HOOVER INSTITUTION
ECONOMICS WORKING PAPERS

On the Role of Learning, Human Capital, and Performance Incentives for Wages*

Braz Camargo† Fabian Lange‡ Elena Pastorino§

Economics Working Paper 22112

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010
June 2022

Performance pay in general amounts to only a small fraction of total pay. In this paper, we show that performance pay is nevertheless important for level and dynamics of wages over the life cycle because of the incentives it indirectly provides for human capital acquisition and because of its impact on the variability of total pay. We articulate this argument in the context of a model that combines three key mechanisms for wage growth and dispersion: employer learning about workers' ability, human capital acquisition, and performance incentives. We use this model to account for the experience profile of wages, their dispersion, and their composition in terms of fixed and variable (performance) pay. The model admits an analytical decomposition of performance pay into four terms that capture (i) the trade-off between risk and incentives characteristic of settings of moral hazard; (ii) the insurance that firms provide against the wage risk due to the uncertainty about ability; (iii) incentives for effort arising from this uncertainty (career concerns); and (iv) incentives for effort generated by the prospect of human capital acquisition. We prove the model is identified under standard assumptions. Despite its parsimony, the model fits the data very well, including the empirical finding that performance pay as a share of total pay first increases and then decreases with experience. This feature of performance pay, which we are the first to document, runs contrary to the prediction of standard models of performance incentives that the ratio of performance pay to total pay increases with experience, especially at the end of the life cycle. Our estimates imply that effort to produce output augments human capital. Also, human capital acquisition and insurance against uncertainty about ability are quantitatively the main determinants of performance pay. Career-concerns incentives, on which the theoretical literature has focused, and the strength of the contemporaneous trade-off between risk and incentives—the primary determinant of variable pay in static moral-hazard models—are instead much less relevant. Importantly, we find that through the cumulative impact of effort on the job on human capital acquisition and the contribution of variable pay to the variance of total pay, performance incentives are a crucial source of wage growth and dispersion over the life cycle.

Keywords: Uncertainty, Learning about Ability, Human Capital Acquisition, Dynamic Moral Hazard,Wage Growth, Wage Dispersion, Inequality, Life Cycle, Identification, Structural Estimation

---

* We thank Peter Arcidiacono, Flavio Cunha, Jan Eeckhout, Cristina Gualdani, Jonathan Heathcote, Erik Hurst, Lance Lochner, Luigi Pistaferri, Chris Taber, and Gabriel Ulyssea as well as participants at various seminars and conferences for their comments and suggestions. Braz Camargo gratefully acknowledges financial support from CNPq, Fabian Lange from the Canada Research Chairs Program, and Elena Pastorino from the Stanford Institute for Economic Policy Research (SIEPR) and the National Science Foundation (NSF). This paper is dedicated to the memory of Eddie Lazear, whose encouragement, mentorship, and generosity we will never forget.

† Sao Paulo School of Economics–FGV. E-mail: braz.camargo@fgv.br.
‡ McGill University. E-mail: fabolange@gmail.com.
§ Stanford University. E-mail: epastori@stanford.edu.

1 Introduction

What accounts for the growth of wages over the life cycle? Why do differences in wages among workers increase with experience in the labor market? Since Becker [1962] and Mincer [1974], economists have proposed models of investment in human capital to explain the dynamics of wages over the life cycle (Heckman et al. [1998]). Many have also emphasized the role of uncertainty for wage inequality (Cunha et al. [2005], Cunha and Heckman [2016], Lochner and Shin [2014], and Lochner et al. [2018]) and the importance of firms and workers learning about workers' productivity for the growth and dispersion of wages as experience accumulates (Farber and Gibbons [1996]).

