997): "The Career Decisions of Young Men," *Journal of Political Economy* 105(3), 473–522.

**Lazear, E.** (1992): "The Job as a Concept," in *Performance Measurement, Evaluations, and Incentives*, W. Bruns, Ed., Boston: Harvard Business School Press, 183–215.

**Lazear, E.** (2009): "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy* 117(5), 914–940.

**Lazear, E., and S. Rosen** (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy* 89(5), 841–864.

**Lazear, E., and K.L. Shaw** (2007): "Personnel Economics: The Economist's View of Human Resources," *Journal of Economic Perspectives* 21(4), 91–114.
**Lazear, E., K.L. Shaw, and C.T. Stanton** (2015): "The Value of Bosses," *Journal of Labor Economics* 33(4), 823–861.
**Lindsay, B.G.** (1995): "Mixture Models: Theory, Geometry and Applications," *NSF-CBMS Regional Conference Series in Probability and Statistics*, 5.
**Lluis, S.** (2005): "The Role of Comparative Advantage and Learning in Wage Dynamics and Intrafirm Mobility: Evidence from Germany," *Journal of Labor Economics* 23(4), 725–767.
**Lochner, L., Y. Park, and Y. Shin** (2018): "Wage Dynamics and Returns to Unobserved Skill," NBER Working Paper 24220.
**Lochner, L., and Y. Shin** (2014): "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks," NBER Working Paper 20068.
**MacDonald, G.** (1982): "A Market Equilibrium Theory of Job Assignment and Sequential Accumulation of Information," *American Economic Review* 72(5), 1038–1055.
**Miller, R.A.** (1984): "Job Matching and Occupational Choice," *Journal of Political Economy* 92(6), 1086–1120.
**Moscarini, G.** (2005): "Job Matching and the Wage Distribution," *Econometrica* 73(2), 481–516.
**Murphy, K.J.** (1992): "Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance," in *Performance Measurement, Evaluations, and Incentives*, W. Bruns, Ed., Boston: Harvard Business School Press, 37–62.
**Nagypál, É.** (2007): "Learning by Doing vs. Learning about Match Quality: Can We Tell Them Apart?," *Review of Economic Studies* 74(2), 537–566.
**Neal, D.** (2017): "Life Cycle Wage Dynamics and Labor Mobility," *Journal of Political Economy* 125(6), 1853–1862.
**Neal, D., and S. Rosen** (2000): "Theories of the Distribution of Earnings," in *Handbook of Income Distribution*, Volume 1, A.B. Atkinson and E. Bourguignon, Eds., Amsterdam: North-Holland, 379–427.
**Oyer, P., and S. Shaefer** (2011): "Personnel Economics: Hiring and Incentives," in *Handbook of Labor Economics*, Volume IVb, O. Ashenfelter and D. Card, Eds., Amsterdam: North-Holland, 1769–1823.
**Papageorgiou, T.** (2014): "Learning Your Comparative Advantages," *Review of Economic Studies* 81(3), 1263–1295.
— (2018): "Large Firms and Within Firm Occupational Reallocation," *Journal of Economic Theory* 174, 184–223.
**Prendergast, C.** (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature* 37(1), 7–63.
**Prescott, E. C., and M. Visscher** (1980): "Organization Capital," *Journal of Political Economy* 88(3), 446–461.
**Pries, M., and R. Rogerson** (2005): "Hiring Policies, Labor Market Institutions, and Labor Market Flows," *Journal of Political Economy* 113(4), 811–839.
**Rosen, S.** (1986): "Prizes and Incentives in Elimination Tournaments," *American Economic Review* 76(4), 701–715.
**Rubinstein, Y., and Y. Weiss** (2006): "Post Schooling Wage Growth: Investment, Search and Learning," in *Handbook of the Economics of Education*, Volume 1, E.A. Hanushek and F. Welch, Eds., Amsterdam: North-Holland, 1–67.
**Rust, J.** (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, Volume 4, R. Engle and D. McFadden, Eds., Amsterdam: North-Holland, 3081–3143.
**Sanders, C.** (2016): "Skill Uncertainty, Skill Accumulation, and Occupational Choice," mimeo.
**Sauer, R.M.** (1998): "Job Mobility and the Market for Lawyers," *Journal of Political Economy* 106(1), 147–171.
**Shi, Z., T. Zheng, and J. Han** (2014): "Identifiability of Multivariate Logistic Mixture Models," *ArXiv e-prints*.
**Song, J., D.J. Price, F. Guvenen, N. Bloom, and T. von Wachter** (2019): "Firming Up Inequality," *Quarterly Journal of Economics* 134(1), 1–50.
**Topel, R.** (1991): "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy* 99(1), 145–176.
**Topel, R., and M.P. Ward** (1992): "Job Mobility and the Careers of Young Men," *Quarterly Journal of Economics* 107(2), 439-479.
**Waldman, M.** (1984): "Job Assignments, Signalling, and Efficiency," *RAND Journal of Economics* 15(2), 255–267.
— (1996): "Asymmetric Learning and the Wage/Productivity Relationship," *Journal of Economic Behavior & Organization* 31(3), 419–429.
— (2013): "Theory and Evidence in Internal Labor Markets," in *Handbook of Organizational Economics*, R. Gibbons and J. Roberts, Eds., Princeton, NJ: Princeton University Press, 520–572.

Table 1: Percentage Distribution of Managers across Levels by Tenure

| | Separation | | Level 1 | | Level 2 | | Level 3 | |
|---|---|---|---|---|---|---|---|---|
| Tenure | Data | Model | Data | Model | Data | Model | Data | Model |
| 1 | 0.0 | 0.0 | 100.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 14.5 | 14.5 | 45.6 | 45.7 | 39.9 | 39.8 | 0.0 | 0.0 |
| 3 | 27.7 | 26.5 | 16.8 | 17.2 | 46.8 | 47.3 | 8.7 | 8.9 |
| 4 | 37.9 | 37.1 | 7.6 | 8.1 | 29.2 | 29.2 | 25.3 | 25.6 |
| 5 | 46.1 | 45.3 | 4.6 | 5.3 | 18.1 | 18.3 | 31.1 | 31.2 |
| 6 | 52.0 | 51.5 | 2.9 | 3.4 | 12.3 | 12.6 | 32.8 | 32.5 |
| 7 | 57.6 | 56.9 | 2.1 | 2.7 | 7.7 | 8.3 | 32.6 | 32.1 |

Table 2: Percentage of High Ratings at Levels 1 and 2

| | Level 1 | | Level 2 | |
|---|---|---|---|---|
| Tenure | Data | Model | Data | Model |
| 1 | 52.7 | 51.7 | - | - |
| 2 | 34.8 | 34.9 | 58.4 | 56.1 |
| 3 | 19.6 | 21.2 | 43.9 | 42.7 |
| 4 | 11.8 | 11.9 | 26.2 | 30.4 |
| 5 | 2.4 | 6.2 | 18.7 | 20.4 |
| 6 | 3.7 | 3.2 | 12.5 | 13.0 |
| 7 | 0.0 | 1.6 | 13.0 | 8.0 |

Table 3: Percentage Wage Distribution by Level and Tenure*

| | | Between \$20K and \$40K | | Between \$40K and \$60K | | Between \$60K and \$80K | |
|---|---|---|---|---|---|---|---|
| Level | Tenure | Data | Model | Data | Model | Data | Model |
| Level 1 | 1 | 59.1 | 57.5 | 40.5 | 42.0 | 0.4 | 0.5 |
| | 2 | 54.4 | 57.4 | 44.8 | 41.7 | 0.8 | 0.8 |
| | 3 | 55.6 | 56.6 | 44.4 | 42.2 | 0.0 | 1.2 |
| | 4 | 53.8 | 56.7 | 46.2 | 41.9 | 0.0 | 1.4 |
| | 5 | 64.1 | 68.1 | 35.9 | 31.1 | 0.0 | 0.7 |
| | 6 | 69.2 | 69.5 | 30.8 | 29.8 | 0.0 | 0.7 |
| | 7 | 75.0 | 70.6 | 25.0 | 28.5 | 0.0 | 0.8 |
| Level 2 | 2 | 35.1 | 34.3 | 63.3 | 63.8 | 1.6 | 1.8 |
| | 3 | 31.3 | 36.1 | 65.7 | 61.7 | 2.9 | 2.2 |
| | 4 | 36.3 | 36.9 | 60.3 | 60.5 | 3.4 | 2.5 |
| | 5 | 37.1 | 37.5 | 59.6 | 59.8 | 3.3 | 2.7 |
| | 6 | 42.4 | 38.1 | 53.3 | 59.0 | 4.2 | 2.9 |
| | 7 | 41.3 | 38.8 | 55.8 | 58.2 | 2.9 | 3.0 |
| Level 3 | 3 | 2.8 | 8.2 | 84.9 | 82.4 | 12.3 | 9.3 |
| | 4 | 4.5 | 10.3 | 85.3 | 80.5 | 10.1 | 9.2 |
| | 5 | 5.3 | 11.3 | 84.2 | 79.1 | 10.5 | 9.5 |
| | 6 | 6.1 | 12.4 | 84.4 | 77.6 | 9.5 | 9.9 |
| | 7 | 4.5 | 13.4 | 81.1 | 76.4 | 14.4 | 10 |

*Mean of wages across tenures: $39,584 at Level 1, $43,179 at Level 2, and $48,963 at Level 3. Standard deviation of wages across tenures: $6,924 at Level 1, $7,377 at Level 2, and $7,270 at Level 3.

Table 4A: Estimates of Prior Distribution*

| Parameters | Type 1 ($i = 1$) | Type 2 ($i = 2$) | Type 3 ($i = 3$) | Type 4 ($i = 4$) |
|---|---|---|---|---|
| Prior: $\phi_{i1}$ | -0.672 | -0.484 | -0.141 | 0.435 |
| | (0.022) | (0.021) | (0.017) | (0.022) |
| Mass: $q_i$ | 0.155 | 0.211 | 0.313 | 0.321 |
| | (0.017) | (0.030) | (0.076) | (NA) |

*I estimate $\phi_{i1} = \ln[p_{i1}/(1-p_{i1})]$, which ranges over the real line and implies that $p_{11} = 0.338$, $p_{21} = 0.381$, $p_{31} = 0.465$, and $p_{41} = 0.607$ with an average initial prior 0.473 and standard deviation 0.102. Asymptotic standard errors in parentheses.

Table 4B: Estimates of Probability of High Output*

| Parameters | Level 1 ($k = 1$) | Level 2 ($k = 2$) | Level 3 ($k = 3$) |
|---|---|---|---|
| High Ability: $\alpha_k$ | 0.514 | 0.5437 | 0.5435 |
| | (0.062) | (0.006) | (0.007) |
| Low Ability: $\beta_k$ | 0.456 | 0.491 | 0.490 |
| | (0.014) | (0.013) | (0.010) |

*Asymptotic standard errors in parentheses.

Table 5A: Estimates of Type-Specific Intercept in the Wage Equation*

| Parameters | Type 1 ($i = 1$) | Type 2 ($i = 2$) | Type 3 ($i = 3$) | Type 4 ($i = 4$) |
|---|---|---|---|---|
| Level 1: $\omega_{0i1}$ | 8.805 | 9.288 | 9.213 | 8.865 |
| | (0.005) | (0.005) | (0.011) | (0.013) |
| Level 2: $\omega_{0i2}$ | 8.969 | 9.359 | 9.281 | 8.945 |
| | (0.004) | (0.004) | (0.009) | (0.012) |
| Level 3: $\omega_{0i3}$ | 9.534 | 9.813 | 9.738 | 9.418 |
| | (0.008) | (0.004) | (0.007) | (0.011) |

*Asymptotic standard errors in parentheses.

Table 5B: Estimates of Coefficients on Age, Education, and Tenure in the Wage Equation

| Parameters | Value | St. Error |
|---|---|---|
| Levels 1-2: $\omega_1$ (age) | 0.028 | 0.0001 |
| Levels 1-2: $\omega_2$ (age$^2$) | -0.0003 | 0.000002 |
| Levels 1-2: $\omega_3$ (education) | 0.022 | 0.0004 |
| Level 1: $\omega_{412}$ (tenure) | 0.007 | 0.0003 |
| Level 3: $\omega_{13}$ (age) | 0.010 | 0.001 |
| Level 3: $\omega_{23}$ (age$^2$) | -0.0001 | 0.00001 |
| Level 3: $\omega_{33}$ (education) | 0.021 | 0.001 |

Table 5C: Estimates of Coefficients on Prior and Standard Deviations in the Wage Equation*

| Parameters | Type 1 ($i = 1$) | Type 2 ($i = 2$) | Type 3 ($i = 3$) | Type 4 ($i = 4$) |
|---|---|---|---|---|
| Prior: $\omega_{5i}$ | 2.371 | 1.833 | 1.316 | 1.364 |
| | (0.045) | (0.027) | (0.015) | (0.010) |
| Level 1: $\sigma_{Ai1}$ | 0.076 | 0.070 | 0.057 | 0.044 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Level 2: $\sigma_{Ai2}$ | 0.063 | 0.047 | 0.0302 | 0.0302 |
| | (0.001) | (0.001) | (0.0004) | (0.0004) |
| Level 3: $\sigma_{Ai3}$ | 0.047 | (as $i = 1$) | (as $i = 1$) | (as $i = 1$) |
| | (0.0004) | (NA) | (NA) | (NA) |

*Asymptotic standard errors in parentheses.

Table 6: Counterfactual Experiments on Importance of Learning for Wages

| Statistic | Wages | | | | |
|---|---|---|---|---|---|
| | Baseline | No Learning | No Experimentation | Fast Learning at Level 1 | Fast Learning at Level 2 |
| Means by Level | | | | | |
| Level 1 | \$39,584 | \$39,706 | \$39,763 | \$58,271 | \$37,847 |
| Level 2 | 43,179 | 43,070 | 42,600 | 61,451 | 77,503 |
| Level 3 | 48,963 | 48,454 | 48,818 | 44,623 | 24,360 |
| St. Dev. by Level | | | | | |
| Level 1 | \$6,936 | \$6,791 | \$6,902 | \$35,961 | \$8,668 |
| Level 2 | 7,077 | 6,464 | 6,831 | 51,466 | 45,057 |
| Level 3 | 8,046 | 6,534 | 7,971 | 45,784 | 4,281 |
| Cumulative Growth | | | | | |
| Tenure 2 | 4.60% | 3.30% | 0.90% | 39.30% | 8.80% |
| Tenure 3 | 8.9 | 6.8 | 17.6 | 48.5 | 51.5 |
| Tenure 4 | 13.8 | 9.8 | 20.5 | 52.8 | 50.1 |
| Tenure 5 | 15.9 | 11.1 | 21.6 | 55.4 | 50.3 |
| Tenure 6 | 17.5 | 12.9 | 22.1 | 58.1 | 50.1 |
| Tenure 7 | 18.5 | 14.6 | 22.2 | 60.6 | 49.8 |
| Tenure 7 (Balanced Panel) | 19.4 | 15.4 | 23.3 | 62.5 | 51.2 |

*No Learning: $\beta_k = \hat{a}_k$, $k = 1, 2, 3$. Fast Learning at Level $k$: $\alpha_k = 0.99$ and $\beta_k = 0.01$, $k = 1, 2$.

# A Omitted Proofs and Derivations

**Proof of Proposition 1**: There are two immediate implications of equilibrium. First, the worker chooses the firm, $f$, offering the highest value of wages; that is, $W(s_t,\varepsilon_t|f) \geq W(s_t,\varepsilon_t|f')$. Second, the employing firm, $f$, offers wages just sufficient to attract the worker; that is, it achieves the value $\Pi^f(s_t,\varepsilon_t|f)$, subject to $W(s_t,\varepsilon_t|f) \geq W(s_t,\varepsilon_t|f')$, with

$$W(s_t,\varepsilon_t|f) = W(s_t,\varepsilon_t|f') \tag{20}$$

by profit maximization. Thus, the worker is indifferent between working at firm $f$ and working at firm $f'$. I refer to (20) as the worker's equilibrium indifference between the offers of firms $A$ and $C$ or, simply, *worker indifference*.

I first show that a firm's employment and job assignment problem in equilibrium reduces to an autarky-type problem—that is, the problem of choosing the assignment for the worker that maximizes the firm's value of output, $\overline{V}^f(s_t,\varepsilon_t)$, $f = A, C$. The argument is as follows. By (20), it is immediate that maximizing profits for a firm is equivalent to maximizing the sum of its own value and the worker's value: the job that maximizes $\Pi^f(s_t,\varepsilon_t)$ clearly also maximizes $\Pi^f(s_t,\varepsilon_t)+W(s_t,\varepsilon_t|f')$, $f' \neq f$, since firm $f$ takes as given the value of wages implied by the offer of firm $f'$. Combining (6), this result, and optimality for firm $A$ implies that $V^A(s_t,\varepsilon_t)$, defined as $\Pi^A(s_t,\varepsilon_t) + W(s_t,\varepsilon_t)$, can be expressed as

$$\begin{aligned} V^A(s_t,\varepsilon_t) = \max\{V^A(s_t,\varepsilon_t|A), V^A(s_t,\varepsilon_t|C)\} = \max\Big\{&\max_{k\in K^A}\{y_A(s_t,k) + \varepsilon_{Akt} \\ &+\delta\eta_{Ak}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^A(s_{t+1},\varepsilon_{t+1}|s_t,k)dG\Big\}, w_C(s_t,\varepsilon_t)+\delta\eta_{Ck_{Ct}}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^A(s_{t+1},\varepsilon_{t+1}|s_t,k_{Ct})dG\Big\}, \end{aligned} \tag{21}$$

since firm $A$'s current profits are zero and the worker's wage is $w_C(s_t,\varepsilon_t)$ if firm $C$ employs the worker at state $(s_t,\varepsilon_t)$, in which case the state is updated conditional on the job the worker performs at firm $C$. Now consider states at which firm $A$ employs the worker. At these states, firm $A$ must prefer employing to not employing the worker—that is, $\Pi^A(s_t,\varepsilon_t|A) \geq \Pi^A(s_t,\varepsilon_t|C)$—and must also pay a wage sufficiently low to make the worker just indifferent between working for $A$ and working for $C$; that is, $W(s_t,\varepsilon_t|A) = W(s_t,\varepsilon_t|C)$. Otherwise, $A$ could lower its wage offer and still attract the worker. These two facts imply that when $A$ employs the worker, $\Pi^A(s_t,\varepsilon_t|A)+W(s_t,\varepsilon_t|A) \geq \Pi^A(s_t,\varepsilon_t|C)+W(s_t,\varepsilon_t|C)$; that is, $V^A(s_t,\varepsilon_t|A) \geq V^A(s_t,\varepsilon_t|C)$ by the definition of match surplus value.

Next, consider states at which firm $C$ employs the worker. Condition (7) requires that firm $A$ be indifferent between employing and not employing the worker; that is, $\Pi^A(s_t,\varepsilon_t|A) = \Pi^A(s_t,\varepsilon_t|C)$. By the same argument as before, firm $C$ must pay a wage that makes the worker just indifferent between working for $A$ and working for $C$. These observations imply that $V^A(s_t,\varepsilon_t|A) = V^A(s_t,\varepsilon_t|C)$.

