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 2019

The Hoover Institution Economics Working Paper Series allows authors to distribute research for discussion and comment among other researchers. Working papers reflect the views of the authors and not the views of the Hoover Institution.

1 I am indebted to Ken Wolpin and Petra Todd for their advice. I am grateful to Moshe Buchinsky, John Geweke, Mike Keane, and Fabian Lange for their suggestions. Jasper Bagger, John Cochrane, Zvi Eckstein, Eric French, Limor Golan, Lance Lochner, Fabien Postel-Vinay, Theo Papageorgiou, Rob Sauer, and Yasutora Watanabe, and especially Pete Klenow, Eddie Lazear, Jeremy Lise, Bob Miller, Luigi Pistaferri, and Mike Waldman, have offered many valuable comments. I also thank George Baker for kindly sharing the data and Bengt Holmström for his support of the project. Sam Bailey, Ross Batzer, Cristina Gualdani, and Zhen Huo provided excellent research assistance and Joan Gieseke invaluable editorial help.
Careers in Firms: the Role of Learning and Human Capital
Elena Pastorino
Economics Working Paper 19118
February 2019
JEL Codes: D22, D83, J24, J31, J44, J62
Keywords: Careers, Job Assignment, Uncertainty, Learning, Human Capital, Selection, Wage Growth

Elena Pastorino
Hoover Institution,
Stanford University, and Federal Reserve Bank of Minneapolis

Abstract:
Workers experience wage increases not only when they turn over across firms but also when they progress through a firm’s internal hierarchy of jobs. The workers who are promoted to higher-level jobs within a firm, typically after good performance, are likely to enjoy further promotions and wage increases. The workers who are not promoted, however, tend to 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’ abilities and workers acquiring human capital with experience. To date, however, little is known about the relative importance of these two sources of wage growth. Using administrative data on one firm first analyzed by Baker, Gibbs, and Holmstrom (1994a, b), I estimate a structural model in which firms and workers learn about workers’ abilities and workers accumulate human capital with employment. The model accounts for the rich patterns of job and wage mobility in the data. I find evidence for a novel mechanism through which learning about ability shapes wages. Learning affects wages not just directly through the impact of current beliefs about ability on current wages, but also indirectly through promotions by allowing the sorting of workers to jobs over time based on their ability. Through this indirect effect, which is absent from existing empirical models of learning, I estimate that learning about ability is a key determinant of the growth and dispersion of individual wages.

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

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 (see Topel and Ward (1992), Keane and Wolpin (1997), and Buchinsky et al. (2010)), especially within firms as workers advance through a firm’s internal hierarchy of jobs (see Lazear (1992, 2009) and the reviews by Gibbons and Waldman (1999a), Lazear and Shaw (2007), and Waldman (2013)).\(^1\) The wages of workers who are promoted at least once are 28 percent higher after five years but only 7 percent higher for unpromoted workers, because of the combination of wage increases at promotion and higher wages at higher levels of a firm’s job hierarchy (see Baker and Holmström (1995) and, for similar findings, Waldman (2013) and Frederiksen et al. (2017)). Workers who are promoted or receive wage increases once are more likely to receive future promotions and wage increases. Thus, understanding the determinants of job mobility in firms is central to understanding individual wage growth.

Yet, the dynamics of wages in firms cannot be described solely by the progression of workers through a firm’s job hierarchy. For instance, a large fraction of the annual changes in real wages that workers experience in a firm are negative, which leads not just to a sizable variability in individual wages over time but also to a significant dispersion in the wages at any given level of the job hierarchy (see Waldman (2013)). Uncovering the sources of this variability is important in 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 the dynamics of jobs and wages arises from a gradual process of learning about workers’ abilities, which are uncertain when workers enter the labor market, and human capital acquisition with experience in the market (see the review by Rubinstein and Weiss (2006)). Intuitively, as workers discover their talents and accumulate new skills through their success or failure on the job, they eventually settle in the jobs and firms that are the best match, and offer the highest rewards, for their productivity.

Indeed, learning and human capital acquisition have long been recognized as key sources of the dynamics of jobs and wages. To date, however, their contribution to wage growth is unclear (see Neal and Rosen (2000), Rubinstein and Weiss (2006), and Neal (2017).) One reason is that the required measurement exercise is theoretically and empirically demanding. Theoretically, when jobs differ in the output they generate and in the opportunities they offer for information and human capital acquisition, firms assign workers to jobs and workers sort into firms in the face of repeated trade-offs between current and future output and wages. As a result, firms and workers might be willing to experiment—that is, sacrifice current output and wages in order to acquire more precise information about ability. This information acquisition problem is nonstandard whenever ability is at least partly transferable across jobs or firms. Empirically, observed job and wage mobility are the result of a complex dynamic selection process whereby workers match with the jobs and firms that best fit their accumulating information and skills and will determine their future information and skills.

This paper estimates the importance of learning about ability and human capital acquisition for careers within a model that accounts for detailed patterns of job and wage mobility in firms. Specifically, I estimate how workers sort into

---

\(^1\)For wage growth between firms, see the review by Rubinstein and Weiss (2006).
firms and jobs and the extent to which this sorting and the resulting wage growth are 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, crucially, performance of all managers of a U.S. firm first analyzed by Baker, Gibbs, and Holmström (1994a,b), henceforth, BGH. Based on this information, I infer how the organization of production within a firm, and hence the allocation of workers to jobs, affects the acquisition of information about workers’ ability and workers’ human capital and, conversely, how these two investment processes influence the dynamics of jobs and wages in a firm.

Three main findings emerge. First, in contrast to the literature (see Gibbons et al. (2005) and Lluis (2005)), I find the contribution of learning to wages to be sizable: the additional wage growth and dispersion due to learning amount to one-quarter of the cumulative wage growth and dispersion over the first seven years at the firm I study; the rest is attributable to human capital acquisition. I separate the impact of learning on wages into a direct effect and an indirect effect. I estimate that the direct effect, which captures the impact of current beliefs about ability on current wages, is small, consistent with the existing literature that focuses solely on this effect. Unlike the existing literature, by estimating the endogenous dynamics of beliefs, human capital, jobs, and wages, I can also measure the indirect effect of learning on wages, due to its impact on the dynamics of promotions within the firm. I find that this indirect effect is large and responsible for almost all of the impact of learning on wages.

Intuitively, the indirect effect of learning on wages operates through the endogenous process by which managers are progressively selected to higher levels of the firm’s job hierarchy. As information about managers’ ability and their human capital accumulate with employment and are revealed by performance, managers who perform well advance to the jobs that are most suited to their ability and acquired skills. Since wages at higher-level jobs are on average much higher than at lower-level ones, this sorting process is key to wage growth with tenure. The combined patterns of promotions, wages, and job performance in my data suggest that this indirect effect is indeed present: wage growth primarily occurs as higher performance leads to promotions to jobs at which managers are paid higher wages. The dynamic selection giving rise to the indirect effect of learning on wages 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 based on observed and unobserved characteristics for careers and educational attainment.

The second main finding is that the differential informativeness of jobs, ignored in models of careers, is a crucial determinant of wage profiles. To see why, note that, on the one hand, according to my estimates, the firm’s lowest job

---

2The BGH data are famous 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).

3Search frictions and performance incentives are other prominent sources of wage growth. Like most of the literature on the careers of non-executives, I focus on learning and human capital acquisition. I find, however, little evidence that performance incentives matter for careers at the firm I study; see Section 8 for details. See Gayle and Miller (2009, 2015) and Margiotta and Miller (2000) on the importance of performance incentives for executive pay, and Gayle et al. (2015) for a model of promotions, turnover, and compensation in the executive labor market that accounts for the relationship between firm size and executive pay. See Sections 7 and 8 for further discussion.

4My approach is analogous to that of Buchinsky et al. (2010), who argue that to estimate returns to tenure and experience accounting for their endogeneity, one needs to estimate not just the wage process but also the process for worker employment and mobility decisions. Similarly, to estimate the importance of learning not just for wages accounting for the endogeneity of employment, job tenure, and beliefs, one needs to estimate not just the wage process but also the process for employment and job assignment decisions and for the evolution of beliefs.

level is most informative about ability, but wages at this level are the lowest. The informational benefit of this job level, at which managers are hired, relative to the remaining ones implies that managers are employed at this level even for beliefs about ability at which they would produce more at higher levels. This force thus *depresses* wages in *early* tenures relative to the case of equally informative jobs. On the other hand, learning about ability is faster at this lowest 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 naturally *increases* wages in *later* tenures relative to the equally informative case. Hence, the differential informativeness of jobs not only explains why wage returns may be small in early tenures but also accounts for the markedly convex relationship between wages and job levels, or tenure, in firms (see Waldman (2013)).

The differential informativeness of jobs is also important for evaluating the impact of learning on wages. As argued, the firm acquires information about managers by first assigning them to jobs that are especially informative about ability, even if they contribute little to output. Since these jobs pay low wages, learning results in a *delay* in promotions to higher-paying jobs. Thus, it may take time for learning to have a positive effect on wages through promotions. To measure the impact of learning on wages, it is then critical to capture its *cumulative* effect on job assignment. As this cumulative effect arises from the sorting of managers to jobs, which is also affected by human capital, a priori the impact of learning on wages could be small even once its indirect effect on promotions is taken into account. Interestingly, I find that, through this effect, learning has a large impact on wages.

The third main finding is that wages systematically differ from output. This evidence suggests that the connection between the so-called labor markets “internal” and “external” to firms is less obvious than implied by perfect competition, which is commonly assumed in models of careers, and is key to the dynamics of wages. In particular, because of imperfect competition among firms, wages are less sensitive than output to changes in beliefs about ability that do not entail changes in job levels, which accounts for why wage growth accrues primarily through promotions.

Formally, I consider a model of the 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, GW1, GW2, or GW. Firms operate technologies with constant returns to scale in which labor is the only input. Production in firms is organized in distinct *jobs*, which differ in the output they generate and the information they provide about ability. Thus, unlike in well-known learning models of the labor market (see Jovanovic (1979), Miller (1984), and Flinn (1986)), learning is correlated across jobs and firms, and the speed of learning differs across them. For instance, performance at entry-level jobs, which usually entail simple tasks of limited value to a firm, may provide a more precise signal about ability than at higher-level jobs, which normally involve more complex tasks of greater value to a firm. When employed, workers also stochastically accumulate human capital, which is job- and firm-specific to varying degrees. Firms compete for workers in jobs and wages, which, as firms may differ in their technologies, generally differ from expected output.

---

6 Differences in information across jobs further help to rationalize the wage increases paid at promotion even to experienced workers whose ability is less uncertain, which is a puzzle for learning models (see Gibbons and Waldman (2006)). Intuitively, when promotions entail switching from more to less informative jobs, these premia compensate workers for the *discrete* loss in information resulting from their new assignments, and so can be large even for small changes in uncertainty about ability, as is the case for workers with several years of labor market experience.

7 See Prescott and Visscher (1980) and Holmström and Tirole (1989) for early intuitions on this output-information trade-off. With ability
Crucial to the empirical exercise of interest is identifying the process of learning and human capital acquisition. The BGH data contain the discrete outcome of the yearly evaluation of each manager’s performance, which provides direct evidence of the information, or *signals*, at each job that firms and managers use to learn about managers’ ability; see BGH on this point. I show that information on performance, even when noisy, is sufficient to identify the distribution of signals about managers’ ability at each job and thus the law of motion of beliefs about ability. Joint information on performance and wages identifies the distribution of the initial prior beliefs about ability when managers enter the firm, which, together with that of signals, completely describes the learning process. Repeated information on job assignment and performance identifies the process of managers’ output at the firm, which pins down the human capital process.

I estimate the model by maximum likelihood using eight years of observations on managers entering the firm between 1970 and 1979, imposing all of the model restrictions. The model captures well the rich patterns of jobs and wages described above as well as finer moments of the data, including the tenure profile of job-to-job transitions at the firm and separations and the distribution of performance and wages at the main job levels in each tenure.

The estimates shed light on several features of the process of information acquisition at the firm, which are central to the impact of learning on wages. First, uncertainty about ability at entry in the firm is substantial. Second, learning about ability is gradual. Yet, despite the estimated slow speed of learning, by comparing the estimated wage growth to that arising in a counterfactual scenario in which learning is precluded, I find that learning contributes to more than 25 percent of wage growth and dispersion over the first seven years at the firm. Absent learning, wages would increase much more slowly, owing to the lower speed of promotions, and would be significantly less dispersed. Third, since I estimate learning to be gradual, uncertainty about ability is highly persistent and significantly reduces the pace of transition to higher levels, relative to the counterfactual case in which ability is learned in one period. Thus, both the differential informativeness of jobs and persistent uncertainty about ability are responsible for a substantial compression of wage growth in early tenures and in this way account for almost all of the observed convexity of wages in job levels and tenure.

As for the human capital process at the firm, I estimate that acquired human capital does not have much of a direct effect on wages, in that in any period, it has a limited impact on a manager’s wage in the same period. Yet, human capital accounts for most of the wage growth at the firm. Intuitively, like learning, human capital acquisition primarily affects wages through its indirect effect on the dynamics of job assignment, for two reasons. First, managers acquire skills over time that make them more productive at higher levels. Thus, the human capital process induces the firm to promote managers to jobs for which they are progressively better suited and that also pay higher wages, because of their greater contribution to firms’ output. Second, human capital acquisition, albeit stochastic, increases a manager’s productivity on average and so makes demotions less likely for any given decrease in beliefs about ability after low performance.

Comparing average output, as inferred from job assignments and performance, and wages further reveals that their correlation is low, which implies that the labor market where the firm hires its managers is imperfectly competitive. In

---

correlated across jobs or firms, firms and workers face a multi-armed *dependent* bandit problem, which implies worker turnover across jobs or firms not just after bad performance, as implied by an independent bandit problem, but also after good performance, as in the data. See Jovanovic (1979), Miller (1984), and Flinn (1986) for influential applications of the independent bandit problem to turnover across firms and occupations.
particular, I find that learning has a large direct effect on output and so on job assignment, despite having a small direct effect on wages. Namely, since output is sensitive to ability, particularly at higher job levels, managers revealed to be of high ability are promoted to these levels where output and wages are higher. As a result, wage growth crucially arises from a better sorting of managers to jobs: learning has a large impact on this growth since it stimulates output growth through improved job assignment. I show that abstracting from imperfect labor market competition would lead to a downward bias in the estimate of the importance of learning for output growth and so for wage growth.

The rest of the paper is organized as follows. Section 2 examines the data, Section 3 describes the model, Section 4 considers identification and model specification, Section 5 presents the model estimates, Section 6 contains the counterfactual exercises, Section 7 reviews the literature, Section 8 discusses alternative models, and Section 9 concludes. See Appendices A and B for omitted proofs and details and the Supplementary Appendix (S.A.) for additional material.

2 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 between 1969 and 1988. I use information on managers’ year of entry, age, education, job level, salary, and job performance rating. I exclude bonus information since it is unavailable before 1981 and, when available, is missing for no fewer than 45.8 percent of managers in each tenure with higher percentages in early tenures (up to 100 percent). See Ekinci et al. (forthcoming) for an analysis of bonuses in the BGH data. The firm’s job hierarchy consists of eight levels, Level 1 through Level 8 (Chairman-CEO), but is clearly divided into two distinct parts: Levels 1 to 3 and 4, which contain nearly all observations (97.6 percent), and Levels 5 to 8, which amount to few observations on top management positions. I use BGH’s Levels 1 and 2, and aggregate Levels 3 to 8 into a single one, henceforth referred to as Level 3. The size of each level varies considerably across years: managers at Level 1 range from 400 to 1,078, at Level 2 from 375 to 1,254, and at Level 3 from 496 to 3,199. Performance ratings 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 rating observations across the sample years are 1 or 2. Hence, I combined ratings of 2 through 5 into a single measure denoted by “0,” thus obtaining a binary classification of high and low performance. Like BGH, I consider ratings as year-end variables. I now turn to describe the main features of the data that will inform the model. See BGH for an extensive descriptive analysis of careers at the firm.

Sample Characteristics. I focus on 1,426 managers who enter managerial positions between 1970 and 1979 at Level 1 with 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 job-level information missing. No such manager has either age or education information
missing or is older than 45 years at entry. Because of the high separation rate from the firm each year (see Table 1), I restrict attention to managers’ first eight years at the firm. I consider entrants at Level 1 between 1970 and 1979 so as to be able to: a) compare my results with those of BGH, as most of their analysis also concerns entrants at the firm at Level 1 over these years; b) observe the greatest number of job transitions for a manager, since demotions are (almost) never observed; and c) avoid excessive right-censoring of the careers of later entrants. For completeness, I also estimated the model on a larger sample that includes entrants at the original Levels 1 through 4. This larger sample displays features that are very similar to those of the sample of entrants at Level 1, and the estimates of key parameters based on it are almost indistinguishable from those reported below; see the S.A. for details.

**Job Assignments.** Table 1 displays two main features of separations and job assignments at the firm. First, separation rates are high in all tenures. By the seventh year, over half of the managers hired at Level 1 have separated. Second, the percentage of managers assigned to Level 1 rapidly decreases with tenure, whereas the percentage of managers assigned to Levels 2 and 3 first increases and then decreases with tenure. Table 2 converts these distributions into level-specific hazard rates of separation, retention at the same level, and promotion to the next level by tenure. Three further features emerge. First, separation hazards are approximately constant with tenure in the firm at each level. Second, promotion hazards initially increase and then decrease with tenure. Third, promotions are by one level, and demotions never occur—demotions are rare in the BGH data, as is common in most firm-level data.

These patterns discipline key features of the model proposed below. First, they are consistent with a gradual process of learning and stochastic human capital acquisition whereby managers are assigned to higher levels of the job hierarchy over time, as their ability is learned and their human capital accumulates, or they separate from the firm. If, on average, human capital accumulation increases a manager’s productivity, then acquired human capital offsets the adverse impact of low performance on beliefs about ability over time, so that managers are retained at a level, rather than demoted to a lower level, even after repeated low performance.

Second, this learning and stochastic human capital process can account for the nonmonotone tenure profile of promotion rates in two ways, which the model will incorporate. First, as the best managers are promoted out of a level, the remaining ones naturally face worse promotion prospects, which explains the eventually declining probability of promotion for managers who are continually assigned to a level. Second, a job-specific component of human capital can also generate the observed promotion profiles. Suppose, say, that acquired human capital makes a manager more productive on average at a given level but not at any other. Then, the probability of promotion out of that level eventually decreases with tenure, as the greater human capital a manager acquires there makes the manager better suited to that level.

**Performance Ratings.** Table 3 shows two patterns about performance 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. As

---

10The 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 exclude entrants in 1969 since it is unclear from the data in which year managers observed in 1969 entered.

11Ratings at Level 3 are missing for no fewer than 35 percent of individuals assigned to (the original) Levels 3 and higher, with much higher proportions of missing values in lower tenures. Moreover, the distribution of ratings at Levels 3 and higher in the data differs from the distribution
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 support the idea that ability and acquired human capital affect performance on the job and are more valuable at higher levels. In this interpretation, managers who receive high ratings earlier in their tenure in a level are characterized, on average, by higher priors about their ability and higher human capital. Thus, they are naturally the first assigned to higher levels, where higher ability and human capital have a greater impact on output. This selection process leads managers with higher ability and human capital to progressively advance to higher job levels, and so explains the higher average performance of managers at higher levels than at lower levels. By the same process, average performance is lower for unpromoted than for promoted managers, and deteriorates with tenure in a level.

**Wages.** Table 4 displays the distribution of wages at the firm (in 1988 U.S. dollars) by level and tenure, and Table 5 reports statistics on the distribution of wage changes by tenure. Three features emerge from Tables 4 and 5. First, wages are on average higher at higher levels—the percentage of managers earning wages above $40K increases with the job level—and the spread of the level distribution of wages increases with the level—the standard deviation of wages reported in the note to Table 4 is higher at higher levels. Since managers at higher levels tend to receive higher performance ratings, as discussed, managers at higher levels perform better on average and are paid more. These features are consistent with a sorting process based on revealed ability and acquired human capital whereby more productive managers are assigned to higher levels over time and receive higher wages. This evidence also supports the idea that wages become more dispersed across managers as their realized human capital increasingly differs.

Second, negative wage changes are quite frequent—over 20 percent in each tenure—as consistent with the presence of learning and stochastic human capital acquisition. Intuitively, since low performance lowers beliefs about ability and implies lower levels of realized human capital, it reduces the perceived 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 simply governed by a manager’s progression through the job ranks of the firm’s hierarchy.

**Case for Integrated Model.** In the BGH data, as is common in firm-level data, wages increase with tenure, job level, and performance; promotions and wage increases occur overwhelmingly after good performance on the job and are correlated over time; demotions (almost) never occur despite wage decreases being common after low performance; and wage dispersion at each level is substantial. (See BGH for details on the serial correlation of promotions and wage increases and the dependence of wages on 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 and new skills while employed. Crucially, I assume that ability is correlated across jobs and firms to account for managers switching jobs after good and bad performance—if ability, instead, were independent, job transitions would only occur after bad performance by the so-called stay-on-a-winner property of (bandit) models of learning with

---

The estimation of the parameters of the distribution of performance ratings at Level 3 thus proved problematic. For these reasons, I do not use information on performance ratings at Level 3, and so omit statistics on ratings at this level. See Section 4.2 for details.
independent ability. Also, if higher ability and acquired skills are more valuable at higher levels, then managers sort through a firm’s jobs over time, and possibly turn over across firms, as information and human capital accumulate.

I interpret jobs as affecting the rate at which information about ability and human capital are acquired. Hence, firms and managers typically face a trade-off between current output and the precision of the information they can gather about ability. This trade-off explains the often large wage increases paid at promotion as compensating managers for switching from more to less informative jobs. More generally, it helps to rationalize the nonmonotone growth of wages with tenure.

Such a framework combining learning and stochastic human capital acquisition can naturally account for the features of the data discussed so far. On the one hand, learning in a model with job assignment can explain not just mobility across jobs and firms, but, unlike a model with only random productivity shocks, can also explain the observed serial correlation in wage increases, promotions, and performance; see also GW and Waldman (2013). The learning process further leads to wage increases and promotions after high performance but also to wage decreases and demotions after low performance. On the other hand, acquired human capital gives rise to promotions and wage increases, on average, by augmenting managers’ productivity and so accounts for the low frequency of demotions relative to promotions.

Intuitively, then, for any given frequency of wage decreases, the lower the frequency of demotions relative to promotions over time, the larger the role of human capital acquisition. Similarly, for any given frequency of wage increases and promotions, the higher the serial correlation in wage increases and promotions, the larger the role of unobserved ability and so the scope for learning—see the analogy with the standard runs test for the non-randomness of data.

Overall, the patterns of promotions, performance ratings, and wages at the firm are consistent with the idea that learning affects wages not only directly, as managers perceived to be of higher ability and with higher human capital are paid more on average, but also indirectly through its impact on job assignment. That is, high performance ratings, and so higher beliefs about ability, are associated with a greater chance of promotions, which, in turn, lead to higher wages. This is the channel through which the indirect effect of learning on wages operates in the model, which I describe next.

3 A Model of Career Paths

I describe here the labor market, human capital and output technologies, firms and workers’ information, the timing of decisions, and equilibrium. Briefly, the labor market consists of firms that produce output using workers’ effective labor or human capital as the only input. Workers differ in their ability and human capital, which depends on ability, and evolves with employment, stochastically. Since ability is unobserved to all, the realized process of human capital, which can be inferred from output, provides information about workers’ ability. See Appendix A for omitted details.

**Firms and Workers.** I consider a labor market for an occupation (managers) in which two firms, \( f = A, C \), compete for workers. Time is discrete and indexed by \( t \geq 1 \). Firms and workers discount the future by the factor \( \delta \in (0, 1) \). Firms produce a homogeneous good sold in a perfectly competitive market at a price normalized to one. I will occasionally refer to firm \( A \) as *my firm* (the firm in my data) and to firm \( C \) as firm \( A \)’s competitor. Each firm \( f \) consists of a set \( K_f \) of jobs for a total of \( N \) jobs per firm. Each job is defined by a constant returns to scale technology in labor, the only
input to production. I interpret a set of jobs as a grouping of productive tasks: different firms may group the same tasks into different jobs according to their specific organization of production. For instance, tasks are known to be grouped by function, such accounting or marketing, in some firms and by product in others. As in GW, a worker is characterized by one of two ability levels, \( \theta \in \{ \alpha, \beta \} \), unobserved to all (including the worker), and his human capital. Each ability level corresponds to a vector of probabilities of success at each job \( k \) of each firm \( f \), so that \( \alpha = \{ \alpha_{fk} \} \) and \( \beta = \{ \beta_{fk} \} \), \( \alpha_{fk} \geq \beta_{fk} \); the event “success” is described below. I refer to a worker of ability \( \alpha \) as a high-ability worker and a worker of ability \( \beta \) as a low-ability worker. I assume that the value of not working in the market considered is sufficiently low that a worker is employed each period.

**Human Capital and Output.** Output at any job depends on the amount of effective labor that a worker supplies, referred to as human capital. A worker in any period \( t \) is endowed with three observed human capital stocks: the human capital acquired before entry in the market, \( H_1 \); the human capital acquired with general experience at the firms in the market, \( H_2^g = \{ H_{A_{kt}}^g, H_{C_{kt}}^g \}_k \); and the human capital acquired with specific experience at the firms’ jobs, \( H_2^s = \{ H_{A_{kt}}^s, H_{C_{kt}}^s \}_k \). General and job-specific human capital accumulate stochastically with employment through a process of learning-by-doing, which depends on a worker’s ability, employing firm, and assigned job. The production functions of general (\( j = g \)) and job-specific (\( j = s \)) human capital at job \( k \) of firm \( f \) in \( t \) are

\[
H_{fkt}^j = \alpha_{fkt} + \beta_{fkt} \varepsilon_{fkt}^j,
\]

where \( \alpha_{fkt} \) and \( \beta_{fkt} \) are observed by all. The term \( \ln A_{fkt}^j \) is stochastic and its distribution depends on a worker’s ability \( \theta \) and market experience \( t \). \( i_{k\tau}^j \) is the investment in period \( \tau \leq t - 1 \), and \( \beta_{fkt} = \ln \varepsilon_{fkt}^j \) is a zero-mean i.i.d. human capital or productivity shock capturing any temporary unexpected change in human capital unrelated to ability or human capital. The investment \( i_{k\tau}^j \) in general capital equals the constant 7 if a worker is employed at any job of either firm in period \( \tau \) and zero otherwise. The investment \( i_{k\tau}^s \) in job-specific capital equals the constant 7 if a worker is employed at job \( k_{\tau} \) of either firm in \( \tau \) and zero otherwise. Thus, general human capital accumulates with time spent in the market, whereas job-specific human capital accumulates with time spent at each job. Investments, \( \alpha_{fkt}^j \) and \( \beta_{fkt}^j \) are observed by all.

When a worker of ability \( \theta \) is employed at job \( k \) of firm \( f \) in \( t \), \( \alpha_{fkt}^j \) equals \( \alpha_{fkt} \) with probability \( p_{fkt}(\theta) \) and \( \beta_{fkt}^j \) otherwise. The probability \( p_{fkt}(\theta) \) equals \( \alpha_{fk} \) for a high-ability worker and \( \beta_{fk} \) for a low-ability worker. When a worker is not employed at job \( k \) of firm \( f \) in \( t \), \( \alpha_{fkt}^j \) equals the constant \( a_t \). Here the specific value of \( a_t \) is irrelevant as it only matters that realizations of \( \alpha_{fkt}^j \) at jobs not performed do not depend on ability and thus are uninformative about it. Note that ability influences the evolution of human capital by affecting the probabilities of the realizations of \( \alpha_{fkt}^j \) rather than its support. The support of \( \alpha_{fkt}^j \), though, depends on a worker’s experience in the market. As \( \alpha_{fk} \geq \beta_{fk} \), the human

---

12 By allowing job technologies to differ across firms, this setup nests the case of perfect competition when firms have the same technology, of a differentiated duopoly when firms have different technologies, or even of a virtual monopoly when, say, A’s technology is more efficient than C’s.

13 This human capital process generalizes that in Bagger et al. (2014) to a setting with multi-job firms and learning about ability. In Bagger et al. (2014), log human capital is \( h_t = a + g(t) + \varepsilon_t \), where \( a \) is a fixed known worker productivity parameter, \( g(t) \) is a trend capturing human capital accumulation with experience, and \( \varepsilon_t \) is a zero-mean i.i.d. shock. I abstract from differences in \( g(t) \) across firms in light of the assumptions maintained in the empirical analysis; see the S.A. for this extension. See Appendix B for examples of laws of motion of human capital of the form

\[
H_{fkt}^j = C_{fkt}^j ([1 - \sigma^f] H_{fkt-1}^j + \varepsilon_{fkt-1}^j) + \varepsilon_{fkt-1}^j \]

with \( C_{fkt}^j \) stochastic, which can be expressed in cumulative form as in (1).
capital of a high-ability worker is on average higher than that of a low-ability worker. Denote by $\overline{h}_{fkt}$ and $\overline{h}_{fkt}$ the human capital at the end of $t$ of a worker employed at job $k$ of firm $f$ after $\overline{w}_{fkt}$ (“success”) and $\overline{w}_{fkt}$ (“failure”).

Normalizing each worker’s labor supply to one, a worker’s (log) output at job $k$ of firm $f$ in $t$ is a function of the worker’s initial and acquired human capital given by

$$y_{fkt} = b_{fkt}(h_1) + c^g_{fkt}(h^g_{fkt}) + c^s_{fkt}(h^s_{fkt}) - \phi_{fkt}(k_{t-1}),$$

where $c^s_{fkt}(h^s_{fkt}) = c^s_{fkt} + c^s_{fkt}h^s_{fkt}$. Both $b_{fkt}(\cdot)$ and $c^s_{fkt}(\cdot)$ are indexed by $f$ to allow for different uses of human capital across firms; see Lazear (2009) for this notion of firm-specific human capital. The term $\phi_{fkt}(k_{t-1})$, where $k_{t-1}$ denotes a worker’s job in $t-1$, is a cost of employing workers or switching them across jobs, which accounts for any capacity (“slot”) constraint or fixed cost of employment that may limit a firm’s ability to employ or assign workers to jobs.\(^{14}\)

Denote output at the end of $t$ by $\overline{y}_{fkt}$ if $(\overline{h}^g_{fkt}, \overline{h}^s_{fkt})$ or, equivalently, $(\overline{w}^g_{fkt}, \overline{w}^s_{fkt})$ are realized and by $\underline{y}_{fkt}$ if $(h^g_{fkt}, h^s_{fkt})$ or, equivalently, $(w^g_{fkt}, w^s_{fkt})$ are realized. Since both human capital and output are observed, the realized values $(a^g_{fkt}, a^s_{fkt})$ can be inferred from output. Thus, interpreting $(a^g_{fkt}, a^g_{fkt})$, rather than $(h^g_{fkt}, h^s_{fkt})$ or $y_{fkt}$, as the period signal about ability is without loss. I refer to $(\overline{w}^g_{fkt}, \overline{w}^s_{fkt})$ and $(\overline{a}^g_{fkt}, \overline{a}^s_{fkt})$, respectively, as high and low performance: $a^g_{fkt}$ and $a^s_{fkt}$ are assumed to be perfectly correlated and specified as Bernoulli just for simplicity and consistency with the one-dimensional and essentially binary measure of performance in the data. As the probability of high performance differs across firms’ jobs, firms and workers learn about a worker’s ability over time from a worker’s history of jobs and performance. The speed of learning is governed by the frequency of high performance for high- and low-ability workers, $\alpha_{fkt}$ and $\beta_{fkt}$. (If, instead, the support of $a^s_{fkt}$ depended on $\theta$, ability could be learned in one period.)

**Expected Output.** To understand firms’ employment and assignment decisions, it is useful to define a worker’s expected output given the prior $p_t$. Note that at the beginning of $t$, a worker’s ability and human capital are summarized by $(\theta, \kappa_t)$ with $\kappa_t = (h_1, \{\overline{w}_{fkt}i\}_{i=1}^{t-1})$. Denote expected output conditional on $(\theta, \kappa_t)$ by $E(y_{fkt}|\theta, \kappa_t)$ and expected output conditional on $(p_t, \kappa_t)$ by $y_f(p_t, \kappa_t, k) = E(y_{fkt}|\beta, \kappa_t) + [E(y_{fkt}|\alpha, \kappa_t) - E(y_{fkt}|\beta, \kappa_t)]p_t$, where, by (1) and (2), one can show

$$y_f(p_t, \kappa_t, k) = b_{fkt}(h_1) + \sum_j c^j_{fkt}E(a^j_{fkt}|\beta) + c^j_{fkt}E(a^j_{fkt}|\alpha) - \phi_{fkt}(k_{t-1})\sum_j c^j_{fkt}[E(a^j_{fkt}|\alpha) - E(a^j_{fkt}|\beta)].$$

A high-ability worker at firm $f$ has an absolute advantage at job $k$ if $e_{fkt} \geq 0$ and a comparative advantage at job $k'$ over job $k$ if $E(y_{fkt}|\alpha, \kappa_t) / E(y_{fkt}|\alpha, \kappa_t) \geq E(y_{fkt}|\beta, \kappa_t) / E(y_{fkt}|\beta, \kappa_t)$.\(^{15}\)

**Information and Belief Updating.** At the beginning of the first period, firms and workers share a common initial prior belief $p_1$ that a given worker is of high ability. Although here I consider the case of a single prior, in estimation I allow

\(^{14}\)The analysis here applies unchanged to the case in which output is subject to shocks independent of human capital, $z_{fkt}$, so that $y_{fkt} = b_{fkt}(h_1) + c^g_{fkt} + c^s_{fkt} + c^s_{fkt}h^s_{fkt} + c^s_{fkt}h^s_{fkt} - \phi_{fkt}(k_{t-1}) + z_{fkt}$, if, for instance, $e^j_{fkt} = e^j_{fkt} + z_{fkt}/c^j_{fkt}$ replaces $e^j_{fkt}$ in (1) for one $j$.