Another source of the persistent variation in wages across workers and over time is variable or *performance* pay (Lemieux et al. [2009], Bloom and Van Reenen [2010], Lazear and Shaw [2007, 2011, 2018], and Waldman [2012]), namely, the feature that wages may be contingent on how well workers perform on the job. Variable pay, though, typically amounts to less than 5% of overall pay and, for most workers, does not represent a major component of pay at any point during the life cycle (Frederiksen et al. [2017]). Accordingly, variable pay has received much less attention in the study of the dynamics of wages (Rubinstein and Weiss [2006]). Acquiring new human capital, however, often requires effort, and workers' effort to invest in human capital can either substitute for the effort expended to produce output, as in models of on-the-job training (Ben-Porath [1967]), or complement it, as in models of learning-by-doing (Heckman et al. [2003]). Hence, by influencing workers' effort to produce output, performance pay, although small, may also affect how much human capital workers accumulate and, correspondingly, how rapidly wages grow with experience. In this paper, we argue that performance incentives are important for the dynamics of wages because they indirectly support the accumulation of human capital and, through variable pay, amplify the variability of wages. Thus, performance pay can have profound impacts on the distribution of wages over the life cycle.

We formalize this point by proposing a tractable model of the labor market that combines human capital acquisition, in the form of both on-the-job training and learning-by-doing, uncertainty and employer learning about workers' ability, and incentives for performance. By doing so, we achieve four objectives. First, our model provides a unitary framework to investigate how human capital acquisition, uncertainty and learning about ability, and performance incentives jointly shape the dynamics of wages and their fixed and variable components. Specifically, the model allows us to analytically decompose the ratio of performance pay to total pay at any experience into the contribution of distinct terms that capture the basic forces we nest and so to determine the conditions under which alternative life-cycle profiles of performance pay emerge. Second, through this decomposition, we show that variable pay provides a rich source of information that can be used to identify the primitive determinants of the process of human capital acquisition—including the degree of complementarity between effort spent producing output and human capital—the process of learning about ability, and performance incentives under standard assumptions.[1] Third, our model resolves

[1]Heckman et al. [2003] discuss the challenges of distinguishing between different models of human capital formation and the usefulness of policy-induced variation to this purpose. In our arguments, we exploit instead the life-cycle variation in the ratio of performance pay to total pay.

an empirical failure that we document for existing models of learning and incentive provision over the life cycle. Such models imply that relative to total pay, performance pay should *increase* with labor market experience, whereas we show that the data strongly suggest that it eventually *decreases*. Finally, using our estimated model, we demonstrate that through its impact on human capital acquisition and on the variance of total pay, performance pay plays a critical role for the growth and dispersion of wages over the life cycle.

Our model builds on the literature on learning and incentives (Hölmstrom [1999]). In particular, we draw on Gibbons and Murphy [1992], who characterize performance pay in settings in which firms are uncertain about workers' ability and rely on performance pay to incentivize effort. The idea behind these "career-concerns" models is simple. When firms gradually learn about workers' ability through their output, workers anticipate that good performance on the job favorably influences potential employers' perceptions about their ability, and so it has a positive effect on their future wages. Accordingly, concerns about the market expectation of their ability—"career concerns"—stimulate workers to expend effort and so can substitute for explicit incentives for performance. We add to this framework another dimension of career concerns: by exerting effort on the job, workers not only affect their current output but also their future human capital. Hence, workers face implicit incentives for effort arising both from their concerns about the market perceptions of their ability and from a desire to invest in human capital.[2] These two types of implicit incentives, together with explicit incentives from performance pay, determine workers' effort and human capital, which affect the level, growth, and dispersion of wages and their fixed and variable components over the life cycle.

Formally, we model the labor market as consisting of homogeneous risk-neutral firms and heterogeneous risk-averse workers, whose ability is unknown to all and subject to persistent shocks. Each period, employed workers exert effort, which contributes to a worker's output and human capital. A worker's effort and human capital are observed only by the worker. By contrast, a worker's output or *performance* in a period, which is a noisy measure of the worker's ability, effort, and human capital, is publicly observed and so provides a signal about the worker's ability that firms and workers use to learn about ability over time. Firms compete for workers by offering wage contracts that allow for variable pay contingent on a worker's output.[3]

We characterize equilibrium wages in this framework and decompose the ratio of performance pay to total pay—the "piece rate" of the wage contract that measures the sensitivity of pay to performance—into four terms. These terms reflect fundamental life-cycle forces, are readily interpretable, and take the form of simple functions of the model primitives.[4] The first term captures the standard trade-off between risk and incentives familiar from static models of moral hazard (Hölmstrom [1979]) and how this trade-off changes over time as uncertainty about ability

[2] As consistent with the literature, in what follows we reserve the term "career concerns" to describe the implicit incentives for effort that arise from workers' desire to affect the market perceptions of their abilities.