Since $V^A(s_t,\varepsilon_t) = \max\{V^A(s_t,\varepsilon_t|A), V^A(s_t,\varepsilon_t|C)\}$, this argument implies that at all states,

$$V^A(s_t,\varepsilon_t) = V^A(s_t,\varepsilon_t|A) = \max_{k\in K^A}\Big\{y_A(s_t,k) + \varepsilon_{Akt}+\delta\eta_{Ak}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^A(s_{t+1},\varepsilon_{t+1}|s_t,k)dG\Big\}. \tag{22}$$

Thus, $V^A(s_t,\varepsilon_t)$ reduces to the value of firm $A$'s problem of choosing the job that maximizes the value of its output, $\overline{V}^A(s_t,\varepsilon_t)$, defined recursively as

$$\overline{V}^A(s_t,\varepsilon_t) = \max_{k\in K^A}\Big\{y_A(s_t,k) + \varepsilon_{Akt}+\delta\eta_{Ak}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} E\overline{V}^A(s_{t+1},\varepsilon_{t+1}|s_t,k)dG\Big\}.$$

A similar argument applies to firm $C$, so that $V^C(s_t,\varepsilon_t)$ also reduces to $\overline{V}^C(s_t,\varepsilon_t)$. Hence,

$$V^f(s_t,\varepsilon_t) = V^f(s_t,\varepsilon_t|f) = \overline{V}^f(s_t,\varepsilon_t), \quad f = A, C. \tag{23}$$

I now turn to deriving (8). As just proved, when firm $A$ employs the worker at state $(s_t,\varepsilon_t)$,

$$V^A(s_t,\varepsilon_t|A) = \Pi^A(s_t,\varepsilon_t|A) + W(s_t,\varepsilon_t|A) \geq V^A(s_t,\varepsilon_t|C) = \Pi^A(s_t,\varepsilon_t|C) + W(s_t,\varepsilon_t|C).$$

Since $\Pi^C(s_t,\varepsilon_t|A) = \Pi^C(s_t,\varepsilon_t|C)$ by (7), it follows that $V^A(s_t,\varepsilon_t|A) \geq V^A(s_t,\varepsilon_t|C)$ further implies

$$\Pi^A(s_t,\varepsilon_t|A) + W(s_t,\varepsilon_t|A) + \Pi^C(s_t,\varepsilon_t|A) \geq \Pi^A(s_t,\varepsilon_t|C) + W(s_t,\varepsilon_t|C) + \Pi^C(s_t,\varepsilon_t|C). \tag{24}$$

Similarly, as just argued, when firm $C$ employs the worker at state $(s_t, \varepsilon_t)$,

$$V^A(s_t, \varepsilon_t|C) = \Pi^A(s_t, \varepsilon_t|C) + W(s_t, \varepsilon_t|C) = \Pi^A(s_t, \varepsilon_t|A) + W(s_t, \varepsilon_t|A) = V^A(s_t, \varepsilon_t|A), \quad (25)$$

whereas it must be that $\Pi^C(s_t, \varepsilon_t|C) \geq \Pi^C(s_t, \varepsilon_t|A)$ for firm $C$, since it is the employing firm. Thus, (25) implies

$$\Pi^A(s_t, \varepsilon_t|C) + W(s_t, \varepsilon_t|C) + \Pi^C(s_t, \varepsilon_t|C) \geq \Pi^A(s_t, \varepsilon_t|A) + W(s_t, \varepsilon_t|A) + \Pi^C(s_t, \varepsilon_t|A). \quad (26)$$

Then, (24) and (26) prove that the employing firm is selected in equilibrium to maximize the sum of the values of the two firms and the worker, conditional on the two firms' job offers, since by definition, $\Pi^f(\cdot)$ and $W(\cdot)$ are the values of profits and wages, respectively, at the equilibrium job offers $k_{ft} = k_f(s_t, \varepsilon_t)$, $f = A, C$. Equivalently, the employing firm is determined by the policy of the program with value (8), where $S(\cdot) = \Pi^A(\cdot) + W(\cdot) + \Pi^C(\cdot)$. $\square$

**Proof of Proposition 2**: Note that the indifference condition in (20) can be rewritten as

$$w_A(s_t, \varepsilon_t) = W(s_t, \varepsilon_t|C) - \delta\eta_{Ak_{At}}(\kappa_t) \int_{\varepsilon_{t+1}} EW(\cdot|s_t, k_{At})dG. \quad (27)$$

Suppose, without loss, that $A$ employs the worker. Then, (7) can be expressed as

$$\begin{aligned} W(s_t, \varepsilon_t|C) &= V^C(s_t, \varepsilon_t|C) - \delta\eta_{Ak_{At}}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} E\Pi^C(\cdot|s_t, k_{At})dG \\ &= \max_{k\in K^C}\left\{y_C(s_t, k) + \varepsilon_{Ckt} + \delta\eta_{Ck}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^C(\cdot|s_t, k)dG\right\} - \delta\eta_{Ak_{At}}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} E\Pi^C(\cdot|s_t, k_{At})dG, \end{aligned}$$

where the second equality follows from (23). By substituting this expression into (27), using that $EV^C(\cdot|s_t, k_{At}) = E\Pi(\cdot|s_t, k_{At}) + EW^C(\cdot|s_t, k_{At})$ by definition of $V(\cdot)$, and combining terms, (27) can be rewritten as

$$w_A(s_t, \varepsilon_t) = \max_{k\in K^C}\left\{y_C(s_t, k) + \varepsilon_{Ckt} + \delta\eta_{Ck}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^C(\cdot|s_t, k)dG\right\} - \delta\eta_{Ak_{At}}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^C(\cdot|s_t, k_{At})dG.$$

By the properties of Gumbel distributions, the first term in this latter expression can be expressed as

$$\max_{k\in K^C}\left\{y_C(s_t, k) + \varepsilon_{Ckt} + \delta\eta_{Ck}(\kappa_t)\textstyle\int_{\varepsilon_{t+1}} EV^C(\cdot)dG\right\} = \max_{k\in K^C}\{v^C(s_t, k) + \varepsilon_{Ckt}\} = \ln\textstyle\sum_{k\in K^C} \exp\{v^C(s_t, k)\} + \varepsilon_{Ct},$$

where $\varepsilon_{Ct}$ is a zero-mean Gumbel-distributed productivity shock and $v^C(s_t, k)$ is given by[48]

$$v^C(s_t, k) = y_C(s_t, k) + \delta\eta_{Ck}(\kappa_t)E\left(\ln\textstyle\sum_{k'\in K^C} \exp\{v^C(s_{t+1}, k')\}|s_t, k\right).$$

By these observations, I can then rewrite $w_A(s_t, \varepsilon_t)$ as

$$\begin{aligned} w_A(s_t, \varepsilon_t) &= \ln\left(\frac{\exp\{v^C(s_t, k_{At})\}\sum_{k\in K^C} \exp\{v^C(s_t, k)\}}{\exp\{v^C(s_t, k_{At})\}}\right) - \delta\eta_{Ak_{At}}(\kappa_t)E\left(\ln\textstyle\sum_{k'\in K^C}\exp\{v^C(s_{t+1}, k')\}|s_t, k_{At}\right) \\ &+ \varepsilon_{Ct} = v^C(s_t, k_{At}) - \ln\Pr(k_{Ct} = k_{At}|f_t = C, s_t) - \delta\eta_{Ak_{At}}(\kappa_t)E\left(\ln\textstyle\sum_{k'\in K^C}\exp\{v^C(s_{t+1}, k')\}|s_t, k_{At}\right) + \varepsilon_{Ct}, \quad (28) \end{aligned}$$

where the second equality uses that $\Pr(k_{Ct} = k_{At}|f_t = C, s_t) = \exp\{v^C(s_t, k_{At})\}/\sum_k \exp\{v^C(s_t, k)\}$. If the law of motion of the state is the same across the two firms at each job and $\eta_{Ck}(\kappa_t) = \eta_{Ak}(\kappa_t)$ for each $k$, it follows that

$$\delta\eta_{Ck_{At}}(\kappa_t)E\left(\ln\textstyle\sum_{k'\in K^C}\exp\{v^C(s_{t+1}, k')\}|s_t, k_{At}\right) = \delta\eta_{Ak_{At}}(\kappa_t)E\left(\ln\textstyle\sum_{k'\in K^C}\exp\{v^C(s_{t+1}, k')\}|s_t, k_{At}\right), \quad (29)$$

where the left side of (29) is the continuation value portion of $v^C(s_t, k_{At})$. The claim follows since $w_A(s_t, \varepsilon_t)$ in (28) reduces to $y_C(s_t, k_{At}) - \ln\Pr(k_{Ct} = k_{At}|f_t = C, s_t) + \varepsilon_{Ct}$. An analogous argument applies if $C$ employs the worker. $\square$

**Proof of Lemma 1**: Assume without loss that $\sigma_{A1k} > \sigma_{A2k} > \sigma_{A3k} > \sigma_{A4k}$ for each $k \geq 1$. Here I show that the weights and component distributions of the mixture distributions of wages at firm $A$ of managers of each possible skill type, conditional on their histories at the firm, are identified at each level and tenure up to their labeling with respect to

[48] In general, if $\varepsilon_1$ and $\varepsilon_2$ are independent Gumbel distributed with parameters $(0, \mu)$, then $y_{\max} = \max\{y_1 + \varepsilon_1, y_2 + \varepsilon_2\}$ is Gumbel distributed with parameters $(\ln(e^{\mu y_1} + e^{\mu y_2})/\mu, \mu)$, mean $\ln(e^{\mu y_1} + e^{\mu y_2})/\mu + \gamma/\mu$, and variance $\pi^2/6\mu^2$. Thus, $y_{\max}$ can be expressed as $y_{\max} = y^* + \varepsilon^*$, where $y^* = \ln(e^{\mu y_1} + e^{\mu y_2})/\mu$ and $\varepsilon^*$ is Gumbel distributed with parameters $(0, \mu)$.

$R_{A1},\ldots,R_{At-1}$. To start, note that $f(w_{A2}|L_{A1}=1, L_{A2}, h_1)$ can be expressed as

$$f(w_{A2}|L_{A1}=1, L_{A2}, h_1) = \textstyle\sum_{R_{A1},i} f(w_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1),$$

which is a mixture density with weights $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$. By Proposition 1 in Al-Hussaini and Ahmad (1981) and Theorem 1 in Shi et al. (2014), referred to as the *logistic mixture results*, $\{f(w_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\}$ are identified for each $R_{A1}$, $L_{A2}$, $h_1$, and $i$, and $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$ are identified for each $L_{A2}$ and $h_1$. For later, it is useful to express these mixture weights as

$$\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1) = \frac{\Pr(L_{A2}|L_{A1}=1, R_{A1}, h_1, i)\Pr(R_{A1}|L_{A1}=1, h_1, i)\Pr(i|L_{A1}=1)}{\Pr(L_{A2}|L_{A1}=1, h_1)}, \tag{30}$$

using that $\Pr(i|L_{A1}=1, h_1) = \Pr(i|L_{A1}=1)$, which is identified by the argument in the main text preceding the statement of the lemma. Once the probabilities $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$ are identified, they can be labeled by $i$ as follows. Since the standard deviations of the component densities satisfy $\sigma_{A1k} > \sigma_{A2k} > \sigma_{A3k} > \sigma_{A4k}$ for each $k$, by comparing their standard deviations, it is possible to determine the skill type $i$ to which each density in $\{f(w_{A2}|L_{A1}= 1, R_{A1}, L_{A2}, h_1, i)\}$ corresponds. Specifically, for any given $L_{A2}$ and $h_1$, there exist eight such densities, given the two possible values of $R_{A1}$ and the four possible values of $i$, but all of these densities have only four possible standard deviations—that is, one for each $i$. Then, each pair of probabilities in $\{\Pr(R_{A1}=1, i|L_{A1}=1, L_{A2}, h_1), \Pr(R_{A1}= 0, i|L_{A1}=1, L_{A2}, h_1)\}$, for given $L_{A2}$, $h_1$, and $i$, is associated with only two component densities that have the same standard deviations, so that determining the value of $i$ in these probabilities is immediate based on their associated component densities. Further, summing each pair of probabilities in $\{\Pr(R_{A1}=1, i|L_{A1}=1, L_{A2}, h_1), \Pr(R_{A1}= 0, i|L_{A1}=1, L_{A2}, h_1)\}$ for given $L_{A2}$, $h_1$, and $i$ yields the probabilities $\{\Pr(i|L_{A1}=1, L_{A2}, h_1)\}$, which, in turn, can be associated with the probabilities in $\{\Pr(i|L_{A1}=1)\}$, $i$ by $i$, through the standard deviations of the corresponding component densities. This result will prove useful in the proof of Proposition 4, which rests on linking the probabilities $\{\Pr(i|L_{A1}=1,\ldots,L_{At}=1, h_1)\}$ across tenures by type.

By a similar logic, I now prove that $f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)$ is an identified mixture and derive additional results about the form of its weights that will prove useful for the propositions that follow. Namely, the wage density at Level $L_{A2}$ for managers with initial human capital $h_1$ and recorded performance $R^o_{A1}$ is

$$f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, h_1) = \textstyle\sum_{R_{A1},i} f(w_{A2}|L_{A1}=1, R^o_{A1}, R_{A1}, L_{A2}, h_1, i)\Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1).$$

By the logistic mixture results above, the densities $\{f(w_{A2}|L_{A1}=1, R^o_{A1}, R_{A1}, L_{A2}, h_1, i)\}$, whose elements are independent of $R^o_{A1}$, are identified for each $R_{A1}$, $L_{A2}$, $h_1$, and $i$, and the mixture weights $\{\Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)\}$ are identified for each $R^o_{A1}$, $L_{A2}$, and $h_1$. Note that these mixture weights can be expressed as

$$\Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1) = \frac{\Pr(L_{A2}|L_{A1}=1, R^o_{A1}, R_{A1}, h_1, i)\Pr(R^o_{A1}, R_{A1}|L_{A1}=1, h_1, i)\Pr(i|L_{A1}=1)}{\Pr(R^o_{A1}, L_{A2}|L_{A1}=1, h_1)}, \tag{31}$$

where $\Pr(L_{A2}|L_{A1}=1, R^o_{A1}, R_{A1}, h_1, i)$ is independent of $R^o_{A1}$. Thus, the same component densities $\{f(w_{A2}|L_{A1}= 1, R_{A1}, L_{A2}, h_1, i)\}$ can be recovered from both $f(w_{A2}|L_{A1}=1, L_{A2}, h_1)$ and $f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)$.

An analogous argument applies to $t=3$. Note that the wage density at Level $L_{A3}$ for managers with initial human capital $h_1$ assigned to Level $L_{A2}$ in $t=2$ is

$$f(w_{A3}|L_{A1}=1, L_{A2}, L_{A3}, h_1) = \textstyle\sum_{R_{A1},R_{A2},i} f(w_{A3}|\cdot, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1, i)\Pr(R_{A1}, R_{A2}, i|\cdot, L_{A2}, L_{A3}, h_1). \tag{32}$$

By the logistic mixture results above, the densities $\{f(w_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1, i)\}$ are identified for each $R_{A1}$, $L_{A2}$, $R_{A2}$, $L_{A3}$, $h_1$, and $i$, and the mixture weights $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$ are identified for each $L_{A2}$, $L_{A3}$, and $h_1$. Observe that these latter mixture weights can be expressed as

$$\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1) = \frac{\begin{array}{c}\Pr(L_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_1, i)\Pr(R_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\\ \cdot\Pr(L_{A2}|L_{A1}=1, R_{A1}, h_1, i)\Pr(R_{A1}|L_{A1}=1, h_1, i)\Pr(i|L_{A1}=1)\end{array}}{\Pr(L_{A2}, L_{A3}|L_{A1}=1, h_1)}. \tag{33}$$

As before, by comparing the standard deviations of the component densities $\{f(w_{A3}|\cdot, h_1, i)\}$ of this mixture, it is possible to determine the skill type $i$ to which each component density corresponds. Note also that summing the four probabilities in $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$ for given $L_{A2}$, $L_{A3}$, $h_1$, and $i$ yields the probabilities $\{\Pr(i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$, which can then be uniquely paired with the probabilities in $\{\Pr(i|L_{A1}=1)\}$ and $\{\Pr(i|L_{A1}=1, L_{A2}, h_1)\}$ through the standard deviations of the corresponding component densities.

Proceeding analogously, note that the wage density at Level $L_{A3}$ in $t=3$ for managers with initial human capital $h_1$ assigned to Level $L_{A2}$ in $t=2$ and with recorded performance $R^o_{A2}$ at that level is

$$\begin{aligned} f(w_{A3}|L_{A1}=1, L_{A2}, R^o_{A2}, L_{A3}, h_1) = \textstyle\sum_{R_{A1},R_{A2},i} & f(w_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R^o_{A2}, R_{A2}, L_{A3}, h_1, i) \\ & \cdot \Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, R^o_{A2}, L_{A3}, h_1), \end{aligned}$$

with $f(w_{A3}|L_{A1}=1,R_{A1},L_{A2},R^o_{A2},R_{A2},L_{A3},h_1,i)$ independent of $R^o_{A2}$, $\Pr(R_{A1},R_{A2}, i|\cdot,h_1)$ given by

$$\Pr(R_{A1}, R_{A2}, i|\cdot, h_1) = \frac{\begin{array}{c}\Pr(L_{A3}|L_{A1}=1,R_{A1},L_{A2},R^o_{A2},R_{A2},h_1,i) \Pr(R^o_{A2},R_{A2}|L_{A1}=1,R_{A1},L_{A2},h_1,i) \\ \cdot \Pr(L_{A2}|L_{A1}=1,R_{A1},h_1,i) \Pr(R_{A1}|L_{A1}=1,h_1,i) \Pr(i|L_{A1}=1)\end{array}}{\Pr(L_{A2},R^o_{A2},L_{A3}|L_{A1}=1,h_1)}, \quad (34)$$

and $\Pr(L_{A3}|L_{A1}=1,R_{A1},L_{A2},R^o_{A2},R_{A2},h_1,i)$ independent of $R^o_{A2}$. By the logistic mixture results above, the densities $\{f(w_{A3}|\cdot,h_1,i)\}$ are identified for each $R_{A1}$, $L_{A2}$, $R_{A2}$, $L_{A3}$, $h_1$, and $i$, and the mixture weights $\{\Pr(R_{A1}, R_{A2}, i|\cdot, h_1)\}$ are identified for each $L_{A2}$, $R^o_{A2}$, $L_{A3}$, and $h_1$. Then, the same set of densities $\{f(w_{A3}|L_{A1}=1,R_{A1},L_{A2},R_{A2},L_{A3},h_1,i)\}$ can be recovered from $f(w_{A3}|L_{A1}=1, L_{A2}, L_{A3}, h_1)$ and $f(w_{A3}|L_{A1}=1, L_{A2}, R^o_{A2}, L_{A3}, h_1)$.

This argument can easily be extended to the remaining tenure years to complete the proof of the lemma. □

**Proof of Proposition 3**: The proof consists of two parts. In Part I, I show that the parameters of the distribution of classification error in recorded performance are identified from the weights of the mixture distributions of wages of managers of each possible skill type, which are identified by Lemma 1. In Part II, I show that once the distribution of classification error is identified, the probability masses of the distribution of performance at each Level $k \geq 1$ for each manager ability, $\{\alpha_k, \beta_k\}$, are identified from repeated observations on performance ratings at each level; see Observation 2 in the proof of Proposition 4 for an alternative argument.

*Part I: Identification of classification error parameters.* Consider identifying $\rho_0$, $\rho_1$, and $\rho_2(1)$ in (11). Fix $R^o_{A1}$, $L_{A2}$, and $h_1$. Note first that the conditional probability of recorded performance given true performance in $t=1$, $\Pr(R^o_{A1}|L_{A1}=1, R_{A1}, h_1, i)$, is independent of $h_1$ and $i$ by (11). Recall that $\Pr(L_{A2}|L_{A1}=1, R^o_{A1}, R_{A1}, h_1, i)$ is independent of $R^o_{A1}$ and express $\Pr(R^o_{A1}, R_{A1}|L_{A1}=1, h_1, i)$ as the product $\Pr(R^o_{A1}|L_{A1}=1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1}=1, h_1, i)$. The probability $\Pr(R^o_{A1}|L_{A1}=1, R_{A1})$ is identified for each $R^o_{A1}$ and $R_{A1}$ from the ratio of the identified expression on the right side of (31) multiplied by the associated probability $\Pr(R^o_{A1}, L_{A2}|L_{A1}=1, h_1)$,

$$\Pr(L_{A2}|L_{A1}=1, R_{A1}, h_1, i) \Pr(R^o_{A1}|L_{A1}=1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1}=1, h_1, i) \Pr(i|L_{A1}=1),$$

to the identified expression on the right side of (30) multiplied by the associated probability $\Pr(L_{A2}|L_{A1}=1, h_1)$,

$$\Pr(L_{A2}|L_{A1}=1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1}=1, h_1, i) \Pr(i|L_{A1}=1). \quad (35)$$

To compute this ratio, expressions (30) and (31) need to be paired. To see that no ambiguity arises, recall that (30) and (31) result from simple manipulations of $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$ and $\{\Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)\}$, which are, respectively, the weights of the mixture distributions $f(w_{A2}|L_{A1}=1, L_{A2}, h_1)$ and $f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)$ identified by Lemma 1. As shown in the proof of Lemma 1, the components of these mixtures can be correctly labeled with respect to $i$. Recall also from the proof of Lemma 1 that the same component densities $\{f(w_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\}$ can be recovered from both $f(w_{A2}|L_{A1}=1, L_{A2}, h_1)$ and $f(w_{A2}|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)$. Since mean wages by skill type (and human capital) are injective functions of the prior $p_{it}$ whenever $\delta$ or $\{\eta_{kt}\}$ are not too large, these densities can be ordered by their means.[49] Then, each probability in $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$ can be uniquely

[49] To see $Ew_{Ait}(k)$ is an injective function of $p_{it}$, let, by contradiction, $p'_{it} > p_{it}$, but $Ew_{Ait}(k) = Ew'_{Ait}(k)$ so that $\ln \Pr(L_{Ct}=k|f_t=C, s'_{it}) - \ln \Pr(L_{Ct}=k|f_t=C, s_{it}) = y_C(s'_{it}, k) - y_C(s_{it}, k)$, $s_{it}=(p_{it},\kappa_t,i)$, or $\ln \sum_{k'} \exp\{\overline{v}^C(s'_{it}, k')\} - \delta\eta_{kt} E(\ln \sum_{k'} \exp\{\overline{v}^C(s'_{it+1}, k')\}|s'_{it}, k) = \ln \sum_{k'} \exp\{\overline{v}^C(s_{it}, k')\} - \delta\eta_{kt} E(\ln \sum_{k'} \exp\{\overline{v}^C(s_{it+1}, k')\}|s_{it}, k)$ with $\overline{v}^C(s_{it}, k)$ defined analogously to $\overline{v}^A(s_{it}, k)$ in (15). But this equality cannot hold if $\delta$ or $\{\eta_{kt}\}$ are small enough, since $\ln \sum_{k'} \exp\{\overline{v}^C(\cdot, \kappa_t, i, k')\}$ strictly increases with $p_{it}$.

paired with the corresponding probability in $\{\Pr(R_{A1}, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)\}$ given $R^o_{A1}$, $L_{A2}$, $h_1$, and $i$, since any correct pair of probabilities is associated with the same component wage density in $\{f(w_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\}$. Thus, the probabilities $\{\Pr(R^o_{A1}|L_{A1}=1, R_{A1})\}$ are identified for any $R^o_{A1}$ and $R_{A1}$.