\(^{15}\)This setup allows for a more flexible impact of ability and human capital on output than GW, where $y_{fkt} = b(h_1) + b_k + c_kh_{fkt}$ and human capital, $h_{fkt} = \theta f(x_t) + \varepsilon_{fkt}$, is a function of ability and experience $x_t$, with $\varepsilon_{fkt}$ learning noise. Here human capital and output parameters can vary across firms and jobs, the speed of learning differs across jobs as $\alpha_{fkt}$ and $\beta_{fkt}$ vary with $k$, human capital accumulates with experience at firms and their jobs, and productivity shocks are present in addition to learning noise.
for different priors for different workers. Since performance is observed by all, learning about a worker’s ability based on a worker’s performance on the job is symmetric as in GW: firms and workers share the same information, and thus the same prior, in any $t$. I interpret this symmetric learning setting as a plausible approximation to the labor market in which my firm, a professional service one, hires its managers. Resumes, references, recommendation letters, and other information gathered by human resource departments can accurately convey a manager’s performance in previous jobs; see Oyer and Shaefer (2011). I also formulate this assumption for comparability with leading models of learning and comparative advantage, such as GW and Gibbons et al. (2005). The assumption that performance or, equivalently, human capital is commonly observed is less restrictive than it may seem: firms other than a worker’s current employer can use information on wages to infer a worker’s performance; see Section 8.

At the end of any period $t$ after production occurs, firms and workers update beliefs about a worker’s ability according to Bayes’ rule, which, for given $p_t$, leads to two possible values of $p_{t+1}$.

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

(4)

after high ($H$) or low ($L$) performance is realized, respectively. Since ability is correlated across jobs and firms, learning about ability is correlated across them. As the probability of high performance for a worker of either ability varies across jobs, the informativeness of a job, and so the implied speed of learning about ability, differ across jobs, ranging from no learning if $\alpha_{fk} = \beta_{fk}$ to complete learning after just one period of employment if $\alpha_{fk} = 1$ and $\beta_{fk} = 0$. By the Blackwell criterion of informativeness (see Blackwell (1951)), I define job $k'$ to be more informative than job $k$ if the posterior beliefs reached after performance is observed at job $k'$ at the end of a period second-order stochastically dominate those reached after performance is observed at job $k$. Note that jobs with the same expected output can be differentially informative about ability: as $\alpha_{fk}$ and $\beta_{fk}$ vary, expected output $y_f(p_t, \kappa_t, k) = [(\alpha_{fk} - \beta_{fk})p_t + \beta_{fk}](\bar{y}_{fkt} - y_{fkt}) + y_{fkt}$ can be kept constant by adjusting $\bar{y}_{fkt}$ and $y_{fkt}$. Thus, low output-variability jobs (with $\bar{y}_{fkt}$ close to $y_{fkt}$) can be more informative than high output-variability ones (with $\bar{y}_{fkt}$ much larger than $y_{fkt}$) or vice versa.

Separations. As information and human capital accumulate, workers’ evolving absolute and comparative advantage lead naturally to endogenous separations between workers and firms. I also account for separations unrelated to ability or human capital by allowing for exogenous separations: I assume a worker leaves the market at the end of each $t$ with probability $1 - \eta_{fkt}(\kappa_t)$, which depends on the worker’s employing firm $f$, job $k$, and human capital component $\kappa_t$.

Timing. At the start of $t$, productivity shocks are realized. Next, firms simultaneously submit offers to workers, which consist of a wage and job assignment for the period. Then, each worker decides which offer to accept. Next, the offered wage is paid, performance and so output are realized, beliefs about ability are updated, and separation shocks are realized. Note that since firms commit to the period offers they make, the timing of wage payments in a period is immaterial. I refer to $y_f(p_t, \kappa_t, k) + \varepsilon_{fkt}$ where $\varepsilon_{fkt} = \sum_j c_{fkt}^j \varepsilon_{fkt}^j$ 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. In this component game, the events in $t$ are $(\varepsilon_t, w_t, k_t, h_t, v_t, \zeta_t)$: $\varepsilon_t = \{\varepsilon_{fkt}\}$ denotes the vector of all productivity shocks; $(w_t, k_t) = \{w_{ft}, k_{ft}\}$
denotes the vector of each firm’s wage and job offer; \( l_t = \{ l_{ft} \} \) denotes the vector of the worker’s decisions to accept \((l_{ft} = 1)\) or reject \((l_{ft} = 0)\) each offer; \( u_t \) is an indicator for whether realized performance is high \((u_t = 1)\) or low \((u_t = 0)\); and \( \zeta_t \) is an indicator for whether the separation shock is realized and the component game ends \((\zeta_t = 1)\) or not \((\zeta_t = 0)\).

**Equilibrium.** I focus on robust Markov perfect equilibria. The state firms face each period when they make 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, \( \kappa_t = (h_1, \{ i_{kt} \}_{t=1}^{\tau-1}) \) summarizes acquired 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 \( P_{fHk}(p_t) \) 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

\[
W(s_t, \varepsilon_t, w_t, k_t) = \max_{\{ l_f \}} \sum_f l_f \left\{ w_f + \delta \eta_{fkt}(\kappa_t) \int_{\varepsilon_{t+1}} EW(s_{t+1}, \varepsilon_{t+1}, w_{t+1}, k_{t+1}|s_t, k_f, l_t) dG \right\}, \tag{5}
\]

where \( (w_{t+1}, k_{t+1}) \) is the future set of offers and \( G \) is the cumulative distribution function of future productivity shocks, \( \varepsilon_{t+1} \). The expectation \( EW(\cdot) \) is over performance at the accepted job, \( k_{ft} \). Conditional on the choice of job \( k_{ft} \), the beginning-of-period state next period is then \((s_{t+1}, \varepsilon_{t+1}) = (P_{fHk}(p_t), \kappa_{t+1}, \varepsilon_{t+1}) \) or \((s_{t+1}, \varepsilon_{t+1}) = (P_{fLk}(p_t), \kappa_{t+1}, \varepsilon_{t+1}) \) if high or low performance is realized. Note that in evaluating an offer, the worker weighs the offered wage, \( w_{ft} \), against the prospect of information and human capital accumulation at the offered job, \( k_{ft} \), as well as the risk of exogenous separation. Given the worker’s and the competitor’s strategies, firm \( f \)’s strategy, \( f \neq j \), satisfies

\[
\Pi^f(s_t, \varepsilon_t) = \max_{w, k} \left\{ l_{ft} \left\{ y_f(s_t, k) + \varepsilon_{fkt} - w + \delta \eta_{fkt}(\kappa_t) \int_{\varepsilon_{t+1}} E\Pi^f(s_{t+1}, \varepsilon_{t+1}|s_t, k)dG \right\} \right. \\
+ l_{jt} \delta \eta_{jkt}(\kappa_t) \int_{\varepsilon_{t+1}} E\Pi^j(s_{t+1}, \varepsilon_{t+1}|s_t, k_{jt})dG \right\}, \tag{6}
\]

where \( l_{jt} = l_j(s_t, \varepsilon_t, w_t, k_t) \) is the worker’s acceptance decision about firm \( j \)’s offer in \( t; k_{jt} \) is the job offered by firm \( j \) in \( t \); and \( E\Pi^f(\cdot) \) is the expectation over performance at the end of \( t \) at the accepted job, either job \( k \) if firm \( f \) employs the worker or job \( k_{jt} \) if firm \( j \) employs him. In choosing the job to offer, a firm trades off the worker’s conditional expected output, \( y_f(s_t, k) + \varepsilon_{fkt} \), against the value of the information conveyed by performance and of the worker’s future human capital, discounted by the probability of the worker’s exit from the market. By the second line in (6), a firm takes into account the option value of not employing the worker in the current period and attracting him in some future period, which arises from the information revealed by performance and the human capital the worker acquires at the competitor.

---

\(^{16}\)Productivity shocks stochastically change the value of a job to a firm across periods and workers so that the variability in firms’ job offers over time and among workers is not just due to learning about ability and human capital acquisition. These shocks also conveniently make a firm’s job offer, which occurs before performance is realized, random from the point of view of the econometrician, conditional on a worker’s ability, prior, and accumulated human capital, and so make maximum likelihood estimation feasible.
From (5) and (6), it is immediate that workers and firms actively learn or \textit{experiment} in that they contemplate sacrificing current wages and output to acquire more information about ability. However, unlike in a standard bandit problem in which a decision maker repeatedly chooses among alternatives with uncertain \textit{independent} rewards to learn about their distribution, here the arms of the bandit—the jobs of all firms—are \textit{dependent}. Namely, payoffs in the form of each job’s wages and output depend on the worker’s unobserved ability, which is correlated across jobs and firms.

I require a Markov perfect equilibrium to be \textit{robust} by assuming that the losing firm’s offer is such that the losing firm would not incur a loss if the worker deviated from the equilibrium path and accepted its offer. Formally, if firm $f$ does not employ the worker in period $t$ but firm $j$ does, then firm $f$’s rejected offer must make it indifferent between not employing and employing the worker. That is,

$$
\delta_{jk} = \max_{y} \left\{ yf(s_{t}, k) + \varepsilon_{jkt} - w + \delta_{jk} \right\} \int \delta_{jk}(\kappa_{t}) \pi_{\ell}(\varepsilon_{t+1} | s_{t}, k) dG_{\ell} \right\}.
$$

where the left side of (7) is firm $f$’s value if firm $j$ employs the worker and the right side is firm $f$’s value if firm $f$ employs the worker in $t$. In shorthand notation, (7) can be expressed as $\Pi^{f}(s_{t}, \varepsilon_{t} | j) = \Pi^{f}(s_{t}, \varepsilon_{t} | f)$. This requirement rules out uninteresting rejected offers by the losing firm in which the left side of (7) is strictly higher than the right side of (7); that is, the losing firm strictly prefers losing to winning the worker. (Offers such that the right side of (7) is strictly higher than the left side are already ruled out by equilibrium, since the losing firm must obtain a higher value of profits by not employing the worker than by employing him.) By pinning down the losing firm’s offer, this requirement implies that paid wages and so equilibrium are unique. This notion generalizes the cautious Markov perfect equilibrium introduced by Bergemann and Välimäki (1996) to a setting in which firms compete not just in wages but also in jobs.

### 3.1 Employment and Job Assignment

With ability correlated across jobs and firms, each firm’s job offer policy and the worker’s acceptance (labor supply) policy solve a \textit{dependent} bandit problem. In particular, firm and worker policies do not have a simple index form as in independent bandit problems. Here I show that equilibrium employment and job assignment can be nonetheless characterized as the solution to two pseudo-planning problems, which are the natural analogues of the stopping problem that leads to an optimal (Gittins) index policy in independent bandit problems. Denote then by $W(s_{t}, \varepsilon_{t})$ the worker’s \textit{value of wages}, which is the worker’s value function expressed using the fact that wage and job offers depend on $(s_{t}, \varepsilon_{t})$ in equilibrium. Denote by $V^{f}(s_{t}, \varepsilon_{t})$ the maximal value of firm $f$’s match surplus at state $(s_{t}, \varepsilon_{t})$ or \textit{match surplus value}, which is the sum of the worker’s value of wages and firm $f$’s value function $\Pi^{f}(s_{t}, \varepsilon_{t})$, referred to as firm $f$’s \textit{value of profits} for simplicity. Denote by $W(s_{t}, \varepsilon_{t} | f')$, $\Pi^{f}(s_{t}, \varepsilon_{t} | f')$, and $V^{f}(s_{t}, \varepsilon_{t} | f')$, respectively, the worker’s value of wages, firm $f$’s value of profits, and firm $f$’s match surplus value conditional on firm $f' = A, C$ employing the worker.

I establish two properties of equilibrium. First, given each firm’s choice of job, the employing firm in equilibrium is the one that generates the largest sum of values to all, namely, $S(s_{t}, \varepsilon_{t}) = \Pi^{A}(s_{t}, \varepsilon_{t}) + W(s_{t}, \varepsilon_{t}) + \Pi^{C}(s_{t}, \varepsilon_{t})$. Equivalently, the employing firm is determined by a planning problem in which the planner’s choice of job is restricted to only the two
jobs that maximize each firm’s value of profits at each state. This first result implies that conditional on a firm’s choice of job, a worker’s choice of firm is efficient. Second, each firm’s choice of job in equilibrium maximizes the value of its own output defined in Proposition 1, $V_f(s_t, \varepsilon_t)$. This second result is an implication of the competition between firms and (7), which induce each firm to correctly internalize the worker’s preference over possible job assignments and so to maximize the value of its output regardless of whether the firm employs the worker or not. Note that these two partial efficiency properties do not imply that the equilibrium is efficient, as neither firm internalizes the impact of its choices on its competitor’s future probability of employing the worker and thus on its competitor’s profits. Yet, by these two results, equilibrium employment and job assignment can be conveniently characterized separately from wages.

**Proposition 1.** Let $f = A, C$. The employing firm in equilibrium is determined by the policy that solves

$$S(s_t, \varepsilon_t) = \max_f \left\{ y_f(s_t, k) + \varepsilon_f k + \delta \eta_f(k) \int_{\varepsilon_{t+1}} ES(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\},$$

where $k = k_f(s_t, \varepsilon_t)$ solves $V_f(s_t, \varepsilon_t) = \max_{k \in K_f} \left\{ y_f(s_t, k) + \varepsilon_f k + \delta \eta_f(k) \int_{\varepsilon_{t+1}} E V_f(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\}$.

### 3.2 Wages

The wage equation implied by the model generalizes the one familiar from static Bertrand competition. In a static Bertrand model, the wage paid by the employing firm is sufficiently high that the competitor cannot match it and obtain positive profits. For example, suppose for simplicity that firms $A$ and $C$ produce $y_{At}$ and $y_{Ct}$ when they employ the worker. 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 his output unless the two firms share the same technology, in which case the perfectly competitive outcome arises with $w_{At} = y_{At} = y_{Ct} = w_{Ct}$.

In the dynamic case considered here, the intuition is similar: the employing firm must pay a wage sufficiently high that the competitor cannot match it and obtain a positive present value of profits—net of the continuation value from not employing the worker—rather than positive static profits as in the static case. In the dynamic case, though, the wage rule is more flexible in that a worker can be paid more or less than his (conditional) expected output. For instance, a promising worker is paid more than his conditional expected output by, say, firm $A$, whenever: $i$) firm $A$ can learn more about the worker by employing him at a job at which the worker is not very productive rather than by observing his performance at firm $C$; and $ii$) this information is valuable. That is, if firm $A$ can improve its future employment and assignment decisions based on the information acquired by employing the worker, then firm $A$’s value of future profits can more than compensate for any static profit loss and justify paying such a high wage. In general, the closer are firms’ technologies, current productivity shocks, and chosen jobs, the closer is the worker’s wage to his conditional expected output. Under the assumptions maintained in estimation that productivity shocks are Gumbel distributed (maximum) and that any job $k$ provides the same opportunity for information acquisition ($\alpha_{Ak} = \alpha_{Ck}$ and $\beta_{Ak} = \beta_{Ck}$) and entails the same risk of exogenous separation ($\eta_{Ak}(\kappa) = \eta_{Ck}(\kappa)$) across firms, paid wages take a simple form, which makes these intuitions transparent. See Section 2 in the S.A. for the general case.
Proposition 2. Let productivity shocks be Gumbel distributed with variance $\pi^2/6$. The wage paid by firm $A$ is

$$w_A(s_t, \varepsilon_t) = y_C(s_t, k_{At}) - \ln Pr(k_{Ct} = k_{At}|f_t = C, s_t) + \varepsilon_{Ct},$$

(9)

if $\alpha_{Ak} = \alpha_{Ck}$, $\beta_{Ak} = \beta_{Ck}$, and $\eta_{Ak}(\kappa_t) = \eta_{Ck}(\kappa_t)$ at each job $k \geq 1$.

By Proposition 2, firm $A$’s wage equals the sum of the worker’s expected output at firm $C$ at the job chosen by $A$, $y_C(s_t, k_{At})$, the log probability that firm $C$ chooses the same job as $A$ (the event $k_{Ct} = k_{At}$) conditional on employing the worker (the event $f_t = C$), and the productivity shock to firm $C’$s match surplus value, $\varepsilon_{Ct}$. Thus, whether a worker is paid by $A$ a wage larger or smaller than his conditional expected output depends on the difference in expected output at the job chosen by $A$ between firms $A$ and $C$, on how likely firm $C$ is to choose the same job as $A$, and on the difference between the productivity shock realized at the job chosen by $A$, $\varepsilon_{Ak_{At}}$, and $\varepsilon_{Ct}$. In particular, if firms $A$ and $C$ had the same technologies and a worker experienced the same productivity shocks at the two firms, then the market would be perfectly competitive and firms would choose the same jobs at all states. In this case, the second term on the right side of (9) would be zero, and a worker’s wage would equal expected output up to productivity shocks as in the static case, $w_A(s_t, \varepsilon_t) = y_A(s_t, k_{At}) + \varepsilon_{Ak_{At}}$, where $y_A(s_t, k_{At}) + \varepsilon_{Ak_{At}}$ plays the role of $y_{At}$ in the above example.

4 Empirical Analysis

Here, I discuss the assumptions maintained in estimation, the identification of the model, and the estimated specification. Since I only observe workers, henceforth, managers, employed at $A$, from now on, I denote by $t = 1$ the year of a manager’s entry in the firm. Thus, $t$ denotes potential tenure in the firm and $h_1 = \ln H_1$ human capital at entry in the firm.

4.1 Preliminaries

I formulate two main assumptions. First, I assume that the discount factor, $\delta$, is known and productivity shocks are Gumbel distributed (maximum) with variance $\pi^2/6$. Second, I assume that firm $C$ has three jobs or levels like firm $A$, my firm, and, because of data limitations, that the probabilities of high performance by manager ability, $\{\alpha_{fjk}, \beta_{fjk}\}$, as well as the probabilities of exogenous separation at each level, $\{\eta_{fkt}\}$ with $\eta_{fkt} = \eta_{fjk}(\kappa_t)$, are common across firms. I refer to these probabilities as $\alpha_k$, $\beta_k$, and $\eta_{kt}$. The identification results below can be extended to the more general case in which the speed of learning and the risk of exogenous separation differ across firms; see Section 2 in the S.A.I discuss next specific assumptions about the distribution of the initial priors, output, performance ratings, and wages.

Initial State. I let managers differ along dimensions that are known to managers and firms, say, from job interviews, but are unobserved to the econometrician. Namely, in the spirit of Heckman (1981), I assume that each manager is of skill type $i = 1, \ldots, I$, which affects the initial prior about a manager’s ability and some of the parameters of the distribution of human capital, output, performance, 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 and $q_i = Pr(i|L_{A1} = 1)$ be

---

17Here, $\varepsilon_{Ct}$ is the random component of the maximum surplus that $C$ can generate by employing the worker in $t$; see the proof of the proposition.
the probability of skill type $i$, where $L_{ft}$ is a manager’s level at firm $f$ in $t$ and corresponds to job $k_f = k_f(s_{it}, \varepsilon_t)$ in the model. A manager’s state at the beginning of $t$ is then $s_{it} = (p_{it}, \kappa_t, i)$, where $\kappa_t = (h_1, \{i^j_{kt}\}_{j=1}^{J} \tau^{-1})$ and $h_1 = (e_1, x_1, y_1)$ is given by education at entry, $e_1$, experience (age) at entry, $x_1$, and year of entry in the firm, $y_1$.

**Expected Output.** To capture that different types of managers may have access to different outside options, I allow firm $A$ to face different competitors for managers of different skill types. Specifically, I allow $a^j_{fkt}$ for firm $C$ to vary with $i$ and denote its two possible values for each $j, k, t$, and $i$ by $\pi^j_{Ckt}(i)$ and $a^j_{Ckt}(i)$. This way firm $A$ faces the same competitor for any given manager, but possibly different competitors for different managers. As for the remaining terms in the human capital function in (1), for simplicity in light of the data, I assume that only the sum of past investments matters for the accumulation of general human capital so $g^g_k(i^g_{kt}, \ldots, i^g_{kt-1}) = g^g_k(t - 1)$. For job-specific human capital, instead, I maintain that only the previous investment matters for its accumulation so $g^k_k(i^g_{kt}, \ldots, i^g_{kt-1}) = g^k_k(i^g_{kt-1})$. See Section 4.3 for a motivation for these forms of limited state dependence and Appendix B for details; the identification results in Section 4.2 can be extended easily to a more general form for $\{g^j_k(\cdot)\}$.\(^{18}\) As for output, I let $b_{fkt}(h_1) = \sum_{j=1}^{J} b_{fj}(k)h_{1j}$, where $\{h_{1j}\}$ are known transformations of the elements of $h_1$; the number of parameters of $b_{fkt}(h_1)$ is the same across firms and job levels just for simplicity. Thus, expected output in (3) becomes

$$y_f(s_{it}, k) = \sum_{j=1}^{J} b_{fj}(k)h_{1j} + d_{fkt}(k_{t-1}, i) + e_{fkt}(i)p_{it}$$

(10)

with $d_{fkt}(k_{t-1}, i)$ and $e_{fkt}(i)$ independent of $i$ for $A$, and the state reduces to $(p_{it}, h_1, t - 1, k_{t-1}, i)$. The parameters $\{b_{fj}(k)\}$, $d_{fkt}(k_{t-1}, i)$, and $e_{fkt}(i)$ in (10) are the human capital and output parameters of interest in the empirical exercise. Their dependence on $t - 1, k_{t-1}$, and $i$ in the estimated specification of the model will be determined by the assumptions on the process of human capital and the output function formulated in Section 4.3.

**Performance Ratings.** It is well known that performance ratings measure performance imperfectly and that the associated error may be systematic. For instance, some supervisors assign uniform ratings to employees regardless of their true performance, which may lead 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)). In my data, ratings at each level are not only skewed toward high performance but also serially dependent and, as mentioned, somewhat declining with tenure. Hence, 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). Formally, denote by $E_0(k, t) = \Pr(R_{At}^0 = 1 | R_{At} = 0, L_{At} = k, t)$ the probability of a recorded high rating, $R_{At}^0 = 1$, when low performance is realized, $R_{At} = 0$, at Level $k$ in tenure $t$: $R_{At}^0$ is the performance rating observed by the econometrician, whereas $R_{At}$ is the true performance observed by firms and managers. Similarly, denote by $E_1(k, t) = \Pr(R_{At}^0 = 0 | R_{At} = 1, L_{At} = k, t)$ the probability of a recorded low rating when high performance is realized at Level $k$ in $t$. I assume that

$$E_0(k, t) = \frac{\exp\{\rho_0 + \rho_2(k)\{t \times 1\{k = 1\} + (t - 1) \times 1\{k = 2\} + (t - 2) \times 1\{k = 3\}\}]}{1 + \exp\{\rho_0 + \rho_2(k)\{t \times 1\{k = 1\} + (t - 1) \times 1\{k = 2\} + (t - 2) \times 1\{k = 3\}\}\}}$$

(11)

\(^{18}\)In Appendix B, I show that these forms for $\{g^j_k(\cdot)\}$ can be derived from laws of motion for human capital that are common and amount to special cases of (1), under the assumption that general capital does not depreciate, whereas job-specific capital fully depreciates after one period.
\[ E_1(k, t) = \frac{1}{1 + \exp\left\{ \rho_0 + \rho_1 + \rho_2(k) [t \times 1\{k = 1\} + (t - 1) \times 1\{k = 2\} + (t - 2) \times 1\{k = 3\}] \right\}}. \]

I allow these error rates to vary across levels, since the evaluation of performance is often job-specific, and with potential tenure in a level, to capture the possibility that performance appraisal may be conducted more thoroughly at certain stages of a manager’s career. Since no manager is observed at Level 2 before \( t = 2 \) or at Level 3 before \( t = 3 \), I let classification error rates at these levels depend on \( t - 1 \) and \( t - 2 \), respectively. I maintain the identifying assumption that the probability of a recorded high rating increases with the probability of high performance, \( \rho_1 > 0 \), which is satisfied by my estimates.

**Wages.** To account for unmeasured aspects of firm and manager behavior as well as recording inaccuracies, I allow wages to be measured with error, which possibly differs across skill types. Let \( \varepsilon_{imt} \) be a zero-mean Gumbel-distributed (minimum) shock with standard deviation \( \sigma_{Aik} \) and define \( \epsilon_{Ait} = \lambda_{ik} \varepsilon_{Clt} - (-\varepsilon_{imt}) \) where \( \lambda_{ik} = \sigma_{Aik} \sqrt{6}/\pi \). Then, the recorded wage of a manager of skill type \( i \) at Level \( L_{Ait} = k \) of firm \( A \) at state \( (s_{it}, \varepsilon_t) \) by (9) is

\[
w_{Ait} = y_C(s_{it}, k) - \ln \Pr(L_{Clt} = k|f_t = C, s_{it}) + \epsilon_{Ait}, \tag{13}\]

where \( \epsilon_{Ait} \) is logistically distributed with standard deviation \( \sigma_{Aik} \), since it is the difference between two Gumbel-distributed (maximum) random variables with standard deviation \( \sigma_{Aik} \). Then, (8)-(9) and (10)-(13) allow me to compute the probability of observing a given sequence of jobs, recorded performance, and wages for a manager conditional on the prior beliefs about the manager’s ability and the manager’s human capital, which are also jointly estimated.

### 4.2 Identification

I first provide an overview of identification and then the formal argument. The parameters to identify are:

**i)** the parameters of classification error, \( (\rho_0, \rho_1, \{\rho_2(k)\}) \), and of the **learning process**, which is completely described by the distribution of performance, \( \{\alpha_k, \beta_k\} \), and of the prior priors, \( \{p_{i1}, q_i\} \), by (4);

**ii)** the parameters of **expected output** at \( A \), \( \{b_{Aij}(k), d_{Akt}(k_{t-1}), e_{Akt}\} \) by (10), and of **exogenous separation**, \( \{\eta_{kt}\};^{19}\)

**iii)** the parameters of **wages at** \( A \), namely, those of (13) including the standard deviations \( \{\sigma_{Aik}\}\).

**Overview.** A challenge is to identify a partial equilibrium model of the labor market from data on one firm, firm \( A \). I can do so from repeated information on wages, job assignments, and performance ratings because of the assumed symmetry in the opportunities for information acquisition across firms and of the type of competition that I consider. This symmetry implies that the law of motion of beliefs about managers’ ability at each job level, and so firms’ expectations about ability, can be recovered just from information on performance ratings at firm \( A \). As for competition, in the Bertrand setting I focus on, a manager’s wage at firm \( A \) in equilibrium only depends on a manager’s prior, human capital, and skill type and on the employment offer of firm \( A \)’s competitor, firm \( C \).\(^{20}\) As (13) shows, \( C \)’s offer is summarized in the wage

\(^{19}\)As apparent from (3) or (10), the human capital parameters \( a_{Aik}^{ij} \) and those of \( g_{ik}^{ij} (\cdot) \) can be recovered from expected output at \( A \) only up to \( d_{Akt}(k_{t-1}) \) and \( e_{Akt} \) without further assumptions. Yet, the tenure profile of \( \{d_{Akt}(k_{t-1})\} \) and \( \{e_{Akt}\} \), the correlation in these profiles across levels, and their variation with previous assignments are informative about the process of human capital acquisition at the firm. See Section 4.3.

\(^{20}\)The employing firm in my model faces the same competitor for a given manager in each period. Thus, unlike in a random search model in which a firm faces multiple potential competitors and so wages depend on the distribution of firm productivity in the market, wages at firm \( A \) depend on the productivity, and so the wage offer, of just one firm, firm \( C \).
equation by its expected output and the probability that $C$ offers the same level as $A$. Although I do not observe firm $C$’s assignment decisions in my data, by exploiting the equilibrium relationship between employment at firms $A$ and $C$ implied by Proposition 1, I can express the probabilities of assignment of any manager skill type to firm $C$’s levels in terms of the assignment probabilities at $A$ and differences in expected output between $A$ and $C$. Since, as I will show, assignment probabilities at $A$ by skill type are identified from observed assignments and ratings, the parameters of these expected output differences and $\{\sigma_{Ai}k\}$ are the only ones under $iii$) that need to be identified and can easily be recovered.

Another important feature of the model is that the distribution of wages at each level of firm $A$, conditional on managers’ histories of observed assignments and performance ratings at the firm, is a finite mixture of logistic distributions by (13): its components are the distribution of wages of managers of each skill type. Finite mixtures of logistic distributions are known to be identified. Then, from the identified weights of these mixtures, I can recover both the probabilities of each skill type and of each skill type’s histories of level assignments and performance ratings at the firm. Combined with repeated information on ratings, these probabilities will prove sufficient to identify the parameters under $i$).  

Lastly, from the identified probabilities of each type’s histories of assignments and performance ratings at the firm, conditional assignment probability can also be identified. From these probabilities and the identified average wages of each skill type, in turn, I can infer the parameters of exogenous separations and expected output at each level of $A$ relative to the expected output at one level of $C$ at one state in each tenure.  

The identification argument consists of four steps. In Step 1, I show that the parameters of the distribution of classification error, $\{\rho_0, \rho_1, \{\rho_2(k)\}\}$, and performance, $\{\alpha_k, \beta_k\}$, are identified. In Step 2, I show that the initial priors, $\{p_{i1}\}$, are identified. Since the probability masses of the initial priors, $\{q_i\}$, are identified by Lemma 1, as I discuss next, these two steps identify the learning process. In Step 3, I show that the parameters of expected output at firm $A$ and exogenous separation are identified. In Step 4, I show that the parameters of wages at $A$ are identified.

**Formal Argument.** A key preliminary result concerns the identification of the distributions of wages of managers of each skill type conditional on their histories at firm $A$. For instance, note that by (13) the density of wages of managers with initial human capital $h_1$ at level $L_{A1}$ in the first year at the firm is a finite mixture of logistic densities,

$$f(w_{A1}|L_{A1}=1, h_1) = \sum_i \Pr(i|L_{A1}=1, h_1)f(w_{A1}|L_{A1}=1, h_1, i),$$

with typical weight $\Pr(i|L_{A1}=1, h_1) = q_i$ and density $f(w_{A1}|L_{A1}=1, h_1, i)$ with standard deviation $\sigma_{Ai1}$. Since finite mixtures of logistic distributions are identified, $\{q_i\}$ and $\{f(w_{A1}|L_{A1}=1, h_1, i)\}$ are identified (see Theorem 1 in Shi et al. (2014)). From the second year on, wages depend on the updated beliefs about managers’ ability based on their performance on the job, which is unobserved to the econometrician. Thus, the density of wages of managers with initial human capital $h_1$ at level $L_{A2}$ in the second year, given their first year rating $R^o_{A1}$, is a logistic mixture of the conditional densities of wages of each skill type $i$ with each possible first period performance $R_{A1}$, namely,  

---

21 A similar argument would apply to any semiparametric finite mixture distribution of wages as long as it is identifiable. See, for instance, Hunter et al. (2007) for the identification of semiparametric finite mixtures with symmetric components.

22 With information on expected output at $C$ provided by wages, no normalization would be needed if separation rates at one level were known.
condition

Over time, managers also differ unobservably in their performance. In Proposition 5, I show that ordering $f$ can be recovered from either $R$ weights each weight by skill type, $L$ at firm $A$ dence on $R$. Thus, up to labeling, the component weights, means, and variances of the mixture distributions in the lemma can be

Lemma 1. Let $H_t = (L_A = 1, \ldots, L_A = t)$, $H_{Rt} = (L_A = 1, R^o_{A1}, \ldots, L_A = t, R^o_{A(t-1)})$, and order $\{\sigma_{Aik}\}$ by $i$ at each level. The mixture distributions of wages conditional on $H_t$, $f(w_A|H_t, L_A, h_1) = \sum_{R_{A1}, \ldots, R_{A(t-1)}} \Pr(R_{A1}, \ldots, R_{A(t-1), i}|H_t, L_A, h_1) f(w_A|H_t, L_A, R_{A1}, \ldots, R_{A(t-1), h_1, i})$, and the mixture distributions of wages conditional on $H_{Rt}$,

$$f(w_A|H_{Rt}, L_A, h_1) = \sum_{R_{A1}, \ldots, R_{A(t-1), i}} \Pr(R_{A1}, \ldots, R_{A(t-1), i}|H_{Rt}, L_A, h_1) f(w_A|H_{Rt}, L_A, R_{A1}, \ldots, R_{A(t-1), h_1, i})$$

at firm $A$ are identified at each level and tenure up to their labeling with respect to $R_{A1}, \ldots, R_{A(t-1)}$.

Thus, up to labeling, the component weights, means, and variances of the mixture distributions in the lemma can be recovered for each skill type and possible history at the firm.

Step 1: Distribution of Classification Error and Performance. As discussed before Lemma 1, omitting the dependence on $L_A$ and $h_1$ for simplicity, the probability $\Pr(R^o_{A1}, R_{A1}, L_{A2}, i)$ is identified from the density of wages at Level $L_{A2}$ in the second year at the firm conditional on $R^o_{A1}, f(w_A|R^o_{A1}, L_{A2})$. By the lemma, $f(w_A|L_{A2})$ is also an identified mixture density with weights $\{\Pr(R_{A1, i}|L_{A2})\}$: multiplying each such weight by the corresponding known probability $\Pr(L_{A2})$ yields that $\Pr(R_{A1, L_{A2}, i})$ is identified. Then, the ratio of $\Pr(R_{A1, L_{A2}, i})$ to $\Pr(R^o_{A1}, R_{A1, L_{A2}, i})$ pins down $\Pr(R^o_{A1}|R_{A1, i})$, which is independent of $i$. 24 By a similar logic, $\Pr(R^o_{A2}|R_{A2})$ is identified as well. From these probabilities, the error parameters $(\rho_0, \rho_1, \rho_2(1))$ are identified by (11)-(12). An analogous argument holds for $\{\rho_2(k)\}_{k \geq 2}$.