[3] See, for instance, Fox [2010] for evidence on the importance of outside offers for worker turnover even in highly regulated labor markets like those in Sweden, which supports our competitive setup.

[4] In our model, variable pay is proportional to performance, so the corresponding factor of proportionality—namely, the contract piece rate—equals both the ratio of (average) performance pay to (average) total pay and the (marginal) sensitivity of pay to performance.

varies over the life cycle. The other three terms capture how uncertainty about ability, and so workers' demand for insurance against the wage risk due to it, learning about ability, and human capital acquisition lead to deviations between the standard static piece rate and the dynamic piece rate implied by our model.

To elaborate, we note that the second and third terms of this decomposition of piece rates are negative, thereby depressing piece rates relative to their static level. The second term describes workers' value of the insurance that a wage contract provides against the uncertainty about ability through lower piece rates. Intuitively, lower piece rates partially insulate workers from the wage risk due to the variability of output and the uncertainty about ability, as they reduce the contemporaneous correlation between pay and performance. But as workers accumulate experience, learn about their ability, and face a shorter working horizon, they also bear lower risk and thereby demand less insurance. Thus, this term decreases in magnitude over time, leading to an increase in the relative importance of performance pay over the life cycle. The third term corresponds to the career-concerns component identified by Gibbons and Murphy [1992]. As discussed, career concerns substitute for explicit incentives but tend to weaken over time as ability is revealed and the working horizon shortens. Accordingly, this term is negative as well and eventually declines in absolute value, also contributing to an increase in the ratio of performance pay to total pay with experience.

The final term of our decomposition is instead positive and proportional to the difference between the social and private marginal returns to effort in terms of additional human capital. The private-return component of this term represents the expected change in wages due to the impact of a marginal increase in the effort to produce output on future human capital. Because of workers' risk aversion, this private marginal return tends to be smaller than the social one in absolute value, since firms reduce the variability of wages by offering lower piece rates than those they would offer if workers were risk neutral.[5] But the larger the difference between these social and private returns, the larger (respectively, lower) piece rates firms offer so as to encourage workers to exert more (respectively, less) effort to produce output, if the efforts to produce output and acquire human capital are complements (respectively, substitutes). Thus, this fourth term tends to positively contribute to piece rates when the effort to produce output augments human capital, as our estimates confirm. As experience accumulates, though, and acquiring human capital becomes less important, this term declines in magnitude and adds progressively less to piece rates. When the degree of human capital acquisition is initially small or uncertainty about ability is initially large but rapidly decreases over time, we prove that this fourth term dominates over the second half of the life cycle, whereas uncertainty and learning about ability dominate over the first half, leading to the hump-shaped profile of piece rates that we document in the data.

We show that the model is identified from panel data on wages and their fixed or variable components. Specifically, we establish that the model primitives can be recovered from the life-cycle profile of piece rates, mean wages, and the covariance structure of wages up to usual level normalizations.[6] Crucially, the life-cycle profile of piece rates is

[5] As we show, private and social returns would be equalized if workers were risk neutral, despite the presence of learning about ability.