Observe now that $\Pr(R^o_{A1}|L_{A1}=1, R_{A1}=1)$ can be distinguished from $\Pr(R^o_{A1}|L_{A1}=1, R_{A1}=0)$ given $R^o_{A1}$ by comparing $1-E_1(1,t)$ and $E_0(1,t)$ from (11), since $\rho_1>0$ implies

$$\Pr(R^o_{At}=1|L_{A1}=1,\dots,L_{At},R_{At}=1)>\Pr(R^o_{At}=1|L_{A1}=1,\dots,L_{At},R_{At}=0) \tag{36}$$

for the same $L_{At}$ and years of experience at firm $A$. In particular, $\Pr(R^o_{A1}|L_{A1}=1, R_{A1}=1)$ can be distinguished from $\Pr(R^o_{A1}|L_{A1}=1, R_{A1}=0)$ by (36), as claimed. Hence, the associated probabilities $\Pr(R_{A1}=1, i|L_{A1}=1, L_{A2}, h_1)$ from (30) and $\Pr(R_{A1}=1, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)$ from (31) can be distinguished from the corresponding probabilities $\Pr(R_{A1}=0, i|L_{A1}=1, L_{A2}, h_1)$ and $\Pr(R_{A1}=0, i|L_{A1}=1, R^o_{A1}, L_{A2}, h_1)$ for each $R^o_{A1}$, $L_{A2}$, $h_1$, and $i$.

Proceeding similarly, the ratio of the identified expression on the right side of (34) multiplied by the associated probability $\Pr(L_{A2}, R^o_{A2}, L_{A3}|L_{A1}=1, h_1)$ to the corresponding identified expression on the right side of (33) multiplied by the associated probability $\Pr(L_{A2}, L_{A3}|L_{A1}=1, h_1)$ for given $R_{A1}$, $L_{A2}$, $R^o_{A2}$, $R_{A2}$, $L_{A3}$, $h_1$, and $i$ identifies $\Pr(R^o_{A2}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_1, i) = \Pr(R^o_{A2}|L_{A1}=1, L_{A2}, R_{A2})$, since $\Pr(L_{A3}|L_{A1}=1,R_{A1},L_{A2},R^o_{A2},R_{A2},$-$h_1,i)$ is independent of $R^o_{A2}$. To see how these ratios can be correctly computed, note that the same component densities $\{f(w_{A3}|L_{A1}=1,R_{A1},L_{A2},R_{A2},L_{A3},h_1,i)\}$ are recovered from both $f(w_{A3}|L_{A1}=1, L_{A2}, L_{A3}, h_1)$ and $f(w_{A3}|$-$L_{A1}=1, L_{A2}, R^o_{A2}, L_{A3}, h_1)$. Thus, through these component densities uniquely identified by their means, the mixture weights $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, R^o_{A2}, L_{A3}, h_1)\}$ can be matched to the corresponding weights $\{\Pr(R_{A1}, R_{A2},$-$i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$. By (36), no label ambiguity arises for $\Pr(R^o_{A2}|L_{A1}=1, L_{A2}, R_{A2})$ with respect to $R_{A2}$.

The results that $\{\Pr(R^o_{A1}|L_{A1}=1, R_{A1})\}$ and $\{\Pr(R^o_{A2}|L_{A1}=L_{A2}=1, R_{A2})\}$ are identified imply that $E_0(1,t)$ and $E_1(1,t)$, $t=1,2$, are identified. By (11), knowledge of $E_0(1,1)$, $E_1(1,1)$, and of $E_0(1,2)$ or $E_1(1,2)$ is sufficient to identify $(\rho_0, \rho_1, \rho_2(1))$. In particular, either $\Pr(R^o_{A1}=1|L_{A1}=1, R_{A1}=1)$ and $\Pr(R^o_{A1}=1|L_{A1}=1, R_{A1}=0)$ or $\Pr(R^o_{A1}=0|L_{A1}=1, R_{A1}=1)$ and $\Pr(R^o_{A1}=0|L_{A1}=1, R_{A1}=0)$, together with either $\Pr(R^o_{A2}=1|L_{A1}=L_{A2}=1, R_{A2})$ or $\Pr(R^o_{A2}=0|L_{A1}=L_{A2}=1, R_{A2})$ for given $R_{A2}$, are sufficient to identify $(\rho_0, \rho_1, \rho_2(1))$.

Now, consider identifying $\rho_2(2)$ for Level 2. Proceeding analogously, it is immediate that $E_0(2,2)$ or $E_1(2,2)$ are also identified from $\Pr(R^o_{A2}|L_{A1}=1, L_{A2}=2, R_{A2})$ for given $R^o_{A2}$ and $R_{A2}$, which implies that $\rho_2(2)$ is also identified from either $\Pr(R^o_{A2}=1|L_{A1}=1, L_{A2}=2, R_{A2})$ or $\Pr(R^o_{A2}=0|L_{A1}=1, L_{A2}=2, R_{A2})$ for given $R_{A2}$, once $\rho_0$ and $\rho_1$ are identified. Lastly, consider identifying $\rho_2(3)$ for Level 3. By extending this argument in the natural way, it is easy to show that $\rho_2(3)$ is identified from the distribution of wages of managers in the fourth year of tenure assigned to Level 3 in $t=3$ with either high or low recorded performance in $t=3$.

*Part II: Identification of* $\{\alpha_k, \beta_k\}$. Recall that $q_{i3}(h_1)=\Pr(i|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)$. The logic of the proof consists of first showing that the distribution of performance ratings at Level 1 in three consecutive years—namely, the probabilities $\Pr(R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)$, $\Pr(R^o_{A2}=1, R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)$, and $\Pr(R^o_{A1}=1, R^o_{A2}=1, R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)$—identify the three corresponding probabilities of high performance—namely, $\nu_1(h_1)=(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1$, $\nu_2(h_1)=(\alpha_1^2-\beta_1^2)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2$, and $\nu_3(h_1)=(\alpha_1^3-\beta_1^3)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^3$—once the distribution of classification error in performance at Level 1 is identified. The proof then proceeds to recover $(\alpha_1, \beta_1, \sum_i q_{i3}(h_1)p_{i1})$ for a given $h_1$ as the unique solution to the system

$$\left\{\begin{array}{l}(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1=\nu_1(h_1)\\ (\alpha_1^2-\beta_1^2)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2=\nu_2(h_1)\\ (\alpha_1^3-\beta_1^3)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^3=\nu_3(h_1)\end{array}\right. . \tag{37}$$

A similar argument can be used to recover $(\alpha_2, \beta_2)$ and $(\alpha_3, \beta_3)$. If $\alpha_k=\beta_k$ for some $k$, then two periods of observations on performance ratings at Level $k$ are sufficient to identify $\alpha_k$. The rest of the proof is articulated in five steps.

*Step 1:* $\nu_1(h_1)$ *is identified for each* $h_1$. To start, note that

$$\begin{aligned}&\Pr(R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)=\textstyle\sum_i \Pr(i|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)\\ &\quad\cdot\textstyle\sum_{R_{A3}}\Pr(R^o_{A3}=1, R_{A3}|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1, i)\\ &=\textstyle\sum_i q_{i3}(h_1)\{\Pr(R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, R_{A3}=1)\Pr(R_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1, i)\\ &\quad+\Pr(R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, R_{A3}=0)\Pr(R_{A3}=0|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1, i)\},\end{aligned} \tag{38}$$

using that $\Pr(R^o_{At}|L_{A1},\ldots,L_{At},R_{At},h_1,i)$ is independent of $h_1$ and $i$ by (11). Also,

$$\Pr(R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)=\textstyle\sum_{R_{A1},R_{A2}}\Pr(R_{A1},R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i),$$

and $\Pr(R_{A1},R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)$ is independent of $h_1$. Using that $\sum_i q_{i3}(h_1)=1$ and the definition of classification error rates, by straightforward algebra, I obtain from (38) that

$$\begin{aligned}\Pr(R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1)&=[(\alpha_1-\beta_1)\textstyle\sum_i q_{i3}(h_1)p_{i1}+\beta_1][1-E_1(1,3)]\\&\quad+\{1-[(\alpha_1-\beta_1)\textstyle\sum_i q_{i3}(h_1)p_{i1}+\beta_1]\}E_0(1,3).\end{aligned}\tag{39}$$

Since the probability on the left side of (39) is known from the data and $E_1(1,3)$ and $E_0(1,3)$ are identified by Part I, it follows that $\nu_1(h_1)=(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1$ is identified by (39) for each $h_1$.

*Step 2: $\nu_2(h_1)$ is identified for each* $h_1$. To this end, note that

$$\begin{aligned}&\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1)\\&=\textstyle\sum_i\Pr(i|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1)\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)\\&=\textstyle\sum_i q_{i3}(h_1)\sum_{R_{A2},R_{A3}}[\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,R_{A2},L_{A3}=1,R_{A3},h_1,i)\\&\quad\cdot\Pr(R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)]\\&=\textstyle\sum_i q_{i3}(h_1)\sum_{R_{A2},R_{A3}}[\Pr(R^o_{A3}=1|L_{A1}=1,L_{A2}=1,R^o_{A2}=1,R_{A2},L_{A3}=1,R_{A3},h_1,i)\\&\quad\cdot\Pr(R^o_{A2}=1|L_{A1}=1,L_{A2}=1,R_{A2},L_{A3}=1,R_{A3},h_1,i)\Pr(R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)]\end{aligned}\tag{40}$$

with $\Pr(R_{A2},R_{A3}|\cdot,h_1,i)$ independent of $h_1$ and $\Pr(R^o_{A2}|L_{A1}=1,L_{A2}=1,R_{A2},L_{A3},R_{A3},h_1,i)$ of $L_{A3},R_{A3},h_1$, and $i$ as

$$\begin{aligned}\Pr(R^o_{A2}|L_{A1}=1,L_{A2}=1,R_{A2},L_{A3},R_{A3},h_1,i)&=\frac{\Pr(L_{A1}=1,L_{A2}=1,R^o_{A2},R_{A2},L_{A3},R_{A3},h_1,i)}{\Pr(L_{A1}=1,L_{A2}=1,R_{A2},L_{A3},R_{A3},h_1,i)}\\&=\frac{\begin{array}{c}\Pr(L_{A3},R_{A3}|L_{A1}=1,L_{A2}=1,R^o_{A2},R_{A2},h_1,i)\\\cdot\Pr(R^o_{A2}|L_{A1}=1,L_{A2}=1,R_{A2},h_1,i)\Pr(L_{A1}=1,L_{A2}=1,R_{A2},h_1,i)\end{array}}{\Pr(L_{A3},R_{A3}|L_{A1}=1,L_{A2}=1,R_{A2},h_1,i)\Pr(L_{A1}=1,L_{A2}=1,R_{A2},h_1,i)}=\Pr(R^o_{A2}|L_{A1}=1,L_{A2}=1,R_{A2}),\end{aligned}$$

where the first two equalities follow by simple manipulations and the last equality follows by canceling terms, since $\Pr(L_{A3},R_{A3}|L_{A1}=1,L_{A2}=1,R^o_{A2},R_{A2},h_1,i)$ is independent of $R^o_{A2}$ and $\Pr(R^o_{A2}|L_{A1}=1,L_{A2}=1,R_{A2},h_1,i)$ is independent of $h_1$ and $i$. Using the fact just established and that $\Pr(R^o_{A3}=1|L_{A1}=1,L_{A2}=1,R^o_{A2},R_{A2},L_{A3}=1,R_{A3},h_1,i)$ is similarly independent of $R^o_{A2}$, $R_{A2}$, $h_1$, and $i$ by (11), I obtain

$$\begin{aligned}&\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)=\textstyle\sum_{R_{A2},R_{A3}}\Pr(R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,R_{A3})\\&\quad\cdot\Pr(R^o_{A2}=1|L_{A1}=1,L_{A2}=1,R_{A2})[\Pr(R_{A1}=1,R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,i)\\&\quad+\Pr(R_{A1}=0,R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,i)],\end{aligned}$$

as $\Pr(R_{A1},R_{A2},R_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)$ is independent of $h_1$. Note also that $\Pr(R^o_{A2},R^o_{A3}|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1,i)$ is independent of $h_1$. Using the definition of error rates in (11), it is easy to show that

$$\begin{aligned}\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,i)&=\underbrace{[1-E_1(1,3)][1-E_1(1,2)]}_{B_2}[\alpha_1^2p_{i1}+\beta_1^2(1-p_{i1})]\\&+\underbrace{\{E_0(1,3)[1-E_1(1,2)]+[1-E_1(1,3)]E_0(1,2)\}}_{C_2}[\alpha_1p_{i1}+\beta_1(1-p_{i1})-\alpha_1^2p_{i1}-\beta_1^2(1-p_{i1})]\\&+\underbrace{E_0(1,3)E_0(1,2)}_{D_2}[\alpha_1^2p_{i1}+\beta_1^2(1-p_{i1})+1-2\alpha_1p_{i1}-2\beta_1(1-p_{i1})],\end{aligned}$$

where $B_2$, $C_2$, and $D_2$ are known constants so that

$$\begin{aligned}&\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1)=\textstyle\sum_i q_{i3}(h_1)\Pr(R^o_{A2}=1,R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,i)\\&=B_2[(\alpha_1^2-\beta_1^2)\textstyle\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2]+C_2\{(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1-[(\alpha_1^2-\beta_1^2)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2]\}\\&\quad+D_2\{(\alpha_1^2-\beta_1^2)\textstyle\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2+1-2[(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1]\}.\end{aligned}\tag{41}$$

Using that the probability on the left side of (41) is known from the data and $\nu_1(h_1) = (\alpha_1 - \beta_1)\sum_i q_{i3}(h_1)p_{i1} + \beta_1$ is identified by Step 1, it follows that $\nu_2(h_1) = (\alpha_1^2 - \beta_1^2)\sum_i q_{i3}(h_1)p_{i1} + \beta_1^2$ is also identified.

*Step 3: $\nu_3(h_1)$ is identified for each $h_1$.* Observe that

$$\Pr(R^o_{A1}=1, R^o_{A2}=1, R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1) = \sum_i q_{i3}(h_1)\sum_{R_{A1},R_{A2},R_{A3}}\Pr(R^o_{A1}=1, R^o_{A2}=1, R^o_{A3}=1|L_{A1}=1, R_{A1}, L_{A2}=1, R_{A2}, L_{A3}=1, R_{A3})\Pr(R_{A1}, R_{A2}, R_{A3}|L_{A1}=1, L_{A2}=1, L_{A3}=1, i),$$

and so there exist known constants, $B_3$, $C_3$, $D_3$, and $E_3$, which are known functions of the classification error rates at Level 1 in tenures 1, 2, and 3, such that

$$\begin{aligned}
&\Pr(R^o_{A1} = 1, R^o_{A2} = 1, R^o_{A3} = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) \\
&\quad = B_3\sum_i q_{i3}(h_1)[\alpha_1^3 p_{i1} + \beta_1^3(1 - p_{i1})] + C_3\sum_i q_{i3}(h_1)[\alpha_1^2(1 - \alpha_1)p_{i1} + \beta_1^2(1 - \beta_1)(1 - p_{i1})] \\
&\qquad + D_3\sum_i q_{i3}(h_1)[\alpha_1(1-\alpha_1)^2 p_{i1} + \beta_1(1-\beta_1)^2(1-p_{i1})] + E_3\sum_i q_{i3}(h_1)[(1-\alpha_1)^3 p_{i1} + (1-\beta_1)^3(1-p_{i1})],
\end{aligned}$$

or equivalently,

$$\begin{aligned}
&\Pr(R^o_{A1}=1, R^o_{A2}=1, R^o_{A3}=1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1) = B_3[(\alpha_1^3 - \beta_1^3)\textstyle\sum_i q_{i3}(h_1)p_{i1} + \beta_1^3] \\
&+C_3\left\{[(\alpha_1^2 - \beta_1^2)\textstyle\sum_i q_{i3}(h_1)p_{i1} + \beta_1^2] - [(\alpha_1^3-\beta_1^3)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^3]\right\} \\
&+D_3\left\{[(\alpha_1^3-\beta_1^3)\textstyle\sum_i q_{i3}(h_1)p_{i1} + \beta_1^3] + [(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1] - 2[(\alpha_1^2-\beta_1^2)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2]\right\} \\
&+E_3\left\{1-3[(\alpha_1-\beta_1)\textstyle\sum_i q_{i3}(h_1)p_{i1} + \beta_1]+3[(\alpha_1^2-\beta_1^2)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^2]-[(\alpha_1^3-\beta_1^3)\sum_i q_{i3}(h_1)p_{i1}+\beta_1^3]\right\}.
\end{aligned}$$

So, $\nu_3(h_1) = (\alpha_1^3 - \beta_1^3)\sum_i q_{i3}(h_1)p_{i1} + \beta_1^3$ is identified as $\Pr(R^o_{A1} = 1, R^o_{A2} = 1, R^o_{A3} = 1|\cdot, h_1)$ is known from the data, and $\nu_1(h_1) = (\alpha_1 - \beta_1)\sum_i q_{i3}(h_1)p_{i1} + \beta_1$ and $\nu_2(h_1) = (\alpha_1^2 - \beta_1^2)\sum_i q_{i3}(h_1)p_{i1} + \beta_1^2$ are identified by Steps 1-2.

*Step 4: $(\alpha_1, \beta_1, \sum_i q_{i3}(h_1)p_{i1})$ are identified.* Fix $h_1$. I now show that the system of equations in (37) admits a unique solution for $(\alpha_1, \beta_1, \sum_i q_{i3}(h_1)p_{i1})$ if $\alpha_1 > \beta_1$. To start, using that $\alpha_1^2 - \beta_1^2 = (\alpha_1 + \beta_1)(\alpha_1 - \beta_1)$ and $\alpha_1^3 - \beta_1^3 = (\alpha_1 - \beta_1)[(\alpha_1 + \beta_1)^2 - \alpha_1\beta_1] = (\alpha_1 + \beta_1)(\alpha_1^2 - \beta_1^2) - \alpha_1\beta_1(\alpha_1 - \beta_1)$, I can rewrite (37) as

$$\begin{cases}
(\alpha_1 - \beta_1)\sum_i q_{i3}(h_1)p_{i1} = \nu_1(h_1) - \beta_1 \\
(\alpha_1 + \beta_1)[\nu_1(h_1) - \beta_1] = \nu_2(h_1) - \beta_1^2 \Leftrightarrow \nu_1(h_1)(\alpha_1 + \beta_1) - \alpha_1\beta_1 = \nu_2(h_1) \\
(\alpha_1 + \beta_1)[\nu_2(h_1) - \beta_1^2] - \alpha_1\beta_1[\nu_1(h_1) - \beta_1] = \nu_3(h_1) - \beta_1^3 \Leftrightarrow \frac{\nu_2(h_1)(\alpha_1+\beta_1)}{\nu_1(h_1)} - \alpha_1\beta_1 = \frac{\nu_3(h_1)}{\nu_1(h_1)}
\end{cases}. \tag{42}$$

By subtracting the third equation from the second equation in (42), side by side, I obtain

$$\alpha_1 + \beta_1 = [\nu_1(h_1)\nu_2(h_1) - \nu_3(h_1)]/[\nu_1^2(h_1) - \nu_2(h_1)]. \tag{43}$$

Substituting this last expression into the second equation in (42) gives

$$\alpha_1\beta_1 = \nu_1(h_1)(\alpha_1 + \beta_1) - \nu_2(h_1) = [\nu_2^2(h_1) - \nu_1(h_1)\nu_3(h_1)]/[\nu_1^2(h_1) - \nu_2(h_1)]. \tag{44}$$

Using (43), (44), that $\alpha_1 - \beta_1 = \sqrt{(\alpha_1 - \beta_1)^2} = \sqrt{(\alpha_1 + \beta_1)^2 - 4\alpha_1\beta_1}$, and $\alpha_1 > \beta_1$ yields that

$$\alpha_1 - \beta_1 = \sqrt{(\alpha_1+\beta_1)^2 - 4\alpha_1\beta_1} = \sqrt{\frac{1}{\nu_1^2(h_1)-\nu_2(h_1)}\left\{\frac{[\nu_1(h_1)\nu_2(h_1)-\nu_3(h_1)]^2}{\nu_1^2(h_1)-\nu_2(h_1)} + 4\nu_1(h_1)\nu_3(h_1) - 4\nu_2^2(h_1)\right\}}. \tag{45}$$

By summing (43) and (45) side by side, it follows that

$$\alpha_1 = \frac{\nu_1(h_1)\nu_2(h_1)-\nu_3(h_1)}{2[\nu_1^2(h_1)-\nu_2(h_1)]} + \frac{1}{2}\sqrt{\frac{1}{\nu_1^2(h_1) - \nu_2(h_1)}\left\{\frac{[\nu_1(h_1)\nu_2(h_1)-\nu_3(h_1)]^2}{\nu_1^2(h_1)-\nu_2(h_1)} + 4\nu_1(h_1)\nu_3(h_1) - 4\nu_2^2(h_1)\right\}},$$

which, substituted into (43) or (44), provides a similar expression for $\beta_1$. So, $(\alpha_1, \beta_1)$ are identified. Plugging the expressions for $\alpha_1$ and $\beta_1$ into the first equation of (42) gives an expression for $\sum_i q_{i3}(h_1)p_{i1}$ that depends only on $\nu_1(h_1)$, $\nu_2(h_1)$, and $\nu_3(h_1)$.Thus, $\sum_i q_{i3}(h_1)p_{i1}$ is also identified for a given $h_1$ and, by (39), for each $h_1$ from $\Pr(R^o_{A3} = 1|L_{A1}=1, L_{A2}=1, L_{A3}=1, h_1)$ by varying $h_1$, once $\alpha_1$, $\beta_1$, and the distribution of classification error are identified.