As for performance, note that the distribution of ratings of managers continually assigned to Level $k$ is a binomial mixture of the distribution of performance of managers of high and low ability with parameters $(\alpha_k, \beta_k)$, subject to classification error. Once classification error is identified, this mixture is identified from three periods of observations on ratings at Level $k$. To see how, consider managers at Level 1 up to $t$. Their probability of high performance in $t$ is

$$\Pr(R_{At} = 1|L_A = \ldots = L_{At} = h_1) = \left[ \sum_i q_{it}(h_1)p_{i1} \right] \alpha_1 + \left[ 1 - \sum_i q_{it}(h_1)p_{i1} \right] \beta_1$$

with $q_{it}(h_1) = \Pr(i|L_A = \ldots = L_{At} = h_1)$, whereas their probability of a high rating in $t$ is

$$\Pr(R^o_{At} = 1|L_A = \ldots = L_{At} = h_1) = E_0(1, t) + [1 - E(1, t) - E_1(1, t)] \Pr(R_{At} = 1|L_A = \ldots = L_{At} = h_1).$$

24Since interchanging the labels of the component densities produces the same mixture, I maintain that $\{\sigma_{Aik}\}$ are ordered by $i$ at each level. Over time, managers also differ unobservably in their performance. In Proposition 5, I show that ordering $\{\sigma_{Aik}\}$ by $i$ at each level and the condition $\rho_1 > 0$ for the identification of classification error in ratings are sufficient to resolve any label ambiguity with respect to performance.

24Note that no label ambiguity arises when this ratio is computed. Since $f(w_A|R^o_{A1}, R_{A1, L_{A2}, i})$ is independent of $R^o_{A1, \{f(w_A|R_{A1, L_{A2}, i})\}}$ can be recovered from either $f(w_A|R_{A1, L_{A2}})$ or $f(w_A|L_{A2})$, since these two distributions are finite mixtures with identical components. So, each weight $\Pr(R_{A1, i}|R^o_{A1, L_{A2}})$ can be paired with the corresponding weight $\Pr(R_{A1, i}|L_{A2})$ through the associated densities.
Suppose that \( t = 3 \). Note that once the parameters of classification error are identified, the classification error rates \( E_0(1,t) \) and \( E_1(1,t) \) are identified. Hence, the probability of high performance, \( \Pr(R_{A3} = 1|L_{A1} = L_{A2} = L_{A3} = 1, h_1) \), is identified from that of a high rating, \( \Pr(R^0_{A3} = 1|L_{A1} = L_{A2} = L_{A3} = 1, h_1) \), by (16). By a similar argument, the probability of high performance in \( t \leq 3 \) consecutive periods at Level 1 can be shown to satisfy

\[
\Pr(R_{A1} = \ldots = R_{At} = 1|L_{A1} = L_{A2} = L_{A3} = 1, h_1) = \left[ \sum_i q_3(h_1) p_{i1} \right] \alpha_1^t + \left[ 1 - \sum_i q_3(h_1) p_{i1} \right] \beta_1^t
\]

(17) and is identified from the probability of a high rating in \( t \leq 3 \) consecutive periods, \( \Pr(R^0_{A1} = \ldots = R^0_{At} = 1|L_{A1} = L_{A2} = L_{A3} = 1, h_1) \). Condition (17) evaluated in \( t \leq 3 \) yields a system of three equations in three unknowns, \( \alpha_1, \beta_1 \), and the average prior \( \sum_i q_3(h_1) p_{i1} \), with a unique solution if \( \alpha_1 > \beta_1 \). Similarly, the performance ratings of managers promoted to Levels 2 and 3 and assigned to these levels for two more years identify \( \{\alpha_k, \beta_k\} \) for \( k = 2, 3 \).

**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 firm A in \( t = 1, 2 \) with either a high or low performance rating in \( t = 1, 2 \). The parameter \( \rho_2(2) \) is identified from the distribution of wages in \( t = 3 \) of managers at Level 2 in \( t = 2 \) with either a high or low performance rating in \( t = 2 \), and \( \rho_2(3) \) is identified from the distribution of wages in \( t = 4 \) of managers at Level 3 in \( t = 3 \) with either a high or low performance rating in \( t = 3 \). The probabilities of high performance \( \{\alpha_k, \beta_k\} \) at Level \( k \geq 1 \) are identified from three years of repeated observations on the distribution of ratings at the corresponding level.

It is important that the distribution of wages be an identified mixture of the distribution of wages of managers of each skill type and that the distribution of performance be an identified mixture of the distribution of performance of managers of each ability. I specify these distributions as finite mixtures of logistic and binomial distributions, respectively. However, these finite mixtures would be identified under many other parametric and semiparametric assumptions.

**Step 2: Initial Priors.** The four initial priors \( \{p_{i1}\} \) are identified as follows. Recall that \( \alpha_1, \beta_1 \), and the average prior \( \sum_i q_3(h_1) p_{i1} \) are identified from the first three years of observations on ratings at Level 1 by Step 1. From (15) and (16), it follows that \( \sum_i q_{it}(h_1) p_{i1} \) is also identified in \( t \geq 4 \), once \( \alpha_1, \beta_1 \), and classification error in performance at Level 1 are identified. Then, the identified average priors \( \sum_i q_{it}(h_1) p_{i1} \) between \( t = 3 \) and \( t = 6 \) provide a system of four linearly independent equations, \( q_{1t}(h_1) p_{i1} + \ldots + q_{4t}(h_1) p_{i4} = \sum_i q_{it}(h_1) p_{i1} \), for each such tenure, in the four unknowns \( \{p_{i1}\} \), where \( q_{it}(h_1) = \Pr(i|L_{A1} = \ldots = L_{At} = 1, h_1) \) is a known coefficient by Lemma 1. Thus, \( \{p_{i1}\} \) are identified.

**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 ratings at Level 1 of A between \( t = 3 \) and \( t = 6 \).

**Step 3: Expected Output at My Firm and Exogenous Separations.** These parameters are identified under two further assumptions. Because the sample only covers the first eight years of managers at firm A, the first assumption, (A1),

---


26Similar information at Levels 2 or 3 would also identify \( \{p_{i1}\} \). Once \( \{p_{i1}\} \) are identified, the proof of Proposition 4 can be adapted to obtain an alternative argument to identify \( \{\alpha_k, \beta_k\}_{k \geq 2} \); see Observations 1 and 2 there. A different argument to identify three of the four priors \( \{p_{i1}\} \) builds on Hu and Shum (2012) and can also be used to identify any three objects out of one \( \{\alpha_1\} \), one \( \{\beta_1\} \), and \( \{p_{i1}\} \), if the remaining probabilities of high performance and priors as well as the probability masses of skill types conditional on past assignments are identified; see the S.A.
imposes that exogenous separation rates are constant from $t = 7$ on. (A1) also introduces four normalizations on the intercept parameters of expected output at $A$ in (10) that cannot be identified because neither multi-level promotions (the first normalization) nor demotions (the last three normalizations) are ever observed. Then, the parameters left to identify in addition to $\{b_{Aj}(k)\}$ are: $\eta_{it}$ in $t \geq 1$, $(d_{A1t}(L1), e_{A1t}, d_{A2t}(L1), e_{A2t}, \eta_{22t})$ in $t \geq 2$, $(d_{A2t}(L2), d_{A3t}(L2), e_{A3t}, \eta_{33t})$ in $t \geq 3$, and $d_{A3t}(L3)$ in $t \geq 4$, where $Lk$ stands for $k_{t-1} = k$. (A1) further imposes a support and rank condition on

$$y_A(s_{it}, k) = \sum_{j=1}^{J} b_{Aj}(k)h_{1j} + d_{Akt}(k_{t-1}) + e_{Akt}p_{it}, \quad (18)$$

which provides a system of linear equations in $(\{b_{Aj}(k)\}, d_{Akt}(k_{t-1}), e_{Akt})$ if $\{p_{it}\}$ are known, which I use to recover the $J$ parameters $\{b_{Aj}(k)\}$ for each $k$, $d_{Akt}(k_{t-1})$ for each possible $k$, $t$, and $k_{t-1}$, and $e_{Akt}$ for each possible $k$ and $t$.

Assumption 1 (A1). Assume that $\eta_{kt} = \eta_k, t \geq 7$, that $d_{A3t}(L1)$, $d_{A1t}(L2)$, $d_{A2t}(L3)$, and $d_{A1t}(L3)$ are known, and that there exist $J$ equations for each $k$, one equation for each possible $k$, $t$, and $k_{t-1}$, and one more equation for each possible $k$ and $t$ defined by (18) that are linearly independent.

Since managers separating from firm $A$ are never reemployed in the data, I assume $A$’s profits are zero once it loses a manager, which is consistent with firm $A$ incurring a cost to rehiring managers, $c_r(\varepsilon_{At})$, that possibly varies with realized productivity. I maintain that $A$ incurs such a cost in assumption (A2), which captures any capacity (“slot”) constraints that may further limit $A$’s employment and assignment decisions. The implications of (A2) for the value $S(s_{it}, \epsilon_t)$ in (8) are derived in the proof of the next lemma. In what follows, the dependence of the state on the identity of the “incumbent” firm, that is, the firm employing a manager in the previous period, is suppressed for simplicity.

Assumption 2 (A2). Firm $A$ incurs a cost $c_r(\varepsilon_{At}) \geq 0, \varepsilon_{At} = \{\varepsilon_{Akt}\}$, to rehiring separated managers in any period.

Lemma 2. There exist $\zeta_r(\varepsilon_{At})$ in each $t$ such that a manager separated from $A$ is never rehired if $c_r(\varepsilon_{At}) \geq \zeta_r(\varepsilon_{At})$.

I first establish that assignment probabilities at $A$ are identified. To see how, omit their dependence on $L_{A1}$ and $h_1$ for simplicity, and consider $\Pr(L_{A2}|R_{A1}, i)$. Recall that the probability $\Pr(R_{A1}, L_{A2}, i)$ is identified by Step 1 from the mixture wage density $f(w_{A2}|L_{A2})$, that $\{\alpha_k, \beta_k, p_{i1}\}$ are identified by Steps 1 and 2 so that $\Pr(R_{A1} = 1|i) = \alpha_k p_{i1} + \beta_k (1 - p_{i1})$ and the complementary probability are identified, and that $\Pr(i) = q_i$ is identified by Lemma 1. Then, $\Pr(L_{A2}|R_{A1}, i)$, which can be expressed as $\Pr(L_{A2} = k|s_{it}, k \geq 1, i)$ is identified. This argument can be extended to any $t$.

Once assignment probabilities are identified, the parameters of interest are identified as follows. For simplicity, ignore exogenous separations. As shown in Lemma 2, since separations from $A$ are permanent, the equilibrium assignment problem characterized in Proposition 1 reduces to a simpler dynamic discrete choice problem, $S(s_{it}, \epsilon_t) = \max_{f, k} \{\zeta^f(s_{it}, k) + \epsilon_{ft}\}$ with $\zeta^f(s_{it}, k) = y_f(s_{it}, k) + \delta E[\ln(\sum_k e^{A(s_{it}, k)} + \sum_k e^{C(s_{it}, k)})|s_{it}, k]$. The log difference between $\Pr(L_{Ct} = k|s_{it})$ and $\Pr(L_{At} = k|s_{it})$ equals the difference $\zeta^C(s_{it}, k) - \zeta^A(s_{it}, k)$, which, since the law of motion of the state at each job level is common across firms, in turn reduces to the difference $y_C(s_{it}, k) - y_A(s_{it}, k)$. Then, using the relationship between the probabilities $\Pr(L_{Ct} = k|s_{it})$ and $\Pr(L_{Ct} = k|f_t = C, s_{it})$, that the probability

---

27By the sample structure, the required number of equations is smaller than that of all combinations of $k$, $t$, and $k_{t-1}$. For example, recovering $d_{Akt}(k_{t-1})$ requires one equation with $k_{t-1} = 1$ for $k = 1, 2, t \geq 2$, one with $k_{t-1} = 2$ for $k = 2, 3, t \geq 3$, and one with $k_{t-1} = 3$ for $k = 3, t \geq 4$. 

of separation from \( A \) satisfies \( \Pr(L_{At} = 0|s_{it}) = \Pr(L_{Ct} > 0|s_{it}) \) without exogenous separations, and that average wages by skill type at each level are given by \( Ew_{Ait} = y_C(s_{it}, k) - \ln \Pr(L_{Ct} = k|f_t = C, s_{it}) \) by (13), it is easy to show that

\[
y_A(s_{it}, k) = Ew_{Ait} - \ln[1 - \Pr(L_{At} > 0|s_{it})] + \ln \Pr(L_{At} = k|s_{it}).
\]  

(19)

Since \( Ew_{Ait} \) is identified by Lemma 1 and \( \Pr(L_{At} = k|s_{it}) \), \( k \geq 1 \), and so \( \Pr(L_{At} > 0|s_{it}) = \sum_k \Pr(L_{At} = k|s_{it}) \) are identified as argued, \( y_A(s_{it}, k) \) is identified by (19) and so are \( \{b_{Aj}(k)\} \), \( d_{Akt}(k_t - 1) \), and \( e_{Akt} \) from it by (18) and (A1).

A similar argument applies with exogenous separations under a location normalization in each identified as argued,

Proposition 5. Assume (A1)-(A2), that \( y_C(s_{it}, 1) \) is known at one \( s_{it} \) with \( k_{t-1} = 1 \) in \( t \geq 2 \), and that \( y_A(s_{it}, 1) = y_C(s_{it}, 1) \) at one \( s_{it} \) with \( k_{t-1} = 1 \) in \( t \geq 2 \). There exists \( c_r(\varepsilon_{At}) \) such that if \( c_r(\varepsilon_{At}) \geq c_r(\varepsilon_{At}) \), then \( \{b_{Aj}(k)\} \), \( \{\eta_{it}\}_{it=1} \), \( \{d_{A1}(L1), e_{A1}, d_{A2}(L1), e_{A2}\}_{t=2} \), \( \{\eta_{it}\}_{it=2} \), \( \{d_{A2}(L2), d_{A3}(L2), e_{A3}\}_{t=3} \), \( \{\eta_{it}\}_{it=3} \), and \( \{d_{A3}(L3)\}_{t=4} \) are identified from observed assignments and wages at firm \( A \) provided that \( \{p_{it}\} \) and \( \{\alpha_k, \beta_k\} \) are identified.

In the proof of this result, I show that \( \rho_1 > 0 \) resolves the label indeterminacy in Lemma 1, since it implies that the weights of the wage mixtures in the lemma can be ordered by \( R_{A1}, \ldots, R_{At-1} \) conditional on assignments and ratings.

Step 4: Wages at My Firm. Since mean wages, \( \{Ew_{Ait}\} \), and standard deviations, \( \{\sigma_{Ait}\} \), from (13) are identified for each manager skill type and job level by Lemma 1, the parameters of mean wages are the only ones left to recover and are identified as follows. Denote the difference in expected output at Level \( k \) between firms \( A \) and \( C \) by

\[
\gamma(s_{it}, k) = y_C(s_{it}, k) - y_A(s_{it}, k).
\]  

(20)

It is easy to show that the conditional probability of assignment to Level \( k \) at firm \( C \) in (13) can be expressed in terms of \( \gamma(s_{it}, 1), \gamma(s_{it}, k), \) the analogous conditional probability of assignment to Level \( k \) at \( A \), and the ratio of the conditional probability of assignment to Level 1 at \( C \) to that at \( A \). An implication of this fact is that I can express \( Ew_{Ait} \) as

\[
Ew_{Ait} = y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) + \ln \left[ \frac{\Pr(L_{At} = 1|f_t = A, s_{it})}{\Pr(L_{Ct} = 1|f_t = C, s_{it}) \Pr(L_{At} = k|f_t = A, s_{it})} \right],
\]  

(21)

which is the sum of a term that depends on firms’ expected output, \( y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) \), and conditional assignment probabilities at firm \( A \), identified by Proposition 5, and to Level 1 of firm \( C \). The conditional probability of assignment to Level 1 of firm \( C \) can be shown to be identified from the corresponding probability at \( A \) and \( \{\eta_{kt}\} \) up to the difference in expected output at Level 1 between the two firms. Up to this difference, all probabilities on the right side of (21) are then identified and so is the term \( y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) \), which is a linear function of the prior by (10) and (20) that I denote by \( \sum_j \varphi_j(k) + \sum_i \varphi_{k_{t-1}}(k_{t-1}, i) + \varphi_{k_{t-2}}(i) p_{it} \). Its parameters, \( \{\varphi_j(k)\}, \varphi_{k_{t-1}}(k_{t-1}, i), \) and \( \varphi_{k_{t-2}}(i) \), are simple transformations of the firms’ expected output parameters at Levels 1 and \( k \) and can be identified from (21).

Proposition 6. If there exist \( J \) equations for each \( k \), one equation for each possible \( k, t, k_{t-1}, \) and \( i \), and one more equation for each possible \( k, t, \) and \( i \) defined by (21) that are linearly independent, the difference \( y_C(s_{it}, 1) - y_A(s_{it}, 1) \)

28Namely, \( \Pr(L_{Ct} = 1|f_t = C, s_{it}) = e^{\gamma(s_{it}, 1)} \Pr(L_{At} > 0|s_{it}) \Pr(L_{At} = 1|f_t = A, s_{it})/[\eta_{t-1} - \Pr(L_{At} > 0|s_{it})] \) with \( \eta_{t-1} \) identified as argued. In the special case in which \( \gamma(s_{it}, k) \) is independent of \( k \), \( \Pr(L_{At} = k|f_t = f, s_{it}) \) is equal across firms so \( y_C(s_{it}, k) \) is identified.
is known at the corresponding states, and \{\eta_{kt}\}, \{p_{i1}\}, and \{\alpha_k, \beta_k\} are identified, then the parameters of (21) are identified from observed assignments and wages at firm A.

Note that if expected output at Level 1 of C is normalized at the required states, then the difference in expected output at this level between the two firms is known—a condition of Proposition 6—once the expected output parameters at A and separation rates are identified by Proposition 5. I do so in estimation to satisfy the conditions of Propositions 5 and 6.

4.3 Estimated Specification

Here I present the specification of human capital, output, and wages. See Appendix B for details.

Expected Output: Human Capital and Output. Consider first output. Let \{y_1 = m\} be a dummy for a manager’s year of entry in firm A, \(y_1\). I specify \(b_{f_k}(h_1)\) in (2) as common across firms and given by \(\sum_m b_{yn}1\{y_1 = m\} + b^*_k e_1 + b^*_k x_1 + b^*_k x_1^2\), where \(x_1\) captures general human capital acquired before entry in the firm. The effect of job-specific human capital acquired before entry is subsumed in \(by_m\), under the alternative assumptions that either all managers perform the same job before employment at A or that this job-specific capital is not transferable to the market considered. This specification of \(b_{f_k}(h_1)\) is consistent with the sensitivity of wages at any job level of the firm to managers’ experience at entry, given tenure, previous assignments, and performance, as discussed by BGH. I specify the fixed cost of employment in (2) as \(\phi_{f_k} + \phi_f(k_{l-1})\). Consider now human capital. By specifying \(g^j_k(\cdot)\) in (1) as \((t - 1)\eta_g\) for general capital \((j = g)\) and \(\eta_{sht-1}\) for job-specific capital \((j = g)\) independent of \(k\), the human capital functions in (1) become

\[
\begin{align*}
\eta_{fkt}^g &= a^g_{jkt}(i) + (t - 1)\eta_g + \epsilon^g_{fkt} \quad \text{and} \quad \eta_{sht}^g = a^s_{jkt}(i) + \eta_{sht-1} + \epsilon^s_{fkt}.
\end{align*}
\]

(22)

As shown in Appendix B, these forms for \(\{g^j_k(\cdot)\}\) can be derived from standard laws of motion of human capital after they are recursively solved to be expressed in cumulative form as in (1), under the assumption that general human capital does not depreciate, whereas job-specific human capital fully depreciates after one period; see also footnote 13. Whereas the process for \(\eta_{fkt}^g\) is a familiar form of learning-by-doing, the one for \(\eta_{sht}^g\) is motivated by parsimony and consistency with the data. Assuming that job-specific human capital amounts to experience at the job level in the previous period allows me to incorporate parsimoniously yet flexibly the benefit of training that may occur at previous or current levels \((c^g_{f_k} + c^s_{f_k} \eta_{sht-1} > 0 \quad \text{for} \quad k \geq k_{l-1}\) as well as the potential loss of job-specific capital upon promotion \((c^g_{f_k} + c^s_{f_k} \eta_{sht-1} = 0 \quad \text{for} \quad k > k_{l-1}\)\). This formulation is consistent with evidence that training often occurs right before a promotion or is effectively a prerequisite for it; see Cobb-Clark and Dunlop (1999). Baker et al. (1994b) argue that training takes place at the firm, since managers are primarily rewarded for good performance with salary increases rather than bonuses, which supports the idea that these raises reward permanent rather than transitory increases in productivity. For simplicity and due to data limitations, I let \(c^g_{f_k} \eta_{sht-1}\) and \(\phi_f(k_{l-1})\) for firm C differ only up to a constant.

With \(\alpha^j_{fkt}(i) = a^j_{f_k}(i) + \alpha^j_{fkt}\), and \(\alpha^j_{fkt}(i)\) independent of \(i\) for A as discussed after (10), expected output becomes

\[
y_f(s_{it}, k) = b_{f_k}(h_1) + C^g_{f_k} + \beta_k [c^g_{f_k} a^g_{fkt}(i) + c^s_{f_k} a^s_{fkt}(i)] + \epsilon^g_{fkt} \eta_g \times (t - 1) + c^s_{f_k} \eta_{sht-1} - \phi_{f_k} - \phi_f(k_{l-1})
\]

\[
d_{fkt}(k_{l-1}, i)
\]
by using (2), (10), (22), the form of $c^j_{fkt}(i)$, and by taking expectations of $a^j_{fkt}(i)$, which equals $\pi^j_{fkt}(i)$ with probability $\alpha_k$ or $\beta_k$ depending on a manager’s ability and $d^j_{fkt}(i)$, which I set to zero, otherwise. In light of Propositions 5 and 6, I consider Level 2 of firm $C$ as the reference level, as it is the only level into which and out of which managers are promoted, and set its parameters to zero. I also conserve on parameters by imposing a form of symmetry in that $d_{Akt}(Lk) = -d_{Akt}(Lk')$, $k' > k$, where $d_{Akt}(Lk')$ can be interpreted as a demotion cost and is normalized by (A1), and by letting $d_{A2t}(L2) = d_{C2t}(L2)$. By a strategy analogous to that in Cameron and Heckman (2001), for transitions with relatively few observations, I did not estimate any of the associated parameters and set them equal to their values at the reference level. For transitions with a similar number of observations, I maintained that their parameters are equal.

This formulation captures different degrees of transferability of human capital across levels. I define human capital acquired with experience at firm $f$ to be task-general if the tenure profiles of the parameters $\{d_{fkt}(k_{t-1}, i), c_{fkt}\}$ at different levels are positively correlated; see Gibbons and Waldman (2006) for a similar definition. This notion of task generality captures the idea that a manager may become more productive at all levels with experience at a firm. I also define human capital acquired with experience at Level $k$ of firm $f$ to be task-general if it is at least as productive at Level $k' \neq k$ as the human capital acquired with experience at Level $k'$ in that $d_{fkt}(Lk, i) \geq d_{fkt}(Lk', i)$, and task-specific otherwise. This notion of task generality captures the idea that experience at a given level may increase productivity at other levels as much as it does experience at these other levels, if not more. An instance of task specificity is when the human capital acquired, say, at Level 3 of $A$ primarily affects output at this level, so $d_{A3t}(L3) > d_{A3t}(Lk)$, $k = 1, 2$.

Wages. Recall that the average wage of managers of skill type $i$ at Level $k$ of firm $A$ at state $s_{it}$ can be expressed as in (21). An analogous argument applies when Level 2, instead of Level 1, of firm $C$ is the reference level. In this case, conditional assignment probabilities to Level 2, rather than Level 1, at the two firms appear on the right side of (21), and the term $\gamma(s_{it}, 1)$ is replaced by $\gamma(s_{it}, 2)$. By the assumed human capital and output process, it can be shown that the linear function of the prior $y_{Ct}(s_{it}, k) + \gamma(s_{it}, 2) - \gamma(s_{it}, k)$ specializes to $\omega_{0ik} + \omega_{1k}x_{1} + \omega_{2k}x_{1}^2 + \omega_{3k}e_{1} + \sum_{m} \omega_{ym}1\{y_{1} = m\} + \omega_{4kt} \times (t - 1) + (\omega_{5i} + \omega_{5kt})p_{it}$. See Appendix B for the simple algebra of the mapping between the parameters $\Omega = (\omega_{0ik}, \omega_{1k}, \omega_{2k}, \omega_{3k}, \omega_{ym}, \omega_{4kt}, \omega_{5i}, \omega_{5kt})$ and those of human capital and output, which is invariant to the counterfactual exercises below. Then, I can rewrite the average wage in (21) as

$$Ew_{Ait} = \omega_{0ik} + \omega_{1k}x_{1} + \omega_{2k}x_{1}^2 + \omega_{3k}e_{1} + \sum_{m} \omega_{ym}1\{y_{1} = m\} + \omega_{4kt} \times (t - 1) + (\omega_{5i} + \omega_{5kt})p_{it}$$

$$+ \ln \frac{Pr(L_{At} = 2|f_{t} = A, s_{it})}{Pr(L_{Ct} = 2|f_{t} = C, s_{it}) Pr(L_{At} = k|f_{t} = A, s_{it})}.$$  

(24)

All parameters are readily interpretable: $\omega_{1k}$, $\omega_{2k}$, and $\omega_{3k}$ measure the (market) value of human capital at entry in the firm; $\omega_{ym}$ allows for business cycle conditions at the time of hiring, as captured by the year of entry in the firm, to persistently affect the value of a manager to all firms, which BGH find important; and $\omega_{4kt}$ and $\omega_{5i} + \omega_{5kt}$ measure the

---

29$Pr(L_{Ct} = 2|f_{t} = C, s_{it})$ can be computed by a formula analogous to that in footnote 28 and is known up to $y_{C}(s_{it}, 2) - y_{A}(s_{it}, 2)$ and $\{\eta_{kt}\}$. 

---
value of ability and the deterministic component of the human capital acquired at the firm. Since the parameters \( \{ \omega_{5kt} \} \) did not vary much with \( k \) or \( t \) in the range of values relevant in estimation, I estimated only \( \omega_{5t} \) in the baseline version of the model; see Section 5 for the results when \( \omega_{5kt} \) is included. Per Lemma 1, I assume \( \sigma_{A1k} \geq \sigma_{A2k} \geq \sigma_{A3k} \geq \sigma_{A4k} \).

5 Estimation Results

Here, I discuss the estimates of the model parameters, which are all significant at the 1 percent level, obtained by full-information full-solution maximum likelihood with \( \delta = 0.95 \). I also derive the implications of the estimates for wage growth at the firm and show that popular methods in the literature to assess the impact of learning on wages measure only part of its impact and are inapplicable when jobs differ in their informativeness. As Tables 1-4 show, the estimated model successfully captures the patterns of managers’ separations from the firm, their transitions across job levels, which are nonlinear and nonmonotone in tenure, and the distribution of wages and performance at the main levels in each tenure. I discuss additional results, model fit, and goodness-of-fit test results in Appendix B. Section 5 in the S.A. reports estimates from a larger sample that includes managers entering the firm at levels higher than Level 1.\(^{30}\)

Uncertainty and Learning. Three findings emerge from Tables 6A and 6B. First, the distribution of the initial priors in Table 6A, \( \{ p_{i1}, q_{i} \} \), implies a large degree of uncertainty about ability and of dispersion in information at entry in the firm. The average initial prior probability that a newly hired manager is of high ability is 0.473, but initial priors range from 0.338 for the first skill type to 0.607 for the fourth skill type with a standard deviation of 0.102 (see the table note).

Second, the estimates of the parameters \( \{ \alpha_k, \beta_k \} \) in Table 6B reveal that Level 1 is the most informative and thus plays an important screening role by allowing the firm to gather information about managers’ ability. Indeed, by the Blackwell criterion, the ordering of job levels by their informativeness from lowest to highest is Levels 2, 3, and 1—the differences \( \alpha_2 - \alpha_1 \) and \( \beta_2 - \beta_1 \) are significant at the 1 percent level. Interestingly, the ordering of levels by the probability of high performance of a manager of high or low ability from lowest to highest is Levels 1, 3, and 2 and so is opposite from their ordering by informativeness. The greater informativeness of Level 1 implies that managers are paid a premium at promotion to Level 2, which compensates managers for the lower informativeness of the new level. Note that \( \alpha_k \geq \beta_k \) further implies that the distribution of output at any level of a manager of high ability first-order stochastically dominates that of a manager of low ability. Since the expected output of a manager of high ability is then higher than that of a manager of low ability at any level, the firm faces a clear trade-off between a manager’s output and the information it can acquire about ability when deciding on a manager’s assignment. For instance, although employment at Level 1 generates the most information, a manager has the lowest probability of success at this level.

Third, despite learning being gradual, its speed varies substantially across job levels and manager skill types. The difference in learning speed across levels is key to the shape of wage profiles at the firm, as the counterfactual experiments below will show. To measure the speed of learning, I compute the number of consecutive years of high performance

\(^{30}\)Although I estimate a richer specification on this larger sample, the estimates of the key parameters across the two samples are very close. Model fit, especially in terms of the predicted wage distribution at Level 3, is improved, though.
that are necessary to infer with a 90 percent chance that a manager is of high ability. To see that this speed varies across
levels, note that 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 Level 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 that of Nagypál (2007) on the impact of
learning on labor market turnover. From her Figure 7, the convergence of beliefs, reflected in the tenure profile of match
quality and output, occurs past the tenth year of tenure. This result is also in line with the large estimated degree of
uncertainty about ability at entry in the firm, even for managers with several years of labor market experience. (See
Appendix B for a discussion of the estimated parameters of classification error in performance in Table 6C.)

**Expected Output at Firm** A. The estimates of the expected output parameters \{d_{Akt}(k_{t-1}), e_{Akt}\} are reported in Table
6D. Their size, expressed in thousands, reflects the normalization of \(d_{A11} (\cdot)\) to 1,000 and \(e_{A11}\) to zero. All parameters
not reported are assumed equal to their corresponding value at Level 2 of firm \(C\), the reference level, which is set to zero
as discussed in Section 4.3. See Appendix B for details on this parameterization strategy and below for the estimates of
\{\(b_{k}(h_{1})\}\), which are given by the estimates of \((\omega_{1k}, \omega_{2k}, \omega_{3k}, \{\omega_{ym}\})\) as explained in Appendix B where (24) is derived.

Three features emerge from Table 6D. First, the parameters \(e_{A2t}\) and \(e_{A3t}\) display an approximately hump-shaped profile with tenure. Namely, the parameters \(e_{A2t}\) start at zero, peak at \(t = 2\) and \(t = 4\), and essentially flatten out thereafter. A similar hump-shaped pattern holds for the parameters \(e_{A3t}\), which peak at \(t = 4\). The resulting positive correlation in the tenure profiles of the parameters \(e_{Akt}\) at Levels 2 and 3 suggests that human capital acquired with tenure is task-general between the two levels by the definition in Section 4.3. Recall that \(y_{A}(s_{it}, k)\) from (3) is given by
\[
E(y_{A|kt}|\beta, \kappa_t) + [E(y_{A|kt}|\alpha, \kappa_t) - E(y_{A|kt} |\beta, \kappa_t)]p_{it}
\]
and that \(e_{Akt} = E(y_{A|kt}|\alpha, \kappa_t) - E(y_{A|kt} |\beta, \kappa_t)\) captures the output return to high ability. The tenure profiles of \(e_{Akt}\) imply that this return tends to increase at Levels 2 and 3 with tenure.

Second, human capital acquired at Level 3 is task-specific and implies that demoting a manager from Level 3 to a
lower level is costly for the firm, which helps to explain the absence of demotions. That is, managers assigned to Level
3 in \(t − 1\) and \(t\) produce more in \(t\) than they would if assigned to Levels 1 or 2 in \(t − 1\) and promoted to Level 3 in \(t\).
Further, managers assigned to Level 3 in \(t − 1\) and \(t\) produce more in \(t\) than they would if assigned to Level 3 in \(t − 1\) and
demoted to Levels 1 or 2 in \(t\) at intermediate to high priors, when the risk of demotion is indeed largest.

Third, the tenure profile of these parameters suggests that high-ability managers have an absolute advantage at Level
2 over 1 as well as, since \(e_{Akt} > e_{Akt}\) for \(k' > k\), a comparative advantage at Level 2 over 1 and, at intermediate to high
 tenures, at Level 3 over 2. Overall, these estimates imply that expected output is higher at higher levels in mid to high
tenures. Thus, higher ability and acquired human capital tend to make managers more productive at higher levels, which

---

31Recall from Section 4.3 that the more task-general the human capital accumulated at Level \(k\) is, the larger the difference \(d_{A11}(Lk) - d_{A21}(Lk')\), \(k' \neq k\). is. Now, with \(d_{A31}(L1)\) and \(d_{A21}(L3)\) normalized to zero by (A1) and \(d_{A31}(L2) = -d_{A21}(L3)\), the differences \(d_{A31}(L3) - d_{A31}(L1)\) and \(d_{A31}(L2)\) are large and positive in \(4 \leq t \leq 7\), which implies that human capital acquired at Level 3 is highly task-specific.