[6] We normalize the first derivative of the effort cost function and the mean of worker ability at entry in the labor market. We rely on outside information to pin down workers' rate of time preference.

pinned down by the ratio of (mean) variable pay to (mean) total pay in each year of experience under the assumption of free entry of firms in the labor market, which we maintain. We show that these arguments also apply to the case of unobserved heterogeneity in any of the model primitives—except for workers' rate of time preference—including the process of workers' human capital acquisition, workers' degree of risk aversion, and the process that governs learning about ability. These identification results extend to the case of general (semi-parametric) human capital production functions, provided that information on worker performance is available in addition to information on wages, as is the case for many firm-level data sets, including those we use.[7]

We estimate the model by minimum distance, using the well-known Baker-Gibbs-Hölmstrom data (Baker et al. [1994a] and Baker et al. [1994b]), BGH hereafter, on supervisory workers (managers) of a large U.S. firm in a service industry using information on wages and their variable (performance pay) component. We document that in the BGH data, performance pay as a fraction of total pay first increases and then decreases with experience. We confirm that this same hump-shaped pattern of performance pay is present in other firm-level data as well as in the Panel Study of Income Dynamics (PSID). These findings directly contradict the prediction of career-concerns models that relative to total pay, performance pay becomes more important over time, especially at the end of the life cycle (Gibbons and Murphy [1992]). The data thus reject the basic career-concerns model as a theory of performance pay over the life cycle. Our estimated model, on the contrary, successfully matches not only the observed hump-shaped pattern of performance pay relative to total pay but also the experience profile of mean wages and of the variance of wages.

Our estimates suggest that individuals differ in their ability at entry in the labor market and that uncertainty about it persists throughout the life cycle—in fact, because of accumulating shocks to ability, it increases with experience, despite firms and workers learning about ability over time. Correspondingly, we estimate that the insurance against the wage risk due to this uncertainty, which wage contracts provide through low piece rates, is the primary factor depressing performance pay. From an asset pricing perspective, the intuition for the relatively low level of performance pay is simple. To reward effort, performance pay must be high whenever output is high and so news about ability and future pay are positive. But then workers employed under performance-pay contracts effectively hold a portfolio of state-contingent claims to output, whose value comoves with a worker's perceived ability. Specifically, this portfolio pays out more in good times—when output and so signals about ability are high—and less in bad times—when output and so signals about ability are low. However, risk-averse investors prefer, and are willing to pay a premium for, assets that diversify their risk. Accordingly, workers demand wage contracts that reduce the wage risk generated by the variability of beliefs about their ability as firms and workers learn about it. As a result, performance pay tends to be low to partially shield workers against the risk in lifetime wages induced by the uncertainty about their ability. This argument confirms and extends the early intuition of Harris and Hölmstrom [1982] on the role of the

[7]See Margiotta and Miller [2000], Gayle and Miller [2009, 2015], Perrigne and Vuong [2011], and Golan et al. [2015] on the identification and estimation of static and dynamic moral hazard models. In contrast to these authors, we consider a model with uncertainty and learning about productivity and persistent shocks to it, and we rely on the experience profile of wages and their variable component for its identification.

dynamic insurance provided by wage contracts for the evolution of wages with experience in a richer framework that incorporates moral hazard, explicit performance incentives, and human capital acquisition.[8]

Although variable pay represents a small fraction of total pay, we find that performance incentives nonetheless play a key role in shaping both the sensitivity of pay to performance and the growth and dispersion of wages over the life cycle. Specifically, by relying on our decomposition of piece rates, we show that insurance against the wage risk due to the uncertainty about ability and human capital acquisition are quantitatively the most important determinants of the estimated sensitivity of pay to performance. By contrast, career-concerns incentives, on which the theoretical literature has focused, and the strength of the contemporaneous trade-off between risk and incentives—a key determinant of variable pay in static moral-hazard models—are empirically much less relevant.

Interestingly, we estimate that performance incentives are critical to life-cycle wage growth because they encourage workers to exert effort, which contributes to output and, over time, to the accumulation of human capital. In particular, our findings imply that workers' effort to produce output is complementary to their effort to acquire human capital, which supports the notion that human capital is acquired through a learning-by-doing process. Since the variance of performance pay amounts to a large fraction of the variance of total pay, especially over the first half of the life cycle, performance incentives are also crucial for wage dispersion. Specifically, we estimate that performance incentives provided through variable pay account for more than 30% of wage growth, once the cumulative impact of effort on human capital accumulation is taken into account, and for no less than 44% of the variability of wages over the first 30 years of labor market experience in our data. Thus, performance incentives are central to the life-cycle profile of wages and their dispersion. To the best of our knowledge, these estimates are new to the literature.