*Step 5: $\{\alpha_k, \beta_k\}_{k>1}$ are identified.* The argument in the previous steps can easily be adapted to show that $\{\alpha_k, \beta_k\}_{k>1}$ are also identified. In particular, information on the performance ratings of managers promoted to Level 2 in the second year at the firm and assigned to Level 2 for at least two more years allows one to identify $\alpha_2$ and $\beta_2$. Information on the performance ratings of managers promoted to Level 2 in the second year at the firm, then promoted to Level 3 in the third year and assigned to Level 3 for at least two more years allows one to identify $\alpha_3$ and $\beta_3$. □

**Proof of Proposition 4**: Here, I prove that the support of the initial prior, $\{p_{i1}\}$, is identified from the distribution of performance ratings at Level 1 between $t=3$ and $t=6$. It is easy to show that the probability of high performance in $t$, conditional on the assignment to Level 1 up to $t$, for a manager with initial human capital $h_1$ is

$$\nu_{1t}(h_1)=\Pr(R_{At}=1|L_{A1}=1,\dots,L_{At}=1,h_1)=(\alpha_1-\beta_1)\textstyle\sum_i \Pr(i|L_{A1}=1,\dots,L_{At}=1,h_1)p_{i1}+\beta_1, \quad (46)$$

where $q_{it}(h_1)=\Pr(i|L_{A1}=1,\dots,L_{At}=1,h_1)$. It is immediate to see that $\nu_{1t}(h_1)$ is identified in each $t$ from $\Pr(R^o_{At}=1|L_{A1}=1,\dots,L_{At}=1,h_1)$ by an argument analogous to that in Part II of the proof of Proposition 3 used to derive (39) and argue that $\nu_1(h_1)=(\alpha_1-\beta_1)\sum_i q_{i3}(h_1)p_{i1}+\beta_1$ is identified from $\Pr(R^o_{A3}=1|L_{A1}=1,L_{A2}=1,L_{A3}=1,h_1)$, once error rates at Level 1 are identified. Recall also from Part II of the proof of Proposition 3 that $\alpha_1$, $\beta_1$, and $\{\sum_i q_{i3}(h_1)p_{i1}\}$ are identified for each $h_1$. So, $\{\sum_i q_{it}(h_1)p_{i1}\}$ are identified for each $h_1$ also in $t\geq 4$ by (46).

By the proof of Lemma 1, each probability in $\{\Pr(i|L_{A1}=1,\dots,L_{At},h_1)\}$ is identified and can be uniquely paired with the corresponding probability in $\{\Pr(i|L_{A1}=1,h_1)\}$ for any $t$. I can then express $\sum_i q_{it}(h_1)p_{i1}=[\nu_{1t}(h_1)-\beta_1]/(\alpha_1-\beta_1)$, which is derived from (46), in matrix form from $t=3$ to $t=6$ as

$$\underbrace{\begin{bmatrix} q_{13}(h_1) & q_{23}(h_1) & q_{33}(h_1) & q_{43}(h_1) \\ q_{14}(h_1) & q_{24}(h_1) & q_{34}(h_1) & q_{44}(h_1) \\ q_{15}(h_1) & q_{25}(h_1) & q_{35}(h_1) & q_{45}(h_1) \\ q_{16}(h_1) & q_{26}(h_1) & q_{36}(h_1) & q_{46}(h_1) \end{bmatrix}}_{Q(h_1)} \begin{bmatrix} p_{11} \\ p_{21} \\ p_{31} \\ p_{41} \end{bmatrix} = \begin{bmatrix} [\nu_{13}(h_1)-\beta_1]/(\alpha_1-\beta_1) \\ [\nu_{14}(h_1)-\beta_1]/(\alpha_1-\beta_1) \\ [\nu_{15}(h_1)-\beta_1]/(\alpha_1-\beta_1) \\ [\nu_{16}(h_1)-\beta_1]/(\alpha_1-\beta_1) \end{bmatrix},$$

where $q_{it}(h_1)=\Pr(L_{A1}=1,\dots,L_{At}=1,h_1,i)/\Pr(L_{A1}=1,\dots,L_{At}=1,h_1)$, $t>1$, satisfies

$$q_{it}(h_1)=\underbrace{\frac{\Pr(L_{At}=1|L_{A1}=1,\dots,L_{At-1}=1,h_1,i)}{\Pr(L_{At}=1|L_{A1}=1,\dots,L_{At-1}=1,h_1)}}_{\pi_{it}(h_1)}\cdot\underbrace{\Pr(i|L_{A1}=1,\dots,L_{At-1}=1,h_1)}_{q_{it-1}(h_1)}.$$

The desired result is obtained if no two rows (or columns) of $Q(h_1)$ are linearly dependent. By contradiction, suppose, for instance, that the first and second rows of $Q(h_1)$ are linearly dependent. By definition, this is the case if there exist two constants $\psi_1$ and $\psi_2$, not both zero, such that

$$\psi_1\times(q_{13}(h_1),q_{23}(h_1),q_{33}(h_1),q_{43}(h_1))'+\psi_2\times(q_{14}(h_1),q_{24}(h_1),q_{34}(h_1),q_{44}(h_1))'=0.$$

This latter condition, by using the recursion $q_{it}(h_1)=\pi_{it}(h_1)q_{it-1}(h_1)$, can be expressed in matrix form as

$$\psi_1\begin{bmatrix} q_{13}(h_1) \\ q_{23}(h_1) \\ q_{33}(h_1) \\ q_{43}(h_1) \end{bmatrix}+\psi_2\begin{bmatrix} q_{14}(h_1) \\ q_{24}(h_1) \\ q_{34}(h_1) \\ q_{44}(h_1) \end{bmatrix}=\psi_1\begin{bmatrix} q_{13}(h_1) \\ q_{23}(h_1) \\ q_{33}(h_1) \\ q_{43}(h_1) \end{bmatrix}+\psi_2\begin{bmatrix} \pi_{14}(h_1)q_{13}(h_1) \\ \pi_{24}(h_1)q_{23}(h_1) \\ \pi_{34}(h_1)q_{33}(h_1) \\ \pi_{44}(h_1)q_{43}(h_1) \end{bmatrix}=\begin{bmatrix} [\psi_1+\psi_2\pi_{14}(h_1)]q_{13}(h_1) \\ [\psi_1+\psi_2\pi_{24}(h_1)]q_{23}(h_1) \\ [\psi_1+\psi_2\pi_{34}(h_1)]q_{33}(h_1) \\ [\psi_1+\psi_2\pi_{44}(h_1)]q_{43}(h_1) \end{bmatrix}=0.$$

For this condition to be satisfied with $q_{i3}(h_1)>0$ for each $i$, it must be $\psi_1=-\psi_2\pi_{i4}(h_1)$ for each $i$, which can hold only if $\psi_1$ and $\psi_2$ are zero, as $\pi_{i4}(h_1)$ varies with $i$ as well as with $h_1$. (For instance, the numerator of $\pi_{i4}(h_1)$ is a smooth function of the prior about manager ability, and the prior varies with $i$.) Thus, the first and second rows of $Q(h_1)$ are linearly independent. Repeating this argument for all other rows yields that $Q(h_1)$ has full rank, so $\{p_{i1}\}$ are identified.

**Observation 1**: With $\{\Pr(i|L_{A1}=1,\dots,L_{At},h_1)\}$ identified by Lemma 1 and $\{\alpha_k,\beta_k\}$ and classification error rates at each level identified by Proposition 3, an analogous argument would apply with four periods of observations on performance ratings at Levels 2 or 3.

**Observation 2**: With $\{\Pr(i|L_{A1}=1,\dots,L_{At},h_1)\}$ identified by Lemma 1 and $(\alpha_1,\beta_1)$ and classification error rates at Level 1 identified by Proposition 3, a similar argument establishes that $(\alpha_2, \beta_2)$ and $(\alpha_3, \beta_3)$ are identified from the analogue of (46) for *i*) managers observed at Levels 2 and 3, respectively, for two periods; or *ii*) managers with two different values of $h_1$ observed at Levels 2 and 3, respectively, for one period. For all these managers, average priors $\sum_i q_{it}(h_1)p_{i1}$ for any $h_1$ are identified by Proposition 3. $\square$

**Proof of Proposition 5**: The argument consists of two parts. In Part I, I show that assignment probabilities at firm $A$ are identified for each skill type at each level and tenure, conditional on past assignments and true and recorded performance. In Part II, I show that exogenous separation rates are identified. The rest of the proof is in the main text.

*Part I: Identification of conditional level assignment probabilities at firm A by skill type*. Here I show that the conditional assignment probabilities $\Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_1, i)$, $\Pr(L_{A3}|L_{A1} = 1, R_{A1}, L_{A2}, R_{A2}, h_1, i)$, and so on are identified at each level and tenure from *i*) the weights of the mixture distributions of wages of managers of each skill type at the corresponding levels and tenures conditional on their histories at the firm, which are identified by Lemma 1; *ii*) the probabilities of true performance for each manager ability, $\{\alpha_k, \beta_k\}$, which are identified by Proposition 3; and *iii*) the initial priors $\{p_{i1}\}$, which are identified by Proposition 4.

Recall first from the proof of Part I of Proposition 3 that it is immediate to distinguish the probability $\Pr(R_{A1} = 1, i|L_{A1} = 1, L_{A2}, h_1)$ from the probability $\Pr(R_{A1} = 0, i|L_{A1} = 1, L_{A2}, h_1)$, defined in (30), for any given $L_{A2}$, $h_1$, and $i$. Also, the probabilities $\{\Pr(i|L_{A1} = 1, \ldots, L_{At}, h_1)\}$ are identified by Lemma 1 for any $t$. Since the product $\Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1} = 1, h_1, i) \Pr(i|L_{A1} = 1)$ in (35) is identified as argued in the proof of Proposition 3 for each possible $L_{A2}$, $R_{A1}$, $h_1$, and $i$, and no label ambiguity arises with respect to either $R_{A1}$ or $i$, I need only show that $\Pr(R_{A1}|L_{A1}=1, h_1, i)$ is identified to establish that $\Pr(L_{A2}|L_{A1}=1, R_{A1}, h_1, i)$ is identified too.

To this end, note that $\Pr(R_{A1} = 1|L_{A1} = 1, h_1, i) = (\alpha_1 - \beta_1)p_{i1} + \beta_1$, $\Pr(R_{A2} = 1|L_{A1} = 1, R_{A1}, L_{A2}, h_1, i) = (\alpha_{L_{A2}} - \beta_{L_{A2}})p_{i2} + \beta_{L_{A2}}$, where $p_{i2}$ is the updated prior from $p_{i1}$ after $R_{A1}$ is realized at Level 1, and so on for the remaining tenures. Recall that these probabilities are independent of $h_1$. The fact that $\{\alpha_k, \beta_k\}$ and $\{p_{i1}\}$ are identified by Propositions 3 and 4, respectively, implies that $\{p_{it}\}$ are known for any given history of assignments and true performance by Bayes's rule; see (4). Thus, the probabilities $\{\Pr(R_{A1}|L_{A1}=1, h_1, i)\}$, $\{\Pr(R_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\}$, and so on are identified. Now fix $i$ and $R_{A1} = 1$ in the product $\Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1} = 1, h_1, i)$-$\Pr(i|L_{A1} = 1)$ for given $L_{A2}$ and $h_1$. Then, the ratio of this product to the product $[(\alpha_1 - \beta_1)p_{i1} + \beta_1] \Pr(i|L_{A1} = 1)$ identifies $\Pr(L_{A2}|L_{A1}=1, R_{A1}=1, h_1, i)$ for any $L_{A2}$, $h_1$, and $i$. A similar argument applies to $R_{A1}=0$.

By analogous steps, it is possible to show that $\Pr(L_{A3}|L_{A1} = 1, R_{A1}, L_{A2}, R_{A2}, h_1, i)$ is identified from the identified wage mixture weight $\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1)$ for each $R_{A1}$, $R_{A2}$, and $i$ given $L_{A2}$, $L_{A3}$, and $h_1$ by (33). Specifically, the product of each probability on the right side of (33) and the associated probability $\Pr(L_{A2}, L_{A3}|L_{A1} = 1, h_1)$ gives the identified expression

$$\begin{aligned}&\Pr(L_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_1, i) \Pr(R_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\\ &\quad \cdot \Pr(L_{A2}|L_{A1}=1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1}=1, h_1, i) \Pr(i|L_{A1}=1),\end{aligned} \tag{47}$$

whose labeling with respect to $i$ for any given $R_{A1}$, $L_{A2}$, $R_{A2}$, $L_{A3}$, and $h_1$ follows from the labeling of (33) with respect to $i$. Once these probabilities are correctly labeled with respect to $R_{A1}$ and $R_{A2}$, taking the ratio of (47) with, say, $R_{A1}=1$ and $R_{A2}=1$ to the product of probabilities $[(\alpha_{L_{A2}} - \beta_{L_{A2}})p_{i2} + \beta_{L_{A2}}] \Pr(L_{A2}|L_{A1}=1, R_{A1} = 1, h_1, i)[(\alpha_1 - \beta_1)p_{i1} + \beta_1] \Pr(i|L_{A1}=1)$ identifies $\Pr(L_{A3}|L_{A1}=1, R_{A1}=1, L_{A2}, R_{A2}=1, h_1, i)$ for each $L_{A2}$, $L_{A3}$, $h_1$, and $i$. An analogous argument holds for the remaining values of $R_{A1}$ and $R_{A2}$.

Hence, what is left to show is that the probabilities in (47) derived from the mixture weights $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$ of the identified wage mixture density in (32) can be correctly labeled with respect to $R_{A1}$ and $R_{A2}$. To correctly label any such mixture weight and so the probability $\Pr(L_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_1, i)$ with respect to $R_{A2}$, it is sufficient to proceed as in Part I of the proof of Proposition 3 and infer the value of $R_{A2}$ by pairing each mixture weight in $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$ as per (33) of the mixture density $f(w_3|L_{A1}=1, L_{A2}, L_{A3}, h_1)$ with the corresponding mixture weight in $\{\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, R^o_{A2}, L_{A3}, h_1)\}$ as per (34) of the mixture density $f(w_3|L_{A1} = 1, L_{A2}, R^o_{A2}, L_{A3}, h_1)$. Such pairings of mixture weights are possible since the wage mixtures $f(w_3|L_{A1} = 1, L_{A2}, L_{A3}, h_1)$ and $f(w_3|L_{A1} = 1, L_{A2}, R^o_{A2}, L_{A3}, h_1)$ have identical component densities, which are ordered by their means. From the ratios of these paired mixture weights, I can recover $\{\Pr(R^o_{A2}=1|L_{A1}=1, L_{A2}, R_{A2})\}$ for each $L_{A2}$ and $R_{A2}$. Since, as shown in (36), $\{\Pr(R^o_{A2} = 1|L_{A1} = 1, L_{A2}, R_{A2})\}$ are unambiguously ordered with respect to $R_{A2}$ given $L_{A2}$ when $\rho_1 > 0$ by (11), it is then possible to determine the value of $R_{A2}$ of each probability in $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$.

To correctly label any such mixture weight and so the probability $\Pr(L_{A3}|L_{A1}=1,R_{A1},L_{A2},R_{A2},h_1,i)$ with respect to $R_{A1}$, I proceed similarly. Namely, I infer the value of $R_{A1}$ from the associated probability of a recorded high rating in $t=1$ conditional on $R_{A1}$, $\Pr(R^o_{A1}=1|L_{A1}=1,R_{A1})$, which, by (11), is unambiguously ordered with respect to $R_{A1}$ when $\rho_1>0$. To do so, I first need to associate $\Pr(R_{A1},R_{A2},i|L_{A1}=1,L_{A2},L_{A3},h_1)$ with the corresponding probability $\Pr(R_{A1},R_{A2},i|L_{A1}=1,R^o_{A1},L_{A2},L_{A3},h_1)$ to recover the probability $\Pr(R^o_{A1}|L_{A1}=1,R_{A1})$ by an argument analogous to that in the proof of Part I of Proposition 3. Specifically, recall from Lemma 1 that the density of wages at Level $L_{A3}$ in $t=3$ for managers with initial human capital $h_1$ assigned to Level $L_{A2}$ in $t=2$, given by $f(w_{A3}|L_{A1}=1,L_{A2},L_{A3},h_1)$ in (32), is an identified mixture with weights given by (33). Similarly, the density of wages at Level $L_{A3}$ in $t=3$ for managers with initial human capital $h_1$ assigned to Level $L_{A2}$ in $t=2$ and with recorded performance $R^o_{A1}$ in $t=1$ is an identified mixture,

$$f(w_{A3}|L_{A1}=1,R^o_{A1},L_{A2},L_{A3},h_1)=\textstyle\sum_{R_{A1},R_{A2},i}f(w_{A3}|L_{A1}=1,R^o_{A1},R_{A1},L_{A2},R_{A2},L_{A3},h_1,i)$$
$$\cdot\Pr(R_{A1},R_{A2},i|L_{A1}=1,R^o_{A1},L_{A2},L_{A3},h_1),$$

where $f(w_{A3}|L_{A1}=1,R^o_{A1},R_{A1},L_{A2},R_{A2},L_{A3},h_1,i)$ is independent of $R^o_{A1}$ and

$$\Pr(R_{A1},R_{A2},i|L_{A1}=1,R^o_{A1},L_{A2},L_{A3},h_1)=\frac{\begin{array}{c}\Pr(L_{A3}|L_{A1}=1,R_{A1},L_{A2},R_{A2},h_1,i)\Pr(R_{A2}|L_{A1}=1,R_{A1},L_{A2},h_1,i)\\ \cdot\Pr(L_{A2}|L_{A1}=1,R_{A1},h_1,i)\Pr(R^o_{A1},R_{A1}|L_{A1}=1,h_1,i)\Pr(i|L_{A1}=1)\end{array}}{\Pr(R^o_{A1},L_{A2},L_{A3}|L_{A1}=1,h_1)},\tag{48}$$

since $\Pr(L_{A3}|\cdot,i)$, $\Pr(R_{A2}|\cdot,i)$, and $\Pr(L_{A2}|\cdot,i)$ are independent of $R^o_{A1}$. With $f(w_{A3}|L_{A1}=1,R^o_{A1},R_{A1},L_{A2},R_{A2},L_{A3},h_1,i)$ independent of $R^o_{A1}$, the component densities of the wage mixtures $f(w_{A3}|L_{A1}=1,L_{A2},L_{A3},h_1)$ and $f(w_{A3}|L_{A1}=1,R^o_{A1},L_{A2},L_{A3},h_1)$ are identical for any $R^o_{A1}$—given $L_{A2}$, $L_{A3}$, and $h_1$—and are ordered by their means. Then, the weights in (33) and (48) can be uniquely matched through their corresponding component densities.