32This result follows from \(d_{A31}(L3)\) being large and positive in \(4 \leq t \leq 7\) and \(d_{A31}(L3)\) and \(d_{A21}(L3)\) being normalized to zero by (A1).

33Comparative advantage at Level \(k' > k\) over \(k\) can be expressed as \(e_{fjkt}(\kappa_{t})/e_{fkt}(\kappa_{t}) \geq [b_{fjt}(h_{1}) + d_{fkt}(\kappa_{t})]/[b_{fjt}(h_{1}) + d_{fkt}(\kappa_{t})]\) by the
definition after (3). Since \(b_{k}(h_{1}) + d_{Akt}(k_{t-1})\) tends to decrease with the job level, a sufficient condition for comparative advantage is \(e_{Akt} > e_{Akt}\).
contribute more to firm output than lower levels as tenure progresses.

Retention and Assignment Policy at Firm A. The policy that emerges by simulating the model at the estimated values of the parameters of learning and expected output implies that, on average, managers at higher levels have higher ability, higher acquired human capital, and higher but less disperse beliefs—as learning takes place, the beliefs about the ability of retained managers, gradually assigned to higher levels, become more precise. 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 that helps to improve their matching 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. This selection process will prove to be the key channel through which learning affects the dynamics of wages at the firm.

As for separations, the estimated exogenous separation rates in Table 6E are close to the observed rates in Table 2, 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 a primary determinant of learning, and approximately constant with tenure, which is a key determinant of human capital acquisition; see 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 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 those who separate, as I discuss below.

Wages. The estimates in Tables 6F-6I are similar to those in the literature in terms of the implied return to age at entry in the firm, the size of wage increases at promotion, the convexity of wages in job levels and tenure, and the magnitude of wage growth on the job. Hence, along these dimensions, the firm I study is comparable to those studied in other work. In particular, my estimates of wage growth are consistent with those from more representative data, such as those by Buchinsky et al. (2010) from the Panel Study of Income Dynamics and by Song et al. (2016) from U.S. social security data. These latter authors estimate wage growth with tenure from a much larger U.S. administrative matched employer-employee dataset, both for a year comparable to my sample, 1982, and for a more recent one. They find that the median wage of employees with two to four years of tenure and with five or more years of tenure are, respectively, $29,763 and $45,010 in 1982. Since my sample consists solely of managerial employees, the wages I observe are naturally higher but are comparable to those in Song et al. (2016): the analogous figures from my sample are $42,493 and $49,544.

The estimates reveal several features of the wage process at the firm. Table 6G implies that human capital acquired before entry in the firm, as captured by age and education at entry, has a significant impact on wages at the firm. 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 by (24), given that \( \omega_1 + 2\omega_2(29.71) = 0.0102 \) with \( \omega_{1k} = \omega_1 \) and \( \omega_{2k} = \omega_2 \) common across Levels 1 and 2. Since \( \omega_3 = 0.022 \), the effect of an additional year of schooling is a 2.2 percent wage increase. At Level 3, the corresponding figures are 0.4 percent, since \( \omega_{13} + 2\omega_{23}(29.71) = 0.0041 \), and 2.1 percent, as \( \omega_{33} = 0.021 \). These estimates of the effects of age and education are comparable to those in the literature. 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 in a large multi-firm sample of U.S. executives between 1981 and 1988, a window of observation that overlaps with mine. Not surprisingly, as I focus on highly educated workers, my estimate of the effect of education on wages is smaller than theirs. The estimated year-of-entry effects in Table 6H are consistent with the recession years between 1973 and 1982, which depressed the wages of entrants in the firm in those years relative to those of earlier entrants—with \( m \) ranging from 0 (1970) to 9 (1979), I let \( \omega_{ym} = 0 \) for \( 0 \leq m \leq 3 \), which are the reference years, and set \( \omega_{y4} = \omega_{y5} \) in accordance with the severity of those recession years.

The estimates of the intercepts \( \{ \omega_{0ik} \} \) in Table 6F imply that promotions lead to sizable permanent wage increases, which are higher at higher levels: a promotion from Level 1 to 2 implies an increase in annual wages of $781, whereas a promotion from Level 2 to 3 implies an increase of $5,723.\(^{34}\) Also, average wages are convex in job levels as in the data, that is, average wages increase more than linearly with the job level, which reflects the higher expected output of higher-ability managers at higher levels; see the first column of Table 7A for the model and the note to Table 4 for the data. These results are qualitatively and quantitatively 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 coefficients on tenure in Table 6G are quite small.\(^{35}\) This finding suggests that the deterministic component of the human capital acquired at the firm, in contrast to the human capital acquired before entry, has a small direct effect on wages. Human capital acquired at the firm, though, has an important indirect effect on wages through its impact on promotions, as illustrated by the counterfactual analysis below.

In terms of managers’ unobserved characteristics, the coefficients \( \{ \omega_{5i} \} \) on the prior in Table 6I are positive and significant, which implies that average wages increase with the prior. (Recall that the coefficient on the prior is \( \omega_{5i} + \omega_{5kt} \) by (24), but \( \omega_{5kt} \) is estimated as part of the robustness exercise reported below.) This finding suggests that managers of high ability are more valued by the market than managers of low ability. Note also that the standard deviation of the wage disturbance, \( \sigma_{Ai} \), tends to decrease with the level. Despite this pattern, the model implies that the standard deviation of wages is higher at higher levels as in the data: the estimated standard deviation of wages at Levels 1, 2, and 3, pooled across tenures, is $6,936 at Level 1, $7,077 at Level 2, and $8,046 at Level 3; see the note to Table 4 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.\(^{36}\)

**Wage Growth: A Decomposition.** The model implies an increase in average wages of 19.4 percent over the first seven years at the firm, or 18.5 percent in the unbalanced panel that includes separating managers, which corresponds to an average yearly growth rate of 3.2 percent. This magnitude of annual wage growth on the job is consistent with Topel

---

\(^{34}\)These wage increases are calculated by first averaging across manager skill types these level-specific intercepts of mean log wages and then computing the differences in the values resulting from converting these averages from logs to levels.

\(^{35}\)I specify \( \omega_{4kt} = \omega_{4t2} I(t < 4) + \omega_{4t4} I(t \geq 4) \) with \( \omega_{4t4} = -\omega_{4t2} \) to conserve on parameters and match the much smaller fraction of managers assigned to Level 1 after the third year of tenure. Since \( \{ \omega_{4kt} \} \) proved negligible at Levels \( k = 2, 3 \) and impossible to estimate with any precision, they are omitted from the baseline specification but included in the robustness exercise. See the sensitivity analysis below.

\(^{36}\)The result in Kahn and Lange (2014) that the dispersion of wage residuals primarily reflects heterogeneous productivity (human capital) innovations depends on their assumption that all jobs are identical. In their model, only the productivity process can account for the increasing dispersion of wages with experience at an increasing rate. Here, instead, an increase in the variance of wages with experience, even at an increasing rate, can occur as managers with higher ability and human capital advance to higher levels, at which wages are higher and more dispersed.
(1991, Table 2) and the lower end of the range estimated by Buchinsky et al. (2010, Figure 2). As for its determinants, note that by (24) the average wage of a manager of skill type i at Level k of the firm at state s_it can be expressed as
\[ Ew_{it} = \omega(i, h_1, k) + \omega_{4kt} \times (t-1) + \omega_{5kt}, \]
where \( \omega(i, h_1, k) \) captures the sum of the first five terms and of the last one on the right side of (24). (The parameter \( \omega_{5kt} \) is estimated as part of the robustness exercise but its value is small, as noted below. Also, in estimation, the contribution of the last term on the right side of (24) is negligible.) Then, wage growth can be decomposed into the contribution of three 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; and iii) \( \omega_{5kt} \), which captures the direct effect of learning. These last two terms describe the process of stochastic human capital acquisition at the firm net of the productivity (human capital) shocks. Over the first seven years at the firm, the term \( \omega(i, h_1, k) \) accounts for more than 98 percent of the increase in wages. Then, it would seem that learning accounts for a trivial percentage of wage growth.

This inference, however, is incorrect since job assignment is endogenous and itself determined by the process of learning and human capital acquisition at the firm: the term \( \omega(i, h_1, k) \) depends on the assigned level \( k \), which changes over time as information and human capital accumulate. Formally, the parameter \( \omega_{5i} \) captures only the direct marginal contemporaneous effect of beliefs, \( p_{it} \), on wages conditional on job assignment. Measuring, instead, the average cumulative effect of beliefs on wages requires explicitly taking into account how learning affects job assignment and human capital acquisition over time.

More precisely, in the simplest possible notation, to assess how a change in current beliefs \( p_{it} \), given the current level \( k_t \), affects future wages, one needs to account for how this change affects beliefs \( p_{it+1} = p(p_{it}, k_t) \), human capital \( \kappa_{t+1} = \kappa(p_{it}, k_t) \), and thus job assignment \( k_{t+1} = k(p_{it+1}, \kappa_{t+1}) \) next period, which all affect next period wages \( w_{t+1} = w(p_{it+1}, \kappa_{t+1}, k_{t+1}) \) as well as future beliefs, human capital, and job assignments, which in turn determine future wages. Since wages differ greatly across levels, the belief process can have a large effect on wage growth through its indirect effect on the path of promotions, despite its direct effect on wages being negligible. As the counterfactual analysis in Section 6 shows, learning leads managers who are progressively revealed to be of high ability to advance to higher levels, where they are paid higher wages. This effect accounts for almost all of the impact of learning on wages.

**Learning and Wages: Comparison with Literature.** The result that the direct effect of learning on wages is small or insignificant is consistent with the findings in Gibbons et al. (2005), Lluis (2005), and Hunnes (2012). The approach of these papers differs from mine in three respects. First, these papers infer the effect of learning on wages from a single equation, namely, the wage equation \( w_t = w(p_{it}, \kappa_t, k_t) \) in the simplified notation of the previous paragraph, and they measure this effect based on the direct, contemporaneous derivative of wages with respect to beliefs, here \( \omega_{5i} + \omega_{5kt} \). This derivative, however, captures only part of the effect of beliefs on wages. A change in \( p_{it} \) affects current as well as future wages, both directly and indirectly through job assignment. Thus, even abstracting from the effect of ability and so learning on human capital acquisition, two further dynamic equations need to be estimated to measure the impact of learning on wages, in addition to the ones for wages and human capital: one equation for the evolution of human capital, another for the evolution of job assignment.
of beliefs, \( p_{it+1} = p(p_{it}, k_t) \), and one for job assignment, \( k_{t+1} = k(p_{it+1}, \kappa_{t+1}) \), to account for how \( p_{it} \) affects the next period prior \( p_{it+1} \) and job assignment \( k_{t+1} \), which, together with human capital \( \kappa_{t+1} \), determine next period wages \( w_{it+1} = w(p_{it+1}, \kappa_{t+1}, k_{t+1}) \) as well as future priors, human capital, assignments, and wages.\(^{37}\)

Second, the analysis in these papers rests critically on the assumption that jobs are equally informative about ability. When jobs differ in their informativeness, the methods in these papers cannot be applied even just to estimate the direct effect of learning on wages. Third, these papers assume that the labor market is perfectly competitive so that all parameters of output governing job assignment can be recovered from wages, since average wages equal average output.

Here, I discuss the second difference. I elaborate on the first and third differences below when presenting, respectively, the counterfactual exercises and the evidence on the firm’s monopsony power implied by my estimates.

To illustrate this second difference, I review Lluis (2005), which is the paper closest in spirit to mine since it estimates the model of GW, and show that although its approach based on Gibbons et al. (2005) is valid when jobs are equally informative, it is inapplicable when jobs differ in their informativeness. Lluis (2005) assumes workers can be of high or low ability, and that initial beliefs about ability and performance are normally distributed. All jobs are assumed to be equally informative about ability. Using the “rank dummies” \( D_{ikt} \) for worker \( i \)’s rank (job, occupation, or sector) \( k \) in \( t \), worker \( i \)’s wage in rank \( k \) in \( t \) is given by the worker’s expected output up to measurement error, \( w_{ikt} = \sum_{k=1}^{K} D_{ikt}[d_k + X_{ikt}\beta_k + c_k\theta_{ikt}f(x_{ikt})] + \mu_{ikt} \), where \( X_{ikt} \) captures worker characteristics, \( \theta_{ikt} \) is the mean of the posterior beliefs about the worker’s ability, \( f(\cdot) \) is the human capital function, \( x_{ikt} \) is labor market experience, and \( \mu_{ikt} \) is measurement error independent of rank. Comparative advantage arises when \( \{\beta_k, c_k\} \) vary by rank.

Lluis compares estimates from two versions of this wage equation when coefficients can vary by rank abstracting from human capital acquisition \( f(x_{ikt}) = 1 \). The first version is obtained under the assumption of complete information about ability so that \( \theta_{ikt} \) reduces to a time-invariant fixed effect term, \( \theta_i \). The second version is obtained under the assumption of incomplete information about ability so that ability is unobserved to both firms and workers, and learning takes place. In this case, the parameter \( c_k = \partial E(w_{ikt}|X_{ikt}, x_{ikt}, \{D_{ikt}\})/\partial \theta_{ikt} \) captures the marginal contemporaneous effect of beliefs on the wage of a worker with characteristics \( (X_{ikt}, x_{ikt}) \) in rank \( k \) in \( t \). In either version of the wage equation, the comparative advantage hypothesis implies that job assignment is endogenous: since rank assignment is determined by \( \theta_i \) or \( \theta_{ikt} \), both \( \theta_i \) and \( \theta_{ikt} \) are correlated with the dummies \( \{D_{ikt}\} \) in the wage equation. Nonetheless, in the incomplete information case, evaluating the wage equation in \( t \) and \( t - 1 \) yields expressions for \( \theta_{ikt} \) and \( \theta_{ikt-1} \) that can be substituted into the martingale condition for beliefs, \( \theta_{ikt} = \theta_{ikt-1} + u_{ikt} \), where \( u_{ikt} \) is an i.i.d. shock, to obtain

\[
\sum_{k} D_{ikt}c_kf(x_{ikt}) \quad = \quad \sum_{k} D_{ikt}d_k + \sum_{k} D_{ikt}X_{ikt}\beta_k + w_{ikt-1} \quad - \quad \sum_{k} D_{ikt-1}d_k - \sum_{k} D_{ikt-1}X_{ikt-1}\beta_k \\
+ \mu_{ikt} + \sum_{k} D_{ikt}c_kf(x_{ikt}) - \sum_{k} D_{ikt-1}c_kf(x_{ikt-1}).
\]

In the complete information case, a similar argument applies based on the condition \( \theta_{ikt} = \theta_{ikt-1} = \theta_i \) and an expression

\(^{37}\)I am grateful to Thomas Lemieux for fruitful conversations on this topic.
analogous to (25) can be obtained just without the term \( u_{it} \). Observe that (25) is a random coefficient panel data model with unobserved individual time-varying effects, in which rank, \( D_{ikt} \), is endogenous since it is correlated with \( e_{it} \). Yet, (25) can be estimated by nonlinear instrumental variables by using interactions between rank assignment in \( t-1 \) and \( t-2 \), \( D_{ikt-1} \) and \( D_{ikt-2} \), as instruments for \( D_{ikt} \). To see why past ranks provide natural instruments, note that \( D_{ikt-1} \) and \( D_{ikt-2} \) are uncorrelated with either \( u_{it} \), since they are functions of \( u_{it-1} \) and \( u_{it-2} \) that are serially uncorrelated by assumption, or \( \mu_{it} \), since \( \mu_{it} \) is pure noise.

When learning is assumed present, Lluis finds that the parameters of (25) are less precisely estimated and most of the standard errors double in magnitude, compared to the case when learning is assumed absent. With learning, the estimated output parameters are also found difficult to reconcile with economic intuition. Namely, as the author explains (p. 751), none of the tests of equality of the slope coefficients \( \{\beta_k, c_k\} \) can reject the null hypothesis of their being equal, implying no evidence of comparative advantage, in which case workers’ job mobility is purely random. The author concludes that these results do not provide much support for the presence of learning.\(^{38}\) Thus, the author’s econometric approach is valid but the estimates imply that learning has a small and insignificant effect on wages.

When jobs are differentially informative, though, the marginal effect of beliefs on wages measured by \( c_k \) as well as the other output parameters cannot be estimated with this approach. In this case, beliefs are updated differently depending on the assigned rank \( D_{ikt} \), so the martingale condition \( \theta^{it}_{kt} = \theta^{it}_{kt-1} + u_{it} \) becomes \( \theta^{it}_{kt} = \theta^{it}_{kt-1} + \sum_k D_{ikt} u_{ikt} \) and the first term in the second line of (25) is no longer \( u_{it} \) but rather \( \sum_k D_{ikt} u_{ikt} \). Hence, a worker’s previous rank \( D_{ikt-1} \) is now correlated with the error term \( e_{it} \), since \( \sum_k D_{ikt} u_{ikt} \) is one of the components of \( e_{it} \) and the covariance between \( \sum_k D_{ikt} u_{ikt} \) and \( D_{ikt-1} \) is nonzero. Thus, past ranks no longer constitute appropriate instruments, which implies that this approach is invalid. This endogeneity is difficult to correct because current rank is correlated with previous ranks and wages, since they all depend on ability.

**Monopsony Power.** Here I argue that the estimates do not support the notion that the market where the firm hires its managers is perfectly competitive and that this finding implies that if output parameters were inferred from wages rather than job assignments, the estimated effect of beliefs on output, and so of learning on wages, would be downward biased.

To see that the estimates imply that the assumption of perfect competition is rejected by the data, note, intuitively, that if firm \( A \) behaved as a perfect competitor in the market for managers, the correlation between estimated average output, \( y_A(s_{it}, k) \), and wages, \( Ew_{Ait} \), would be close to one at each level, \( k \). According to my estimates, it is much lower.\(^{39}\) Indeed, when competition is imperfect, equilibrium leads to systematic deviations between output and wages.

Since firm \( A \) just needs to match the offer of the competitor to employ a manager, wages reflect a manager’s value to the competitor of firm \( A \) rather than to firm \( A \), and so provide only a lower bound on a manager’s value to firm \( A \).

---

\(^{38}\)By replicating the analysis of Lluis (2005) on my data—the analysis in Hunnes (2012) is similar but based on different data—I also found that the impact of learning on wages is insignificant; details are available upon request. These papers, though, just attempt to detect whether ability is uncertain rather than measuring the importance of learning about ability for wages, when ability is indeed uncertain.

\(^{39}\)Over the first seven years of tenure in the firm, the correlation between average output and wages is 0.478 at Level 1, 0.380 at Level 2, and 0.521 at Level 3 (by excluding the third tenure year in the case of Level 3; otherwise the correlation is much lower). At the estimated values of the parameters, managers capture larger fractions of the surplus with the firm in later tenures.
understand the impact of imperfect competition on parameters and estimates at the simplest level, consider a version of the model without human capital, productivity shocks, costs of employing and switching workers across jobs, or skill types, and with equally informative jobs. In this case, expected output at job \( k \) for a manager with prior \( p_t \) is 
\[ y_A(p_t, k) = d_Ak + e_Ak p_t \] at firm \( A \) and 
\[ y_C(p_t, k) = d_Ck + e_Ck p_t \] at firm \( C \). The wage equation for managers employed at firm \( A \) is then 
\[ w_A(p_t) = y_A(p_t, k) = d_Ak + e_Ak p_t. \] In this setup, assuming perfect competition is equivalent to imposing that the two firms’ technologies are identical in that \( d_Ak = d_Ck \) and \( e_Ak = e_Ck \). In this case, the wage equation \( w_A(p_t) = d_Ck + e_Ck p_t \), say, augmented with measurement error, could be used to directly estimate firm \( A \)’s technology. When the assumption of perfect competition does not apply, though, this wage equation is a misspecified estimating equation for the parameters of the technology of the employing firm, firm \( A \).

In fact, given the large estimated direct effect of beliefs on output, captured by the coefficients on the prior \( \{ e_Akt \} \) in (18), and the small estimated direct effect of beliefs on wages, captured by the coefficients on the prior \( \{ \omega_{5i} \} \) in (24), if output parameters were inferred from wages rather than job assignments, the estimated effect of beliefs on output would be substantially downward biased. For instance, the estimate of \( e_{A21} \) is 59.21, whereas the largest estimate of \( \omega_{5i} \) is 2.37.\(^{40}\) This bias, in turn, would cause an underestimate of the importance of ability for output, especially at Levels 2 and 3, and thus of the role of learning for job assignment and wages.

The reason for this bias is as follows. Over time, learning induces managers of high ability to be assigned to higher job levels that contribute more to output. Namely, \( e_{Akt} \geq 0 \) implies that high-ability managers are more productive than low-ability ones at any level, whereas \( e_{Akt'} \geq e_{Akt}, k' > k \), in mid to high tenures implies that high-ability managers are relatively more productive than low-ability ones at higher levels in those tenures. Given the difference in magnitude between \( \{ e_{Akt} \} \) and \( \{ \omega_{5i} \} \), inferring output parameters from wages would lead to understating the importance of ability for output and so of learning for output growth, which is due to the improved matching of managers to jobs by their ability that learning induces. Understating the importance of learning for output growth through promotions, in turn, would lead to understating the impact of learning on wage growth, since promotions result in the assignment to levels where wages, on average, are much higher despite \( \{ \omega_{5i} \} \) being constant across levels. Indeed, as shown below, both learning and human capital have a mostly indirect effect on wage growth through their impact on promotions.\(^{41}\)

**Sensitivity Analysis.** Tables 6J1-6J6 show that the estimates are robust to the main assumptions underlying the specification of the wage process. Specifically, to assess whether \( e_{Ckt}(i) \) from (10) can be treated as independent of a manager’s level and human capital acquired with tenure and so specified as \( \omega_{5i} \) without loss, I also estimated a more general version of (24). This version includes the tenure coefficient \( \omega_{4k} \) for \( k > 1 \) (see footnote 35) and the prior coefficient \( \omega_{5kt} \) specified as \( \omega_{5kt} = \omega_{5k} + (t - 1) \times \omega_{5k} + \omega_{5} x_1 \) with additional parameters \( \omega_{41}, \omega_{43}, \omega_{52}, \omega_{53}, \omega_{54}, \omega_{55}, \omega_{56}, \omega_{57}, \omega_{58}, \omega_{59}, \omega_{510}, \omega_{511}, \omega_{512}, \omega_{513}, \omega_{514}, \omega_{515} \). Note from the tables that the estimates of the parameters that are also in the baseline model are fairly similar to those from the

\(^{40}\)Note that the largest possible value of \( \omega_{5i} + \omega_{5kt} \) is 2.328 + 0.312, where 0.312 = max\{6ω3, ω4, ω5, ω6, ω7\}; see Tables 6J4 and 6J6. So this bias is large even once the impact of \( \omega_{5kt} \) is accounted for.

\(^{41}\)In Kahn and Lange (2014), the condition that wages equal (expected) output is required by their identification strategy to distinguish learning from an exogenous productivity process, which proxies human capital acquisition in their model. Their model allows for error in performance, which, however, somewhat weakens the authors’ ability to distinguish learning from human capital acquisition, as they remark in footnote 32.
baseline model, except for the expected 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, in particular the slope coefficient on tenure, $\{\omega_{5k}^g\}$, and experience, $\omega_{5k}^x$, compared to $\{\omega_{5i}^g\}$. 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 the results. Similarly, experience at entry does not have much of an impact on the coefficient on $p_{it}$ through $\omega_{5k}^x$.

6 The Role of Learning, Human Capital, and Uncertainty

Here I assess counterfactually the role of learning, human capital acquisition, and persistent uncertainty about ability for the profiles of jobs and wages at the firm. See the derivation of the wage equation in Appendix B for details.

Learning. To evaluate the impact of learning on the dynamics of jobs and wages in the baseline model, I compare its predictions with those of the version of the model in which learning is absent. In the no learning version of the model, jobs are assumed to be uninformative about ability; namely, beliefs are updated with $\beta_k = \alpha_k$ for all $k$.\footnote{When $\beta_k = \alpha_k$, the model is simulated 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. The same approach is adopted for the remaining experiments that involve $\{\alpha_k, \beta_k\}$. Although I focus here on the baseline model, analogous results are obtained for the model considered in the sensitivity analysis.}

Table 7A shows that learning generates faster wage growth and greater wage dispersion with tenure. 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 ($19.4-15.4)/15.4 = 26.0$ percent). The additional wage dispersion due to learning is 23 percent ($(8,046-6,534)/16,534 = 23.1$ percent) at Level 3, with somewhat lower percentages at lower levels. It turns out that job assignment is a key channel through which learning affects wages. Table 7B 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 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 on occupational mobility examined by Papageorgiou (2014). Overall, learning implies more rapid promotions for managers who are progressively revealed to be of high ability and thus leads to higher wage growth, since wages at higher levels are higher. Hence, the total effect of learning on wages is sizable through its impact on promotions, although its direct effect on wages is small, as discussed above.

Human Capital Acquisition. In the previous experiment of no learning, the progression of managers across levels is due to the human capital they acquire at the firm. Acquired human capital improves managers’ expected output over time, 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 wages. Over time, acquired human capital is also important to prevent demotions and thus reduce the wage loss that managers experience after low performance. These findings on the importance of human capital from the previous experiment are confirmed when the model is simulated assuming human capital acquisition is absent. In this latter case, almost no promotion occurs, when promotions occur they are followed by demotions, and virtually no wage
growth accrues. Hence, acquired human capital has a sizable impact on wages. Together with the finding that human capital acquired at the firm has a limited direct effect on wages, as implied by the small estimated wage coefficients on tenure, \( \{\omega_{4kt}\} \), these results imply that human capital, like learning, primarily affects wages through its impact on managers’ job paths.

**Learning by Experimentation.** Here I focus on one exercise, referred to as *no experimentation*, in which all jobs are assumed to be equally informative 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 when beliefs are updated, whereas all other parameters are unchanged. Results for the analogous experiments of equal informativeness as Levels 2 and 3 are reported in Tables A.11-A.13 in the S.A. Table 7C shows that, with no experimentation, nearly all managers who do not separate from the firm are quickly assigned to Level 3. In particular, after the first two years of tenure, the proportion of managers assigned to Level 1 is very small, even though learning takes place slowly over time. Experimentation proves to be quantitatively important not just for job mobility but also for wage growth. As Table 7A shows, rapid promotions are accompanied by rapid wage growth. By the third year of tenure, without experimentation 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 than their cumulative wage growth during the first seven years in the baseline model. After that, wage growth is fairly flat. Thus, the differential informativeness of jobs is critical to account 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. Overall, cumulative wage growth during the first seven years at the firm would be 20 percent higher (20.1 = (23.3 − 19.4)100/19.4) without experimentation.

**Persistent Uncertainty.** Since I estimate learning to be a slow process, uncertainty about ability is highly persistent over time. To assess the role of this persistence, I conduct two experiments. 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 7A shows the implications for wages and Table 7B 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. 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. Further, 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, and thus helps to account for the convexity of wages in job levels and tenure. Indeed, the wage growth that managers would experience over two years, if they could sort to the best jobs given their ability and human capital just after one year, is twice as large as the cumulative wage growth that managers experience over the first seven years at the firm in the baseline model. As these tables show, the case of *fast learning at Level 2* is similar (see Appendix B for a discussion).
7 Related Literature

The model considered here shares four features with existing models of careers in firms. First, I follow Rosen (1982), Waldman (1984), and GW by allowing higher-ability workers to have a comparative advantage at higher-level jobs of a firm’s hierarchy. Second, as in Jovanovic (1979), Harris and Holmström (1982), MacDonald (1982), Miller (1984), Farber and Gibbons (1996), Jovanovic and Nyarko (1997), and GW, firms and workers learn symmetrically about ability. In contrast to Jovanovic (1979), Miller (1984), and Flinn (1986), though, unobserved ability has a common component across jobs and firms. Third, workers can improve their productivity over time by stochastically acquiring human capital on the job that can be task- and firm-specific to varying degrees, unlike GW and in the spirit of Keane and Wolpin (1997) and Sauer (1998). Fourth, output is separable across workers. See Kremer (1993), Kremer and Maskin (1996), Davis (1997), Ferrall (1997), and Ferrall et al. (2009) for work with complementarity among workers; see also Appendix B.

Learning about worker productivity in the labor market or in different occupations has been empirically investigated by Miller (1984), Flinn (1986), Berkovec and Stern (1991), Nagypál (2007), Groes et al. (2010), Antonovics and Golan (2012), and Sanders (2016). In particular, Antonovics and Golan (2012) provide descriptive evidence on the role of learning through experimentation for wages and occupational mobility. Pries and Rogerson (2005) consider a model of turnover with learning and matching frictions. Papageorgiou (2014) analyzes a search model of occupational mobility with learning that allows for correlation in ability across occupations and documents the importance of comparative advantage for occupational sorting. Papageorgiou (2018) enriches this framework with occupational mobility within firms to account for the relationship between firm size and wages in a calibrated version of this augmented model.\footnote{Unlike in my framework, in his model all tasks are ex ante identical and equally informative and workers do not acquire human capital. The paper focuses on a worker’s task choice, under the assumptions that a worker never returns to a previously tried task and wages are determined by Nash bargaining. As a result, wages are unaffected by the dynamic considerations that influence job assignment here.}

My work is complementary to that of Gayle et al. (2015). Their model integrates job assignment and sorting within a model of moral hazard and human capital accumulation to study how information frictions affect the equilibrium assignment of executives to firms and their pay. They measure the relative importance of monitoring costs, effort on the job, and the demand for entrepreneurial talent for the firm size-pay premium. In their model, the match between a firm and an executive changes endogenously over time through human capital accumulation. My model integrates job assignment and sorting within a model of learning about ability and human capital accumulation to study how the acquisition of information about ability and of new skills affects the dynamics of jobs and wages. I measure the relative importance of learning and human capital acquisition for individual job and wage mobility. In my model, the match between a firm and a manager changes endogenously over time through learning and human capital accumulation. In my model too, jobs have an investment value, which arises from both human capital acquisition and the impact of current assignments on future information. Interestingly, the result in Gayle et al. (2015) that output signals are more informative about effort at lower-level jobs is consistent with my finding that information acquired at lower-level jobs is more precise.

As for learning about ability in firms, see Garicano and Van Zandt (2012) on the role of organizations for acquir-
ing and processing information. Chiappori et al. (1999) provide evidence of learning and downward wage rigidity for executives of a French state-owned firm. Following Gibbons et al. (2005), who study sectoral and inter-industry wage differentials, Lluis (2005) and Hunnes (2012) assess the importance of comparative advantage and learning for worker mobility in firms and across occupations. Using information on wages and performance, Kahn and Lange (2014) document that learning and stochastic productivity changes are important for the variance of wages in the BGH data.\footnote{Lochner et al. (2018) identify the role of changes in the returns to unobserved skills, in the variance of unobserved skills, and in the variance of transitory non-skill shocks for the increase in U.S. residual wage inequality from the 1980s onward. Within a competitive assignment model of the labor market, they estimate that both demand and supply factors contributed to a downward trend in returns to skills. See Lochner and Shin (2014) on the importance of unobserved skills for the evolution of log earnings residuals. See Ghosh (2007) for a model of learning and human capital acquisition in which worker turnover across firms stems from disutility shocks to continuing employment with the same firm.} The authors consider a perfectly competitive model of the labor market that abstracts from job assignment, in which workers are randomly assigned to firms and jobs. My results imply that the nonrandom selection of managers to jobs through promotions as well as to firms is central to the growth of wages with tenure and their dispersion in the BGH firm.

8 Discussion

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 discuss alternative explanations, focusing on the potential role of asymmetric information and incentives. I discuss the role of complementarities in production in Appendix B.

Asymmetric Learning. I have assumed that firms and managers share the same information about ability. To see that this assumption is not implausible 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 performed by Devaro and Waldman (2012) on the BGH data does not find conclusive evidence of asymmetric information. Although the assumption of symmetric learning may seem strong, when wages are observable to a firm’s competitors and the initial prior is common, an outside firm can infer from a manager’s wage the performance signal observed by the employing firm and the manager at the end of the previous period. Hence, observing wages is analogous to observing performance. For related work that relaxes the assumption of symmetric information, see Waldman (1984), Greenwald (1986), Ricart i Costa (1988), Bernhardt and Scoones (1993), Bernhardt (1995), Waldman (1996), and Ekinci et al. (forthcoming).\footnote{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 common assumptions in the empirical literature.}

Performance Incentives. The model abstracts from the possibility that firms may induce workers to exert variable amounts of 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 the bonus information in my data, available only between 1981 and 1989, 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 source of observed pay than the sources I focus on.\footnote{Since bonuses are higher at higher levels and learning contributes to wage growth primarily through its impact on promotions, the effect of learning on wages that I estimate excluding bonuses can be conjectured to be a lower bound on the effect of learning on total compensation.}

**Tournaments.** A firm may also provide incentives for performance through the implicit promise of promotion. In tournament models (see 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 merely determined by luck. On the contrary, tournament models with heterogeneous workers and sorting are difficult to empirically distinguish from learning models (see Rosen (1986)). For instance, both tournament and learning models with sorting predict correlation in promotion rates and wage increases over time. Yet, one aspect of my data that is at odds with the idea that tournaments are key to the dynamics of jobs and wages is the high frequency of wage decreases. A tournament model does not naturally lead to these 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. Yet, the immediate wage premium from a promotion is only around 7 percent. They interpret this finding as evidence against the hypothesis that the wage structure at the firm reflects a tournament. A full analysis of the role of learning, human capital, and incentives for wage growth in the presence of sorting is left to future work.