A lesson from our work is that common statistical decompositions of wage dispersion (as in Abowd et al. [1999]) can be misleading. The variance of wages is often decomposed into that of a "worker" effect, a "firm" effect, and a residual. These terms are usually interpreted to capture, respectively, differences in ability among workers, in firms' attributes, including output risk, and in other unmeasured factors. In our framework, a large portion of the observed variation in wages is attributable to the dispersion in workers' ability, as is often found. But our model also predicts that performance pay declines with the uncertainty about ability. As a consequence, if it were possible to eliminate differences in ability among workers altogether, then our model would imply that the variance of wages would substantially *increase*, rather than decrease, at any level of experience.[9] Intuitively, without uncertainty about ability, wage contracts would feature much higher piece rates, since workers would no longer demand insurance against the risk induced by this uncertainty. Higher piece rates, in turn, would amplify any residual productivity risk, leading, on balance, to much greater wage dispersion. This simple exercise thus illustrates the importance of

[8]This result does not rely on an implausibly high degree of risk aversion, as our estimated coefficient of worker risk aversion falls within the range of existing estimates. Rather, this finding stems from uncertainty about ability amplifying wage risk. See Section 7 for details.

[9]According to our model, *small* decreases in uncertainty about ability for given piece rates lead to a *lower* variance of wages, but *large* decreases in uncertainty about ability may well lead to a *higher* variance of wages.

accounting for the endogeneity of the wage structure, in terms of the fixed and variable pay composition of wages, to the degree of risk and uncertainty in the labor market when assessing the role of different sources of wage dispersion.[10]

In terms of the debate on inequality, this result implies that a trade-off may exist between ex-ante wage risk due to the uncertainty about workers' ability at entry in the labor market and ex-post wage risk due to the variability in wages induced by performance pay. Specifically, lower dispersion in initial ability—achieved, for instance, through better schooling—may induce firms to offer wages more sensitive to performance. Hence, groups of workers that are more homogeneous in terms of their initial skills might end up experiencing more, rather than less, wage inequality.

Related Literature and Outline. Our paper is related to multiple strands of literature, including work on (*i*) assessing the importance of human capital acquisition for wage growth (Heckman et al. [1998], Heckman et al. [2003], Gladden and Taber [2009], and Sanders and Taber [2012]); (*ii*) distinguishing the impact on wage dispersion of uncertainty and heterogeneity among workers (Cunha et al. [2005] and Cunha and Heckman [2016]); (*iii*) measuring the role of uncertainty and learning about ability for job mobility and wages (Miller [1984], Kahn and Lange [2014], and Pastorino [forthcoming]); and (*iv*) estimating human capital functions (Cunha and Heckman [2008], Cunha [2011], and Cunha et al. [2010]) and moral hazard models (Margiotta and Miller [2000], Gayle and Miller [2009, 2015], Perrigne and Vuong [2011], and Golan et al. [2015]).[11] In particular, whereas we focus on the life-cycle dynamics of wages and their components for (supervisory) workers, Golan et al. [2015] analyze how moral hazard and human capital acquisition determine wages to account for the relationship between firm size and executive pay. In their work, executives acquire general and firm-specific human capital, and they choose among jobs and firms that differ in their pecuniary and non-pecuniary attributes. In our baseline model, we refrain from studying the assignment of workers to jobs, but we incorporate unobserved worker ability, allow for persistent shocks to it, and consider a richer moral-hazard problem with multiple possible effort levels for workers so as to capture the varying labor supply and human capital investment choices over the life cycle.[12] We show in Section 9 that our model can be extended to incorporate job mobility and promotions. Also, Golan et al. [2015] rely on bond prices to recover executives' preferences. Our model is instead identified just from data on wages and their fixed or variable components.