Finally, the value of $R_{A1}$ in the mixture weights in (33) and (48) can be determined as follows. First, compute the product of the term on the right side of (48) and the corresponding $\Pr(R^o_{A1},L_{A2},L_{A3}|L_{A1}=1,h_1)$. Next, compute the product of the term on the right side of (33) and the corresponding $\Pr(L_{A2},L_{A3}|L_{A1}=1,h_1)$. Then, take the ratio of these two products—paired by the means of their associated component densities—for given $R_{A1}$, $R_{A2}$, $i$, $R^o_{A1}$, $L_{A2}$, $L_{A3}$, and $h_1$, which gives $\Pr(R^o_{A1}|L_{A1}=1,R_{A1},h_1,i)$ that is independent of $h_1$ and $i$ by (11). Thus, by comparing $\Pr(R^o_{A1}|L_{A1}=1,R_{A1})$ for given $R^o_{A1}$ across the two possible values of $R_{A1}$ and using (36), it is possible to infer the value of $R_{A1}$ in $\Pr(R^o_{A1}|L_{A1}=1,R_{A1})$ and so in $\Pr(R_{A1},R_{A2},i|L_{A1}=1,L_{A2},L_{A3},h_1)$. This argument can be extended to the remaining tenures. It also proves that the component densities of the mixture distributions of wages at firm $A$ of managers of each skill type, conditional on their histories of level assignments (and performance ratings) at the firm, and their mixture weights can be correctly labeled with respect to $R_{At},\ldots,R_{At-1}$ at each level and tenure.

*Part II: Identification of exogenous separation rates*. Observe that $\eta_{k_{t-1}t-1}$, which is the probability that a manager employed in $t-1$ by $A$ at Level $k_{t-1}$ is still in the market at $t$, can be recovered as the ratio of the probability of employment at firms $A$ or $C$ at a state relative to the probability of employment at $A$ in the previous period at Level $k_{t-1}$. Since the probabilities of employment and assignments at $A$, as argued in Part I, are identified, I will proceed by showing how it is possible to pin down the probability of employment at $C$ at such a state and so recover $\eta_{k_{t-1}t-1}$.

To start, since assignment probabilities at $A$—namely, $\Pr(L_{A2}|L_{A1}=1,R_{A1},h_1,i)$, $\Pr(L_{A3}|L_{A1}=1,R_{A1},L_{A2},R_{A2},h_1,i)$, and so on—are identified by Part I, the probability of retention at $A$ of managers of each skill type is also identified for any history of level assignments, true performance, and initial level of human capital as $\Pr(L_{At}>0|L_{A1}=1,R_{A1},\ldots,L_{At-1},R_{At-1},h_1,i)=\sum_{k\geq 1}\Pr(L_{At}=k|L_{A1}=1,R_{A1},\ldots,L_{At-1},R_{At-1},h_1,i)$. Denote more compactly by $\Pr(L_{At}=k,f_t=A|s_{it})$ and $\Pr(f_t=A|s_{it})$ the probability of employment at Level $k$ of $A$ and of employment at $A$ at state $s_{it}$, respectively, conditional on a manager's survival in the market up to $t$. By Proposition 1 and the assumption of Gumbel productivity shocks, I can express $S(s_{it},\varepsilon_t)$ in Proposition 1 as $S(s_{it},\varepsilon_t)=\max_f\{\varsigma(s_{it},k_{ft},f)+\varepsilon_{fk_{ft}t}\}$, where $k_{ft}=k_f(s_{it},\varepsilon_t)$ achieves the value $\overline{V}^f(s_{it},\varepsilon_t)$. Hence,

$$\varsigma(s_{it},k_{ft},f)=y_f(s_{it},k_{ft})+\delta\eta_{k_{ft}t}E\left(\ln\textstyle\sum_{f'=A,C}\exp\{\varsigma(s_{it+1},k_{f't+1},f')\}|s_{it},k_{ft}\right),$$

so that $\Pr(f_t=A|L_{At}=k,L_{Ct}=k',s_{it})=\exp\{\varsigma(s_{it},k,A)\}/[\exp\{\varsigma(s_{it},k,A)\}+\exp\{\varsigma(s_{it},k',C)\}]$ conditional on a manager's survival in the market, where $\{f_t = A\}$ is the event that a manager is employed by firm $A$, and

$$\Pr(L_{At}=k|f_t=A,s_{it}) = \frac{\sum_{k'\geq 1}\Pr(f_t=A|L_{At}=k,L_{Ct}=k',s_{it})\Pr(L_{At}=k,L_{Ct}=k'|s_{it})}{\Pr(f_t=A|s_{it})}$$
$$= \frac{\Pr(L_{At}=k|s_{it})\sum_{k'\geq 1}\Pr(f_t=A|L_{At}=k,L_{Ct}=k',s_{it})\Pr(L_{Ct}=k'|s_{it})}{\sum_{k\geq 1}\Pr(L_{At}=k|s_{it})\sum_{k'\geq 1}\Pr(f_t=A|L_{At}=k,L_{Ct}=k',s_{it})\Pr(L_{Ct}=k'|s_{it})},$$

where the last equality follows from the independence of productivity shocks across firms and the law of total probability. An implication of Proposition 1 is that $\Pr(L_{At}=k|f_t=A,s_{it}) = \Pr(L_{At}=k|f_t=C,s_{it})$, so that

$$\Pr(L_{At}=k|s_{it}) = \Pr(L_{At}=k,f_t=A|s_{it}) + \Pr(L_{At}=k,f_t=C|s_{it})$$
$$= \Pr(L_{At}=k|f_t=A,s_{it})\left[\Pr(f_t=A|s_{it})+\Pr(f_t=C|s_{it})\right] = \Pr(L_{At}=k|f_t=A,s_{it}).$$

Thus, $\Pr(L_{At}=k|s_{it})=\Pr(L_{At}=k|f_t=f,s_{it})$ and $\Pr(L_{Ct}=k|s_{it})=\Pr(L_{Ct}=k|f_t=f,s_{it})$, $f=A,C$. Observe that

$$\Pr(L_{At}=k,f_t=A|s_{it}) = \textstyle\sum_{k'\geq 1}\Pr(f_t=A|L_{At}=k,L_{Ct}=k',s_{it})\Pr(L_{At}=k,L_{Ct}=k'|s_{it})$$
$$= \textstyle\sum_{k'\geq 1}\Pr(f_t=A|L_{At}=k,L_{Ct}=k',s_{it})\Pr(L_{At}=k|s_{it})\Pr(L_{Ct}=k'|s_{it})$$
$$= \sum_{k'\geq 1}\frac{\Pr(L_{At}=k|s_{it})\Pr(L_{Ct}=k'|s_{it})}{1+\exp\{\varsigma(s_{it},k',C)-\varsigma(s_{it},k,A)\}} = \sum_{k'\geq 1}\frac{\Pr(L_{At}=k|s_{it})\Pr(L_{Ct}=k'|s_{it})}{1+\exp\{\varsigma(s_{it},k',A)-\varsigma(s_{it},k,A)+\gamma(s_{it},k')\}}, \quad (49)$$

where $\gamma(s_{it},k')$ is defined as $y_C(s_{it},k')-y_A(s_{it},k')$ and the last equality follows from the law of motion of the state being equal across the two firms by job level. Similarly,

$$\Pr(L_{Ct}=k,f_t=C|s_{it}) = \sum_{k'\geq 1}\frac{\Pr(L_{At}=k'|s_{it})\Pr(L_{Ct}=k|s_{it})}{1+\exp\{\varsigma(s_{it},k',A)-\varsigma(s_{it},k,A)-\gamma(s_{it},k)\}}. \quad (50)$$

Recall by (12) that $Ew_{Ait}(k)=y_C(s_{it},k)-\ln\Pr(L_{Ct}=k|f_t=C,s_{it})$ holds at all $s_{it}$. Select a state $s_{it}$ in $t$ and set $y_C(s_{it},k)$ equal to $\tilde{y}(s_{it})$, defined as $Ew_{Ait}(k)+\ln\Pr(L_{At}=k|f_t=A,s_{it})$, at such an $s_{it}$ for all $k$, which implies that $\Pr(L_{At}=k|f_t=A,s_{it})=\Pr(L_{Ct}=k|f_t=C,s_{it})$ and so $\Pr(L_{At}=k|s_{it})=\Pr(L_{Ct}=k|s_{it})$ at such an $s_{it}$ for all $k$. If, in addition, $\gamma(s_{it},k)=0$ for all $k$ at such a state, then $\Pr(f_t=A|s_{it})=\Pr(f_t=C|s_{it})$ at such a state by (49) and (50). Hence, $\eta_{k_{t-1}t-1}$ can be recovered from the probability of employment at $A$ or $C$ at such $s_{it}$ of managers of skill type $i$ employed at Level $k_{t-1}$ of $A$ at a state $s_{it-1}$, from which the chosen $s_{it}$ can be reached, and the probability of employment of these managers at Level $k_{t-1}$ of $A$ at such a state $s_{it-1}$. □

**Proof of Proposition 6**: That $C$'s expected match surplus value can be recovered from wages at $A$ follows immediately by expressing $Ew_{Ait}(k)$ as $Ew_{Ait}(k)=\ln\sum_{k'}\exp\{v^C(s_{it},k')\}-\delta\eta_{kt}E(\ln\sum_{k'}\exp\{v^C(s_{it+1},k')\}|s_{it},k)$ by (12) and (28). Then, the functional equation $\ln\sum_{k'}\exp\{v^C(s_{it},k')\}=Ew_{Ait}(k)+\delta\eta_{kt}E(\ln\sum_{k'}\exp\{v^C(s_{it+1},k')\}|s_{it},k)$ identifies $C$'s expected match surplus value $\ln\sum_{k'}\exp\{v^C(s_{it},k')\}$ at each state. The rest of the proof shows how two level normalizations at each state are sufficient to further recover $C$'s expected output at each level. I consider here normalizations in the spirit of those common in dynamic discrete choice models. Specifically, they rely on knowledge of expected output at one job level of $C$ or of sums or differences of $C$'s expected output at two levels.

Consider case a). Assume that $y_C(s_{it},2)$, and so $\Pr(L_{Ct}=2|f_t=C,s_{it})$ by (12), and $y_C(s_{it},1)+y_C(s_{it},3)$ are known at two values of $p_{it}$ for each $\kappa_t$ and $i$. (The case in which $y_C(s_{it},2)$ and either $y_C(s_{it},1)+y_C(s_{it},2)$ or $y_C(s_{it},2)+y_C(s_{it},3)$ are known at two values of $p_{it}$ for each $\kappa_t$ and $i$ is trivial, since expected output at two job levels of $C$ is then identified at these states, and so at all other states, together with the probabilities $\{\Pr(L_{Ct}=k|f_t=C,s_{it})\}_k$. Then, expected output at the remaining level $k$ is identified as $y_C(s_{it},k)=Ew_{Ait}(k)+\ln\Pr(L_{Ct}=k|f_t=C,s_{it})$.) Since

$$Ew_{Ait}(1)+Ew_{Ait}(3)=y_C(s_{it},1)+y_C(s_{it},3)-\ln[\Pr(L_{Ct}=1|f_t=C,s_{it})\Pr(L_{Ct}=3|f_t=C,s_{it})],$$

knowledge of $y_C(s_{it},1)+y_C(s_{it},3)$ implies that $\Pr(L_{Ct}=1|f_t=C,s_{it})\Pr(L_{Ct}=3|f_t=C,s_{it})$ is identified from

$$\ln[\Pr(L_{Ct}=1|f_t=C,s_{it})\Pr(L_{Ct}=3|f_t=C,s_{it})] = \underbrace{y_C(s_{it},1)+y_C(s_{it},3)-Ew_{Ait}(1)-Ew_{Ait}(3)}_{\ln m_{13}(s_{it})}. \quad (51)$$

Using this condition and that the probabilities $\{\Pr(L_{Ct}=k|f_t=C, s_{it})\}_k$ sum to one, evaluated at one of the two chosen values of $s_{it}$, I obtain a system of two equations in the two unknowns $\Pr(L_{Ct}=1|f_t=C, s_{it})$ and $\Pr(L_{Ct}=3|f_t=C, s_{it})$,

$$\begin{cases} \Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=3|f_t=C, s_{it}) \equiv xz = m_{13}(s_{it}) \\ \Pr(L_{Ct}=1|f_t=C, s_{it}) + \Pr(L_{Ct}=3|f_t=C, s_{it}) \equiv x + z = 1 - \Pr(L_{Ct}=2|f_t=C, s_{it}) \equiv m_2(s_{it}) \end{cases},$$

which admits two solutions,

$$\begin{gathered} \left(x=[m_2(s_{it}) - \sqrt{m_2^2(s_{it}) - 4m_{13}(s_{it})}]/2, z=[m_2(s_{it}) + \sqrt{m_2^2(s_{it}) - 4m_{13}(s_{it})}]/2\right) \\ \left(x'=[m_2(s_{it}) + \sqrt{m_2^2(s_{it}) - 4m_{13}(s_{it})}]/2, z'=[m_2(s_{it}) - \sqrt{m_2^2(s_{it}) - 4m_{13}(s_{it})}]/2\right), \end{gathered}$$

and, correspondingly, two (symmetric) solutions for $y_C(s_{it}, 1)$ and $y_C(s_{it}, 3)$. This multiplicity can be resolved by checking whether $\Pr(L_{Ct} = 1|f_t = C, s_{it})$ is smaller or greater than $\Pr(L_{Ct} = 3|f_t = C, s_{it})$ or, equivalently, by using (12) and verifying whether the difference $2y_C(s_{it}, 1) - Ew_{Ait}(1)$ is smaller or greater than the value $y_C(s_{it}, 1) + y_C(s_{it}, 3) - Ew_{Ait}(3)$, which is identified by the assumptions of the case. For instance, if $2y_C(s_{it}, 1) - Ew_{Ait}(1)$, computed using the first solution, is smaller than $y_C(s_{it}, 1) + y_C(s_{it}, 3) - Ew_{Ait}(3)$, then the first solution is the correct one; if not, the second solution is. Alternatively, one could check whether $2y_C(s_{it}, 1) - Ew_{Ait}(1)$, computed using the second solution, is larger than $y_C(s_{it}, 1) + y_C(s_{it}, 3) - Ew_{Ait}(3)$, in which case the second solution is the correct one; if not, the first solution is. By repeating this argument at the second state $s'_{it}$ at which $y_C(s'_{it}, 2)$ and $y_C(s'_{it}, 1) + y_C(s'_{it}, 3)$ are known, it follows that $y_C(s_{it}, k) = d_{Ck}(\kappa_t, i) + e_{Ck}(\kappa_t, i)p_{it}$ and $y_C(s'_{it}, k) = d_{Ck}(\kappa_t, i) + e_{Ck}(\kappa_t, i)p'_{it}$ are identified for each $k$, from which $d_{Ck}(\kappa_t, i)$ and $e_{Ck}(\kappa_t, i)$ can be recovered. This argument can be repeated for all $\kappa_t$ and $i$. Hence, $\{y_C(s_{it}, k)\}_{k=1,3}$ are identified. The remaining case under a) has been considered in the main text.

Consider case b). Suppose first that $y_C(s_{it}, 1) + y_C(s_{it}, 2)$ and $y_C(s_{it}, 1) + y_C(s_{it}, 3)$ are known at two states $s_{it}$ with distinct values of $p_{it}$ and the same values of $\kappa_t$ and $i$. Given that

$$Ew_{Ait}(1) + Ew_{Ait}(2) = y_C(s_{it}, 1) + y_C(s_{it}, 2) - \ln[\Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=2|f_t=C, s_{it})],$$

knowledge of $y_C(s_{it}, 1) + y_C(s_{it}, 2)$ implies that $\Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=2|f_t=C, s_{it})$ is identified from

$$\ln[\Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=2|f_t=C, s_{it})] = \underbrace{y_C(s_{it}, 1) + y_C(s_{it}, 2) - Ew_{Ait}(1) - Ew_{Ait}(2)}_{\ln m_{12}(s_{it})}. \tag{52}$$

As noted, knowledge of $y_C(s_{it}, 1) + y_C(s_{it}, 3)$ implies that $\Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=3|f_t=C, s_{it})$ is identified by (51). Using these conditions and that the probabilities $\{\Pr(L_{Ct}=k|f_t=C, s_{it})\}_k$ sum to one, evaluated at one of the two chosen values of $s_{it}$, I obtain a system of three equations in the three unknowns $\{\Pr(L_{Ct}=k|f_t=C, s_{it})\}_k$,

$$\begin{cases} \Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=2|f_t=C, s_{it}) \equiv xy = m_{12}(s_{it}) \\ \Pr(L_{Ct}=1|f_t=C, s_{it})\Pr(L_{Ct}=3|f_t=C, s_{it}) \equiv xz = m_{13}(s_{it}) \\ \Pr(L_{Ct}=1|f_t=C, s_{it}) + \Pr(L_{Ct}=2|f_t=C, s_{it}) + \Pr(L_{Ct}=3|f_t=C, s_{it}) \equiv x + y + z = 1 \end{cases},$$

which admits two solutions,

$$\begin{gathered} \left(x=\tfrac{1-\sqrt{1-4[m_{12}(s_{it})+m_{13}(s_{it})]}}{2}, y=\tfrac{2m_{12}(s_{it})}{1-\sqrt{1-4[m_{12}(s_{it})+m_{13}(s_{it})]}}, z=\tfrac{2m_{13}(s_{it})}{1-\sqrt{1-4[m_{12}(s_{it})+m_{13}(s_{it})]}}\right) \\ \left(x'=\tfrac{1+\sqrt{1-4[m_{12}(s_{it})+m_{13}(s_{it})]}}{2}, y'=\tfrac{2m_{12}(s_{it})}{1+\sqrt{1-4[m_{12}(s_{it})+m_{13}(s_{it})]}}, z'=\tfrac{2m_{13}(s_{it})}{1+\sqrt{1-4[m_{12}(s_{it})+m_{13}(s_{it})]}}\right), \end{gathered}$$

the first one with $\Pr(L_{Ct}=1|f_t=C, s_{it})$ smaller than $1/2$ and the second one with $\Pr(L_{Ct}=1|f_t=C, s_{it})$ greater than $1/2$. But this multiplicity can be resolved by using (52) and (51) and checking whether $m_{12}(s_{it})$ or $m_{13}(s_{it})$, which are both identified by the assumptions of the case, is bounded above by $y/2$ and $z/2$, respectively, or below by $y'/2$ and $z'/2$, respectively. Specifically, if $m_{12}(s_{it})$ is smaller than $y/2$ or $m_{13}(s_{it})$ is smaller than $z/2$, then the first solution is the correct one. If $m_{12}(s_{it})$ is instead greater than $y'/2$ or $m_{13}(s_{it})$ is greater than $z'/2$, then the second solution is the correct one. By repeating this argument at the second state $s'_{it}$ at which $y_C(s'_{it}, 1) + y_C(s'_{it}, 2)$ and $y_C(s'_{it}, 1) + y_C(s'_{it}, 3)$ are known, it follows that $y_C(s_{it}, k) = d_{Ck}(\kappa_t, i) + e_{Ck}(\kappa_t, i)p_{it}$ and $y_C(s'_{it}, k) = d_{Ck}(\kappa_t, i) + e_{Ck}(\kappa_t, i)p'_{it}$ are identified for each $k$, from which $d_{Ck}(\kappa_t, i)$ and $e_{Ck}(\kappa_t, i)$ can be recovered. This argument can be repeated for all $\kappa_t$ and $i$. Thus, $\{y_C(s_{it}, k)\}$ are identified.