### 9 Conclusion

This paper estimates a model that integrates learning about ability, human capital acquisition, and job assignment to account for rich patterns of individual careers in firms. My estimates call for a reassessment of the importance of learning for wages. Most of the literature argues that the impact of learning on wages is small but, by construction, uncovers only the direct effect of learning, which is due to the impact of current beliefs about ability on current wages. By explicitly modeling 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 affects wages indirectly through the dynamic selection process that leads managers who are progressively revealed to be of high ability to be promoted to higher levels of a firm’s job hierarchy. Since wages at higher job levels are higher because of their greater contribution to output, learning critically contributes to wage growth by stimulating promotions. I estimate that this indirect effect of learning on wages is large and accounts for almost all of the impact of learning on wages. As with any model-based analysis, these results should be subject to further verification, especially since they are based on data on one firm. Yet, these findings attest to the potential of learning and human capital models to account for salient features of careers.
References


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

\[
W(s_t, \varepsilon_t | f) = W(s_t, \varepsilon_t | f')
\]  

(26)

by profit maximization. Thus, the worker is indifferent between working at firm \( f \) or working at firm \( f' \). I refer to (26) 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, of choosing the assignment for the worker that maximizes the firm’s value of output, \( \nabla_f(s_t, \varepsilon_t), f = A, C \). The argument is as follows. By (26), 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 \( V^A(s_t, \varepsilon_t) \) can be expressed as

\[
V^A(s_t, \varepsilon_t) = \max \{ V^A(s_t, \varepsilon_t | A), V^A(s_t, \varepsilon_t | C) \} = \max \left\{ \max_{k \in K^A} \left\{ y_A(s_t, k) + \varepsilon_A k t + \delta \eta_A k(t) \cdot \int_{\varepsilon_{t+1}}^V A(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\}, w_C(s_t, \varepsilon_t) + \delta \eta_C k C(t) \int_{\varepsilon_{t+1}}^V A(s_{t+1}, \varepsilon_{t+1} | s_t, k C) dG \right\},
\]

(27)

since firm \( A \)’s current profits are zero, whereas 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 \( C \). Consider now states at which firm \( A \) employs the worker. At these states, firm \( A \) must prefer employing the worker to not employing him, 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 \( A \) or 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, \( \Pi_A(s_t, \varepsilon_t | A) = W(s_t, \varepsilon_t | A) = \Pi_A(s_t, \varepsilon_t | C) = W(s_t, \varepsilon_t | C) \). That is, \( V^A(s_t, \varepsilon_t | A) = V^A(s_t, \varepsilon_t | C) \). By the same argument as before, firm \( C \) must pay a wage that makes the worker just indifferent between working for \( A \) or 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,

\[
V^A(s_t, \varepsilon_t) = \Pi_A(s_t, \varepsilon_t | A) + W(s_t, \varepsilon_t | A) = \Pi_A(s_t, \varepsilon_t | C) + W(s_t, \varepsilon_t | C).
\]

(28)

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, \( \nabla^A(s_t, \varepsilon_t) \), defined recursively as

\[
\nabla^A(s_t, \varepsilon_t) = \max \limits_{k \in K^A} \left\{ y_A(s_t, k) + \varepsilon_A k t + \delta \eta_A k(t) \int_{\varepsilon_{t+1}}^V A(s_{t+1}, \varepsilon_{t+1} | s_t, k) dG \right\}.
\]

A similar argument applies to firm \( C \) so that \( V^C(s_t, \varepsilon_t) \) also reduces to \( \nabla^C(s_t, \varepsilon_t) \). Thus,

\[
\nabla_f(s_t, \varepsilon_t) = \nabla_f(s_t, \varepsilon_t | f) = \nabla^f(s_t, \varepsilon_t), \quad f = A, C.
\]

(29)

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

\[
\Pi_A(s_t, \varepsilon_t | A) + W(s_t, \varepsilon_t | A) = \Pi_A(s_t, \varepsilon_t | C) + W(s_t, \varepsilon_t | C).
\]

(30)

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

\[
\Pi_A(s_t, \varepsilon_t | A) + W(s_t, \varepsilon_t | A) + \Pi_A(s_t, \varepsilon_t | A) \geq \Pi_A(s_t, \varepsilon_t | C) + W(s_t, \varepsilon_t | C) + \Pi_A(s_t, \varepsilon_t | C).
\]

(31)
Similarly, as just argued, when firm $C$ employs the worker at state $(s_t, \varepsilon_t)$,

$$V^A(s_t, \varepsilon_t|C) = \Pi^A(s_t, \varepsilon_t|C) + W(s_t, \varepsilon_t|C) = \Pi^A(s_t, \varepsilon_t|A) + W(s_t, \varepsilon_t|A) = V^A(s_t, \varepsilon_t|A), \quad (32)$$

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, (32) implies

$$\Pi^A(s_t, \varepsilon_t|C) + W(s_t, \varepsilon_t|C) + \Pi^C(s_t, \varepsilon_t|C) \geq \Pi^A(s_t, \varepsilon_t|A) + W(s_t, \varepsilon_t|A) + \Pi^C(s_t, \varepsilon_t|A). \quad (33)$$

Then, (31) and (33) 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 $k_f = k_f(s_t, \varepsilon_t)$, $f = A, C$. Equivalently, the employing firm is determined by the policy of the program with value (8), where $S(\cdot) = \Pi^A(\cdot) + W(\cdot) + \Pi^C(\cdot)$.

**Proof of Proposition 2:** Note that the indifference condition in (26) can be rewritten as

$$w_A(s_t, \varepsilon_t) = W(s_t, \varepsilon_t|C) - \delta \eta_{Ak,t} \int_{\varepsilon_{t+1}} EW(\cdot|s_t, k_A) dG. \quad (34)$$

When $A$ employs the worker, (7) can be expressed as

$$W(s_t, \varepsilon_t|C) = \Pi^C(s_t, \varepsilon_t|C) - \delta \eta_{Ak,t} \int_{\varepsilon_{t+1}} \Pi^C(\cdot|s_t, k_A) dG$$

$$= \max_{k \in \mathbb{K}} \left\{ y_C(s_t, k) + \varepsilon_{Ckt} + \delta \eta_{Ck}(\kappa_t) \int_{\varepsilon_{t+1}} EV^C(\cdot|s_t, k) dG \right\} - \delta \eta_{Ak,t} \int_{\varepsilon_{t+1}} \Pi^C(\cdot|s_t, k_A) dG,$$

where the second equality follows from (29). By substituting this expression into (34), using that $EV^C(\cdot|s_t, k_A) = EW(\cdot|s_t, k_A) + \Pi^C(\cdot|s_t, k_A)$ by definition of $V(\cdot)$, and combining terms, (34) can be rewritten as

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

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

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

where $\varepsilon_{Ct}$ is a zero-mean Gumbel productivity shock and $v_C(s_t, k)$ is given by

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

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

$$w_A(s_t, \varepsilon_t) = \ln \left( \frac{\exp\{v_C(s_t, k_A)\} \sum_{k \in \mathbb{K}} \exp\{v_C(s_t, k)\}}{\exp\{v_C(s_t, k_A)\}} \right) - \delta \eta_{Ak,t} \int_{\varepsilon_{t+1}} EV^C(\cdot|s_t, k) dG$$

$$= v_C(s_t, k_A) - \ln \Pr(k_{Ct} = k_A|f_t = C, s_t) - \delta \eta_{Ak,t} \int_{\varepsilon_{t+1}} EV^C(\cdot|s_t, k) dG$$

$$+ \varepsilon_{Ct}, \quad (35)$$

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

$$\delta \eta_{Ck,t}(\kappa_t) E\left( \ln \sum_{k' \in \mathbb{K}} \exp\{v_C(s_{t+1}, k')\}|s_t, k \right) = \delta \eta_{Ak,t}(\kappa_t) E\left( \ln \sum_{k' \in \mathbb{K}} \exp\{v_C(s_{t+1}, k')\}|s_t, k \right), \quad (36)$$

where the left side of (36) is the continuation value portion of $v_C(s_t, k_A)$. The desired result then follows, since

---

47 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 $(\ln(e^{y_1} + e^{y_2})/\mu, \mu)$, mean $\ln(e^{y_1} + e^{y_2})/\mu + \gamma/\mu$, and variance $\pi^2/6\mu^2$. Thus, $y_{max}$ can be expressed as $y_{max} = y^* + \varepsilon^*$, where $y^* = \ln(e^{y_1} + e^{y_2})/\mu$ and $\varepsilon^*$ is Gumbel distributed with parameters $(0, \mu)$. 

43
Assume without loss that \( w \sim \mathcal{N}(s, \varepsilon) \) in (35) reduces to \( y_C(s, k_A) = \ln \Pr(k_C = k_A | f_t = C, s_t) + \varepsilon_C \).

\[ \ln \Pr(k_C = k_A | f_t = C, s_t) + \varepsilon_C. \]

**Proof of Lemma 1:** Assume without loss that \( \sigma_{Ak} > \sigma_{A2k} > \sigma_{A3k} > \sigma_{A4k} \) 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 \( R_A, \ldots, R_{A(t-1)} \). To start, note that \( f(w_{A2} | L_{A1} = 1, L_{A2}, h_1) \) can be expressed as

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

which is a mixture density with weights \( \{ \Pr(R_{A1}, i | L_{A1} = 1, L_{A2}, h_1) \} \). By Theorem 1 and 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 \( (R_{A1}, L_{A2}, h_1, i) \), and \( \{ \Pr(R_{A1}, i | L_{A1} = 1, L_{A2}, h_1) \} \) are identified for each \( (L_{A2}, h_1) \). For later, it is useful to express these mixture weights as

\[
\Pr(R_{A1}, i | L_{A1} = 1, L_{A2}, h_1) = \frac{\Pr(L_{A2} | L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1} | L_{A1} = 1, h_1, i) \Pr(i | L_{A1} = 1)}{\Pr(L_{A2} | L_{A1} = 1, h_1)},
\]

(37)

since \( \Pr(i | L_{A1} = 1, h_1) = \Pr(i | L_{A1} = 1) \), which is identified by the argument in the main text preceding the statement of the lemma. Once the set of probabilities \( \{ \Pr(R_{A1}, i | L_{A1} = 1, L_{A2}, h_1) \} \) are identified, it is possible to label these probabilities by \( i \) as follows. Since the standard deviations of the component densities satisfy \( \sigma_{Ak1} > \sigma_{Ak2} > \sigma_{Ak3} > \sigma_{Ak4} \) for each \( k \), it is possible to determine the skill type \( i \) that each density in \( \{ f(w_{A2} | L_{A1} = 1, R_{A1}, L_{A2}, h_1, i) \} \) corresponds to, for given \( L_{A2} \) and \( h_1 \), by comparing the standard deviations of these densities. Specifically, for any given \( L_{A2} \) and \( h_1 \), there are 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 for each \( L_{A2} = k \), that is, \( \{ \sigma_{Ak} \}_i \). Then, each pair of probabilities in \( \{ \Pr(R_{A1} = 1, i | L_{A1} = 1, L_{A2}, h_1), \Pr(R_{A1} = 0, i | L_{A1} = 1, L_{A2}, h_1) \} \) gives the probabilities \( \{ \Pr(i | L_{A1} = 1, L_{A2}, h_1) \} \), which, in turn, can be associated with the corresponding probabilities in \( \{ \Pr(i | L_{A1} = 1) \} \) 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 \( \{ \Pr(i | L_{A1} = 1, \ldots, L_{A(t-1)} = 1, h_1) \} \) across tenures by type.

By a similar logic, I now prove that \( f(w_{A2} | L_{A1} = 1, R_{A1}, L_{A2}, h_1) \) 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 capital \( h_1 \) and recorded performance \( R_{A1} \) is

\[
f(w_{A2} | L_{A1} = 1, R_{A1}, L_{A2}, h_1) = \sum_{R_{A1}, i} f(w_{A2} | L_{A1} = 1, R_{A1}, L_{A2}, h_1, i) \Pr(R_{A1}, i | L_{A1} = 1, R_{A1}, L_{A2}, h_1).
\]

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

\[
\Pr(R_{A1}, i | L_{A1} = 1, R_{A1}, L_{A2}, h_1) = \frac{\Pr(L_{A2} | L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1} | L_{A1} = 1, h_1, i) \Pr(i | L_{A1} = 1)}{\Pr(R_{A1} | L_{A1} = 1, h_1)},
\]

(38)

where \( \Pr(L_{A2} | L_{A1} = 1, R_{A1}, h_1, i) \) is independent of \( R_{A1} \). 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 \( f(w_{A2} | L_{A1} = 1, R_{A1}, L_{A2}, h_1) \).

A nearly identical 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

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

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

$$\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1) = \frac{\Pr(L_{A3}|L_{A1} = 1, R_{A1}, L_{A2}, R_{A2}, h_1, i) \Pr(R_{A2}|L_{A1} = 1, R_{A1}, L_{A2}, h_1, i) \cdot \Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1} = 1, h_1, i) \Pr(i|L_{A1} = 1)}{\Pr(L_{A2}, L_{A3}|L_{A1} = 1, h_1)}.$$

As before, by comparing the standard deviations of the component wage densities of this mixture, it is possible to determine the skill type $i$ to which each component density in $\{f(w_{A3}|\cdot, h_1, i)\}$ 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 so those in $\{\Pr(i|L_{A1} = 1, L_{A2}, h_1)\}$ through the standard deviations of the associated densities of the corresponding wage mixtures.

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 $t = 2$ with recorded performance $R_{A2}^o$ at that level is

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

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

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 classification error in recorded performance are identified from the identified mixture weights of the distributions of wages at each level and tenure of managers of each possible skill type for any possible history of level assignments and true and recorded performance at firm $A$. In Part II, I show that once 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) and (12). Fix $R_{A1}^o, L_{A2},$ and $h_1$. Note first that the conditional probability of recorded performance given true performance in $t = 1$,

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

is independent of $h_1$ and $i$ by (11) and (12). The probability $\Pr(R_{A1}^o|L_{A1} = 1, R_{A1})$ is identified for each $R_{A1}^o$ and $R_{A1}$ from the ratio of the identified expression on the right side of (38) multiplied by the associated probability $\Pr(R_{A1}^o, L_{A2}|L_{A1} = 1, h_1),$

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

where I have used that $\Pr(L_{A2}|L_{A1} = 1, R_{A1}^o, R_{A1}, h_1, i)$ is independent of $R_{A1}^o$ and expressed $\Pr(R_{A1}^o, R_{A1}|\cdot)$ as the product of conditional probabilities, to the corresponding identified expression on the right side of (37) multiplied by the
associated probability $\Pr(L_{A2}|L_{A1}=1, h_1)$,

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

To compute this ratio, I need to ensure that no label ambiguity arises when pairing expressions (37) and (38). To see that no such ambiguity arises, recall that (37) and (38) are simple manipulations of $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$ and $\{\Pr(R_{A1}, i|L_{A1}=1, R'_{A1}, L_{A2}, h_1)\}$, which are, respectively, the weights of the mixture distributions $f(w_{A2}|L_{A1}=1, L_{A2}, h_1)$ and $f(w_{A2}|L_{A1}=1, R'_{A1}, L_{A2}, h_1)$ identified by Lemma 1 and, as proved there, can be correctly labeled with respect to $i$. Recall also from the proof of that lemma 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 $f(w_{A2}|L_{A1}=1, R'_{A1}, L_{A2}, h_1)$. Since mean wages by skill type are injective functions of the prior $p_{it}$ for any $\delta$ and $\eta_{it}$ not too large, these densities can be distinguished by their means.\textsuperscript{48} Then, each probability in $\{\Pr(R_{A1}, i|L_{A1}=1, L_{A2}, h_1)\}$ can be uniquely paired with the corresponding one in $\{\Pr(R_{A1}, i|L_{A1}=1, R'_{A1}, L_{A2}, h_1)\}$, given $R'_{A1}, L_{A2}, h_1$, and $i$, since any such pair of probabilities is associated with the same component wage density in $\{f(w_{A2}|L_{A1}=1, R_{A1}, L_{A2}, h_1, i)\}$. Hence, the set of probabilities $\{\Pr(R'_{A1}|L_{A1}=1, R_{A1})\}$ are identified for any given $R'_{A1}$ and $R_{A1}$.

To see how $\Pr(R'_{A1}|L_{A1}=1, R_{A1})$ can be distinguished from $\Pr(R_{A1}|L_{A1}=1, R_{A1}=0)$ for given $R'_{A1}$, observe that $\rho_1 > 0$ implies that, for given $L_{A1}$,

$$
\Pr(R'_{A1}|L_{A1}=1, L_{A1}, R_{A1} = 0) > \Pr(R'_{A1}|L_{A1}=1, L_{A1}, R_{A1} = 0),
$$

which is immediate by comparing $1 - E(1, t)$ and $E(0, t)$ from (11) and (12). Then, $\Pr(R'_{A1}|L_{A1}=1, R_{A1} = 1)$ can be distinguished from $\Pr(R'_{A1}|L_{A1}=1, R_{A1} = 0)$ for any given $R'_{A1}$, as claimed. By this fact, it is also immediate that the associated probabilities $\Pr(R_{A1}=1, i|L_{A1}=1, L_{A2}, h_1)$ in (37) and $\Pr(R_{A1}=1, i|L_{A1}=1, R'_{A1}, L_{A2}, h_1)$ in (38) can be distinguished from the corresponding probabilities $\Pr(R_{A1}=0, i|L_{A1}=1, L_{A2}, h_1)$ and $\Pr(R_{A1}=0, i|L_{A1}=1, R'_{A1}, L_{A2}, h_1)$ for each $R'_{A1}, L_{A2}, h_1$, and $i$.

Proceeding similarly, the ratio of the identified expression on the right side of (41), multiplied by the associated probability $\Pr(L_{A2}, R'_{A2}, L_{A3}|L_{A1}=1, h_1)$, to the corresponding identified expression on the right side of (40), multiplied by the associated probability $\Pr(L_{A2}, R_{A2}, L_{A3}|L_{A1}=1, h_1)$, for given $R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1$, and $i$, identifies $\Pr(R'_{A2}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_1, i) = \Pr(R'_{A2}|L_{A1}=1, L_{A2}, R_{A2})$ by using that $\Pr(L_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_1) = \Pr(L_{A3}|L_{A1}=1, R_{A1}, L_{A2}, L_{A2}, h_1)$ is independent of $R'_{A2}$. Since the same set of densities $\{f(w_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1, i)\}$ is recovered from both $f(w_{A3}|L_{A1}=1, L_{A2}, R_{A2}, L_{A3}, h_1)$ and $f(w_{A3}|L_{A1}=1, L_{A2}, R_{A2}, L_{A3}, h_1)$, it follows that, through these densities, the mixture weights $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1)\}$ can be matched to the corresponding weights $\{\Pr(R_{A1}, R_{A2}, i|L_{A1}=1, L_{A2}, L_{A3}, h_1)\}$. Again, these probabilities can be unambiguously paired: given $i$, different combinations of values of $R_{A1}$ and $R_{A2}$ lead to different posteriors and therefore to different mean wages by skill type—otherwise, it must be that the corresponding densities and weights are equal—since mean wages are injective functions of the prior $p_{it}$. Thus, each weight in (41) can be associated with the appropriate weight in (40) through the corresponding identical component density. By (43), no label ambiguity for $\Pr(R'_{A2}|L_{A1}=1, L_{A2}, R_{A2})$ arises with respect to $R_{A2}$.

The results that $\{\Pr(R'_{A1}|L_{A1}=1, R_{A1})\}$ and $\{\Pr(R'_{A2}|L_{A1}=1, L_{A2}=2, R_{A2})\}$ are identified imply that $E_0(1, t)$ and $E_1(1, t)$, $t = 1, 2$, are identified. By (11) and (12), 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 $\Pr(R'_{A1}=1|L_{A1}=1, R_{A1}=1)$ and $\Pr(R'_{A1}=1|L_{A1}=1, R_{A1}=0)$ or $\Pr(R_{A1}=1|L_{A1}=1, R_{A1}=1)$ and $\Pr(R_{A1}=0|L_{A1}=1, R_{A1}=0)$, together with either $\Pr(R_{A2}=1|L_{A1}=L_{A2}=1, R_{A2})$ or $\Pr(R_{A2}=0|L_{A1}=L_{A2}=1, R_{A2})$ for given $R_{A2}$, are sufficient to identify $(\rho_0, \rho_1, \rho_2(1))$.

Now, consider identifying $\rho_2(2)$ at Level 2. Proceeding analogously, it is immediate that $E_0(2, 2)$ or $E_1(2, 2)$ are also identified from $\Pr(R'_{A2}|L_{A1}=1, L_{A2}=2, R_{A2})$ for given $R_{A2}$, which implies that $\rho_2(2)$ is also identified from either $\Pr(R_{A2}=1|L_{A1}=1, L_{A2}=2, R_{A2})$ or $\Pr(R_{A2}=0|L_{A1}=1, L_{A2}=2, R_{A2})$ for given $R_{A2}$, once $\rho_0$ and $\rho_1$ are identified. Lastly, consider identifying $\rho_2(3)$ at Level 3. By extending this argument in the natural way, it is easy to

\textsuperscript{48}To see that $E_{w, at}$ is an injective function of the prior $p_{it}$, suppose, by contradiction, that $p_{it} > p_{it}$ but $E_{w, at} = E_{w, at}^{'}$, which implies that

$$
\ln \Pr(L_{ct} = k|\hat f_s = C, s_{it}) = \ln \Pr(L_{ct} = k|f_s = C, s_{it}) = \gamma C_s, k = \gamma C_{s_{it}, k}
$$

or, equivalently, $\ln \sum_k e^{C_s(k)} - \beta_{s_{it}} E[\ln \sum_k e^{C_s(k)}|s_{it}, k] = \ln \sum_k e^{C_s(k)} - \beta_{s_{it}} E[\ln \sum_k e^{C_s(k)}|s_{it}, k]$. This latter equality cannot hold if $\delta$ or $\eta_{it}$ is small enough, since $\ln \sum_k e^{C_s(k)}$ is a strictly increasing function of the prior.
show that $\rho_2(3)$ is identified from the distribution of wages of managers in the fourth year of tenure in firm $A$ 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\}**. Consider ($\alpha_1, \beta_1$). Recall that $q_3(h_1) = \Pr(i|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)$. The logic of the proof consists of first showing that the observed distributions of ratings at Level 1 in three consecutive years, namely, the probabilities $\Pr(R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)$, $\Pr(R_{A2}^o = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)$, and $\Pr(R_{A2}^o = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)$, identify the three corresponding probabilities of high performance, namely, $\nu_1(h_1) = (\alpha_1 - \beta_1) \sum_i q_3(h_1)p_{i1} + \beta_1$, $\nu_2(h_1) = (\alpha_1^2 - \beta_1^2) \sum_i q_3(h_1)p_{i1} + \beta_1^2$, and $\nu_3(h_1) = (\alpha_1^3 - \beta_1^3) \sum_i q_3(h_1)p_{i1} + \beta_1^3$, once classification error in performance at Level 1 is identified. The proof then proceeds to recover ($\alpha_1, \beta_1; \sum_i q_3(h_1)p_{i1}$) for a given $h_1$ as the unique solution to the system

$$
\begin{cases}
(\alpha_1 - \beta_1) \sum_i q_3(h_1)p_{i1} + \beta_1 = \nu_1(h_1), \\
(\alpha_1^2 - \beta_1^2) \sum_i q_3(h_1)p_{i1} + \beta_1^2 = \nu_2(h_1), \\
(\alpha_1^3 - \beta_1^3) \sum_i q_3(h_1)p_{i1} + \beta_1^3 = \nu_3(h_1).
\end{cases}
$$

(44)

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. The rest of the proof is articulated in five steps.

**Step 1**: $\nu_1(h_1) = (\alpha_1 - \beta_1) \sum_i q_3(h_1)p_{i1} + \beta_1$ is identified for each $h_1$. Using that $\Pr(R_{A1}^o|L_{A1}, \ldots, L_{A3}, R_{A1}, h_1, i)$ is independent of $h_1$ and $i$ by (11) and (12), it follows that

$$
\Pr(R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) = \sum_i \Pr(i|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)
$$

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

$$
= \sum_i q_3(h_1)\{\Pr(R_{A3} = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, R_{A3} = 1) \Pr(R_{A3} = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i)
$$

$$
+ \Pr(R_{A3} = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, R_{A3} = 0) \Pr(R_{A3} = 0|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i)\}
$$

(45)

where

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

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 $\sum_i q_3(h_1) = 1$ and the definition of classification error rates, by straightforward algebra I obtain from (45) that

$$
\Pr(R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) = \left[ (\alpha_1 - \beta_1) \sum_i q_3(h_1)p_{i1} + \beta_1 \right] [1 - E_1(1, 3)]
$$

$$
+ \left[ 1 - (\alpha_1 - \beta_1) \sum_i q_3(h_1)p_{i1} - \beta_1 \right] E_0(1, 3).
$$

(46)

Since the probability on the left side of (46) is known from the data, and $E_1(1, 3)$ and $E_0(1, 3)$ are identified by Part I, it follows that $\nu_1(h_1) = (\alpha_1 - \beta_1) \sum_i q_3(h_1)p_{i1} + \beta_1$ is identified by (46) for each $h_1$.

**Step 2**: $\nu_2(h_1) = (\alpha_1^2 - \beta_1^2) \sum_i q_3(h_1)p_{i1} + \beta_1^2$ is identified for each $h_1$. To this purpose, note that

$$
\Pr(R_{A2}^o = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)
$$

$$
= \sum_i \Pr(i|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) \Pr(R_{A2}^o = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i)
$$

$$
= \sum_i q_3(h_1) \sum_{R_{A2}, R_{A3}} \Pr(R_{A2} = 1, R_{A3} = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, R_{A2}, h_1, i)
$$

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

$$
= \sum_i q_3(h_1) \sum_{R_{A2}, R_{A3}} \Pr(R_{A2} = 1|L_{A1} = 1, L_{A2} = 1, R_{A2}, L_{A3} = 1, R_{A3}, h_1, i)
$$

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

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

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

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

(47)

with $\Pr(R_{A2}, R_{A3}, h_1, i)$ independent of $h_1$ and $\Pr(R_{A2}^o|L_{A1} = 1, L_{A2} = 1, R_{A2}, L_{A3}, R_{A3}, h_1, i)$ independent of $L_{A3}$,
\[ R_{A3}, h_1, \text{ and } i, \text{ since} \]
\[
\Pr(R_{A2}^o|L_{A1} = 1, L_{A2} = 1, R_{A2}, L_{A3}, R_{A3}, h_1, i) = \frac{\Pr(L_{A1} = 1, L_{A2} = 1, R_{A2}^o, R_{A2}, L_{A3}, R_{A3}, h_1, i)}{\Pr(L_{A1} = 1, L_{A2} = 1, R_{A2}, L_{A3}, R_{A3}, h_1, i)}
\]
\[
\Pr(L_{A3}, R_{A3}|L_{A1} = 1, L_{A2} = 1, R_{A2}^o, R_{A2}, h_1, i) = \frac{\Pr(R_{A2}^o|L_{A1} = 1, L_{A2} = 1, R_{A2}, h_1, i)}{\Pr(L_{A3}, R_{A3}|L_{A1} = 1, L_{A2} = 1, R_{A2}, h_1, i)} = \Pr(R_{A2}^o|L_{A1} = 1, L_{A2} = 1, R_{A2}),
\]
where the first two equalities follow from simple manipulations and the last equality follows by canceling terms since \( \Pr(L_{A3}, R_{A3}|L_{A1} = 1, L_{A2} = 1, R_{A2}^o, R_{A2}, h_1, i) \) is independent of \( R_{A2}^o \) and \( \Pr(R_{A2}^o|L_{A1} = 1, L_{A2} = 1, R_{A2}, h_1, i) \) is independent of \( h_1 \) and \( i \).

Using the fact just established and the similar fact that \( \Pr(R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, R_{A2}^o, R_{A2}, L_{A3} = 1, R_{A3}, h_1, i) \) is independent of \( R_{A2}^o, R_{A2}, h_1, i \) by (11) and (12), I obtain
\[
\Pr(R_{A2} = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i) = \sum_{R_{A2}, R_{A3}} \Pr(R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, R_{A3})
\]
\[
\cdot \Pr(R_{A2} = 1|L_{A1} = 1, L_{A2} = 1, R_{A2})[\Pr(R_{A1} = 1, R_{A2}, R_{A3}|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i)
\]
\[
+ \Pr(R_{A1} = 0, R_{A2}, R_{A3}|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, i)],
\]
where I used the fact that \( \Pr(R_{A1}, R_{A2}, R_{A3}|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i) \) is independent of \( h_1, i \), and expressed \( \Pr(R_{A2}, R_{A3}|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, i) \) as \( \Pr(R_{A1} = 1, R_{A2}, R_{A3}|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, i) \) by the law of total probability. Note also that \( \Pr(R_{A2}^o, R_{A3}^o|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1, i) \) is independent of \( h_1 \). By simple manipulations and using the definition of classification error rates in (11) and (12), it is easy to show that
\[
\Pr(R_{A2} = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, i) = \left[ 1 - E_1(1, 3)[1 - E_1(1, 2)] \left[ \alpha_1^2 p_{11} + \beta_1^2 (1 - p_{11}) \right] \right]
\]
\[
+ \left[ E_0(1, 3)[1 - E_0(1, 2)] + [1 - E_0(1, 3)] E_0(1, 2) \right] \left[ \alpha_1 p_{11} + \beta_1 (1 - p_{11}) - \alpha_1^2 p_{11} - \beta_1^2 (1 - p_{11}) \right]
\]
\[
+ E_0(1, 3) E_0(1, 2) [\alpha_1^2 p_{11} + \beta_1^2 (1 - p_{11}) + 1 - 2\alpha_1 p_{11} - 2\beta_1 (1 - p_{11})],
\]
where \( B_2, C_2, \text{ and } D_2 \) are known constants so that
\[
\Pr(R_{A2} = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) = \sum_i q_{i3}(h_1) \Pr(R_{A2} = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, i)
\]
\[
= B_2(\alpha_1^2 - \beta_1^2) \sum_i q_{i3}(h_1)p_{i1} + \beta_1^2) + C_2[(\alpha_1 - \beta_1) \sum_i q_{i3}(h_1)p_{i1} + \beta_1 - (\alpha_1^2 - \beta_1^2) \sum_i q_{i3}(h_1)p_{i1} - \beta_1^2]
\]
\[
+ D_2(\alpha_1^2 - \beta_1^2) \sum_i q_{i3}(h_1)p_{i1} + \beta_1^2 + 1 - 2[(\alpha_1 - \beta_1) \sum_i q_{i3}(h_1)p_{i1} + \beta_1]. \quad (48)
\]
Using the fact that the probability on the left side of (48) 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:** \( \nu_3(h_1) = (\alpha_1^3 - \beta_1^3) \sum_i q_{i3}(h_1)p_{i1} + \beta_1^3 \) is identified for each \( h_1 \). Observe that
\[
\Pr(R_{A1}^o = 1, R_{A2}^o = 1, R_{A3}^o = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)
\]
\[
= \sum_i q_{i3}(h_1) \sum_{R_{A1}, R_{A2}, R_{A3}} \Pr(R_{A1}^o = 1, R_{A2}^o = 1, R_{A3}^o = 1|L_{A1} = 1, R_{A1}, L_{A2} = 1, R_{A2}, L_{A3} = 1, R_{A3})
\]
\[
\cdot \Pr(R_{A1}, R_{A2}, R_{A3}|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, i),
\]
and so there exist known constants, \( B_3, C_3, D_3, \text{ and } E_3 \), which are known functions of the classification error rates at
Level 1 in tenures 1, 2, and 3, such that

\[ \Pr(R_{A1}^0 = 1, R_{A2}^0 = 1, R_{A3}^0 = 1 | L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) = B_3 \sum_i q_{33}(h_1) [\alpha_1 p_{11} + \beta_1^2 (1 - p_{11})] + C_3 \sum_i q_{33}(h_1) [\alpha_1^2 (1 - \alpha_1) p_{11} + \beta_1^2 (1 - \beta_1) (1 - p_{11})] + D_3 \sum_i q_{33}(h_1) [\alpha_1 (1 - \alpha_1)^2 p_{11} + \beta_1 (1 - \beta_1)^2 (1 - p_{11})] + E_3 \sum_i q_{33}(h_1) [(1 - \alpha_1)^3 p_{11} + (1 - \beta_1)^3 (1 - p_{11})], \]

or, equivalently,

\[ \Pr(R_{A1}^0 = 1, R_{A2}^0 = 1, R_{A3}^0 = 1 | L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) = B_3 [(\alpha_1^3 - \beta_1^3) \sum_i q_{33}(h_1) p_{11} + \beta_1^3] + C_3 \left\{ [(\alpha_1^2 - \beta_1^2) \sum_i q_{33}(h_1) p_{11} + \beta_1^2] - [(\alpha_1^2 - \beta_1^2) \sum_i q_{33}(h_1) p_{11} + \beta_1^2] \right\} + D_3 \left\{ [(\alpha_1^3 - \beta_1^3) \sum_i q_{33}(h_1) p_{11} + \beta_1^3] + [(\alpha_1 - \beta_1) \sum_i q_{33}(h_1) p_{11} + \beta_1] - 2 [(\alpha_1^2 - \beta_1^2) \sum_i q_{33}(h_1) p_{11} + \beta_1^2] \right\} + E_3 \left\{ 1 - 3[(\alpha_1 - \beta_1) \sum_i q_{33}(h_1) p_{11} + \beta_1] + 3[(\alpha_1^2 - \beta_1^2) \sum_i q_{33}(h_1) p_{11} + \beta_1^2] - [(\alpha_1^3 - \beta_1^3) \sum_i q_{33}(h_1) p_{11} + \beta_1^3] \right\}. \]

Hence, \( \nu_3(h_1) = (\alpha_1^3 - \beta_1^3) \sum_i q_{33}(h_1) p_{11} + \beta_1^3 \) is identified given that \( \Pr(R_{A1}^0 = 1, R_{A2}^0 = 1, R_{A3}^0 = 1 | h_1) \) is known from the data, and \( \nu_1(h_1) = (\alpha_1 - \beta_1) \sum_i q_{33}(h_1) p_{11} + \beta_1 \) and \( \nu_2(h_1) = (\alpha_1^2 - \beta_1^2) \sum_i q_{33}(h_1) p_{11} + \beta_1^2 \) are identified by Steps 1 and 2.