Much work has emphasized the importance for the wage process of unobserved worker heterogeneity, which is at the heart of our learning and dynamic incentive mechanisms. For instance, Geweke and Keane [2000] provide evidence from the PSID on the role of transitory shocks and individual heterogeneity for the dynamics of wages. Meghir and Pistaferri [2004], whose findings are also based on the PSID, document the importance of idiosyncratic transitory and

[10]See Ackerberg and Botticini [2002] for evidence on the importance of unobserved characteristics of the two sides of a market for the choice of contract form in the case of agricultural contracts between landlords and tenants.

[11]Using information on wages and performance from the BGH data, Kahn and Lange [2014] document that learning and stochastic productivity changes are important for the variance of wages. They also provide evidence that learning continues throughout the life cycle. Pastorino [forthcoming] uses job, wage, and performance information from the BGH data to identify and estimate the relative contribution of learning and human capital acquisition to the dynamics of workers' jobs and wages.

[12]In their framework, only two effort levels are possible, "effort" or "shirking." This feature simplifies issues of incentive compatibility of wage contracts. Our model is identified up to the mean of worker ability at entry in the labor market and the first derivative of the effort cost function. The model in Golan et al. [2015] is identified up to the non-pecuniary utility and human capital acquired upon shirking.

permanent components of the wage process—in particular, unobserved heterogeneity—for the variance of wages. For related evidence from the PSID on the role of returns to unobserved skills, their dispersion, and the dispersion of non-skill shocks for wage inequality, see Lochner and Shin [2014] and Lochner et al. [2018].[13] Dustmann and Meghir [2005] show the importance for the impact of experience on wages of match-specific effects and heterogeneous returns to human capital, in the form of a correlated random-coefficients model. Adda and Dustmann [2020] estimate the contribution of human capital and unobserved ability to wage growth, using a rich model of occupational choice.

Our analysis abstracts from standard search frictions. Our setup, though, allows for parameters that capture other forces that we do not explicitly model, including search frictions, as we discuss in Sections 3 and 9.[14] See Bowlus and Liu [2013] on the role of search frictions and human capital acquisition for wage growth based on a partial-equilibrium model with human capital as in Ben-Porath [1967] and endogenous search effort. We leave the full analysis of the impact of search frictions and performance incentives on the dynamics of wages and their components to future work.

The paper proceeds as follows. We introduce our data in Section 2, in which we document that the life-cycle profile of performance pay relative to total pay is hump-shaped. Section 3 describes the model, Section 4 informally discusses the equilibrium, and Section 5 contains our formal equilibrium analysis. Section 6 establishes the conditions under which the model is identified, Section 7 presents the estimation results, Section 8 explores their implications for the impact of performance incentives on wages, and Section 9 discusses extensions of our model. Section 10 concludes. Appendix A presents the equilibrium derivation, Appendix B contains the proofs of equilibrium properties, and Appendix C provides omitted proofs and details about the identification of the model. Appendix D is a Supplementary Appendix detailing extensions of our model and presenting additional estimation results.

2 Performance Pay over the Life Cycle

We provide evidence on the experience profile of the ratio of variable (performance) pay to total pay using public data from the PSID as well as proprietary data from the personnel records of two firms first studied in three influential studies in the literature on careers—namely, Baker et al. [1994a,b] and Gibbs and Hendricks [2004]. Both the PSID and these firm-level data contain information on workers' fixed pay f_{it} and variable pay v_{it}, which account for a worker i's total pay or *wage* in any period t, $w_{it} = f_{it} + v_{it}$. In models like ours with variable pay proportional to output and free entry of firms in the labor market, $v_{it} = b_t y_{it}$ and average wages equal average output in each period t. Thus, the ratio $\mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$ of average variable pay to average total pay equals the piece rate b_t, which measures the sensitivity of pay to performance. Using these data spanning multiple years, firms, and industries, we document that the importance of performance pay relative to total pay eventually *declines* with labor market experience, contrary to the prediction of existing career-concerns models with explicit performance incentives.

[13] 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. Lochner and Shin [2014] similarly document the importance of unobserved skills for the evolution of log earning residuals.

[14] We thank Audra Bowlus for very helpful suggestions.