Alternatively, suppose that $y_C(s_{it}, 1) - y_C(s_{it}, 2)$ and $y_C(s_{it}, 3) - y_C(s_{it}, 2)$ are known at two states $s_{it}$ with distinct values of $p_{it}$ and the same values of $\kappa_t$ and $i$. Note that $\overline{v}^C(s_{it}, 1) - \overline{v}^C(s_{it}, 2) = Ew_{Ait}(2) - Ew_{Ait}(1) - [y_C(s_{it}, 2) - y_C(s_{it}, 1)]$ and $\overline{v}^C(s_{it}, 3) - \overline{v}^C(s_{it}, 2) = Ew_{Ait}(2) - Ew_{Ait}(3) - [y_C(s_{it}, 2) - y_C(s_{it}, 3)]$ by (12) and the relationship between probabilities and values, $\Pr(L_{ft} = k | f_t = f, s_{it}) = e^{\overline{v}^f(s_{it}, k)} / \sum_{k'} e^{\overline{v}^f(s_{it}, k')}$, already used in (15). From

$$\begin{cases} Ew_{Ait}(1) = y_C(s_{it}, 1) + \ln\left(1 + e^{\overline{v}^C(s_{it},2) - \overline{v}^C(s_{it},1)} + e^{\overline{v}^C(s_{it},3) - \overline{v}^C(s_{it},2) + \overline{v}^C(s_{it},2) - \overline{v}^C(s_{it},1)}\right) \\ Ew_{Ait}(2) = y_C(s_{it}, 2) + \ln\left(e^{\overline{v}^C(s_{it},1) - \overline{v}^C(s_{it},2)} + 1 + e^{\overline{v}^C(s_{it},3) - \overline{v}^C(s_{it},2)}\right) \\ Ew_{Ait}(3) = y_C(s_{it}, 3) + \ln\left(e^{\overline{v}^C(s_{it},1) - \overline{v}^C(s_{it},2) + \overline{v}^C(s_{it},2) - \overline{v}^C(s_{it},3)} + e^{\overline{v}^C(s_{it},2) - \overline{v}^C(s_{it},3)} + 1\right) \end{cases},$$

it then follows that $\{y_C(s_{it}, k)\}$ are identified at the chosen state. By repeating this argument at the second state $s'_{it}$ at which $y_C(s'_{it}, 1) - y_C(s'_{it}, 2)$ and $y_C(s'_{it}, 3) - y_C(s'_{it}, 2)$ are known, it follows that $y_C(s_{it}, k) = d_{Ck}(\kappa_t, i) + e_{Ck}(\kappa_t, i) p_{it}$ and $y_C(s'_{it}, k) = d_{Ck}(\kappa_t, i) + e_{Ck}(\kappa_t, i) p'_{it}$ are identified for each $k$, from which $d_{Ck}(\kappa_t, i)$ and $e_{Ck}(\kappa_t, i)$ can be recovered. By repeating this argument for all $\kappa_t$ and $i$, it follows, as before, that $\{y_C(s_{it}, k)\}$ are identified.

This logic can be extended to any combination of job levels. □

# B Empirical Appendix

See Section 4 in the S.A. for the likelihood function and other omitted details.

**B.1 Human Capital Process.** Abstract first from skill types. Here I show that instances of (1) can be derived from standard laws of motion of human capital in the spirit of Heckman, Lochner, and Taber (1998, equation (I.4)):

$$H^j_{fkt} = C^j_{fkt}[(1 - \sigma^j) H^j_{fkt-1} + z^j_{fkt-1} (H^j_{fkt-1})^{\lambda_j} (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] \quad j = G, S, \tag{53}$$

where $H^j_{fk1} = C^j_{fk1}$, $C^j_{fkt}$ is total factor productivity, $\sigma^j$ is the depreciation rate of human capital, and $\lambda_j \in \{0, 1\}$. The instances of (1) derived here amount to special cases of (16).

***Case 1:*** $\lambda_j = 0$. If $\sigma^j \in [0, 1)$, then (53) becomes $H^j_{fkt} = C^j_{fkt}[(1 - \sigma^j) H^j_{fkt-1} + z^j_{fkt-1} (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}]$. Let $z^j_{fkt} = C^j_{fkt} C^j_{fkt-1} \cdots C^j_{fk1}$ with $z^j_{fk1} = C^j_{fk1}$. It follows that

$$H^j_{fk2} = C^j_{fk2}[(1-\sigma^j) H^j_{fk1} + z^j_{fk1} (i^j_{k_1})^{\mu_{jk_1}}] = (1-\sigma^j) C^j_{fk2} C^j_{fk1} + C^j_{fk2} z^j_{fk1} (i^j_{k_1})^{\mu_{jk_1}} = C^j_{fk2} C^j_{fk1}[1 - \sigma^j + (i^j_{k_1})^{\mu_{jk_1}}],$$
$$H^j_{fk3} = C^j_{fk3}[(1-\sigma^j) H^j_{fk2} + z^j_{fk2} (i^j_{k_2})^{\mu_{jk_2}}] = C^j_{fk3}[C^j_{fk2} C^j_{fk1} (1-\sigma^j)^2 + C^j_{fk2} C^j_{fk1} (1-\sigma^j) (i^j_{k_1})^{\mu_{jk_1}} + C^j_{fk2} C^j_{fk1} (i^j_{k_2})^{\mu_{jk_2}}],$$

and so on. Describe the process for $C^j_{fkt}$ as $A^j_{fkt} \epsilon^j_{fkt} = C^j_{fkt} C^j_{fkt-1} \cdots C^j_{fk1}$, which can be expressed as $\ln(A^j_{fkt} \epsilon^j_{fkt}) = \ln(A^j_{fkt-1} \epsilon^j_{fkt-1}) + \ln C^j_{fkt}$ so that

$$H^j_{fkt} = A^j_{fkt}[(1-\sigma^j)^{t-1} + (1-\sigma^j)^{t-2} (i^j_{k_1})^{\mu_{jk_1}} + \ldots + (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] \epsilon^j_{fkt}. \tag{54}$$

Consider first $H^G_{fkt}$. Since $i^G_{k_{t-1}} = \bar{i}$ for managers in the market in $t-1$, if $a^G_{fkt} = \ln A^G_{fkt}$, $\sigma^G = 0$, $\mu_{Gk_\tau} = \mu_G$ for $\tau \geq 1$, $\eta_G = \bar{i}^{\mu_G}$, and $\varepsilon^G_{fkt} = \ln \epsilon^G_{fkt}$, then the law of motion of $H^G_{fkt}$ in (54) can be expressed as a special case of (16):

$$h^G_{fkt} = \ln H^G_{fkt} = \ln A^G_{fkt} + \ln[1 + (t-1)\bar{i}^{\mu_G}] + \ln \epsilon^G_{fkt} \simeq a^G_{fkt} + (t-1)\eta_G + \varepsilon^G_{fkt}.$$

Consider now $H^S_{fkt}$. As $i^S_{k_{t-1}} = \bar{i}_{k_{t-1}}$ for managers at Level $k_{t-1}$ in $t-1$, if $\sigma^S = 1$, then (54) becomes $H^S_{fkt} = A^S_{fkt} (\bar{i}_{k_{t-1}})^{\mu_{Sk_{t-1}}} \epsilon^S_{fkt}$. With $a^S_{fkt} = \ln A^S_{fkt}$, $\eta_{Sk_{t-1}} = \mu_{Sk_{t-1}} \ln \bar{i}_{k_{t-1}}$, and $\varepsilon^S_{fkt} = \ln \epsilon^S_{fkt}$, I obtain a special case of (16):

$$h^S_{fkt} = \ln H^S_{fkt} = \ln A^S_{fkt} + \mu_{Sk_{t-1}} \ln \bar{i}_{k_{t-1}} + \ln \epsilon^S_{fkt} = a^S_{fkt} + \eta_{Sk_{t-1}} + \varepsilon^S_{fkt}. \tag{55}$$

***Case 2:*** $\lambda_j = 1$. If $\sigma^j \in [0, 1)$, then (53) becomes $H^j_{fkt} = C^j_{fkt}[1 - \sigma^j + z^j_{fkt-1} (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] H^j_{fkt-1}$. Thus, for instance, $H^j_{fk2} = C^j_{fk2}[1 - \sigma^j + z^j_{fk1} (i^j_{k_1})^{\mu_{jk_1}}] H^j_{fk1} = C^j_{fk2} C^j_{fk1}[1 - \sigma^j + z^j_{fk1} (i^j_{k_1})^{\mu_{jk_1}}]$ so that

$$H^j_{fkt} = C^j_{fkt} \cdots C^j_{fk1}[1 - \sigma^j + z^j_{fk1} (i^j_{k_1})^{\mu_{jk_1}}] \cdots [1 - \sigma^j + z^j_{fkt-1} (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}].$$

As in the previous case, defining $A^j_{fkt}\epsilon^j_{fkt}=C^j_{fkt}C^j_{fkt-1}\cdots C^j_{fk1}$, it follows that

$$H^j_{fkt}=A^j_{fkt}[1-\sigma^j+z^j_{fk1}(i^j_{k_1})^{\mu_{jk_1}}]\cdots[1-\sigma^j+z^j_{fkt-1}(i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}]\epsilon^j_{fkt}. \tag{56}$$

Consider first $H^G_{fkt}$. Since $i^G_{k_{t-1}}=\bar{i}$ for managers in the market in $t-1$, if $a^G_{fkt}=\ln A^G_{fkt}$, $\sigma^G=0$, $z^G_{fkt-1}$ does not depend on $f$, $k$, or $t-1$, $\mu_{Gk_\tau}=\mu_G$ for $\tau\geq 1$, $\eta_G$ is redefined as $\eta_G=\ln(1+z^G\bar{i}^{\mu_G})$, and $\varepsilon^G_{fkt}=\ln\epsilon^G_{fkt}$, then the law of motion of $H^G_{fkt}$ can be expressed as in Case 1:

$$h^G_{fkt}=\ln H^G_{fkt}=\ln A^G_{fkt}+(t-1)\ln(1+z^G\bar{i}^{\mu_G})+\ln\epsilon^G_{fkt}=a^G_{fkt}+(t-1)\eta_G+\varepsilon^G_{fkt}.$$

Consider now $H^S_{fkt}$. Recall that $i^S_{k_{t-1}}=\bar{i}_{k_{t-1}}$ for managers at Level $k_{t-1}$ in $t-1$. Suppose that $\sigma^S=1$ and $z^S_{fkt-1}=C^S_{fkt-1}\cdots C^S_{fk1}/H^S_{fkt-1}$ so that (53) becomes $H^S_{fkt}=C^S_{fkt}\cdots C^S_{fk1}(\bar{i}_{k_{t-1}})^{\mu_{Sk_{t-1}}}$. With $A^S_{fkt}\epsilon^S_{fkt}=C^S_{fkt}\cdots C^S_{fk1}$, then, $h^S_{fkt}$ can be expressed as in (55). Analogous expressions can be obtained once skill types are introduced. □

**B.2 Derivation of Expected Output.** Expression (17) follows from (10), the fact that $\underline{y}^e_{fkt}(i)=b_k(h_1)+\sum_j[\underline{a}^j_{fkt}+\underline{a}^j_{fk}(i)+g^j_k(\cdot)]$, and the discussion after (16). Specifically,

$$\begin{aligned}y_f(s_{it},k)&=\underline{y}^e_{fkt}(i)+\beta_k[\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)]+(\alpha_k-\beta_k)[\overline{y}_{fkt}(i)-\underline{y}_{fkt}(i)]p_{it}=b_k(h_1)+\textstyle\sum_j[\underline{a}^j_{fkt}+\underline{a}^j_{fk}(i)+g^j_k(\cdot)]\\&\quad+\beta_k\textstyle\sum_j[\overline{a}^j_{fkt}-\underline{a}^j_{fkt}+\overline{a}^j_{fk}(i)-\underline{a}^j_{fk}(i)]+(\alpha_k-\beta_k)\sum_j[\overline{a}^j_{fkt}-\underline{a}^j_{fkt}+\overline{a}^j_{fk}(i)-\underline{a}^j_{fk}(i)]p_{it},\end{aligned} \tag{57}$$

where $d_{fkt}(k_{t-1},i)$ is defined as $\sum_j[\underline{a}^j_{fkt}+\underline{a}^j_{fk}(i)+g^j_k(\cdot)]+\beta_k\sum_j[\overline{a}^j_{fkt}-\underline{a}^j_{fkt}+\overline{a}^j_{fk}(i)-\underline{a}^j_{fk}(i)]$. □

**B.3 Market-Wide Assignment Problem.** I now specialize the environment considered so far and show that if the entry of firms in the labor market considered is free in that firms enter until the cost of entry $c_f$ equals the expected present discounted value of profits after entry, then under the assumption of no recall, the solution to the market-wide assignment problem in Proposition 1 implies a more convenient expression for the surplus value $S(s_{it},\varepsilon_t)$ derived there. Formally:

**Lemma 2.** *Suppose that entry in the labor market is free and that firms A and C have incurred the cost $c_f$ to enter. If $\Pi^f(\cdot|f')=0$ for either firm $f\neq f'$ after separating from a manager it has previously employed (no recall), then*

$$S(s_{it},\varepsilon_t)=\max_f\left\{\max_{k\in K^f}\left\{y_f(s_{it},k)+\varepsilon_{fkt}+\delta\eta_{kt}\int_{\varepsilon_{t+1}}ES(s_{it+1},\varepsilon_{t+1}|s_{it},k)dG\right\}\right\} \tag{58}$$

*along the equilibrium sample paths in which a manager is either employed by A or first employed by A and then by C. When productivity shocks are Gumbel distributed with mean zero and variance $\pi^2/6$, $S(s_{it},\varepsilon_t)$ can be expressed as*

$$S(s_{it},\varepsilon_t)=\max\left\{\max_{k\in K^A}\{\varsigma^A(s_{it},k)+\varepsilon_{Akt}\},\max_{k\in K^C}\{\varsigma^C(s_{it},k)+\varepsilon_{Ckt}\}\right\},$$

*where $\varsigma^f(s_{it},k)=y_f(s_{it},k)+\delta\eta_{kt}E[\ln(\sum_{k'\in K^A}\exp\{\varsigma^A(s_{it+1},k')\}+\sum_{k'\in K^C}\exp\{\varsigma^C(s_{it+1},k')\})|s_{it},k]$. Also, $\overline{V}^f(s_{it},\varepsilon_t)=\max_{k\in K^f}\{\overline{\varsigma}^f(s_{it},k)+\varepsilon_{fkt}\}$, where $\overline{\varsigma}^f(s_{it},k)=y_f(s_{it},k)+\delta\eta_{kt}E(\ln\sum_{k'\in K^f}\exp\{\overline{\varsigma}^f(s_{it+1},k')\}|s_{it},k)$.*

To establish this result, consider first states at which firm $A$ employs a manager. Thus, $S(\cdot)=S(\cdot|A)$, where $S(\cdot|A)=V^A(\cdot|A)+\Pi^C(\cdot|A)$. Since $\Pi^C(\cdot|A)=c_f$ is independent of $k_{At}$ by the assumption of free entry and, by definition, $k_{At}$ maximizes firm $A$'s value of profits as well as, by Proposition 1, $V^A(\cdot|A)$, it follows that $k_{At}$ also maximizes $S(\cdot|A)$. Then, $S(\cdot|A)=\max_k\{y_A(s_{it},k)+\varepsilon_{Akt}+\delta\eta_{kt}\int_{\varepsilon_{t+1}}ES(\cdot)dG\}$. Off the equilibrium path, $S(\cdot|C)=\Pi^A(\cdot|C)+V^C(\cdot|C)$ and, given that $\Pi^A(\cdot|C)=0$ by the assumption of no recall, the choice of $k_{Ct}$ by firm $C$, which, by definition, maximizes $\Pi^C(\cdot|C)$ and, by Proposition 1, $V^C(\cdot|C)$, also maximizes $S(\cdot|C)$. With $\Pi^A(\cdot|A)\geq\Pi^A(\cdot|C)$, $W(\cdot|A)=W(\cdot|C)$, and $\Pi^C(\cdot|A)=\Pi^C(\cdot|C)$ at all states at which $A$ employs a manager in equilibrium by Proposition 1, it follows that

$$\begin{aligned}S(\cdot|A)&=\max_{k\in K^A}\left\{y_A(s_{it},k)+\varepsilon_{Akt}+\delta\eta_{kt}\int_{\varepsilon_{t+1}}ES(s_{it+1},\varepsilon_{t+1}|s_{it},k)dG\right\}\\&\geq S(\cdot|C)=\max_{k\in K^C}\left\{y_C(s_{it},k)+\varepsilon_{Ckt}+\delta\eta_{kt}\int_{\varepsilon_{t+1}}ES(s_{it+1},\varepsilon_{t+1}|s_{it},k)dG\right\}.\end{aligned} \tag{59}$$

Consider now states at which firm $C$ employs the manager along the continuation game of interest, in which the manager is employed by firm $C$ for the first time. With $S(\cdot)=S(\cdot|C)$, where $S(\cdot|C)=\Pi^A(\cdot|C)+V^C(\cdot|C)$, and $\Pi^A(\cdot|C)=0$ by the assumption of no recall, it follows that $k_{Ct}$ not only maximizes firm $C$'s value of profits and, by

Proposition 1, $V^C(\cdot|C)$, but also maximizes $S(\cdot|C)$. Off the equilibrium path, $S(\cdot|A) = V^A(\cdot|A) + \Pi^C(\cdot|A)$, where $\Pi^C(\cdot|A) = c_f$ by the assumption of free entry so that the choice of $k_{At}$ by firm $A$ does not affect $\Pi^C(\cdot|A)$. Given that $\Pi^A(\cdot|C) = \Pi^A(\cdot|A)$, $W(\cdot|C) = W(\cdot|A)$, and $\Pi^C(\cdot|C) \geq \Pi^C(\cdot|A)$ at all states at which $C$ employs a manager in equilibrium by Proposition 1, it follows that $S(\cdot) = S(\cdot|C) \geq S(\cdot|A)$ at such states and the opposite inequality to that in (59) holds. Therefore, $S(s_{it}, \varepsilon_t)$ satisfies (58) at all equilibrium states. Using the properties of Gumbel-distributed productivity shocks, it is easy to show that

$$S(s_{it}, \varepsilon_t) = \max\left\{\max_{k\in K^A}\{\varsigma^A(s_{it}, k) + \varepsilon_{Akt}\}, \max_{k\in K^C}\{\varsigma^C(s_{it}, k) + \varepsilon_{Ckt}\}\right\}, \tag{60}$$

$$\varsigma^f(s_{it}, k) = y_f(s_{it}, k) + \delta\eta_{kt} E\left[\ln\left(\textstyle\sum_{k'\in K^A}\exp\{\varsigma^A(s_{it+1}, k')\} + \sum_{k'\in K^C}\exp\{\varsigma^C(s_{it+1}, k')\}\right)|s_{it}, k\right], \tag{61}$$

$$\overline{V}^f(s_{it}, \varepsilon_t) = \max_{k\in K^f}\{\bar{\varsigma}^f(s_{it}, k) + \varepsilon_{fkt}\}, \tag{62}$$

and $\bar{\varsigma}^f(s_{it}, k) = y_f(s_{it}, k) + \delta\eta_{kt} E(\ln\sum_{k'\in K^f}\exp\{\bar{\varsigma}^f(s_{it+1}, k')\}|s_{it}, k)$ by standard arguments (Rust (1994)). □

**B.4 Assignment Probabilities.** Letting $\gamma(s_{it}, k) = y_C(s_{it}, k) - y_A(s_{it}, k)$ and using that the law of motion of the state at each level is the same across firms, I can express $\varsigma^C(s_{it}, k)$ in Lemma 2 as

$$\varsigma^C(s_{it}, k) = y_A(s_{it}, k) + \gamma(s_{it}, k) + \delta\eta_{kt} E\{\ln[\cdot]\,|s_{it}, k\} = \varsigma^A(s_{it}, k) + \gamma(s_{it}, k). \tag{63}$$

Note that $\Pr(L_{At} = k, f_t = A|s_{it})$, which is the probability of employment at Level $k$ of $A$ for a manager of skill type $i$ at state $s_{it}$ conditional on survival in the market, is given by

$$\Pr(L_{At} = k, f_t = A|s_{it}) = \exp\{\varsigma^A(s_{it}, k)\}/(\textstyle\sum_{k'}\exp\{\varsigma^A(s_{it}, k')\} + \sum_{k'}\exp\{\varsigma^C(s_{it}, k')\}), \tag{64}$$

$k' \geq 1$. Denote by $p_f(s_{it}, k)$ the probability of assignment to Level $k$ conditional on employment at $f = A, C$,

$$p_f(s_{it}, k) = \Pr(L_{ft} = k|f_t = f, s_{it}) = \exp\{\varsigma^f(s_{it}, k)\}/\textstyle\sum_{k'}\exp\{\varsigma^f(s_{it}, k')\}, \tag{65}$$

which is identified for each possible $s_{it}$ and $k$ when $f = A$, since it is equal to the ratio $\Pr(L_{ft} = k, f_t = f|s_{it})/\Pr(f_t = f|s_{it})$, identified by Proposition 5. Note that by the assumption of no recall, $S(s_{it}, \varepsilon_t)$ can be expressed as

$$S(s_{it}, \varepsilon_t) = \max\left\{\max_{k\in K^A}\{\varsigma^A(s_{it}, k) + \varepsilon_{Akt}\}, \max_{k\in K^C}\{\bar{\varsigma}^C(s_{it}, k) + \varepsilon_{Ckt}\}\right\},$$

where $\bar{\varsigma}^C(s_{it}, k) = y_C(s_{it}, k) + \delta\eta_{kt} E(\ln\sum_{k'\in K^C}\exp\{\bar{\varsigma}^C(s_{it+1}, k')\}|s_{it}, k)$ by Lemma 2. By (65), it follows that

$$\ln\textstyle\sum_{k'}\exp\{\bar{\varsigma}^C(s_{it}, k')\} = \bar{\varsigma}^C(s_{it}, 2) - \ln p_C(s_{it}, 2) = y_C(s_{it}, 2) - \ln p_C(s_{it}, 2) + \delta\eta_{2t} E\left[\ln\sum_{k'}\exp\{\bar{\varsigma}^C(s_{it+1}, k')\}|s_{it}, 2\right]$$

is a functional equation with a unique solution. According to Proposition 5, I normalize $y_C(s_{it}, 2)$ to $\ln p_C(s_{it}, 2)$ so that $\ln\sum_{k'}\exp\{\bar{\varsigma}^C(s_{it}, k')\}$ is zero. The probability in (64) forms the basis of the likelihood function of the model.[50] □

**B.5 Human Capital and Output Parameters.** As discussed after (17), these parameters to estimate for $A$ are $\{d_{A1t}(L1)$,-$e_{A1t}, e_{A2t}\}_{t\geq 2}$, $\{d_{A2t}(L2), e_{A3t}\}_{t\geq 3}$, and $\{d_{A3t}(L3)\}_{t\geq 4}$. As noted, the data exhibit high attrition because of the large number of separations in each tenure, and they contain a number of level transitions with few observations. To conserve on parameters, for transitions with no or relatively few observations, I did not estimate any of the associated parameters and normalized their value to zero. Specifically, consider first $\{d_{Akt}(k_{t-1})\}$. I set $d_{A11}(\cdot) = d_{A11}$ to $1,000$ and $e_{A11}$ to zero, since all managers are hired at Level 1, and normalize $d_{A2t}(L2) = 0$, as is consistent with Proposition 5. The combination of rapid promotions to Level 2 and the high separation rate in each tenure leads the fraction of managers at Level 1 to sharply decrease in medium and high tenures—it is less than 8 percent from tenure $t = 4$ on. Thus, I estimate only $d_{A14}(L1)$ and $d_{A15}(L1)$ and maintain that $d_{A1t}(L1) = d_{A15}(L1)$ for $t = 6, 7$. Similarly, I let $d_{A36}(L3) = d_{A35}(L3)$ at Level 3 and estimate only $d_{A3t}(L3)$ for $t = 4, 5, 7$—no manager is observed at Level 3 before the third year of tenure.