Step 4: \( (\alpha_1, \beta_1, \sum_i q_{33}(h_1) p_{11}) \) are identified. Fix \( h_1 \). I now show that the system in (44) admits a unique solution for \( (\alpha_1, \beta_1, \sum_i q_{33}(h_1) p_{11}) \) if \( \alpha_1 > \beta_1 \). To start, using the facts 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 (44) as

\[
\begin{align*}
(\alpha_1 - \beta_1) \sum_i q_{33}(h_1) p_{11} &= \nu_1(h_1) - \beta_1 \\
(\alpha_1 + \beta_1)(\nu_1(h_1) - \beta_1) &= \nu_2(h_1) - \beta_1^2 \iff \nu_1(h_1)(\alpha_1 + \beta_1) - \alpha_1 \beta_1 = \nu_2(h_1) \\
(\alpha_1 + \beta_1)(\nu_2(h_1) - \beta_1^2) - \alpha_1 \beta_1(\nu_1(h_1) - \beta_1) &= \nu_3(h_1) - \beta_1^3 \iff \frac{\nu_2(h_1)(\alpha_1 + \beta_1)}{\nu_1(h_1)} - \alpha_1 \beta_1 = \frac{\nu_3(h_1)}{\nu_1(h_1)}.
\end{align*}
\]

By subtracting the third equation from the second equation in (49), side by side, I obtain

\[ \alpha_1 + \beta_1 = \frac{\nu_1(h_1)v_2(h_1) - \nu_3(h_1)}{\nu_1^2(h_1) - \nu_2(h_1)}. \]

Substituting this last expression into the second equation in (49) gives

\[ \alpha_1 \beta_1 = \nu_1(h_1)(\alpha_1 + \beta_1) - \nu_2(h_1) = \frac{\nu_2^2(h_1) - \nu_1(h_1)\nu_3(h_1)}{\nu_1^2(h_1) - \nu_2(h_1)}. \]

Using (50), (51), the fact 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

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

By summing (50) and (52) side by side, it follows that

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

Substituting this latter expression for \( \alpha_1 \) into (50) or (51) provides an analogous expression for \( \beta_1 \). So, \( \alpha_1 \) and \( \beta_1 \) are identified. Plugging the expressions for \( \alpha_1 \) and \( \beta_1 \) into the first equation of (49) gives an expression for \( \sum_i q_{33}(h_1) p_{11} \) that only depends on \( \nu_1(h_1), \nu_2(h_1), \) and \( \nu_3(h_1) \). Thus, \( \sum_i q_{33}(h_1) p_{11} \) is also identified for a given \( h_1 \) and, by (46), is identified for each \( h_1 \) from \( \Pr(R_{A3}^0 = 1 | L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1) \) by varying \( h_1 \), once \( \alpha_1, \beta_1, \) and classification error are identified.

Step 5: \( (\alpha_k, \beta_k), k \geq 2, \) are identified. The argument in the previous steps can easily be adapted to show that

49
\{\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 show that the support of the initial prior, \(\{p_{i1}\}\), is identified from the distribution of recorded 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 \(t\), is

\[
\nu_{1t}(h_1) = \Pr(R_{At} = 1|L_{A1} = 1, \ldots, L_{At} = 1, h_1) = (\alpha_1 - \beta_1) \sum_i \Pr(i|L_{A1} = 1, \ldots, L_{At} = 1, h_1)p_{i1} + \beta_1
\]

for a manager with initial human capital \(h_1\). It is also easy to see that \(\nu_{1t}(h_1)\) is identified in each \(t\) from the known probability \(\Pr(R_{At} = 1|L_{A1} = 1, \ldots, L_{At} = 1, h_1)\) by an argument analogous to the one in Part II of the proof of Proposition 3 used to derive (46) and to argue that \(\nu_{1t}(h_1) = (\alpha_1 - \beta_1) \sum q_{3t}(h_1)p_{i1} + \beta_1\) is identified from \(\Pr(R_{A3} = 1|L_{A1} = 1, L_{A2} = 1, L_{A3} = 1, h_1)\), once classification error parameters are identified. Recall also from Part II of the proof of Proposition 3 that \(\alpha_1, \beta_1\), and \(\{\sum q_{3t}(h_1)p_{i1}\}\) for each \(h_1\) are identified. Then, \(\{\sum q_{3t}(h_1)p_{i1}\}\) are identified for each \(h_1\) also in \(t \geq 4\) by (53).

By the proof of Lemma 1, each probability in \(\{\Pr(\nu_{1t}(h_1) = 1, \ldots, L_{At} = 1, h_1, i)\}\) is identified and can be uniquely paired with each corresponding probability in \(\{\Pr(\nu_{1t}(h_1) = 1, h_1, i)\}\) for any \(t\). Note now that I can express \(\sum q_{it}(h_1)p_{i1} = \nu_{1t}(h_1) - \beta_1\) / \((\alpha_1 - \beta_1)\), which is derived from (53), in matrix form from \(t=3\) to \(t=6\) as

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

where \(q_{it}(h_1) = \Pr(L_{A1} = 1, \ldots, L_{At} = 1, h_1, i) / \Pr(L_{A1} = 1, \ldots, L_{At} = 1, h_1)\) satisfies

\[
q_{it}(h_1) = \frac{\Pr(L_{At} = 1|L_{A1} = 1, \ldots, L_{At-1} = 1, h_1, i) \Pr(i|L_{A1} = 1, \ldots, L_{At-1} = 1, h_1)}{\Pr(L_{At} = 1|L_{A1} = 1, \ldots, L_{At-1} = 1, h_1)} = \frac{\Pr(L_{At} = 1|L_{A1} = 1, \ldots, L_{At-1} = 1, h_1)}{\Pr(L_{At} = 1|L_{A1} = 1, \ldots, L_{At-1} = 1, h_1)} q_{it-1}(h_1),
\]

\(t > 1\). 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

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

For this condition to hold with \(q_{33}(h_1) > 0\) for each \(i\), it must be that \(\psi_1 = -\psi_2 \pi_{i4}(h_1)\) for each \(i\), which can hold only if \(\psi_1\) and \(\psi_2\) are zero given that \(\pi_{i4}(h_1)\) 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 \( \{ \Pr(i|L_{A1} = 1, \ldots, L_{At}, h_1) \} \) identified by Lemma 1 and \( \{ \alpha_k, \beta_k \} \) and classification error rates at the corresponding levels identified by Proposition 3, a similar argument would apply with four periods of observations on performance ratings at Levels 2 or 3.

**Observation 2:** With \( \{ \Pr(i|L_{A1} = 1, \ldots, L_{At}, h_1) \} \) identified by Lemma 1 and \( \{ \alpha_1, \beta_1 \} \) and classification error rates at Level 1 identified by Proposition 3, \( \{ \alpha_2, \beta_2 \} \) and \( \{ \alpha_3, \beta_3 \} \) are also identified from the analogue of (53) either for managers observed at Levels 2 and 3, respectively, for two periods or for managers with two distinct values of \( h_1 \) observed at Levels 2 and 3, respectively, for one period whose average priors \( \sum_i q_{it}(h_1)p_{i1} \) for different values of \( h_1 \) are identified by the argument in Proposition 3.

**Proof of Lemma 2:** Recall that \( \eta_{kt} = \eta_{fk}(\kappa_i) \). Here, first I prove that if firm \( A \) faces a sufficiently high cost to rehiring a previously separated manager, then its value of profits from any separating manager is \( \Pi^A(\cdot | C) = 0 \). Second, I derive a more convenient expression for \( S(s_{it}, \epsilon_t) \) defined in Proposition 1 in light of this result.

To this purpose, consider first states at which firm \( A \) employs the manager—along the continuation game of interest here, in which the manager has been employed by firm \( A \) at some previous state. With \( S(\cdot | C) = \Pi^A(\cdot | C) + V^C(\cdot | C) \) by
construction, $\Pi^A(\cdot|C) = 0$ as established, and $V^C(\cdot|C) = \nabla^C(\cdot)$ by (29), it follows that $S(\cdot|C) = \nabla^C(\cdot)$. Hence, firm $C$'s assignment decision achieves the value $S(\cdot|C)$. Since employment at firm $C$ is efficient conditional on firm $C$’s assignment decision by Proposition 1, it must be that

$$\nabla^C(\cdot) \geq \max_{k \in K} \left\{ y_A(s_{it}, k) + \varepsilon_{Akt} + \delta \eta_{kt} \int_{\varepsilon_{t+1}} ES(s_{it+1}, \varepsilon_{t+1}|s_{it}, k) dG \right\}. \tag{59}$$

Therefore, at all equilibrium states, $S(s_{it}, \varepsilon_t)$ satisfies

$$S(s_{it}, \varepsilon_t) = \max_f \left\{ \max_{k \in K^f} \left\{ y_f(s_{it}, k) + \varepsilon_{fkt} + \delta \eta_{kt} \int_{\varepsilon_{t+1}} ES(s_{it+1}, \varepsilon_{t+1}|s_{it}, k) dG \right\} \right\}. \tag{60}$$

In particular, since (58) holds at all states in which $A$ employs and (59) holds at all states in which $C$ employs, (56) holds. With Gumbel-distributed productivity shocks, by standard arguments as in Rust (1987, 1994), one can show that

$$S(s_{it}, \varepsilon_t) = \max \left\{ \max_{k \in K^A} \left\{ \zeta^A(s_{it}, k) + \varepsilon_{Akt} \right\}, \max_{k \in K^C} \left\{ \zeta^C(s_{it}, k) + \varepsilon_{Ckt} \right\} \right\}, \tag{61}$$

$$\zeta^f(s_{it}, k) = y_f(s_{it}, k) + \delta \eta_{kt} E \left[ \ln \left( \sum_{k' \in K^A} e^{\zeta^A(s_{it+1}, k')} + \sum_{k' \in K^C} e^{\zeta^C(s_{it+1}, k')} \right) | s_{it}, k \right], \tag{62}$$

$$\nabla^f(s_{it}, \varepsilon_t) = \max_{k \in K^f} \left\{ \zeta^f(s_{it}, k) + \varepsilon_{fkt} \right\}, \tag{63}$$

and $\zeta^f(s_{it}, k) = y_f(s_{it}, k) + \delta \eta_{kt} E \left[ \ln \sum_{k' \in K} e^{\zeta^f(s_{it+1}, k')} | s_{it}, k \right]$, which will prove useful in the proof of Proposition 5.

**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 possible level and tenure, conditional on past assignments and true and recorded performance. In Part II, I show that exogenous separation rates, $\{\eta_{kt}\}$, and the parameters of expected output at firm $A$, namely, $\{b_{A_j}(k), d_{Akt}(k_{-1}), \varepsilon_{Akt}\}$, are identified.

**Part I: Identification of conditional level assignment probabilities at firm $A$ by skill type**. Here I show that the conditional probabilities of level assignment $\Pr(L_{A2}|L_{A1}=1, R_{A1}, h_{1}, i)$, $\Pr(L_{A3}|L_{A1}=1, R_{A1}, L_{A2}, R_{A2}, h_{1}, i)$, and so on are identified at each possible level and tenure from joint information on: 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; and ii) the probabilities of true performance, $\{\alpha_k, \beta_k\}$, which are identified by Proposition 3.

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 (37) for any given $L_{A2}, h_{1}$, and $i$. Also, the probabilities $\{\Pr(i|L_{A1} = 1, \ldots, L_{A1}, h_{1})\}$ are identified by Lemma 1. 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 (42) 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 only need to 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.

To see that this is the case, note that $\Pr(R_{A1} = 1|L_{A1} = 1, h_{1}, i) = (\alpha_1 - \beta_1) p_{i1} + \beta_1$, $\Pr(R_{A2} = 1|L_{A1} = 1, L_{A1}, L_{A2}, h_{1}, i) = (\alpha_{L_{A2}} - \beta_{L_{A2}}) p_{i2} + \beta_{L_{A2}}$, 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 $\{p_{i2}\}$ are known for any given sequence of assignments and true performance by Bayes’ rule; see (4). Thus, the probabilities $\{\Pr(R_{A1}|L_{A1} = 1, h_{1}, i)\}$, $\{\Pr(L_{A2}|L_{A1} = 1, R_{A1}, L_{A2}, h_{1}, i)\}$, and so on are known.

Now fix $i$ and $R_{A1} = 1$ in the product of probabilities $\Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_{1}, i) \Pr(R_{A1}|L_{A1} = 1, h_{1}, i)$ $\Pr(i|L_{A1} = 1)$ 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 (40) starting 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 $h_{1}$. Specifically, the product of each probability on the right side of (40) and the associated probability
\(\Pr(L_{A2}, L_{A3}|L_{A1} = 1, h_1)\) gives the identified expression

\[
\Pr(L_{A3}|L_{A1} = 1, R_{A1}, L_{A2}, R_{A2}, h_1, i) \Pr(R_{A2}|L_{A1} = 1, R_{A1}, L_{A2}, h_1, i) \\
\cdot \Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1} = 1, h_1, i) \Pr(i|L_{A1} = 1),
\]

(64)

which can be correctly labeled with respect to \(i\) for any given \(R_{A1}, L_{A2}, R_{A2}, L_{A3}\), and \(h_1\) by the argument after (40). Once these probabilities are correctly labeled with respect to \(R_{A1}\) and \(R_{A2}\), taking the ratio of (64) with, say, \(R_{A1} = 1\) and \(R_{A2} = 1\) to the product of probabilities \([(\alpha_{L_{A2}} - \beta_{L_{A2}})p_{i2} + \beta_{L_{A2}}] \Pr(L_{A2}|L_{A1} = 1, R_{A1} = 1, h_1, i) [(\alpha - \beta)p_{11} + \beta_1] \Pr(i|L_{A1} = 1)\) 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 combination of values for \(R_{A1}\) and \(R_{A2}\).

Hence, what is left to show is that the probabilities in (64) derived from the mixture weights \(\{\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\}\) of the identified mixture distribution in (39) 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 \(\{\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\}\) with the corresponding one in \(\{\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, R_{A2}^0, L_{A3}, h_1)\}\). Such pairing of mixture weights is possible since these two sets of weights of the wage mixtures \(f(w_3|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\) and \(f(w_3|L_{A1} = 1, L_{A2}, R_{A2}^0, L_{A3}, h_1)\) are associated with the same component densities. From the ratios of these paired mixture weights, I can then recover \(\{\Pr(R_{A2}^0 = 1|L_{A1} = 1, L_{A2}, R_{A2})\}\) for given \(L_{A2}\). Since \(\{\Pr(R_{A2}^0 = 1|L_{A1} = 1, L_{A2}, R_{A2})\}\) are unambiguously ordered with respect to \(R_{A2}\) given \(L_{A2}\) by (11) and (12) with \(p_1 > 0\), as shown in (43), it is thus possible to determine the value of \(R_{A2}\) in each probability in \(\{\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\}\).

To correctly label any mixture weight \(\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\) 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 analogously. Namely, I infer the value of \(R_{A1}\) from the associated probability of a recorded high rating in \(t = 1\) conditional on \(R_{A1}\), \(\Pr(R_{A1}|L_{A1} = 1, R_{A1})\), which is unambiguously ordered with respect to \(R_{A1}\) by (11) and (12) since \(p_1 > 0\). To this purpose, I first need to associate \(\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\) with the corresponding probability \(\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, R_{A1}, L_{A2}, L_{A3}, h_1)\) to recover the probability \(\Pr(R_{A1}^0|L_{A1} = 1, R_{A1})\) by an argument analogous to the one 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 \(h_1\) at level \(L_{A2}\) in \(t = 2\), given by \(f(w_{A3}|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\) in (39), is an identified mixture with weights given by (40). Similarly, the density of wages at Level \(L_{A3}\) in \(t = 3\) for managers with initial human capital \(h_1\) at level \(L_{A2}\) in \(t = 2\) with recorded rating \(R_{A1}^0\) in \(t = 1\) is an identified mixture given by

\[
f(w_{A3}|L_{A1} = 1, R_{A1}^0, L_{A2}, L_{A3}, h_1) = \sum_{R_{A1}, R_{A2}, i} f(w_{A3}|L_{A1} = 1, R_{A1}^0, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1, i) \\
\cdot \Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, R_{A1}^0, L_{A2}, L_{A3}, h_1),
\]

where \(f(w_{A3}|L_{A1} = 1, R_{A1}^0, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1, i)\) is independent of \(R_{A1}\) and

\[
\Pr(R_{A1}, R_{A2}, i|L_{A1} = 1, R_{A1}^0, L_{A2}, L_{A3}, h_1) = \frac{\Pr(L_{A3}|L_{A1} = 1, R_{A1}, L_{A2}, R_{A2}, h_1, i) \Pr(R_{A2}|L_{A1} = 1, R_{A1}, L_{A2}, h_1, i) \\
\cdot \Pr(L_{A2}|L_{A1} = 1, R_{A1}, h_1, i) \Pr(R_{A1}|L_{A1} = 1, h_1, i) \Pr(i|L_{A1} = 1)}{\Pr(R_{A1}^0, L_{A2}, L_{A3}|L_{A1} = 1, h_1)},
\]

(65)

since \(\Pr(L_{A3}|i), \Pr(R_{A2}|i),\) and \(\Pr(L_{A2}|i)\) are independent of \(R_{A1}^0\).

With \(f(w_{A3}|L_{A1} = 1, R_{A1}^0, R_{A1}, L_{A2}, R_{A2}, L_{A3}, h_1, i)\) independent of \(R_{A1}^0\), the component wage densities associated with the two mixtures considered, \(f(w_{A3}|L_{A1} = 1, L_{A2}, L_{A3}, h_1)\) and \(f(w_{A3}|L_{A1} = 1, R_{A1}^0, L_{A2}, L_{A3}, h_1)\), are identical for any \(R_{A1}^0\), given \(L_{A2}, L_{A3}, h_1, R_{A1}, R_{A2},\) and \(i\). Then, the weights in (40) and (65) can be uniquely matched through their corresponding component wage densities for each \(R_{A1}, R_{A2},\) and \(i\), given \(R_{A1}^0, L_{A2}, L_{A3},\) and \(h_1\).

Finally, the value of \(R_{A1}\) in the mixture weights in (40) and (65) can be pinned down as follows. First, compute the product of the term on the right side of (65) and the corresponding \(\Pr(R_{A1}^0, L_{A2}, L_{A3}|L_{A1} = 1, h_1)\). Next, compute the product of the term on the right side of (40) and the corresponding \(\Pr(L_{A2}, L_{A3}|L_{A1} = 1, h_1)\). Then, take the ratio of the two resulting products for given \(R_{A1}, R_{A2}, i, R_{A1}^0, L_{A2}, L_{A3},\) and \(h_1\), which gives \(\Pr(R_{A1}^0|L_{A1} = 1, R_{A1}, h_1, i)\). Recall that this latter probability is independent of \(h_1\) and \(i\) by (11) and (12). Thus, by comparing \(\Pr(R_{A1}^0|L_{A1} = 1, R_{A1})\)
for given $R^0_{A1}$ across the two possible values of $R_A$ and using (43), it is immediate to determine the value of $R_A$ in $Pr(R_{A1}, R_{A2}, t|L_{A1} = 1, L_{A2}, L_{A3}, h_1)$.

This argument can be extended to the remaining tenures and proves that the component distributions of the mixture distribution of wages at firm $A$ of managers of each possible skill type, conditional on their histories of level assignments and performance ratings at firm $A$, and their mixture weights can be correctly labeled with respect to $R_{At-1}, \ldots, R_{At-1}$ at each possible level and tenure.

**Part II: Identification of separation rates and output parameters.** The proof of this part of the argument uses the assignment probabilities at $A$ recovered in Part I. Before proceeding, first note that since conditional 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 any manager skill type is also identified for any sequence of level assignments, (unobserved) performance, and any initial level of human capital, given that such a probability is given by $Pr(L_{At} > 0|L_{A1} = 1, R_{A1}, \ldots, R_{At-1}, h_1, i) = \sum_{k \geq 1} Pr(L_{At} = k|L_{A1} = 1, R_{A1}, \ldots, L_{At-1}, R_{At-1}, h_1, i)$. For any given history at the firm, the probability of separation for each skill type is simply the complementary probability.

I now show how these probabilities are related to the values $\zeta^A(s_{it}, k)$ and $\zeta^C(s_{it}, k)$, $k \geq 1$, derived in Lemma 2, since this mapping is key to identifying the parameters of interest. Specifically, using the definition of $\gamma(s_{it}, k)$ in (20), (62) derived in Lemma 2, and that each Level $k \geq 1$ of firms $A$ and $C$ implies the same law of motion of the state across firms (but not across levels), I can express $\zeta^A(s_{it}, k)$ as

$$\zeta^A(s_{it}, k) = y_C(s_{it}, k) - \gamma(s_{it}, k) + \delta \eta_{kt} E[\ln(\cdot)|s_{it}, k] = \zeta^C(s_{it}, k) - \gamma(s_{it}, k). \tag{66}$$

Note that $Pr(L_{At} > 0; h_1, i)$, denoted more compactly by $Pr(L_{At} > 0|s_{it})$, and $Pr(L_{At} = k|s_{it})$ are given by

$$Pr(L_{At} > 0|s_{it}) = \frac{\eta_{kt-1} \sum_{k' \geq 1} e^{A(s_{it}, k')}}{e^{A(s_{it}, k')} + \sum_{k' \geq 1} e^{C(s_{it}, k')}} \quad \text{and} \quad Pr(L_{At} = k|s_{it}) = \frac{\eta_{kt-1} e^{A(s_{it}, k)}}{e^{C(s_{it}, k')} + \sum_{k' \geq 1} e^{C(s_{it}, k')}}, \tag{67}$$

$t \geq 2, k_{t-1}, k_1$. Denote by $p_f(s_{it}, k)$ the probability of assignment to Level $k$ conditional on employment at $f = A, C$,

$$p_f(s_{it}, k) = Pr(L_{ft} = k|f_t = f, s_{it}) = \frac{e^{C(s_{it}, k)}}{\sum_{k' \geq 1} e^{f(s_{it}, k')}} \tag{68}$$

which is identified for each possible $s_{it}$ and $k$, since it is equal to $Pr(L_{At} = k|s_{it})/Pr(L_{At} > 0|s_{it})$ and this probability ratio is identified by Part I. I now turn to the proof of Part II, which establishes that the parameters of interest are identified in four steps. Recall that $s_{it} = (p_{it}, h_1, t-1, k_{t-1}, i)$.

**Step 1: Separation rates at Level 1.** First, note that the ratio $Pr(L_{Ct} = k|s_{it})/Pr(L_{At} = k|s_{it})$ can be expressed as

$$Pr(L_{Ct} = k|s_{it}) = \frac{\eta_{kt-1} e^{C(s_{it}, k)}}{e^{C(s_{it}, k')} + \sum_{k' \geq 1} e^{C(s_{it}, k')}} = \frac{e^{C(s_{it}, k)}}{e^{C(s_{it}, k')} + \sum_{k' \geq 1} e^{C(s_{it}, k')}} = \frac{e^{C(s_{it}, k)}}{e^{A(s_{it}, k')} + \sum_{k' \geq 1} e^{C(s_{it}, k')}}$$

where the last equality holds by (66), so that $y_C(s_{it}, k) - y_A(s_{it}, k) = \ln Pr(L_{Ct} = k|s_{it}) - \ln Pr(L_{At} = k|s_{it})$. Then,

$$y_A(s_{it}, k) = y_C(s_{it}, k) - \ln Pr(L_{Ct} = k|s_{it}) + \ln Pr(L_{At} = k|s_{it}) + \ln Pr(L_{Ct} > 0|s_{it}) - \ln Pr(L_{Ct} > 0|s_{it})$$

$$= y_C(s_{it}, k) - \ln \left[ \frac{Pr(L_{Ct} = k|s_{it})}{Pr(L_{Ct} > 0|s_{it})} \right] - \ln Pr(L_{Ct} > 0|s_{it}) + \ln Pr(L_{At} = k|s_{it})$$

$$= y_C(s_{it}, k) - \ln p_C(s_{it}, k) - \ln \left[ \frac{Pr(L_{Ct} > 0|s_{it}) + 1 - \eta_{kt-1} - 1 + \eta_{kt-1}}{Pr(L_{At} = k|s_{it})} \right]$$

$$= \eta_{kt-1} - \ln \frac{Pr(L_{At} = k|s_{it})}{Pr(L_{Ct} = k|s_{it})}, \tag{69}$$

where the first equality follows by summing and subtracting $\ln Pr(L_{Ct} > 0|s_{it})$, the second equality follows by collecting and rearranging terms, the third equality follows by using that $p_C(s_{it}, k) = Pr(L_{Ct} = k|s_{it})/Pr(L_{Ct} > 0|s_{it})$ by definition of $p_C(s_{it}, k)$ and by summing and subtracting $1 - \eta_{kt-1}$, and the fourth equality follows by using that $Ew_Ait = y_C(s_{it}, k) - \ln p_C(s_{it}, k)$ by (13), that $Pr(L_{At} = 0|s_{it}) = Pr(L_{Ct} > 0|s_{it}) + 1 - \eta_{kt-1}$ by definition, and by
rearranging terms. When \( k_{t-1} = 1 \), \((69)\) implies that

\[
\eta_{1t} \equiv \Pr(L_{At} > 0 | s_{it}) + \Pr(L_{At} = k | s_{it}) e^{Ew_{At} - y_{A}(s_{it}, k)}.
\]

If \( y_{C}(s_{it}, 1) \) is known at one state with \( k_{t-1} = 1 \) in \( t \geq 2 \) and \( y_{A}(s_{it}, 1) = y_{C}(s_{it}, 1) \) at that state in \( t \geq 2 \), then \( \eta_{1t-1} \), \( t \geq 2 \), is identified, since \( \Pr(L_{At} > 0 | s_{it}) \) and \( \Pr(L_{At} = k | s_{it}) \) are identified by Part I and \( Ew_{At} \) by Lemma 1.

Step 2: Firm A’s output parameters at Level 1. Since \( \eta_{1t-1} \), \( t \geq 2 \), is identified by Step 1, it is immediate to identify \( y_{A}(s_{it}, 1), t \geq 2 \), at each possible \( s_{it} \) with \( k_{t-1} = 1 \) except for one such \( s_{it} \): one point in the support of \( s_{it} \) with \( k_{t-1} = 1 \) in \( t \geq 2 \) identifies \( \eta_{1t-1} \) by Step 1. Specifically, observe that at states with \( k_{t-1} = k = 1 \), \((69)\) implies that

\[
y_{A}(s_{it}, 1) = Ew_{At}(k_{t-1} = 1) - \ln \left[ \eta_{1t-1} \Pr(L_{At} > 0 | s_{it}) \right],
\]

where \( Ew_{At}(k_{t-1} = 1) \) is the average wage of managers of skill type \( i \) at Level 1 at state \( s_{it} \) with \( k_{t-1} = 1 \). Since \( Ew_{At}(k_{t-1} = 1) \) is identified by Lemma 1 and \( \Pr(L_{At} > 0 | s_{it}) \) and \( \Pr(L_{At} = 1 | s_{it}) \) are identified by Part I at each possible \( s_{it} \), it follows that \( y_{A}(s_{it}, 1) \) is identified. Then, the \( J + 2 \) parameters of \( y_{A}(s_{it}, 1) \) in each \( t \geq 2 \) when \( k_{t-1} = 1 \) are identified from \( J + 2 \) equations of the form

\[
\sum_{j=1}^{J} b_{Aj}(1) h_{1j} + d_{A1t}(L1) + e_{A1t} p_{it} = y_{A}(s_{it}, 1),
\]

which exist and are linearly independent by \((A1)\). Then, \( \{b_{Aj}(1)\}, d_{A1t}(L1), \) and \( e_{A1t}, t \geq 2 \), are identified. Recall that \( p_{it} \) is identified, since \( \{a_{k}, \beta_{k}\} \) and \( \{p_{1i}\} \) are identified by Propositions 3 and 4. To see that the required support points for \( p_{it} \) exist, observe that in each \( t \geq 2 \) there exist as many points in the support of \( s_{it} \) with different values of \( p_{it} \) as possible sequences of high and low performance for each skill type, which clearly amount to more than the required values of \( p_{it} \) in each \( t \geq 2 \).

Step 3: Firm A’s output parameters and separation rates at Level 2. It is immediate that \( y_{A}(s_{it}, 2) \) is identified in \( t \geq 2 \) at each possible \( s_{it} \) with \( k_{t-1} = 1 \), by \((69)\)

\[
y_{A}(s_{it}, 2) = Ew_{At}(k_{t-1} = 1) - \ln \left[ \eta_{1t-1} \Pr(L_{At} > 0 | s_{it}) \right],
\]

Here, \( Ew_{At}(k_{t-1} = 1) \) is the average wage of managers of skill type \( i \) at Level 2 at state \( s_{it} \) with \( k_{t-1} = 1 \) and is identified by Lemma 1, \( \eta_{1t-1}, t \geq 2 \), is identified by Step 1, and \( \Pr(L_{At} > 0 | s_{it}) \) and \( \Pr(L_{At} = 2 | s_{it}) \) are identified by Part I. Then, the \( J + 2 \) parameters of \( y_{A}(s_{it}, 2) \) in each \( t \geq 2 \) when \( k_{t-1} = 1 \) are identified from \( J + 2 \) equations of the form

\[
\sum_{j=1}^{J} b_{Aj}(2) h_{1j} + d_{A2t}(L1) + e_{A2t} p_{it} = y_{A}(s_{it}, 2),
\]

which exist and are linearly independent by \((A1)\). So, the parameters \( \{b_{Aj}(2)\}, d_{A2t}(L1), \) and \( e_{A2t}, t \geq 2 \), are identified. Now, evaluating \((69)\) with \( k = 2 \) at two states \( s'_{it} \) and \( s''_{it} \) with \( k_{t-1} = 2, t \geq 3 \), that differ only in their values of \( p_{it} \) gives

\[
\begin{align*}
&\left\{ b_{A2}(h_{1}) + d_{A2t}(L2) + e_{A2t} p'_{it} = Ew'_{At}(k_{t-1} = 2) - \ln \left[ \eta_{2t-1} \Pr(L_{At} > 0 | s'_{it}) \right], \\
&\left\{ b_{A2}(h_{1}) + d_{A2t}(L2) + e_{A2t} p''_{it} = Ew''_{At}(k_{t-1} = 2) - \ln \left[ \eta_{2t-1} \Pr(L_{At} > 0 | s''_{it}) \right],
\end{align*}
\]

where \( Ew'_{At}(k_{t-1} = 2) \) and \( Ew''_{At}(k_{t-1} = 2) \) are evaluated, respectively, at \( s'_{it} \) and \( s''_{it} \), so that

\[
e_{A2t}(p'_{it} - p''_{it}) - Ew'_{At}(k_{t-1} = 2) + Ew''_{At}(k_{t-1} = 2) = \ln \left[ \eta_{2t-1} \Pr(L_{At} > 0 | s''_{it}) \right] \cdot \Pr(L_{At} = 2 | s''_{it}) + \Pr(L_{At} = 2 | s'_{it}) - \Pr(L_{At} > 0 | s''_{it}) \cdot \Pr(L_{At} = 2 | s''_{it}) - \Pr(L_{At} > 0 | s'_{it}) \cdot \Pr(L_{At} = 2 | s'_{it})
\]

55
or, equivalently,
\[
\ln \left[ \frac{\eta_{t-1} - \Pr(L_{At} > 0|s'_{it})}{\eta_{t-1} - \Pr(L_{At} > 0|s''_{it})} \right] = e_{A2t}(p'_it - p''_it) - Ew'_{At}(k_{t-1} = 2) + Ew''_{At}(k_{t-1} = 2) - \ln \left[ \frac{\Pr(L_{At} = 2|s'_{it})}{\Pr(L_{At} = 2|s''_{it})} \right]
\]
with \(A_2(s'_{it}, s''_{it})\) identified. Hence, \(\eta_{t-1}\) is identified in \(t \geq 3\) as
\[
\eta_{t-1} = \frac{\Pr(L_{At} > 0|s'_{it}) - \Pr(L_{At} > 0|s''_{it}) \exp\{A_2(s'_{it}, s''_{it})\}}{1 - \exp\{A_2(s'_{it}, s''_{it})\}}.
\]

Either expression in the system in (72) identifies \(d_{A2t}(L2)\) in each \(t \geq 3\).