Consider now $\{e_{Akt}\}$. At Level 1, I did not estimate any such parameter in $t \geq 3$ because of the small fraction of

[50] In estimation, I compute values and probabilities in $t \geq 8$ under the assumption that managers no longer acquire human capital after the seventh year of tenure. For any given vector of parameter values, I calculate "terminal" values in $t = 8$ for the market-wide employment and assignment problem derived in Lemma 2 as the solution to the corresponding infinite-horizon problem from $t = 8$ on. Given these terminal values, I solve by backward induction the market-wide employment and assignment problem between $t = 1$ and $t = 7$ as a finite-horizon one in each such tenure. Note that the assumption that human capital acquisition eventually tapers off is not implausible: compared with those in earlier tenures, the employment outcomes of managers in $t \geq 8$ in the data display much less variation with tenure, previous assignments, and performance.

managers observed at Level 1 in those tenures. At Level 2, I estimate the parameters $e_{A22}$, $e_{A23}$, $e_{A25}$, and $e_{A26}$, and set $e_{A24} = e_{A22}$ and $e_{A27} = e_{A26}$. Since no manager is assigned to Level 2 at entry, I did not estimate $e_{A21}$. At Level 3, I restrict $e_{A31} = e_{A32} = e_{A33}$, since no manager is at this level until the third year of tenure. As the empirical hazard rates of promotion from Level 1 to 2 and from Level 2 to 3 display very similar features in $t \geq 2$, I allow for common components across the parameters $e_{A2t}$ and $e_{A3t}$ to limit parameter proliferation as follows. First, I specify $e_{A2t} = \varrho_2 + \varrho_{2t}$ for $2 \leq t \leq 4$ and $e_{A3t} = \varrho_{22} + \varrho_{3t}$ for $4 \leq t \leq 6$. The choice of $\varrho_{22}$ as the benchmark parameter is motivated by the high separation rate from Level 2 and the high promotion rate from Level 2 to 3 in each tenure: expected output parameters at Level 2 in low tenures can be more precisely estimated. By this logic, this formulation has led to $e_{A35} = e_{A36} = \varrho_{22}$, since the differences in these parameters proved insignificantly different from zero. Then, I estimate the parameters $e_{A33} = -\varrho_{34}$, $e_{A34} = \varrho_{22} + \varrho_{34}$, $e_{A37}$, and $e_{A38}$ at Level 3. In Table A.4, I display the point estimate and standard error of $e_{A35}$ for completeness: the parameters $e_{A33}$ and $e_{A34}$ are sufficient to pin down $e_{A35}$. $\square$

**B.6 Parameters of Exogenous Separation Rates.** To conserve on parameters, I only allow for variation in $\{\eta_{kt}\}$ across levels and tenures that proves statistically significant, whenever setting these parameters equal across levels or tenures does not affect any other parameter estimate. As a result, the exogenous separation rates I estimate at Levels 1, 2, and 3, respectively, are: $\eta_{11}$, $\eta_{21}$, and $\eta_{31}$ for tenure 1; none for tenure 2, since $\eta_{12} = \eta_{11}$, $\eta_{22} = \eta_{21}$, and $\eta_{32} = \eta_{31}$; $\eta_{13}$ for tenure 3, that is, $\xi_3$, since $\eta_{13} = \eta_{14} + \xi_3$, $\eta_{23} = \eta_{22} + \xi_3$, and $\eta_{33} = \eta_{32}$; $\eta_{14}$ and $\eta_{24}$ for tenure 4, since $\eta_{34} = \eta_{24}$; $\eta_{25}$ for tenure 5, since $\eta_{15} = \eta_{14}$ and $\eta_{35} = \eta_{25}$; $\eta_{26}$ for tenure 6, since $\eta_{16} = \eta_{15}$ and $\eta_{36} = \eta_{26}$; $\eta_{27}$ for tenure 7, since $\eta_{17} = \eta_{16}$ and $\eta_{37} = \eta_{27}$; and none for tenure 8 onward, since $\eta_{1t} = \eta_{17}$, $\eta_{2t} = \eta_{27}$, and $\eta_{3t} = \eta_{2t}$, $t \geq 8$. So, I estimate $(\eta_{11}, \eta_{13}, \eta_{14})$ at Level 1, $(\eta_{21}, \eta_{24}, \eta_{25}, \eta_{26}, \eta_{27})$ at Level 2, and $\eta_{31}$ at Level 3. $\square$

**B.7 Estimated Wage Equation.** Here, I derive the wage equation in (18) using Lemma 2 to express assignment probabilities at $C$ as simple functions of the corresponding identified probabilities at $A$ under the assumption of free entry of firms in the labor market, no-recall of previously employed managers, and a suitable normalization for $\{y_A(s_{it}, 2)\}$. To start, observe that (60) and (61) imply that $\Pr(L_{At} = k, f_t = A|s_{it}) / \Pr(L_{Ct} = k, f_t = C|s_{it}) = \exp\{\varsigma^A(s_{it}, k) - \varsigma^C(s_{it}, k)\} = \exp\{y_A(s_{it}, k) - y_C(s_{it}, k)\}$, as the law of motion of the state is common across firms by level. Thus, by (64) and (65),

$$
\begin{aligned}
\Pr(L_{At} = k, f_t = A|s_{it}) &= \frac{e^{\varsigma^A(s_{it},k) - \varsigma^C(s_{it},k)}}{\sum_{k'} e^{\varsigma^A(s_{it},k') - \varsigma^C(s_{it},k)} + \sum_{k'} e^{\varsigma^C(s_{it},k') - \varsigma^C(s_{it},k)}} \\
&= \frac{e^{y_A(s_{it},k) - y_C(s_{it},k)}}{\sum_{k'} e^{\varsigma^A(s_{it},k') - \varsigma^A(s_{it},k) - \varsigma^C(s_{it},k) + \varsigma^A(s_{it},k)} + \frac{1}{\Pr(L_{Ct}=k|f_t=C,s_{it})}} = \frac{e^{y_A(s_{it},k) - y_C(s_{it},k)}}{\frac{e^{y_A(s_{it},k) - y_C(s_{it},k)}}{\Pr(L_{At}=k|f_t=A,s_{it})} + \frac{1}{\Pr(L_{Ct}=k|f_t=C,s_{it})}},
\end{aligned}
$$

which implies

$$
\frac{\exp\{y_A(s_{it}, k) - y_C(s_{it}, k)\}}{\Pr(L_{At} = k|f_t = A, s_{it})} + \frac{1}{\Pr(L_{Ct} = k|f_t = C, s_{it})} = \frac{\exp\{y_A(s_{it}, k) - y_C(s_{it}, k)\}}{\Pr(L_{At} = k, f_t = A|s_{it})}.
$$

Multiplying both sides of this equality by $\Pr(L_{At} = k|f_t = A, s_{it})$ and exploiting the forms of $\Pr(L_{At} = k|f_t = A, s_{it})$ and $\Pr(L_{At} = k, f_t = A|s_{it})$, I obtain that

$$
\Pr(L_{Ct} = k|f_t = C, s_{it}) = \frac{\Pr(f_t = A|s_{it}) \exp\{y_C(s_{it}, k) - y_A(s_{it}, k)\}}{1 - \Pr(f_t = A|s_{it})} \Pr(L_{At} = k|f_t = A, s_{it}). \tag{66}
$$

Taking the ratio of (66) to the same expression evaluated when $k = 2$,

$$
\frac{\Pr(L_{Ct} = k|f_t = C, s_{it})}{\Pr(L_{Ct} = 2|f_t = C, s_{it})} = \exp\{y_C(s_{it}, k) - y_A(s_{it}, k) - y_C(s_{it}, 2) + y_A(s_{it}, 2)\} \frac{\Pr(L_{At} = k|f_t = A, s_{it})}{\Pr(L_{At} = 2|f_t = A, s_{it})}.
$$

Hence, $\Pr(L_{Ct} = k|f_t = C, s_{it})$ can be expressed as a function of $\Pr(L_{Ct} = 2|f_t = C, s_{it})$ and the corresponding assignment probabilities at firm $A$. As a result, by (12),

$$
\begin{aligned}
Ew_{Ait}(k) &= y_C(s_{it},k) - [y_C(s_{it},k) - y_A(s_{it},k) - y_C(s_{it},2) + y_A(s_{it},2)] - \ln\left[\frac{\Pr(L_{At}=k|f_t=A,s_{it})}{\Pr(L_{At}=2|f_t=A,s_{it})}\right] - \ln \Pr(L_{Ct}=2|\cdot) \\
&= y_A(s_{it},k) + y_C(s_{it},2) - y_A(s_{it},2) + \ln\left[\frac{\Pr(L_{At}=2|f_t=A,s_{it})}{\Pr(L_{Ct}=2|f_t=C,s_{it}) \Pr(L_{At}=k|f_t=A,s_{it})}\right].
\end{aligned} \tag{67}
$$

Note that by (16), (17), (57), and the fact that output at $A$ does not vary with $i$, $y_C(s_{it}, 2) - y_A(s_{it}, 2)$ is given by

$$\underline{y}^e_{C2t}(i)-\underline{y}^e_{A2t}+\beta_2[\overline{y}_{C2t}(i)-\underline{y}_{C2t}(i)-\overline{y}_{A2t}+\underline{y}_{A2t}]+(\alpha_2-\beta_2)[\overline{y}_{C2t}(i)-\underline{y}_{C2t}(i)-\overline{y}_{A2t}+\underline{y}_{A2t}]p_{it}$$
$$=\textstyle\sum_j[\underline{a}^j_{C2t}(i)-\underline{a}^j_{A2t}]+\beta_2\sum_j[\overline{a}^j_{C2t}(i)-\underline{a}^j_{C2t}(i)-\overline{a}^j_{A2t}+\underline{a}^j_{A2t}]+(\alpha_2-\beta_2)\sum_j[\overline{a}^j_{C2t}(i)-\underline{a}^j_{C2t}(i)-\overline{a}^j_{A2t}+\underline{a}^j_{A2t}]p_{it},$$

which yields that

$$y_A(s_{it},k)+y_C(s_{it},2)-y_A(s_{it},2)=b_k(h_1)+\textstyle\sum_j[\underline{a}^j_{Akt}+g^j_k(\cdot)]+\beta_k\sum_j(\overline{a}^j_{Akt}-\underline{a}^j_{Akt})+\sum_j[\underline{a}^j_{C2t}(i)-\underline{a}^j_{A2t}]+\beta_2$$
$$\cdot\textstyle\sum_j[\overline{a}^j_{C2t}(i)-\underline{a}^j_{C2t}(i)-\overline{a}^j_{A2t}+\underline{a}^j_{A2t}]+\{(\alpha_2-\beta_2)\sum_j[\overline{a}^j_{C2t}(i)-\underline{a}^j_{C2t}(i)-\overline{a}^j_{A2t}+\underline{a}^j_{A2t}]+(\alpha_k-\beta_k)\sum_j(\overline{a}^j_{Akt}-\underline{a}^j_{Akt})\}p_{it}.$$

Recalling that $a^j_{fkt}(i)=a^j_{fkt}+a^j_{fk}(i)$, $a^j_{fkt}\in\{\overline{a}^j_{fkt},\underline{a}^j_{fkt}\}$, $a^j_{fk}(i)\in\{\overline{a}^j_{fk}(i),\underline{a}^j_{fk}(i)\}$, and $a^j_{Ak}(i)=0$, I obtain

$$y_A(s_{it},k)+y_C(s_{it},2)-y_A(s_{it},2)=\underbrace{\textstyle\sum_j[\beta_2\overline{a}^j_{C2}(i)-\beta_2\underline{a}^j_{C2}(i)+\underline{a}^j_{C2}(i)]+b_{0k}}_{\omega_{0ik}}+\underbrace{b_k(h_1)-b_{0k}}_{\omega_{1k}x_1+\omega_{2k}x_1^2+\omega_{3k}e_1+\sum_m\omega_{ym}1\{y_1=m\}}$$
$$+\underbrace{\textstyle\sum_j\left[\frac{\underline{a}^j_{Akt}+g^j_k(\cdot)+\beta_k(\overline{a}^j_{Akt}-\underline{a}^j_{Akt})+\underline{a}^j_{C2t}-\underline{a}^j_{A2t}+\beta_2(\overline{a}^j_{C2t}-\underline{a}^j_{C2t}-\overline{a}^j_{A2t}+\underline{a}^j_{A2t})}{t-1}\right](t-1)}_{\omega_{4kt}}$$
$$+\{\underbrace{(\alpha_2-\beta_2)\textstyle\sum_j[\overline{a}^j_{C2}(i)-\underline{a}^j_{C2}(i)]}_{\omega_{5i}}+\underbrace{\textstyle\sum_j[(\alpha_2-\beta_2)(\overline{a}^j_{C2t}-\underline{a}^j_{C2t}-\overline{a}^j_{A2t}+\underline{a}^j_{A2t})+(\alpha_k-\beta_k)(\overline{a}^j_{Akt}-\underline{a}^j_{Akt})]}_{\omega_{5kt}}\}p_{it}.$$

Let $\omega(i,h_1,k)=\omega_{0ik}+\omega_{1k}x_1+\omega_{2k}x_1^2+\omega_{3k}e_1+\sum_m\omega_{ym}1\{y_1=m\}$ and recall $w_{Ait}=Ew_{Ait}(k)+\epsilon_{Ait}$ by (12). Then,

$$Ew_{Ait}(k)=\omega(i,h_1,k)+\omega_{4kt}\times(t-1)+(\omega_{5i}+\omega_{5kt})p_{it}+\ln\left[\frac{\Pr(L_{At}=2|f_t=A,s_{it})}{\Pr(L_{Ct}=2|f_t=C,s_{it})\Pr(L_{At}=k|f_t=A,s_{it})}\right] \quad (68)$$

by (67). I impose that $y_A(s_{it},2)-y_C(s_{it},2)$ is equal to $-\ln[1/\Pr(f_t=A|s_{it})-1]$ so that $\Pr(L_{Ct}=2|f_t=C,s_{it})=\Pr(L_{At}=2|f_t=A,s_{it})$ by (66); see Appendix B.4 for the normalization of $y_C(s_{it},2)$. Then, the last term in (68) simplifies to $-\ln\Pr(L_{At}=k|f_t=A,s_{it})$, which gives (18).

I set the parameters $\omega_{1k}$, $\omega_{2k}$, and $\omega_{3k}$ equal at Levels 1 and 2 and denote their common values by $\omega_1$, $\omega_2$, and $\omega_3$. As for the parameters $\{\omega_{ym}\}$, I set $\omega_{ym}=0$, $0\le m\le 3$, and $\omega_{y4}=\omega_{y5}$, since the difference between these latter two parameters proved insignificant. I specify the coefficient on tenure at Level 1 as $\omega_{41t}=\omega_{412}1\{t<4\}-\omega_{412}1\{t\ge 4\}$, which helps the model account for the progressively greater proportion of managers at Level 1 who are paid relatively low wages in high tenures; the coefficients at higher levels proved impossible to estimate with any precision. I restrict $\{\sigma_{Ai3}\}$ to not vary across skill types, owing to the relatively small number of observations at Level 3 compared with the number of observations at lower levels, especially in high tenures, and denote their common value by $\sigma_{A3}$. Thus, the estimated baseline parameters are $\{\omega_{0i1},\omega_{0i2},\omega_{0i3}\},\omega_1,\ \omega_2,\omega_3,\ \omega_{13},\ \omega_{23},\ \omega_{33},\ \omega_{412},\ \{\omega_{ym}\}^9_{m=5},\ \{\omega_{5i}\},\{\sigma_{Ai1},\sigma_{Ai2}\}$, and $\sigma_{A3}$. See Appendix B.11 for the results of the estimation of a more general version of the model.

When the parameters of interest are modified to evaluate the importance of uncertainty, learning, and human capital acquisition in Section 5, wages are recomputed for the new sequences of states, equilibrium assignments, and their associated probabilities for each manager skill type. Note that the mapping between the wage parameters and the underlying human capital and expected output parameters is invariant across these experiments. Specifically, in assessing the importance of uncertainty and learning, expected output at each job level is kept at its baseline value for each prior value. Only $\{\alpha_k\}$ and $\{\beta_k\}$ are adjusted in the updating of priors to capture that job levels are more or less informative than in the baseline model, depending on the exercise. In the experiment that evaluates the role of human capital acquisition, the parameters of expected output at each job level of $A$ are simply assumed to be constant with tenure. □

**B.8 Estimates of Classification Error.** To limit parameter proliferation and based on model diagnostics and fit, I maintain that $\rho_1=\rho_0$ in (11). As discussed, I estimate classification error parameters only at Levels 1 and 2, because many performance ratings are missing among managers assigned to Level 3. As a result, the estimated parameters of the distribution of performance ratings are $\{\alpha_k,\beta_k\}$, $\rho_0$, and $\{\rho_2(k)\}_{k=1,2}$. Since the probability of a recorded rating is not a linear function of the probability of true performance, as apparent from (11), this error structure leads to bias. The greater in absolute value $\rho_2(k)$ is, the greater the persistence in misreporting is. The estimates in Table A.3 imply that performance is measured with error, and over time, this error is more likely for individuals whose true performance is high. This result is consistent with the idea that the performance of managers who fail is assessed more thoroughly and precisely than the performance of managers who succeed, as is common in many firms, since poor performance may eventually lead to disciplinary actions. □

**B.9 Retention and Assignment Policy at My Firm.** The policy that emerges by simulating the model at the estimated values of the parameters of the process of learning, human capital acquisition, and expected output at the firm implies that, on average, managers at higher levels have higher ability, higher acquired human capital, and higher beliefs that their ability is high. These results support the intuition of BGH, among others, that a firm acts as an information acquisition filter in the labor market by producing information about managers' ability. This information helps improve the matching of managers to jobs (and firms) over time. Within the firm, this process of information acquisition leads to the endogenous selection of managers to higher-level jobs through promotions, which is the key channel through which learning affects wages at the firm. As for separations, the estimated exogenous separation rates in Table A.5 are close to the observed hazard rates of separation in Table A.1, which implies that most separations are exogenous. Crucial features of the data that explain this finding are that separations are largely uncorrelated with performance, which is the primary determinant of learning, and approximately constant with tenure, which is a key determinant of human capital acquisition; see also the discussion in BGH. In particular, there is no evidence that separations mask a tendency for managers to be laid off or move to other firms in response to poor evaluations of performance that would otherwise have led to demotions. That some separations are endogenous is confirmed by the higher wage growth of managers retained throughout the sample years relative to that of those who separate, as discussed. □

**B.10 Goodness of Fit.** To assess the fit of the model to the data, I simulate 4,000 prior realizations per manager, drawn from the estimated distribution of initial priors. Table 1 shows that the model tracks remarkably well the profiles of managers' assignment to Levels 1, 2, and 3, which are nonlinear and nonmonotone in tenure, as well as the tenure pattern of managers' separations. Table A.1 shows that the model also accurately reproduces the feature that outflows from Levels 1 and 2 come from an essentially constant hazard rate of separation and a hazard of promotion that first increases then decreases with tenure. As Table 2 further implies, the model successfully fits the tenure distributions of performance ratings both at Level 1 (with a slight overprediction in the fifth year of tenure) and at Level 2 (except for some discrepancies in the fourth and seventh years of tenure). Lastly, Table 3 shows that the model reproduces quite well the distribution of wages at each level and tenure, except for some discrepancies at Level 3 in the highest tenures. Indeed, the largest such discrepancies are at Level 3 in the sixth and seventh years of tenure. These discrepancies are partly due to the high rate of attrition in the sample. In fact, the fit of the model to the wage data from the larger sample that includes entrants in the firm at Levels 1 to 4 is substantially better in this dimension; see the S.A. (Section 5).

One criterion to formally evaluate model fit is the Pearson's $\chi^2$ goodness-of-fit test. I perform it based on the statistic $s\sum_{r=1}^{R}\{[\zeta(r)-\hat{\zeta(r)}]^2/\hat{\zeta(r)}\}$, where $\zeta(\cdot)$ denotes the empirical density function of a given endogenous variable, $\hat{\zeta(\cdot)}$ denotes the maximum-likelihood estimate of the density function of that variable, $s$ indicates the number of observations, and $R$ indicates the number of categories considered (not taking into account that the parameters of the model are estimated). I compare the observed and predicted distributions of managers across levels, performance ratings at Levels 1 and 2, and wages at each level in each of the first seven years of tenure. The results of the test are as follows. In terms of the distribution of managers across levels and their probability of retention at the firm in each tenure, the test does not reject the model at conventional significance levels in any tenure. As for the hazard rates of separation, retention at a level, and promotion to the next level in each tenure, the test does not reject the model at conventional significance levels, apart from the second, third, fourth, and sixth years of tenure at Level 1 and the second and third years of tenure at Level 2. However, for these tenures, the outcome of the test is very much influenced by the small number of observations at Levels 1 and 2 in high tenures. Regarding the distribution of performance ratings at Levels 1 and 2, the test does not reject the model at conventional significance levels in any tenure. In terms of the level distribution of wages, the test does not reject the model at conventional significance levels, apart from the third year of tenure at Level 2 and the fourth, fifth, sixth, and seventh years of tenure at Level 3. One reason for these results on the distributions of wages is the small fraction of managers at Level 3 in high tenures, as confirmed by the improved fit of the model to the larger sample of entrants in the firm, as mentioned. □

**B.11 Sensitivity Analysis on the Robustness of the Estimates.** To examine whether the estimates are robust to the main assumptions underlying the specification of the wage process, I also estimated a more general version of the wage equation in (18), which includes the tenure coefficient $\omega_{4k}$ for Levels $k>1$ (see footnote 40) and the prior coefficient $\omega_{5kt}$ specified as $\omega_{5kt}=\omega_{5k}+(t-1)\times\omega_{5k}^{g}+\omega_{5}^{x}x_1$ with additional parameters $(\omega_{42},\omega_{43},\omega_{52},\omega_{53},\omega_{51}^{g},\omega_{52}^{g},\omega_{53}^{g},\omega_{5}^{x})$. Note from Tables A.15-A.20 in the S.A. that the estimates of the parameters of this more general version of the model that are common to the baseline model are fairly similar to those estimated for the baseline model, except for the output parameters, which are somewhat larger. Some estimates, though, are less significant, especially $\alpha_3$ and $e_{A38}$, which cannot be estimated with any precision. The estimates of the additional wage parameters are fairly small, particularly the slope coefficients on tenure, $\{\omega^g_{5k}\}$. Thus, allowing the coefficient on $p_{it}$ in the wage equation to vary with the job level and tenure does not have much of an effect on results. □

**B.12 The Role of Persistent Uncertainty.** Recall the experiment of *fast learning at Level 1* discussed in Section 5. In the case of *fast learning at Level 2*, jobs at Level 2 are assumed to be nearly perfectly informative about ability with $\alpha_2 = 0.99$ and $\beta_2 = 0.01$, whereas the other parameters are fixed at their baseline values. See Tables 6 and A.7, respectively, for the results on the distribution of wages at each tenure and managers' job paths at the firm. Similar to the case of fast learning at Level 1, fast learning at Level 2 yields higher wage growth, much larger wage dispersion at Level 2, and faster promotions. Perhaps surprisingly, it also leads to a lower percentage of managers assigned to Level 3 in high tenures and much lower wage dispersion at Level 3 relative to the baseline model. Intuitively, the greater informativeness of Level 2 makes it the firm's preferred assignment over Level 3 for a larger set of states, including higher priors, than in the baseline model. Further, since managers reach Level 3 at higher priors than in the baseline model, the standard deviation of wages at Level 3 is much lower than at Level 2 in this experiment and at Level 3 in the baseline model. □

# C A Discussion of Alternative Explanations

The analysis so far has focused on the role of uncertainty, learning, and human capital acquisition in accounting for the joint dynamics of jobs and wages at the firm in my data. Here I briefly discuss alternative explanations, focusing on the potential role of asymmetric information, performance incentives provided through performance pay and tournaments, and complementarities in production.

**C.1 Asymmetric Learning.** I have assumed that firms and managers share the same information about ability. To see that this assumption is plausible for my data, note first that the analysis in BGH supports the idea that managers' performance is public information. Further, the formal test of asymmetric information that Devaro and Waldman (2012) perform on the BGH data does not find conclusive evidence of asymmetric information. For related work that relaxes the assumption of symmetric information, see Waldman (1984), Bernhardt and Scoones (1993), Bernhardt (1995), Waldman (1996), and Ekinci et al. (2019).[51]

**C.2 Performance Incentives through Performance Pay.** The model abstracts from the possibility that firms may induce workers to exert effort on the job by making pay contingent on performance. In the presence of incentive pay, wages may be linked directly to *current* performance, not just indirectly to past performance through beliefs. But if so, then current wages should be more strongly correlated with current performance than with past performance. As Kahn and Lange (2014) also note, my data do not support the notion that the correlation of wages with current performance exceeds that with past performance. Frederiksen et al. (2017) analyze information on bonus pay in my data, available from 1981, and find some evidence that bonuses may be used to set incentives for performance; see also Gibbs (1995). Bonuses, though, represent a small fraction of total pay: they are paid to 25 percent of managers, mainly at the highest levels, and do not account for a large portion of pay. The median bonus of managers receiving one at (the original) Levels 1-3 is less than 10 percent of salary; for managers at (the original) Level 4, it is less than 15 percent. Hence, this evidence does not suggest that performance incentives are a more critical determinant of pay than the sources I focus on.[52]

**C.3 Performance Incentives through Tournaments.** Firms may also provide incentives for performance through the implicit promise of promotion. In tournament models (Lazear and Rosen (1981)), this mechanism links workers' future promotion and pay to their current effort and performance. Tournament models with homogeneous workers and no sorting are easily distinguishable from learning models, since they imply that all workers exert the same amount of effort, and so the winner of the tournament is determined merely by luck. On the contrary, tournament models with heterogeneous workers and sorting are difficult to empirically distinguish from learning models (Rosen (1986)). For instance, both tournament and learning models with sorting predict serial correlation in promotion rates and wage

[51] A general model in which the information flow to other firms is imperfect and possibly affected by an employing firm's retention and job assignment decisions is beyond the scope of this paper. Allowing for private information about ability, and so different information sets across firms, would also render identification prohibitively difficult in light of my data. I consider the formulation proposed here as a first attempt at measuring the importance of learning for wages under assumptions commonly maintained in the empirical literature on careers.

[52] Since bonuses are higher at higher levels and learning contributes to wage growth primarily through its impact on promotions, the effect of learning on wage growth that I estimate excluding bonuses can be conjectured to be a lower bound on the effect of learning on total pay growth.

increases. Yet, one aspect of my data that is at odds with the idea that tournaments are key to the dynamics of jobs and wages at the firm is the high frequency of (real) wage decreases. A tournament model does not naturally lead to such decreases, whereas a model of learning and stochastic human capital accumulation does. Moreover, Baker and Holmström (1995) document that the wage differential between adjacent job levels at the firm ranges from 18 to 47 percent—wages at each job level are highly dispersed. Yet, the immediate wage premium upon promotion is around only 7 percent. They interpret this finding as evidence against the hypothesis that the wage structure at the firm is governed by a tournament. A full analysis of the role of learning, human capital acquisition, and performance incentives for wages in the presence of sorting is left to future work.

**C.4 Production Complementarities.** I assume that managers are imperfect substitutes in production across job levels, as captured by the dependence of the parameters $d_{fkt}(k_{t-1}, i)$ of expected output on $k_{t-1}$ discussed in Section 3.3. I abstract from explicitly modeling complementarities or capacity constraints for three reasons. First, the goal of my analysis is to investigate whether a model that integrates learning, human capital acquisition, and job assignment can account for the main patterns of individual jobs and wages observed in firms. To this end, I formulate assumptions that make my model comparable to those I integrate into my framework, which ignore these complementarities. Second, the great variability in the size of job levels over the sample years discussed in Section 1 suggests that the firm is not subject to too stringent capacity constraints on the employment or assignment of managers to job levels. Moreover, the correlation of performance or wages across managers in the same cost center, a proxy for production units at the firm, is low, which supports a low degree of complementarity across managers. Third, work that relaxes the assumption of separable worker productivity typically addresses questions that are different from those I focus on here—such as the importance of organizational capital for the growth rate and the size distribution of firms (Prescott and Visscher (1980))—or, because of its complexity, confines the analysis to essentially static environments (Ferrall (1997) and Ferrall et al. (2009)) or stylized cases (Davis (1997)). In contrast to the existing literature, I allow for a fully dynamic interaction between forward-looking firms and workers, which leads firms and workers to face complex intertemporal trade-offs between the opportunity cost of investing in information and human capital and the future benefits of doing so. Given the challenges that the estimation of such a model entails, I consider the assumption of imperfect substitutability of managers in production as a first approximation to developing a tractable yet empirically rich framework for careers in firms. Since the model I propose closely matches the assignment and wage paths of the managers in my data, this framework, although stylized, seems to be promising at capturing key features of careers.

Table A.1: Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)*

| | | Separation | | Retention | | Promotion | |
|---|---|---|---|---|---|---|---|
| Level | Tenure | Data | Model | Data | Model | Data | Model |
| Level 1 | 1 to 2 | 14.5 | 14.5 | 45.6 | 45.7 | 39.9 | 39.8 |
| | 2 to 3 | 14.6 | 14.5 | 36.8 | 20.3 | 48.6 | 55.2 |
| | 3 to 4 | 11.7 | 8.4 | 45.6 | 46.9 | 42.7 | 27.8 |
| | 4 to 5 | 11.9 | 5.0 | 60.6 | 65.0 | 27.5 | 19.9 |
| | 5 to 6 | 9.1 | 5.1 | 62.1 | 64.8 | 28.8 | 21.6 |
| | 6 to 7 | 12.2 | 5.0 | 73.2 | 79.1 | 14.6 | 11.4 |
| Level 2 | 2 to 3 | 16.3 | 13.6 | 61.9 | 55.5 | 21.8 | 10.9 |
| | 3 to 4 | 15.6 | 16.9 | 47.0 | 51.7 | 37.4 | 31.4 |
| | 4 to 5 | 14.7 | 14.2 | 54.8 | 56.9 | 30.5 | 28.9 |
| | 5 to 6 | 12.8 | 12.1 | 60.5 | 62.6 | 26.7 | 25.3 |
| | 6 to 7 | 15.4 | 11.5 | 59.4 | 62.9 | 25.1 | 25.6 |
| Level 3 | 3 to 4 | 10.5 | 12.2 | 89.5 | 87.8 | - | - |
| | 4 to 5 | 12.2 | 14.2 | 87.8 | 85.8 | - | - |
| | 5 to 6 | 10.1 | 12.1 | 89.9 | 87.9 | - | - |
| | 6 to 7 | 10.0 | 11.5 | 90.0 | 88.5 | - | - |

*Promotions are by one level.

Table A.2: Percentage Distribution of Changes in Log Wages by Tenure

| Tenure | Between -0.15 and 0.00 | Between 0.00 and 0.15 | Between 0.15 and 0.30 | Growth Rate |
|---|---|---|---|---|
| 1 to 2 | 22.9 | 69.9 | 7.2 | 5.2 |
| 2 to 3 | 22.6 | 70.4 | 6.6 | 5.1 |
| 3 to 4 | 24.9 | 70.3 | 4.3 | 3.9 |
| 4 to 5 | 23.6 | 70.1 | 5.9 | 2.2 |
| 5 to 6 | 22.5 | 70.5 | 6.9 | 0.7 |
| 6 to 7 | 21.9 | 68.5 | 8.3 | 1.8 |

Table A.3: Estimates of Parameters of Classification Error in Performance Ratings*

| Parameters | Level 1 ($k = 1$) | Level 2 ($k = 2$) |
|---|---|---|
| Base Error: $\rho_0$ | 0.521 | (same as for $k = 1$) |
| | (0.040) | (same as for $k = 1$) |
| Persistence: $\rho_2(k)$ | -0.703 | -0.544 |
| | (0.040) | (0.029) |

*Recall that $\rho_1 = \rho_0$. Asymptotic standard errors in parentheses.

Table A.4: Estimates of Intercept and Slope Parameters of Expected Output*

| Parameters | Value | St. Error |
|---|---|---|
| Level 1 | | |
| $d_{A14}(L1)$ | -14.095 | 0.072 |
| $d_{A15}(L1)$ | -9.592 | 0.065 |
| $d_{A16}(L1)$ | *-9.592* | (same as $d_{A15}(L1)$) |
| $d_{A17}(L1)$ | *-9.592* | (same as $d_{A15}(L1)$) |
| $e_{A12}$ | 59.210 | 0.274 |
| Level 2 | | |
| $e_{A22}$ | 51.269 | 0.111 |
| $e_{A23}$ | 44.202 | 0.179 |
| $e_{A24}$ | *51.269* | (same as $e_{A22}$) |
| $e_{A25}$ | 43.438 | 0.168 |
| $e_{A26}$ | 44.496 | 0.108 |
| $e_{A27}$ | *44.496* | (same as $e_{A26}$) |
| Level 3 | | |
| $d_{A34}(L3)$ | 17.070 | 0.119 |
| $d_{A35}(L3)$ | 4.056 | 0.090 |
| $d_{A36}(L3)$ | *4.056* | (same as $d_{A35}(L3)$) |
| $d_{A37}(L3)$ | 4.561 | 0.054 |
| $e_{A31}$ | *-7.999* | (same as $e_{A33}$) |
| $e_{A32}$ | *-7.999* | (same as $e_{A33}$) |
| $e_{A33}$ | -7.999 | 0.193 |
| $e_{A34}$ | 45.173 | 0.233 |
| $e_{A35}$ | 37.174 | 0.040 |
| $e_{A36}$ | *37.174* | (same as $e_{A35}$) |
| $e_{A37}$ | 43.814 | 0.021 |
| $e_{A38}$ | 40.067 | 0.021 |

*All parameters in the table are expressed in thousands. Parameters whose values are in italics are not estimated.

Table A.5: Estimates of Exogenous Separation Rates*

| Parameters (%) | Value | St. Error |
|---|---|---|
| Level 1 | | |
| $\eta_{11}$ | 14.5 | 0.004 |
| $\eta_{13}$ | 8.3 | 0.001 |
| $\eta_{14}$ | 5.0 | 0.0001 |
| Level 2 | | |
| $\eta_{21}$ | 13.6 | 0.002 |
| $\eta_{24}$ | 14.2 | 0.001 |
| $\eta_{25}$ | 12.1 | 0.001 |
| $\eta_{26}$ | 11.5 | 0.0003 |
| $\eta_{27}$ | 11.1 | 0.0003 |
| Level 3 | | |
| $\eta_{31}$ | 12.2 | 0.002 |

*Exogenous separations occur with probability $1-\eta_{kt}$. Recall that $\eta_{12}=\eta_{11}$, $\eta_{13}=\eta_{14}+\xi_3$, $\eta_{1t}=\eta_{14}, t>4$; $\eta_{22}=\eta_{21}$, $\eta_{23}=\eta_{22}+\xi_3$, $\eta_{2t}=\eta_{27}, t>7$; $\eta_{3t}=\eta_{31}, t=2,3$, $\eta_{3t}=\eta_{2t}, t>3$.

Table A.6: Estimates of Year Dummies in the Wage Equation (Baseline: 1970-1973)*

| Parameters | 1975 ($\omega_{y5}$) | 1976 ($\omega_{y6}$) | 1977 ($\omega_{y7}$) | 1978 ($\omega_{y8}$) | 1979 ($\omega_{y9}$) |
|---|---|---|---|---|---|
| $\omega_{ym}$ | -0.063 | -0.107 | -0.140 | -0.208 | -0.169 |
| | (0.003) | (0.004) | (0.004) | (0.003) | (0.003) |

*Recall that $\omega_{y4}=\omega_{y5}$. Asymptotic standard errors in parentheses.

Table A.7: Counterfactual Experiments on Importance of Learning for Level Assignments

| | Separation | | | | Level 1 | | | | Level 2 | | | | Level 3 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tenure | Base. | No L | Fast L at 1 | Fast L at 2 | Base. | No L | Fast L at 1 | Fast L at 2 | Base. | No L | Fast L at 1 | Fast L at 2 | Base. | No L | Fast L at 1 | Fast L at 2 |
| 1 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 14.5 | 14.5 | 14.5 | 14.5 | 45.7 | 57.7 | 40.5 | 0.0 | 39.8 | 27.8 | 45.0 | 85.4 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | 26.5 | 26.6 | 26.5 | 26.1 | 17.2 | 20.6 | 14.6 | 5.0 | 47.3 | 52.0 | 38.9 | 35.0 | 8.9 | 0.8 | 20.0 | 33.8 |
| 4 | 37.1 | 37.3 | 36.7 | 36.6 | 8.1 | 11.6 | 7.3 | 2.0 | 29.2 | 40.3 | 25.9 | 29.1 | 25.6 | 10.8 | 30.1 | 32.3 |
| 5 | 45.3 | 45.1 | 45.0 | 45.4 | 5.3 | 8.2 | 4.9 | 1.3 | 18.3 | 30.3 | 17.7 | 24.7 | 31.2 | 16.4 | 32.4 | 28.6 |
| 6 | 51.5 | 51.2 | 51.3 | 51.9 | 3.4 | 5.6 | 3.2 | 1.0 | 12.6 | 23.3 | 13.1 | 21.7 | 32.5 | 19.9 | 32.3 | 25.4 |
| 7 | 56.9 | 56.4 | 56.7 | 57.4 | 2.7 | 4.5 | 2.6 | 0.9 | 8.3 | 15.4 | 8.7 | 19.2 | 32.1 | 23.6 | 32.0 | 22.6 |

*Baseline (Base.), No Learning (No L), and Fast Learning at Level $k$ (Fast L at $k$), $k = 1, 2$. No Learning: $\beta_k = \hat{\alpha}_k$, $k=1,2,3$. Fast Learning at Level $k$: $\alpha_k = 0.99$ and $\beta_k = 0.01$, $k=1,2$.

Table A.8: Counterfactual Experiment on Importance of Experimentation for Level Assignments

| | Separation | | Level 1 | | Level 2 | | Level 3 | |
|---|---|---|---|---|---|---|---|---|
| Tenure | Baseline | No Experimentation | Baseline | No Experimentation | Baseline | No Experimentation | Baseline | No Experimentation |
| 1 | 0.0 | 0.0 | 100.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | 14.5 | 14.5 | 45.7 | 84.6 | 39.8 | 0.9 | 0.0 | 0.0 |
| 3 | 26.5 | 26.9 | 17.2 | 6.2 | 47.3 | 18.3 | 8.9 | 48.6 |
| 4 | 37.1 | 36.4 | 8.1 | 2.0 | 29.2 | 7.7 | 25.6 | 53.8 |
| 5 | 45.3 | 45.3 | 5.3 | 1.1 | 18.3 | 3.5 | 31.2 | 50.2 |
| 6 | 51.5 | 51.8 | 3.4 | 0.6 | 12.6 | 2.0 | 32.5 | 45.6 |
| 7 | 56.9 | 57.3 | 2.7 | 0.5 | 8.3 | 1.3 | 32.1 | 40.9 |

*No Experimentation: $\alpha_k = \hat{\alpha}_1$ and $\beta_k = \hat{\beta}_1,\ k=2,3$.