**Step 4: Firm A’s output parameters and separation rates at Level 3.** It is easy to see that \(y_A(s_{it}, 3)\) is identified in \(t \geq 3\) at each possible \(s_{it}\) with \(k_{t-1} = 2\), since by (69)
\[
y_A(s_{it}, 3) = Ew_{At}(k_{t-1} = 2) - \ln \left[ \frac{\eta_{t-1} - \Pr(L_{At} > 0|s_{it})}{\Pr(L_{At} = 3|s_{it})} \right].
\]

Here, \(Ew_{At}(k_{t-1} = 2)\) is the average wage of managers of skill type \(i\) at Level 3 at state \(s_{it}\) with \(k_{t-1} = 2\) and is identified by Lemma 1, \(\eta_{t-1}, t \geq 3\), is identified by Step 3, and \(\Pr(L_{At} > 0|s_{it})\) and \(\Pr(L_{At} = 3|s_{it})\) are identified by Part I. Then, the \(J+2\) parameters of \(y_A(s_{it}, 3)\) in each \(t \geq 3\) when \(k_{t-1} = 2\) are identified from \(J+2\) equations of the form
\[
\sum_{j=1}^{J} b_{A3}(h_1) + d_{A3t}(L3) + e_{A3t}p_{it} = y_A(s_{it}, 3),
\]
which exist and are linearly independent by (A1). Hence, \(b_{A3}(h_1)\), \(d_{A3t}(L3)\), and \(e_{A3t}\), \(t \geq 3\), are identified. Proceeding analogously to Step 3 and evaluating (69) with \(k = 3\) at two states \(s'_{it}\) and \(s''_{it}\) now with \(k_{t-1} = 3, t \geq 4\), that differ only in their values of \(p_{it}\) provides the system
\[
\begin{align*}
\{ & b_{A3}(h_1) + d_{A3t}(L3) + e_{A3t}p_{it}' = Ew_{At}'(k_{t-1} = 3) - \ln \left[ \frac{\eta_{t-1} - \Pr(L_{At} > 0|s_{it})}{\Pr(L_{At} = 3|s_{it})} \right], \\
& b_{A3}(h_1) + d_{A3t}(L3) + e_{A3t}p_{it}'' = Ew_{At}''(k_{t-1} = 3) - \ln \left[ \frac{\eta_{t-1} - \Pr(L_{At} > 0|s_{it})}{\Pr(L_{At} = 3|s_{it})} \right],
\end{align*}
\]

where \(Ew_{At}'(k_{t-1} = 3)\) and \(Ew_{At}''(k_{t-1} = 3)\) are evaluated at the corresponding states. Now, (74) yields that
\[
\ln \left[ \frac{\eta_{t-1} - \Pr(L_{At} > 0|s_{it})}{\eta_{t-1} - \Pr(L_{At} > 0|s_{it})} \right] = e_{A3t}(p_{it}' - p_{it}'') - Ew_{At}'(k_{t-1} = 3) + Ew_{At}''(k_{t-1} = 3) - \ln \left[ \frac{\Pr(L_{At} = 3|s_{it})}{\Pr(L_{At} = 3|s_{it})} \right]
\]
with \(A_3(s'_{it}, s''_{it})\) identified. It then follows that
\[
\eta_{t-1} = \frac{\Pr(L_{At} > 0|s_{it}) - \Pr(L_{At} > 0|s_{it}) \exp\{A_3(s'_{it}, s''_{it})\}}{1 - \exp\{A_3(s'_{it}, s''_{it})\}},
\]
so that \(\eta_{t-1}\) is identified in each \(t \geq 4\). Either expression in the system in (74) identifies \(d_{A3t}(L3)\) in each \(t \geq 4\).

**Proof of Proposition 6:** To start, note that \(p_C(s_{it}, 1)\) defined in (68) can be expressed as
\[
p_C(s_{it}, 1) = \frac{e^{y_C(s_{it}, 1) - y_A(s_{it}, 1)} \Pr(L_{At} > 0|s_{it})}{\eta_{k_{t-1}t-1} - \Pr(L_{At} > 0|s_{it})} p_A(s_{it}, 1).
\]

Thus, \(p_C(s_{it}, 1)\) can be recovered from \(\Pr(L_{At} > 0|s_{it})\), \(\eta_{k_{t-1}t-1}\), and \(p_A(s_{it}, 1)\), which are identified as argued in the proof of Proposition 5, provided that the difference \(y_C(s_{it}, 1) - y_A(s_{it}, 1)\) is known, which is assumed.\(^{49}\) Next, observe
\[\]
\(^{49}\)Clearly, \(p_C(s_{it}, 1)\) is identified by (75) at states with \(k_{t-1} = 1\) as long as they are different from those used in the proof of Proposition 5 to recover \(\eta_{k_{t-1}}\) and \(y_A(s_{it}, 1)\). Note that \(p_A(s_{it}, 1)\) is also known at states reached after Levels 2 or 3 have been assigned, even though no
that \( p_C(s_{it}, k) = e^{C(s_{it}, k)} / \sum_{k'} e^{C(s_{it}, k')} \) can be expressed as

\[
p_C(s_{it}, k) = \frac{p_C(s_{it}, 1) e^{C(s_{it}, k)}}{e^{C(s_{it}, 1)}} = \frac{p_C(s_{it}, 1) e^{A(s_{it}, k) + \gamma(s_{it}, k)}}{e^{A(s_{it}, 1) + \gamma(s_{it}, 1)}} = \frac{p_C(s_{it}, 1) p_A(s_{it}, k) e^{\gamma(s_{it}, k) - \gamma(s_{it}, 1)}}{p_A(s_{it}, 1)},
\]

(76)

where the first and third equalities follow from the definition of \( p_f(s_{it}, k), f = A, C, \) and the second equality holds by the definition of \( \gamma(s_{it}, k) \). By using (76), the average wage of managers of skill type \( i \) assigned to Level \( k \) of firm \( A \) at state \( s_{it} \), \( E w_{Ait} = y_C(s_{it}, k) - \ln p_C(s_{it}, k) \) by (13), can be rewritten as

\[
E w_{Ait} = y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) + \ln \left( \frac{p_A(s_{it}, 1)}{p_C(s_{it}, 1) p_A(s_{it}, k)} \right).
\]

(77)

By Lemma 1 and Part I of the proofs of Propositions 3 and 5, \( \{E w_{Ait}\} \) are identified at each possible state and job level without any label ambiguity. With \( \{p_C(s_{it}, 1)\} \) identified as just argued and \( \{p_A(s_{it}, k)\} \) identified by Part I of the proof of Proposition 5, (77) implies that \( y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) \) is identified and, by (10) and (20), given by

\[
y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) = \sum_j \left[ \left( b_C(j) + b_C(j)(1 - b_A(j)(1 - b_C(j) - b_A(j))(k)) \right) h_{1j} \right.
\]

\[
+ d_C k(l_{k-1}, i) + d_C k(l_{k-1}, i) - d_A k(l_{k-1}) - [d_C k(l_{k-1}, i) - d_A k(l_{k-1})] + e_C k(l_{i}) + e_C k(l_{i}) - e_A k l - e_A k l p_l.
\]

The parameters \( \varphi_j(k), \varphi_d k(l_{k-1}, i), \) and \( \varphi_e k(l_{i}) \) are identified if: \( i \) there exist \( J \) equations for each possible \( k \) to recover \( \varphi_j(k), \) one equation for each possible \( k, t, k_{t-1}, \) and \( i \) to recover \( \varphi_d(k_{t-1}, i), \) and one equation for each possible \( k, t, \) and \( i \) to recover \( \varphi_e(k_{t}, i) \) defined by (77) that are linearly independent; and \( ii \) the difference in expected output at Level \( 1 \) between firms \( A \) and \( C \) is known at all these states.

\[\square\]

**B Empirical Appendix**

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

**Details on Human Capital and Output Process:** Abstract first from skill types. I first 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) of the form

\[
H^j_{fkt} = C^j_{fkt}[(1 - \sigma^j)H^j_{fkt-1} + z^j_{fkt-1}(H^j_{fkt-1})^\lambda_j(i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] = C^j_{fkt}C^j_{fkt-1}(1 - \sigma^j)(i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}
\]

(78)

with \( H^j_{fkt} = C^j_{fkt} \) and \( \lambda_j \in \{0, 1\}, j = g, s \). I then derive the empirical specification of the human capital process in (22).

**Case 1:** \( \lambda_j = 0 \). If \( \sigma^j \in \{0, 1\} \), then (78) becomes \( H^j_{fkt} = C^j_{fkt}[(1 - \sigma^j)H^j_{fkt-1} + z^j_{fkt-1}(i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] \) so that

\[
H^j_{fkt2} = C^j_{fkt2}[(1 - \sigma^j)H^j_{fkt1} + z^j_{fkt1}(i^j_{k_{2}})^{\mu_{jk_{2}}}] = (1 - \sigma^j)C^j_{fkt2}C^j_{fkt1} + C^j_{fkt2}z^j_{fkt1}(i^j_{k_{2}})^{\mu_{jk_{2}}}
\]

\[
H^j_{fkt3} = C^j_{fkt3}[(1 - \sigma^j)H^j_{fkt2} + z^j_{fkt2}(i^j_{k_{3}})^{\mu_{jk_{3}}}] = C^j_{fkt3}C^j_{fkt2}(1 - \sigma^j)(i^j_{k_{3}})^{\mu_{jk_{3}}}
\]

and so on. With \( z^j_{fkt} = C^j_{fkt}C^j_{fkt-1} \cdots C^j_{fkt1} \), it follows that

\[
H^j_{fkt} = C^j_{fkt}C^j_{fkt-1} \cdots C^j_{fkt1}[(1 - \sigma^j)^{t-1} + (1 - \sigma^j)^{t-2}(i^j_{k_{1}})^{\mu_{jk_{1}}} + \ldots + (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] = C^j_{fkt}C^j_{fkt-1} \cdots C^j_{fkt1}.
\]

Thus, letting \( A^j_{fkt}C^j_{fkt} = C^j_{fkt}C^j_{fkt-1} \cdots C^j_{fkt1}, \) that is, \( \ln(A^j_{fkt}C^j_{fkt}) = \ln(A^j_{fkt-1}C^j_{fkt-1}) + \ln C^j_{fkt} \), I obtain

\[
H^j_{fkt} = A^j_{fkt}C^j_{fkt}[(1 - \sigma^j)^{t-1} + (1 - \sigma^j)^{t-2}(i^j_{k_{1}})^{\mu_{jk_{1}}} + \ldots + (i^j_{k_{t-1}})^{\mu_{jk_{t-1}}}] C^j_{fkt}.
\]

(79)
Consider first $H_{fkt}^g$. Since $z_{fkt-1}^g = t$ for managers employed in $t-1$, if $a_{fkt}^g = \ln A_{fkt}^g$, $\sigma^g = 0$, $\mu_{gkt-1} = \mu_g$ for $t > 1$, $\eta_g = \sum_{i=1}^g$, and $\epsilon_{fkt}^g = \ln \epsilon_{fkt}^g$, then the law of motion of $H_{fkt}^g$ can be expressed as

$$h_{fkt}^g = \ln H_{fkt}^g = \ln A_{fkt}^g + \ln[1 + \sum_{i=1}^g (t - 1)] + \ln \epsilon_{fkt}^g \simeq a_{fkt}^g + (t - 1)\eta_g + \epsilon_{fkt}^g.$$  

Consider now $H_{fkt}^s$. Since $z_{fkt-1}^s = t_{k-1}$ for managers employed in $t-1$ at job $k_{t-1}$, if $\sigma^s = 1$ and $\epsilon_{fkt}^s = C_{fkt}^s \cdots C_{fkt}^s (t_{k-1})^{\mu_{skt-1}}$. With $A_{fkt}^s$, $\sigma^s = 1$, and $\epsilon_{fkt}^s = \ln \epsilon_{fkt}^s$, I obtain

$$h_{fkt}^s = \ln H_{fkt}^s = \ln A_{fkt}^s + \mu_{skt-1} \ln t_{k-1} + \ln \epsilon_{fkt}^s \simeq a_{fkt}^s + \eta_{skt-1} + \epsilon_{fkt}^s.$$  

(80)

**Case 2**: $\lambda_j = 1$. In this case, if $\sigma^j \in [0, 1)$, then (78) becomes $H_{fkt}^j = C_{fkt}^j [1 - \sigma^j + \frac{z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}}{\mu_{jk}^s}] H_{fkt-1}^j$. Thus,

$$H_{fkt}^j = C_{fkt}^j [1 - \sigma^j + z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}] H_{fkt-1}^j = C_{fkt}^j [1 - \sigma^j + z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}] \cdots H_{fkt}^j = C_{fkt}^j C_{fkt}^j [1 - \sigma^j + z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}] \cdots [1 - \sigma^j + z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}].$$

As in the previous case, with $A_{fkt}^j = C_{fkt}^j \cdots C_{fkt}^j$, it follows that

$$H_{fkt}^j = A_{fkt}^j [1 - \sigma^j + z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}] \cdots [1 - \sigma^j + z_{fkt-1}^j (i_{k-1}^j)^{\mu_{jk-1}}] \epsilon_{fkt}^j.$$  

(81)

Consider first $H_{fkt}^g$. Since $z_{fkt-1}^g = t$ for managers employed in $t-1$, if $a_{fkt}^g = \ln A_{fkt}^g$, $\sigma^g = 0$, $z_{fkt-1}^g$ does not depend on $f$, $k$, or $t-1$, $\mu_{gkt-1} = \mu_g$ for $t > 1$, and $\epsilon_{fkt}^g = \ln \epsilon_{fkt}^g$, then the law of motion of $H_{fkt}^g$ can be expressed as

$$h_{fkt}^g = \ln H_{fkt}^g = a_{fkt}^g + (t - 1)\eta_g + \epsilon_{fkt}^g.$$  

as in the previous case, provided $\eta_g$ is suitably redefined as $\eta_g = \ln(1 + z_{fkt-1}^g \theta_g)$.

Consider now $H_{fkt}^s$. Recall that $i_{k-1}^s = t_{k-1}$ for managers employed in $t-1$ at job $k_{t-1}$. Suppose that $\sigma^s = 1$ and $z_{fkt-1}^s = C_{fkt-1}^s \cdots C_{fkt}^s / H_{fkt-1}^s$ so that (78) becomes $H_{fkt}^s = C_{fkt}^s \cdots C_{fkt}^s (t_{k-1})^\mu_{skt-1}$. Then, by the same derivations as in the previous case, $h_{fkt}^s$ can be expressed as (80).

Consider now skill types and denote by $a_{fkt}^j(i) = \ln A_{fkt}^j(i)$ the realized performance at Level $k$ of firm $f$ in $t$ of a manager of skill type $i$. By Cases 1 and 2, log human capital in (1) specializes to the expression (22), that is,

$$h_{fkt}^j = \begin{cases} 
  a_{fkt}^j(i) + (t - 1)\eta_g + \epsilon_{fkt}^j & \text{if } j = g \\
  a_{fkt}^j(i) + \eta_{skt-1} + \epsilon_{fkt}^s & \text{if } j = s.
\end{cases}$$  

(82)

**Derivation of Expected Output in (23):** From (2), $c_{fkt}^j h_{fkt}^j = c_{fkt}^j + c_{fkt}^j h_{fkt}^j$, $j = g, s$, and (82), it follows that

$$y_{fkt} = b_{fkt}(h_{fkt}^j) + c_{fkt}^j h_{fkt}^j + c_{fkt}^j h_{fkt}^j - \phi_{fkt}(k_{t-1}) = b_{fkt}(h_{fkt}^j) + c_{fkt}^j + c_{fkt}^j - \phi_{fkt}(k_{t-1})$$

$$+ c_{fkt}^j a_{fkt}^j(i) + c_{fkt}^j a_{fkt}^j(i) + c_{fkt}^j \eta_g \times (t - 1) + c_{fkt}^j \eta_{skt-1} + c_{fkt}^j \epsilon_{fkt}^s + c_{fkt}^j \epsilon_{fkt}^s.$$  

Using the fact that $\phi_{fkt}(k_{t-1}) = \phi_{fkt} + \phi_{fkt}(k_{t-1})$ and taking the expectation of $\{a_{fkt}^j(i)\}$, which take values $\{\bar{a}_{fkt}^j(i)\}$ with probability $\alpha_{fkt}$ or $\beta_{fkt}$ depending on a manager’s ability and zero otherwise, and of $\{\bar{\epsilon}^j_{fkt}\}$, I obtain (23), namely,

$$y_{fkt} = b_{fkt}(h_{fkt}^j) + c_{fkt}^j + c_{fkt}^s + \beta_k [c_{fkt}^j \bar{a}_{fkt}^j(i) + c_{fkt}^j \bar{a}_{fkt}^j(i)] + c_{fkt}^j \eta_g \times (t - 1) + c_{fkt}^j \eta_{skt-1} - \phi_{fkt} - \phi_{fkt}(k_{t-1})$$

$$+ (\alpha_k - \beta_k) [c_{fkt}^j \bar{a}_{fkt}^j(i) + c_{fkt}^j \bar{a}_{fkt}^j(i)] p_{it} = b_{fkt}(h_{fkt}^j) + d_{fkt}(k_{t-1}, i) + e_{fkt}(i) p_{it},$$

with $b_{fkt}(h_{fkt}^j) = \sum_m b_{ym} 1\{y_1 = m\} + b_{y1} e_1 + b_{y1} x_1 + b_{x1} x_1^2$ independent of $f$.

**Assignment Probabilities and Expected Output Parameters:** Here I discuss how I compute assignment probabilities.
at firm $A$ and the parameterization of $A$’s expected output. Note that $\Pr(L_{At} = k | s_{it})$ in (67) can be expressed as

$$
\Pr(L_{At} = k | s_{it}) = \frac{\eta_{kt}^{\ln (s_{it}, k)}}{e^{\ln \sum k' \geq 1 e^{C(s_{it}, k')}} + \sum k' \geq 1 e^{A(s_{it}, k')}} = \frac{e^{\ln \eta_{kt}^{t-1} + \psi_{A}(s_{it}, k) - \ln \sum k' \geq 1 e^{C(s_{it}, k')}}}{1 + \sum k' \geq 1 e^{A(s_{it}, k') - \ln \sum k' \geq 1 e^{C(s_{it}, k')}}}.
$$

(83)

Recall that $\varpi^f(s_{it}, e_t)$ from Proposition 1 is given by $\varpi^f(s_{it}, e_t) = \max_{k \geq 1} \{ \varpi^f(s_{it}, k) + e_{fkt} \}$ by (63), where

$$
\varpi^f(s_{it}, k) = y^f(s_{it}, k) + \delta \eta_{kt} \left[ \ln \sum k' \geq 1 e^{\varpi^f(s_{it+1}, k')} | s_{it}, k \right].
$$

(84)

By (56) and (60), $\psi^C(s_{it}, k) = \varpi^C(s_{it}, k)$, $k \geq 1$. With $p^f(s_{it}, k)$ defined in (68), observe that $\ln \sum k' e^{\psi^C(s_{it}, k')}$ satisfies

$$
\ln \sum k' e^{\varpi^C(s_{it}, k')} = y_C(s_{it}, 2) - \ln p_C(s_{it}, 2) + \delta \eta_{2t} E \left[ \ln \sum k' e^{\varpi^C(s_{it+1}, k')} | s_{it}, 2 \right],
$$

(85)

which is a functional equation that admits a unique solution and is obtained from

$$
\ln \sum k' e^{\varpi^C(s_{it}, k')} = \varpi^C(s_{it}, 2) - \ln p_C(s_{it}, 2) = y_C(s_{it}, 2) + \delta \eta_{2t} E \left[ \ln \sum k' e^{\varpi^C(s_{it+1}, k')} | s_{it}, 2 \right] - \ln p_C(s_{it}, 2),
$$

(86)

by using the definition of $p_C(s_{it}, 2)$ to derive the first equality in (86) and the definition of $\varpi^C(s_{it}, 2)$ to establish the second equality. I rely on information on level assignments to estimate the parameters of expected output at $A$ and exogenous separations. As is standard, estimating firm $A$’s expected output parameters from level transitions based on (83) requires a location normalization at each state. Another location normalization in each $t$ is necessary by Proposition 5 to pin down the exogenous separation rates. For these reasons, I set $\ln p_C(s_{it}, 2) = y_C(s_{it}, 2)$, which implies that $\ln \sum k' e^{\varpi^C(s_{it}, k')} = 0$ by (86), and $y_C(s_{it}, 2) = 0$—recall that Level 2 of firm $C$ is the reference level. The probabilities in (83) form the basis of the likelihood function of the model to estimate firm $A$’s expected output parameters and the exogenous separation rates.50

With Level 2 of $C$ as the reference level, the expected output parameters at $A$ to estimate are

$$
\{(d_{A11(L1)}, e_{A11}, d_{A2L(L1)}, e_{A2L})_{t \geq 2}, \{d_{A2L(L2)}, d_{A3L(L2)}, e_{A3L}\}_{t \geq 3}, \{d_{A3L(L3)}\}_{t \geq 4}\).
$$

Since the data exhibit high attrition because of the large number of separations in each year, discussed in Section 2, and contain a number of level transitions with zero observations, such as demotions or multi-level promotions, which further limits the variability of beliefs, identification is hard to achieve in practice based on the characteristics of the sample. As discussed, to conserve on parameters, for transitions with no or relatively few observations, I did not estimate any of the associated parameters and normalized them to their value at Level 2 of firm $C$. For transitions with approximately the same number of observations, I maintained that the associated parameters are equal.

Specifically, consider first $\{d_{Akt}(k_{t-1})\}_t$. I normalize $d_{A11}(-) = d_{A11}$ to 1, 000 and $e_{A11}$ to zero, since all managers are hired at Level 1. As (83) shows, the separate identification of $\eta_{kt-1}$ and the intercept parameter $d_{Akt}(k_{t-1})$ of the expected output part of $\psi_A(s_{it}, k)$ can be difficult to achieve. Thus, I set $d_{A2L(L2)} = d_{C2L(L2)}$ in order to limit parameter proliferation. By (A1) and the symmetry assumption $d_{Akt}(Lk) = -d_{Akt}(Lk')$, $k' > k$, the parameters left to estimate are $\{d_{A1L(L1)}\}_{t \geq 2}, \{d_{A2L(L2)}\}_{t \geq 3}, \{d_{A3L(L3)}\}_{t \geq 4}$. The combination of rapid promotions to Level 2 and the high separation rate in each $t$ implies that Level 1 has few observations in medium to high tenures. Indeed, the fraction of managers at Level 1 changes dramatically in medium tenures: it is less than 8 percent from tenure $t = 4$ on. Thus, I only estimate $d_{A14(L1)}$ and $d_{A15(L1)}$ and maintain that $d_{A1L(L1)} = d_{A15(L1)}$ for $t = 6, 7$. Similarly, I let $d_{A36(L3)} = d_{A35(L3)}$ at Level 3 and only estimate $d_{A3L(L3)}$ for $t = 4, 5, 7$.

Consider now $\{e_{Akt}\}_t$. At Level 1, I did not estimate any such parameter in $t \geq 3$ because of the much smaller fraction of managers assigned to Level 1 from the third year of tenure on and the very small number of observations

50I compute probabilities and values in $t \geq 8$ under the assumption that managers no longer acquire human capital after $t = 7$. For any given vector of parameter values, I can then obtain “terminal” values in $t = 8$ for the market-wide employment and assignment problem in (56) as the solution to the corresponding infinite horizon problem from $t = 8$. Given these terminal values, I solve the market-wide employment and assignment problem between $t = 1$ and $t = 7$ as a finite horizon one in each such tenure by backward induction. Note that the assumption that human capital acquisition tapers off at some point is not implausible: the employment outcomes of managers in later tenures ($t \geq 8$) display much less variation with tenure and previous assignments and performance compared to those of managers in earlier tenures.
overall at Level 1 in high tenures. At Level 2, I estimate the parameters \( e_{A22}, e_{A23}, e_{A25}, \text{ and } e_{A26} \), 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 observed at this level until the third year of tenure. As the empirical hazard rates of promotion from Levels 1 to 2 and from 2 to 3 display similar qualitative and quantitative features in \( t \geq 2 \), I allow for common components across the parameters \( e_{A22} \) and \( e_{A23} \) to conserve on parameters. That is, I specify \( e_{A21} = e_{22} + e_{23} + e_{24} \) for \( 2 \leq t \leq 4 \) and \( e_{A33} = e_{22} + e_{23} + e_{24} \) for \( 4 \leq t \leq 6 \). The choice of \( e_{22} \) as the benchmark parameter is from the high separation rate from Level 2 and the high promotion rate from Levels 2 to 3 in each tenure: output parameters at Level 2 in low tenures can be estimated more precisely. By this logic, this formulation has led to \( e_{A35} = e_{A36} = e_{22} \), since the differences in these parameters proved insignificantly different from zero. So, at Level 3 I estimate the parameters \( e_{A33} = -e_{34} \)—I assume that \( e_{A33} \) and \( e_{34} \) are relatively negatively to account for the very small fraction of managers (or none) at Level 3 in the first three years at the firm—\( e_{A34} = e_{22} + e_{34}, e_{A37} \), and \( e_{A38} \). In Table 6D, I display the point estimate and standard error of \( e_{A35} \) for completeness: the parameters \( e_{A22}, e_{A33} \), and \( e_{A34} \) are sufficient to pin down \( e_{22}, e_{23}, \) and \( e_{34} \). □

**Parameters of Exogenous Separation Rates:** I specify the probability of exogenous separation as \( \eta_{pk} \). I allow only for variation in \( \eta_{pk} \) 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 are: (1) at the end of period 1: \( \eta_{11}, \eta_{21}, \text{ and } \eta_{31} \); (2) at the end of period 2: none, since \( \eta_{12} = \eta_{11}, \eta_{22} = \eta_{21}, \text{ and } \eta_{32} = \eta_{31} \); (3) at the end of period 3: \( \eta_{13} \text{ or } \eta_3, \eta_{23} = \eta_{22} + \eta_3, \text{ and } \eta_{33} = \eta_{32} \); (4) at the end of period 4: \( \eta_{14} \text{ and } \eta_{24} \), since \( \eta_{34} = \eta_{24} \); (5) at the end of period 5: \( \eta_{15} \), since \( \eta_{15} = \eta_{14} \) and \( \eta_{35} = \eta_{25} \); (6) at the end of period 6: \( \eta_{26}, \text{ since } \eta_{16} = \eta_{15} \text{ and } \eta_{36} = \eta_{26} \); (7) at the end of period 7: \( \eta_{27}, \text{ since } \eta_{17} = \eta_{16} \text{ and } \eta_{37} = \eta_{27} \); and (8) from the end of period 8 onward: none, since \( \eta_{1t} = \eta_{17}, \eta_{2t} = \eta_{27}, \text{ and } \eta_{3t} = \eta_{37} \), \( t \geq 8 \). 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. □

**Derivation of Estimated Wage Equation for Baseline Model and Counterfactuals:** Suppose first that Level 1 of firm \( C \) is the reference level for consistency with Propositions 5 and 6. For illustrative purposes, I start with a more general version of \( b_{fk}(h_1) \) given by \( b_{fk}(h_1) = b_{fk} + \sum_m b_{ym} \{ y_1 = m \} + b_{fk1} e_1 + b_{fk} x_1 + b_{fk} x_1^2 \), which differs across firms, with \( c_{fkt} h_{ft}^2 = \) varying over time. Then, I can express (10) as

\[
y_A(s_{it}, k) = b_{Ak} + \sum_m b_{ym} \{ y_1 = m \} + b_{Ak1} e_1 + b_{Ak} x_1 + b_{Ak} x_1^2 + c_{Ak} + c_{Ak} - \phi_{Ak} - \phi_{At}(k_{t-1}) + \beta_k(c_{Ak} \bar{A}_{Akt} + c_{Ak} \bar{A}_{Akt}) + c_{Ak} \eta_{t} \times (t - 1) + c_{Ak} \eta_{s_{kt-1}} + (\alpha_k - \beta_k) \right) (c_{Ak} \bar{A}_{Akt} + c_{Ak} \bar{A}_{Akt}) \right) p_{it},
\]

\[
y_C(s_{it}, k) = b_{Ck} + \sum_m b_{ym} \{ y_1 = m \} + b_{Ck1} e_1 + b_{Ck} x_1 + b_{Ck} x_1^2 + c_{Ck} + c_{Ck} - \phi_{Ck} - \phi_{Ct}(k_{t-1}) + \beta_k(c_{Ck} \bar{A}_{Ckt} + c_{Ck} \bar{A}_{Ckt}) + c_{Ck} \eta_{t} \times (t - 1) + c_{Ck} \eta_{s_{kt-1}} + (\alpha_k - \beta_k) \right) (c_{Ck} \bar{A}_{Ckt} + c_{Ck} \bar{A}_{Ckt}) \right) p_{it}.
\]

Since \( y_C(s_{it}, k) + (s_{it}, 1) - (s_{it}, 1) = y_A(s_{it}, k) + y_C(s_{it}, 1) - y_A(s_{it}, 1) \) by (10) and (20), it follows that

\[
y_C(s_{it}, k) + (s_{it}, 1) - (s_{it}, 1) = b_{Ak} + \sum_m b_{ym} \{ y_1 = m \} + b_{Ak1} e_1 + b_{Ak} x_1 + b_{Ak} x_1^2 + c_{Ak} + c_{Ak} - \phi_{Ak} - \phi_{At}(k_{t-1}) + \beta_k(c_{Ak} \bar{A}_{Akt} + c_{Ak} \bar{A}_{Akt}) + c_{Ak} \eta_{t} \times (t - 1) + c_{Ak} \eta_{s_{kt-1}} + (\alpha_k - \beta_k) \right) (c_{Ak} \bar{A}_{Akt} + c_{Ak} \bar{A}_{Akt}) \right) p_{it},
\]

Assuming that output parameters are constant with \( t \), namely, \( c_{fkt} j_f = c_{fkt} j_f \) and \( f_f k(k_{t-1}) = f_f k(k_{t-1}) \) for each \( f \) and \( t \), \( c_{Ct} \eta_{s_{kt-1}} = \phi_{Ct}(k_{t-1}) + \bar{A}_{Ct} + \bar{A}_{Ct} \) \( k(k_{t-1}) = \bar{A}_{Ct} + \bar{A}_{Ct} \), \( b_{fk} \) is zero, and all the other parameters of \( b_{fk}(h_1) \) are equal across.
firms, it further follows that

\[ y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) = c_{Ak}^{g} + c_{A1}^{g} + c_{C1}^{g} - c_{A1}^{s} - c_{C1}^{s} - \phi_{Ak} - \phi_{A1} + \beta_1 \left[ c_{A1}^{g} + c_{C1}^{g} \right] + c_{C1}^{s} \]

and

\[ + \left( \sum_{j} \beta_k p_{Ajt}^{K} \right) c_{A1}^{g} + \beta_1 \left( c_{C1}^{g} + c_{C1}^{s} \right) \times (t-1) \]

Since \( w_{Ait} = E w_{Ait} + \epsilon_{Ait} \) by (13) and \( p_f(s_{it}, k) = \Pr(L_{ft} = k | f_t = f, s_{it}) \) by (68), I can rewrite (21) as

\[ w_{Ait} = \omega_{0it}^{k} + \omega_{1it} x_{1}^{2} + \omega_{2it} x_{2}^{2} + \omega_{3it} e_{1} + \sum_{m} \omega_{ym} y_{1} = m + \omega_{Ait}(t-1) + (\omega_{5it} + \omega_{5kt}) p_{it} + \ln \left[ \frac{p_A(s_{it}, 1)}{p_A(s_{it}, 2)} \right] + \epsilon_{Ait}. \]  

Recall that in estimation I consider Level 2 of firm C the reference level so that (21) becomes

\[ E w_{Ait} = y_C(s_{it}, k) + \gamma(s_{it}, 2) - \gamma(s_{it}, k) + \ln \left[ \frac{p_A(s_{it}, 2)}{p_A(s_{it}, 1)} \right], \]  

from which (24) can easily be obtained by an argument analogous to that leading to (88), except for two differences. First, \( y_C(s_{it}, k) + \gamma(s_{it}, 2) - \gamma(s_{it}, k) \) replaces \( y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) \) and can be expressed analogously to \( y_C(s_{it}, k) + \gamma(s_{it}, 1) - \gamma(s_{it}, k) \) in (87). Hence, the parameters of Level 2 rather than Level 1 appear in the corresponding expression. Second, \( p_A(s_{it}, 2) \) rather than \( p_A(s_{it}, 1) \) appears in the expression for \( w_{Ait}, \) as is apparent from (24).  

In estimation, I set the parameters \( \omega_{1it}, \omega_{2it}, \) and \( \omega_{3it}, \) respectively, on \( x_{1}, x_{2}, \) and \( e_{1}, \) equal at Levels 1 and 2. I denote their common values, respectively, by \( \omega_{1}, \omega_{2}, \) and \( \omega_{3}. \) I also assume that \( \{ \sigma_{Ak} \} \) does 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. I denote the common value of \( \{ \sigma_{Ak} \} \) across skill types by \( \sigma_{Ak}. \) So, the estimated wage parameters are \( \{ \omega_{0it}, \omega_{0it}', \omega_{0it}^{3}, \omega_{1}, \omega_{2}, \omega_{3}, \omega_{10}, \omega_{20}, \omega_{30}, \omega_{41}, \omega_{ym} \}_{m=5}, \{ \omega_{5}, \}, \{ \sigma_{Ak}, \sigma_{A2}, \}, \) \( \sigma_{A3}. \)  

In Section 6 when the values of \( \alpha_k, \beta_k, \) and the other parameters of interest are modified to evaluate the importance of uncertainty, learning, and human capital acquisition, wages are recomputed for the new sequences of equilibrium assignments and their associated probabilities for each manager skill type, as consistent with (89). In particular, \( \{ s_{it} \} \) and \( \{ p_A(s_{it}, k) \} \) are simulated anew in any such experiment; see footnote 29 for how \( \{ p_A(s_{it}, 2) \} \) are recomputed in these experiments. Note that the identity \( y_C(s_{it}, k) + \gamma(s_{it}, 2) - \gamma(s_{it}, k) = y_A(s_{it}, k) + y_C(s_{it}, 2) - y_A(s_{it}, 2), \) which follows from (20), implies that in any experiment in which the values of \( \{ \alpha_k, \beta_k \} \) are varied, the mapping between the parameters \( \Omega = \{ \omega_{0it}, \omega_{0it}', \omega_{0it}^{3}, \omega_{1}, \omega_{2}, \omega_{3}, \omega_{ym}, \omega_{Ait}, \omega_{5it}, \omega_{5kt} \} \) and the underlying human capital and output parameters is invariant, since any such experiment involves only changes in the \( \{ p_{it} \} \) process. In the experiment in which the contribution of human capital acquisition is evaluated, the parameters of expected output at firm \( A \) are assumed to be invariant with tenure, and so wages are simulated maintaining that (24) is independent of \( t. \)  

**Parameterization and Estimates of Classification Error:** To limit parameter proliferation and based on model diagnostics and fit, I maintain that \( \rho_1 = \rho_0. \) As discussed in Section 2, 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 \( \{ p_k(k) \}_{k=1,2}. \) Since the probability of a recorded rating is not a linear function of the probability of true performance, as apparent from (11) and (12), 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 6C imply that performance is measured with error, and over time this error is more likely for individuals whose 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

---

51I maintain that \( \{ \sigma_{Ak} \} \) are equal in estimation by imposing that their differences do not exceed a bound chosen small enough, given the assumed ordering \( \sigma_{Ak} \geq \sigma_{Ak}^{i} \) for \( i, i' \), since type-specific standard errors cannot be estimated with any precision at Level 3. In practice, the component densities can also be ordered by the slope of their means in the prior.
Model Fit: In assessing the fit of the model to the data and conducting counterfactual experiments, I simulated 4,000 prior realizations per manager, drawn from the estimated distribution of initial priors. Table 1 shows that the model tracks remarkably well, qualitatively and quantitatively, the profiles of level assignment of managers to Levels 1, 2, and 3, which are nonlinear and nonmonotone in tenure, as well as the tenure pattern of manager separation. Table 2 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 3 further shows, the model successfully fits the patterns of performance ratings both at Level 1 (with slight overpredictions in the third and fifth years of tenure) and at Level 2 (except for some discrepancies in the fourth year of tenure). Lastly, Table 4 shows that the model reproduces quite well the distribution of wages at each level and tenure, except for slight 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, which are partly due to the high rate of attrition in the sample. The fit of the model to the wage data from the larger sample that includes entrants in the firm at all levels is substantially better in this dimension. See Section 5 in the S.A. One criterion to formally evaluate model fit is the Pearson’s $\chi^2$ goodness-of-fit test. I perform this test based on the statistic $s \sum_{r=1}^{R} \left\{ \left( \frac{\hat{\zeta}(r) - \zeta(r)}{\hat{\zeta}(r)} \right)^2 \right\}$, where $\hat{\zeta}(\cdot)$ denotes the empirical density function of a given endogenous variable, $\zeta(\cdot)$ 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. Regarding the distribution of managers across Levels 0 (separation) through 3 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 Levels 2 and 3 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, in these cases 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. As for 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 distribution of wages is the small fraction of managers at Level 3 in high tenures. This intuition is confirmed by the improved fit of the model to the larger sample; see the S.A.

The Role of Persistent Uncertainty: In the fast learning at Level 2 case, 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 7A-7B for results. Similarly 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. This is because the greater informativeness of Level 2 makes it the preferred assignment for a larger set of states, including higher priors, than in the baseline case. Further, since managers reach Level 3 at higher priors than in the baseline case, the standard deviation of wages at Level 3 is much lower than at Level 2 and lower than in the baseline over the first seven years of tenure.

Production Complementarities: I assume that managers are imperfect substitutes in production due to the costs of employing or switching managers across jobs, captured by $\phi_{fkt}(\tilde{k}_{t-1})$, and the cost of rehiring a manager, discussed in Sections 3 and 4.2. The reason for abstracting from explicitly modeling complementarities or capacity constraints is threefold. First, the goal of the 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 the models I integrate into my framework. Second, the great variability of the size of job levels across sample years discussed in Section 2 suggests that my firm is not subject to stringent capacity constraints on employment or assignment. 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 points to a low degree of complementarity across managers. Third, papers that relax the assumption of separable productivity typically address issues that are different from those I focus on here, such as the importance of organizational capital and the growth rate of firms (Prescott and Visscher(1980)), or, because of its complexity, confine the analysis to essentially static environments.
(Ferrall (1997) and Ferrall et al. (2009)) or stylized cases (Davis (1997)). Differently from 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 both information and human capital and the future benefits. 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 an empirically relevant and tractable framework for careers in firms. Since the model I propose closely matches the assignment and wage paths of managers in my data, this framework, although stylized, seems useful in capturing key features of the data.
### Table 1: Percentage Distribution of Managers Across Levels by Tenure

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

### Table 2: Hazard Rates of Separation, Retention at Level, and Promotion (Percentages)

<table>
<thead>
<tr>
<th>Level</th>
<th>Separation</th>
<th>Retention</th>
<th>Promotion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
<td>Data</td>
</tr>
<tr>
<td>Level 1</td>
<td>1 to 2</td>
<td>14.5</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>2 to 3</td>
<td>14.6</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>3 to 4</td>
<td>11.7</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>4 to 5</td>
<td>11.9</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>5 to 6</td>
<td>9.1</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>6 to 7</td>
<td>12.2</td>
<td>5.0</td>
</tr>
<tr>
<td>Level 2</td>
<td>2 to 3</td>
<td>16.3</td>
<td>13.6</td>
</tr>
<tr>
<td></td>
<td>3 to 4</td>
<td>15.6</td>
<td>16.9</td>
</tr>
<tr>
<td></td>
<td>4 to 5</td>
<td>14.7</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>5 to 6</td>
<td>12.8</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>6 to 7</td>
<td>15.4</td>
<td>11.5</td>
</tr>
<tr>
<td>Level 3</td>
<td>3 to 4</td>
<td>10.5</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>4 to 5</td>
<td>12.2</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>5 to 6</td>
<td>10.1</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>6 to 7</td>
<td>10.0</td>
<td>11.5</td>
</tr>
</tbody>
</table>

### Table 3: Percentage of High Ratings at Levels 1 and 2

<table>
<thead>
<tr>
<th>Tenure</th>
<th>Level 1</th>
<th>Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>1</td>
<td>52.7</td>
<td>51.7</td>
</tr>
<tr>
<td>2</td>
<td>34.8</td>
<td>34.9</td>
</tr>
<tr>
<td>3</td>
<td>19.6</td>
<td>21.2</td>
</tr>
<tr>
<td>4</td>
<td>11.8</td>
<td>11.9</td>
</tr>
<tr>
<td>5</td>
<td>2.4</td>
<td>6.2</td>
</tr>
<tr>
<td>6</td>
<td>3.7</td>
<td>3.2</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>1.6</td>
</tr>
</tbody>
</table>
Table 4: Percentage Wage Distributions by Level and Tenure*

<table>
<thead>
<tr>
<th>Level</th>
<th>Tenure</th>
<th>Data</th>
<th>Model</th>
<th>Data</th>
<th>Model</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Between</td>
<td>$20K and $40K</td>
<td>Between</td>
<td>$40K and $60K</td>
<td>Between</td>
<td>$60K and $80K</td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>1</td>
<td>59.1</td>
<td>57.5</td>
<td>40.5</td>
<td>42.0</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>54.4</td>
<td>57.4</td>
<td>44.8</td>
<td>41.7</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>55.6</td>
<td>56.7</td>
<td>46.2</td>
<td>41.9</td>
<td>0.0</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>64.1</td>
<td>68.1</td>
<td>35.9</td>
<td>31.1</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>69.2</td>
<td>69.5</td>
<td>30.8</td>
<td>29.8</td>
<td>0.0</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>75.0</td>
<td>70.6</td>
<td>25.0</td>
<td>28.5</td>
<td>0.0</td>
<td>1.2</td>
</tr>
<tr>
<td>Level 2</td>
<td>2</td>
<td>35.1</td>
<td>34.3</td>
<td>63.3</td>
<td>63.8</td>
<td>1.6</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>31.3</td>
<td>36.1</td>
<td>65.7</td>
<td>61.7</td>
<td>2.9</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>36.3</td>
<td>36.9</td>
<td>60.3</td>
<td>60.5</td>
<td>2.4</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>37.1</td>
<td>37.5</td>
<td>59.6</td>
<td>59.8</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>42.4</td>
<td>38.1</td>
<td>53.3</td>
<td>59.6</td>
<td>4.2</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>41.3</td>
<td>38.8</td>
<td>55.8</td>
<td>58.2</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>Level 3</td>
<td>3</td>
<td>2.8</td>
<td>8.2</td>
<td>84.9</td>
<td>82.4</td>
<td>12.3</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.5</td>
<td>10.3</td>
<td>85.3</td>
<td>80.5</td>
<td>10.1</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5.3</td>
<td>11.3</td>
<td>84.2</td>
<td>79.1</td>
<td>10.5</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6.1</td>
<td>12.4</td>
<td>84.4</td>
<td>77.6</td>
<td>9.5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>4.5</td>
<td>13.4</td>
<td>81.1</td>
<td>76.4</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

*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 5: Percentage Distribution of Changes in Log Wages by Tenure

<table>
<thead>
<tr>
<th>Tenure</th>
<th>Between -0.15 and 0.00</th>
<th>Between 0.00 and 0.15</th>
<th>Between 0.15 and 0.30</th>
<th>Growth Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 2</td>
<td>22.9</td>
<td>69.9</td>
<td>7.2</td>
<td>5.2</td>
</tr>
<tr>
<td>2 to 3</td>
<td>22.6</td>
<td>70.4</td>
<td>6.6</td>
<td>5.1</td>
</tr>
<tr>
<td>3 to 4</td>
<td>24.9</td>
<td>70.3</td>
<td>4.9</td>
<td>3.9</td>
</tr>
<tr>
<td>4 to 5</td>
<td>23.6</td>
<td>70.1</td>
<td>5.9</td>
<td>2.2</td>
</tr>
<tr>
<td>5 to 6</td>
<td>22.5</td>
<td>70.5</td>
<td>6.9</td>
<td>0.7</td>
</tr>
<tr>
<td>6 to 7</td>
<td>21.9</td>
<td>68.5</td>
<td>8.3</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Table 6A: Estimates of Prior Distribution*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type 1 ($i = 1$)</th>
<th>Type 2 ($i = 2$)</th>
<th>Type 3 ($i = 3$)</th>
<th>Type 4 ($i = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior: $\phi_i$</td>
<td>-0.672 (0.022)</td>
<td>-0.484 (0.021)</td>
<td>-0.141 (0.017)</td>
<td>0.435 (0.022)</td>
</tr>
<tr>
<td>Mass: $q_i$</td>
<td>0.155 (0.017)</td>
<td>0.211 (0.030)</td>
<td>0.313 (0.076)</td>
<td>NA</td>
</tr>
</tbody>
</table>

*Asymptotic standard errors in parentheses.

Table 6B: Estimates of Learning Parameters*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Level 1 ($k = 1$)</th>
<th>Level 2 ($k = 2$)</th>
<th>Level 3 ($k = 3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Ability: $\alpha_k$</td>
<td>0.514 (0.062)</td>
<td>0.5437 (0.006)</td>
<td>0.5435 (0.007)</td>
</tr>
<tr>
<td>Low Ability: $\beta_k$</td>
<td>0.456 (0.014)</td>
<td>0.491 (0.013)</td>
<td>0.490 (0.010)</td>
</tr>
</tbody>
</table>

*Asymptotic standard errors in parentheses.

Table 6C: Estimates of Classification Error in Performance Ratings*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Level 1 ($k = 1$)</th>
<th>Level 2 ($k = 2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Error: $\rho_0$</td>
<td>0.521 (0.040)</td>
<td>(same as for $k = 1$)</td>
</tr>
<tr>
<td>Persistence: $\rho_2(k)$</td>
<td>-0.703 (0.040)</td>
<td>-0.544 (0.029)</td>
</tr>
</tbody>
</table>

*Recall that $\rho_1 = \rho_0$. Asymptotic standard errors in parentheses.
Table 6D: Intercept ($d_{Akt}(k_{t-1})$) and Slope ($e_{Akt}$) Parameters of Expected Output*

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

*Recall $d_{A11} = 1,000$. All parameters in the table are expressed in thousands. Parameters whose values are in italics are not estimated.

Table 6E: Estimates of Separation Shocks ($1 - \eta_{kt}$)*

<table>
<thead>
<tr>
<th>Parameters (%)</th>
<th>Value</th>
<th>St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>14.5</td>
<td>0.004</td>
</tr>
<tr>
<td>$\eta_{13}$</td>
<td>8.3</td>
<td>0.001</td>
</tr>
<tr>
<td>$\eta_{14}$</td>
<td>5.0</td>
<td>0.0001</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_{21}$</td>
<td>13.6</td>
<td>0.002</td>
</tr>
<tr>
<td>$\eta_{24}$</td>
<td>14.2</td>
<td>0.001</td>
</tr>
<tr>
<td>$\eta_{25}$</td>
<td>12.1</td>
<td>0.001</td>
</tr>
<tr>
<td>$\eta_{26}$</td>
<td>11.5</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\eta_{27}$</td>
<td>11.1</td>
<td>0.0003</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_{31}$</td>
<td>12.2</td>
<td>0.002</td>
</tr>
</tbody>
</table>

*Recall $\eta_{12} = \eta_{11}$, $\eta_{13} = \eta_{14} + \xi_3$, $\eta_{14} = \eta_{14}$, $t > 4$; $\eta_{22} = \eta_{21}$, $\eta_{23} = \eta_{22} + \xi_3$, $\eta_{24} = \eta_{27}$, $t > 7$; $\eta_{34} = \eta_{31}$, $t = 2, 3$, $\eta_{34} = \eta_{27}$, $t > 3$.

Table 6F: Type-Specific Intercept ($\omega_{0ik}$) of the Wage Equation*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type 1 ($i = 1$)</th>
<th>Type 2 ($i = 2$)</th>
<th>Type 3 ($i = 3$)</th>
<th>Type 4 ($i = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: $\omega_{011}$</td>
<td>8.805 (0.005)</td>
<td>9.288 (0.005)</td>
<td>9.213 (0.011)</td>
<td>8.865 (0.013)</td>
</tr>
<tr>
<td>Level 2: $\omega_{012}$</td>
<td>8.969 (0.004)</td>
<td>9.359 (0.004)</td>
<td>9.281 (0.009)</td>
<td>8.945 (0.012)</td>
</tr>
<tr>
<td>Level 3: $\omega_{013}$</td>
<td>9.534 (0.008)</td>
<td>9.813 (0.004)</td>
<td>9.738 (0.007)</td>
<td>9.418 (0.011)</td>
</tr>
</tbody>
</table>

*Asymptotic standard errors in parentheses.
### Table 6G: Coefficients on Age, Education, and Tenure in the Wage Equation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
<th>St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels 1 and 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age: $\omega_1$</td>
<td>0.028</td>
<td>0.0001</td>
</tr>
<tr>
<td>Age$^2$: $\omega_2$</td>
<td>-0.0003</td>
<td>0.000002</td>
</tr>
<tr>
<td>Education: $\omega_3$</td>
<td>0.022</td>
<td>0.0004</td>
</tr>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenure: $\omega_{412}$</td>
<td>0.007</td>
<td>0.0003</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age: $\omega_{13}$</td>
<td>0.010</td>
<td>0.001</td>
</tr>
<tr>
<td>Age$^2$: $\omega_{23}$</td>
<td>-0.0001</td>
<td>0.00001</td>
</tr>
<tr>
<td>Education: $\omega_{343}$</td>
<td>0.021</td>
<td>0.001</td>
</tr>
</tbody>
</table>

### Table 6H: Year Dummies $\omega_{ym}$ in the Wage Equation (Baseline: 1970-1973)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{ym}$</td>
<td>-0.063</td>
<td>-0.107</td>
<td>-0.140</td>
<td>-0.208</td>
<td>-0.169</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

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

### Table 6I: Coefficients on Prior ($\omega_{2i}$) and Standard Deviations ($\sigma_{Aik}$) in the Wage Equation*

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Type 1 ($i = 1$)</th>
<th>Type 2 ($i = 2$)</th>
<th>Type 3 ($i = 3$)</th>
<th>Type 4 ($i = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior: $\omega_{5i}$</td>
<td>2.371</td>
<td>1.833</td>
<td>1.316</td>
<td>1.364</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.027)</td>
<td>(0.015)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Level 1: $\sigma_{A11}$</td>
<td>0.076</td>
<td>0.070</td>
<td>0.057</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Level 2: $\sigma_{A12}$</td>
<td>0.063</td>
<td>0.047</td>
<td>0.0302</td>
<td>0.0302</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Level 3: $\sigma_{A13}$</td>
<td>0.047</td>
<td>(as $i = 1$)</td>
<td>(as $i = 1$)</td>
<td>(as $i = 1$)</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(NA)</td>
<td>(NA)</td>
<td>(NA)</td>
</tr>
</tbody>
</table>

*Asymptotic standard errors in parentheses.

### Table 6J1: Sensitivity Analysis with Additional Wage Parameters*

<table>
<thead>
<tr>
<th>Prior and Learning</th>
<th>Value</th>
<th>Asymptotic St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi_{11}$</td>
<td>-0.656</td>
<td>0.018</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>-0.494</td>
<td>0.014</td>
</tr>
<tr>
<td>$\phi_{31}$</td>
<td>-0.141</td>
<td>0.014</td>
</tr>
<tr>
<td>$\phi_{41}$</td>
<td>0.440</td>
<td>0.020</td>
</tr>
<tr>
<td>$q_1$</td>
<td>0.133</td>
<td>0.013</td>
</tr>
<tr>
<td>$q_2$</td>
<td>0.220</td>
<td>0.031</td>
</tr>
<tr>
<td>$q_3$</td>
<td>0.335</td>
<td>0.091</td>
</tr>
<tr>
<td>$q_4$</td>
<td>0.312</td>
<td>(NA)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.514</td>
<td>0.052</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.456</td>
<td>0.009</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.544</td>
<td>0.004</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.491</td>
<td>0.011</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.543</td>
<td>79.287</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.490</td>
<td>0.006</td>
</tr>
<tr>
<td>$d_0$</td>
<td>0.515</td>
<td>0.040</td>
</tr>
<tr>
<td>$d_1(1)$</td>
<td>-0.692</td>
<td>0.040</td>
</tr>
<tr>
<td>$d_2(2)$</td>
<td>-0.532</td>
<td>0.028</td>
</tr>
</tbody>
</table>

*I estimate $\phi_{11} = \ln[p_{11}/(1-p_{11})]$, which ranges over the real line.
Table 6J2: Sensitivity Analysis with Additional Wage Parameters

<table>
<thead>
<tr>
<th>Output and Human Capital</th>
<th>Value</th>
<th>Asymptotic St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{A14}(L1)$</td>
<td>-18.132</td>
<td>0.038</td>
</tr>
<tr>
<td>$d_{A15}(L1)$</td>
<td>-35.0721</td>
<td>0.027</td>
</tr>
<tr>
<td>$d_{A16}(L1)$</td>
<td>-35.072</td>
<td>(same as $d_{A15}(L1)$)</td>
</tr>
<tr>
<td>$d_{A17}(L1)$</td>
<td>-35.072</td>
<td>(same as $d_{A15}(L1)$)</td>
</tr>
<tr>
<td>$e_{A12}$</td>
<td>82.342</td>
<td>0.343</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e_{A22}$</td>
<td>74.309</td>
<td>0.067</td>
</tr>
<tr>
<td>$e_{A23}$</td>
<td>66.275</td>
<td>0.102</td>
</tr>
<tr>
<td>$e_{A24}$</td>
<td>74.309</td>
<td>(same as $e_{A22}$)</td>
</tr>
<tr>
<td>$e_{A25}$</td>
<td>61.837</td>
<td>0.101</td>
</tr>
<tr>
<td>$e_{A26}$</td>
<td>62.747</td>
<td>0.070</td>
</tr>
<tr>
<td>$e_{A27}$</td>
<td>62.747</td>
<td>(same as $e_{A26}$)</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{A34}(L3)$</td>
<td>24.104</td>
<td>0.097</td>
</tr>
<tr>
<td>$d_{A35}(L3)$</td>
<td>3.695</td>
<td>0.081</td>
</tr>
<tr>
<td>$d_{A36}(L3)$</td>
<td>3.695</td>
<td>(same as $d_{A35}(L3)$)</td>
</tr>
<tr>
<td>$d_{A37}(L3)$</td>
<td>4.113</td>
<td>0.049</td>
</tr>
<tr>
<td>$e_{A31}$</td>
<td>-12.640</td>
<td>(same as $e_{A33}$)</td>
</tr>
<tr>
<td>$e_{A32}$</td>
<td>-12.640</td>
<td>(same as $e_{A33}$)</td>
</tr>
<tr>
<td>$e_{A33}$</td>
<td>-12.640</td>
<td>0.125</td>
</tr>
<tr>
<td>$e_{A34}$</td>
<td>68.817</td>
<td>0.154</td>
</tr>
<tr>
<td>$e_{A35}$</td>
<td>56.177</td>
<td>0.0290</td>
</tr>
<tr>
<td>$e_{A36}$</td>
<td>56.177</td>
<td>(same as $e_{A35}$)</td>
</tr>
<tr>
<td>$e_{A37}$</td>
<td>62.086</td>
<td>0.017</td>
</tr>
<tr>
<td>$e_{A38}$</td>
<td>64.265</td>
<td>9.378749</td>
</tr>
</tbody>
</table>

*Recall $d_{A11} = 1,000$. All parameters in the table are expressed in thousands. Parameters whose values are in italics are not estimated.

Table 6J3: Sensitivity Analysis with Additional Wage Parameters

<table>
<thead>
<tr>
<th>Exogenous Separation Rates</th>
<th>Value</th>
<th>Asymptotic St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>0.146</td>
<td>0.003</td>
</tr>
<tr>
<td>$\eta_{13}$</td>
<td>0.087</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\eta_{14}$</td>
<td>0.059</td>
<td>0.00004</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_{21}$</td>
<td>0.142</td>
<td>0.001</td>
</tr>
<tr>
<td>$\eta_{24}$</td>
<td>0.143</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\eta_{25}$</td>
<td>0.122</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\eta_{26}$</td>
<td>0.116</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\eta_{27}$</td>
<td>0.111</td>
<td>0.0002</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\eta_{31}$</td>
<td>0.120</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 6J4: Sensitivity Analysis with Additional Wage Parameters

<table>
<thead>
<tr>
<th>Wage Parameters: Intercepts and Slopes</th>
<th>Value</th>
<th>Asymptotic St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept: Type and Level Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{011}$</td>
<td>8.827</td>
<td>0.005</td>
</tr>
<tr>
<td>$\omega_{021}$</td>
<td>9.315</td>
<td>0.005</td>
</tr>
<tr>
<td>$\omega_{031}$</td>
<td>9.281</td>
<td>0.010</td>
</tr>
<tr>
<td>$\omega_{041}$</td>
<td>9.000</td>
<td>0.014</td>
</tr>
<tr>
<td>$\omega_{012}$</td>
<td>8.914</td>
<td>0.004</td>
</tr>
<tr>
<td>$\omega_{022}$</td>
<td>9.310</td>
<td>0.004</td>
</tr>
<tr>
<td>$\omega_{032}$</td>
<td>9.260</td>
<td>0.008</td>
</tr>
<tr>
<td>$\omega_{042}$</td>
<td>8.952</td>
<td>0.012</td>
</tr>
<tr>
<td>$\omega_{013}$</td>
<td>9.407</td>
<td>0.008</td>
</tr>
<tr>
<td>$\omega_{023}$</td>
<td>9.685</td>
<td>0.004</td>
</tr>
<tr>
<td>$\omega_{033}$</td>
<td>9.638</td>
<td>0.007</td>
</tr>
<tr>
<td>$\omega_{043}$</td>
<td>9.343</td>
<td>0.011</td>
</tr>
<tr>
<td>Intercept: Coefficients on Age and Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{1}$</td>
<td>0.029</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\omega_{2}$</td>
<td>-0.0003</td>
<td>0.000003</td>
</tr>
<tr>
<td>$\omega_{3}$</td>
<td>0.021</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\omega_{13}$</td>
<td>0.013</td>
<td>0.001</td>
</tr>
<tr>
<td>$\omega_{23}$</td>
<td>-0.0002</td>
<td>0.000001</td>
</tr>
<tr>
<td>$\omega_{33}$</td>
<td>0.021</td>
<td>0.001</td>
</tr>
<tr>
<td>Intercept: Year-of-Entry Dummies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{y5}$</td>
<td>-0.056</td>
<td>0.003</td>
</tr>
<tr>
<td>$\omega_{y6}$</td>
<td>-0.099</td>
<td>0.004</td>
</tr>
<tr>
<td>$\omega_{y7}$</td>
<td>-0.146</td>
<td>0.003</td>
</tr>
<tr>
<td>$\omega_{y8}$</td>
<td>-0.195</td>
<td>0.003</td>
</tr>
<tr>
<td>$\omega_{y9}$</td>
<td>-0.161</td>
<td>0.003</td>
</tr>
<tr>
<td>Intercept: Coefficient on Tenure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{412}$</td>
<td>0.007</td>
<td>0.0003</td>
</tr>
<tr>
<td>Slope: Coefficient on Prior by Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{51}$</td>
<td>2.328</td>
<td>0.041</td>
</tr>
<tr>
<td>$\omega_{52}$</td>
<td>1.847</td>
<td>0.023</td>
</tr>
<tr>
<td>$\omega_{53}$</td>
<td>1.235</td>
<td>0.014</td>
</tr>
<tr>
<td>$\omega_{54}$</td>
<td>1.224</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 6J5: Sensitivity Analysis with Additional Wage Parameters

<table>
<thead>
<tr>
<th>Wage Standard Deviations by Type and Level</th>
<th>Value</th>
<th>Asymptotic St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{A11}$</td>
<td>0.076</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_{A21}$</td>
<td>0.071</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_{A31}$</td>
<td>0.060</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_{A41}$</td>
<td>0.045</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_{A12}$</td>
<td>0.061</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_{A22}$</td>
<td>0.046</td>
<td>0.001</td>
</tr>
<tr>
<td>$\sigma_{A32}$</td>
<td>0.031</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\sigma_{A42}$</td>
<td>0.028</td>
<td>0.0004</td>
</tr>
<tr>
<td>$\sigma_{A3}$</td>
<td>0.042</td>
<td>0.0003</td>
</tr>
</tbody>
</table>
### Table 6J6: Sensitivity Analysis with Additional Wage Parameters

<table>
<thead>
<tr>
<th>Additional Wage Parameters</th>
<th>Value</th>
<th>Asymptotic St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients on Tenure at Levels 1 and 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{42}$</td>
<td>0.0005</td>
<td>0.0003</td>
</tr>
<tr>
<td>$\omega_{43}$</td>
<td>0.005</td>
<td>0.001</td>
</tr>
<tr>
<td>Coefficients on Prior at Levels 2 and 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{52}$</td>
<td>0.161</td>
<td>0.006</td>
</tr>
<tr>
<td>$\omega_{53}$</td>
<td>0.123</td>
<td>0.029</td>
</tr>
<tr>
<td>Coefficients on Prior $\omega_{5k}^g$ at Levels 1-3 and $\omega_{5k}^x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{51}^g$</td>
<td>0.023</td>
<td>0.003</td>
</tr>
<tr>
<td>$\omega_{52}^g$</td>
<td>0.026</td>
<td>0.005</td>
</tr>
<tr>
<td>$\omega_{53}^g$</td>
<td>0.027</td>
<td>0.008</td>
</tr>
<tr>
<td>$\omega_{5k}^x$</td>
<td>-0.005</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

### Table 7A: Counterfactual Experiments on Importance of Learning for Wages

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Means by Level</th>
<th>St. Dev. by Level</th>
<th>Cumulative Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Learning</td>
<td>No Experimentation</td>
<td>Fast Learning at</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td></td>
<td>Level 1</td>
</tr>
<tr>
<td>Means by Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>$39,584$</td>
<td>$39,706$</td>
<td>$58,271$</td>
</tr>
<tr>
<td>Level 2</td>
<td>43,179</td>
<td>43,070</td>
<td>61,451</td>
</tr>
<tr>
<td>Level 3</td>
<td>48,963</td>
<td>48,454</td>
<td>44,623</td>
</tr>
<tr>
<td>St. Dev. by Level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>$6,936$</td>
<td>$6,791$</td>
<td>$35,961$</td>
</tr>
<tr>
<td>Level 2</td>
<td>7,077</td>
<td>6,464</td>
<td>51,466</td>
</tr>
<tr>
<td>Level 3</td>
<td>8,046</td>
<td>6,534</td>
<td>45,784</td>
</tr>
<tr>
<td>Cumulative Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenure 2</td>
<td>4.60%</td>
<td>3.30%</td>
<td>0.90%</td>
</tr>
<tr>
<td>Tenure 3</td>
<td>8.9</td>
<td>6.8</td>
<td>17.6</td>
</tr>
<tr>
<td>Tenure 4</td>
<td>13.8</td>
<td>9.8</td>
<td>20.5</td>
</tr>
<tr>
<td>Tenure 5</td>
<td>15.9</td>
<td>11.1</td>
<td>21.6</td>
</tr>
<tr>
<td>Tenure 6</td>
<td>17.5</td>
<td>12.9</td>
<td>22.1</td>
</tr>
<tr>
<td>Tenure 7</td>
<td>18.5</td>
<td>14.6</td>
<td>22.2</td>
</tr>
<tr>
<td>Tenure 7 (Balanced)</td>
<td>19.4</td>
<td>15.4</td>
<td>23.3</td>
</tr>
</tbody>
</table>

*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 7B: Counterfactual Experiments on Importance of Learning for Level Assignments

<table>
<thead>
<tr>
<th>Tenure</th>
<th>Fast L at 1</th>
<th>Fast L at 2</th>
<th>Level 1</th>
<th>Fast L at 1</th>
<th>Fast L at 2</th>
<th>Level 2</th>
<th>Fast L at 1</th>
<th>Fast L at 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
</tr>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>14.5</td>
<td>14.5</td>
<td>45.7</td>
<td>57.7</td>
<td>39.8</td>
<td>27.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>26.6</td>
<td>26.6</td>
<td>17.2</td>
<td>20.6</td>
<td>47.3</td>
<td>52.0</td>
<td>38.9</td>
<td>35.0</td>
<td>8.9</td>
</tr>
<tr>
<td>4</td>
<td>37.3</td>
<td>36.7</td>
<td>8.1</td>
<td>11.6</td>
<td>29.2</td>
<td>40.3</td>
<td>25.9</td>
<td>29.1</td>
<td>25.6</td>
</tr>
<tr>
<td>5</td>
<td>45.3</td>
<td>45.1</td>
<td>5.3</td>
<td>8.2</td>
<td>18.3</td>
<td>30.3</td>
<td>17.7</td>
<td>24.7</td>
<td>31.2</td>
</tr>
<tr>
<td>6</td>
<td>51.5</td>
<td>51.3</td>
<td>3.4</td>
<td>5.6</td>
<td>12.6</td>
<td>23.3</td>
<td>13.1</td>
<td>21.7</td>
<td>32.5</td>
</tr>
<tr>
<td>7</td>
<td>56.9</td>
<td>56.7</td>
<td>2.7</td>
<td>4.5</td>
<td>8.3</td>
<td>15.4</td>
<td>8.7</td>
<td>19.2</td>
<td>32.1</td>
</tr>
</tbody>
</table>

*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 7C: Counterfactual Experiment on Importance of Experimentation for Level Assignments

<table>
<thead>
<tr>
<th>Tenure</th>
<th>Fast L at 1</th>
<th>Fast L at 2</th>
<th>Level 1</th>
<th>Fast L at 1</th>
<th>Fast L at 2</th>
<th>Level 2</th>
<th>Fast L at 1</th>
<th>Fast L at 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
<td>No L</td>
<td>Base.</td>
</tr>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>14.5</td>
<td>14.5</td>
<td>45.7</td>
<td>57.7</td>
<td>39.8</td>
<td>27.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>26.6</td>
<td>26.6</td>
<td>17.2</td>
<td>20.6</td>
<td>47.3</td>
<td>52.0</td>
<td>38.9</td>
<td>35.0</td>
<td>8.9</td>
</tr>
<tr>
<td>4</td>
<td>37.3</td>
<td>36.7</td>
<td>8.1</td>
<td>11.6</td>
<td>29.2</td>
<td>40.3</td>
<td>25.9</td>
<td>29.1</td>
<td>25.6</td>
</tr>
<tr>
<td>5</td>
<td>45.3</td>
<td>45.1</td>
<td>5.3</td>
<td>8.2</td>
<td>18.3</td>
<td>30.3</td>
<td>17.7</td>
<td>24.7</td>
<td>31.2</td>
</tr>
<tr>
<td>6</td>
<td>51.5</td>
<td>51.3</td>
<td>3.4</td>
<td>5.6</td>
<td>12.6</td>
<td>23.3</td>
<td>13.1</td>
<td>21.7</td>
<td>32.5</td>
</tr>
<tr>
<td>7</td>
<td>56.9</td>
<td>56.7</td>
<td>2.7</td>
<td>4.5</td>
<td>8.3</td>
<td>15.4</td>
<td>8.7</td>
<td>19.2</td>
<td>32.1</td>
</tr>
</tbody>
</table>

*Baseline (Base.) and No Experimentation (No Experiment.). No Experimentation: \( \alpha_k = \hat{\alpha}_1, \beta_k = \hat{\beta}_1, k = 2, 3 \).