PSID. We focus on the main PSID sample, excluding the poverty, Latino, and immigrant subsamples, and consider male heads of households aged 21 to 65 observed between 1993 and 2013 with valid education information—that is, with more than zero and up to 17 years of education, the largest value. We further restrict attention to those who work more than 45 weeks each year in any industry except for the government and the military, have non-missing positive total labor income, and are not self-employed. The resulting sample consists of more than 24,000 person-year observations. We compute labor market *experience* as potential experience, defined as the difference between an individual's age (minus six) and years of education. We refer to an individual's labor income as the individual's *wage*. We calculate performance pay as the sum of the three measures of variable pay that are available in the PSID from 1993 onward—namely, tips, bonuses, and commissions. Accordingly, we interpret individuals who do not report any tip, bonus, or commission in a year as not receiving performance pay in that year. We exclude observations on performance pay larger than total labor income. In this sample, the average salary is $60,000 (in 2009 dollars), with a standard deviation of $41,000, and the average variable pay is $14,000, with a standard deviation of $46,000.

Figure 1 shows how the sensitivity of pay to performance varies with experience by broad industry categories—manufacturing, transport, services, and the financial, insurance, and real estate (FIRE) industry—for three cohorts of individuals with 10, 15, and 20 years of experience when first observed between 1993 and 1998. Each experience profile is smoothed by taking a five-year moving average. Remarkably, all cohorts exhibit a similar hump-shaped pattern for the sensitivity of pay to performance. Analogous profiles emerge if we divide the sample into workers with and without a college degree.[15] The PSID data thus suggest that the sensitivity of pay to performance increases early in the life cycle, peaks around its middle, and then subsequently declines. This pattern is robust across cohorts, industries, and education groups.

Firm Data. We use data from two large U.S. firms studied in previous work and described in detail by Frederiksen et al. [2017]. As the identities of these firms cannot be disclosed, we refer to them by the names of the authors who first analyzed their data and so refer to them as the Baker-Gibbs-Holmström (BGH) firm and the Gibbs-Hendricks (GH) firm. For both firms, we have information only on white-collar workers—managers (supervisory workers) in the case of the BGH data. The BGH firm operates in a service industry, and the data from it cover the period from 1969 to 1988. Our analysis, however, is limited to the period between 1981 and 1988 because bonus pay, which is the only form of variable pay that managers receive, is not available before 1981. The BGH data contain 36,695 person-year observations and 9,800 unique individuals. Since we have information only about managers at this firm, the average salary is fairly high: $55,000 (in 1988 dollars), with a standard deviation of $31,500. On average, bonus pay accounts for almost $2,000, with a standard deviation of about $7,600. Base salary makes up the remaining $53,000, with a standard deviation of $27,700. The GH data cover the years from 1989 to 1993. We cannot reveal the industry the

[15]Not all individuals in the sample are employed in the four industry categories discussed, but the sample size for the remaining industries is too small to reliably measure experience profiles of performance pay.

firm belongs to. For the GH firm, we have information about 15,648 individuals, for a total of 47,715 person-year observations. As these data pertain to all white-collar employees of the firm, the average salary is intuitively lower than in the BGH data and close to $40,000 (in 1988 dollars), with a standard deviation of $28,000. Bonus pay on average accounts for almost $2,000, with a standard deviation of about $9,300.

The left panels of Figures 2 and 3 report the experience profile of the sensitivity of pay to performance in the BGH and GH data for managers between 21 to 65 years of age. In both firms, performance pay is hump-shaped, peaking after about 20 years in the BGH data and 30 years in the GH data. Similar patterns emerge if we focus on college-educated or non-college-educated workers; see the center and right panels of Figures 2 and 3. Thus, these data, too, reject the main implication of the standard career-concerns model with explicit performance incentives (Gibbons and Murphy [1992]), which predicts that performance pay increases with experience. As we discuss in Subsection 5.2, through the lens of our model, the location of the peak of these piece-rate profiles is informative about the relative importance for wages of learning about ability and human capital acquisition, as well as the speed of these processes.

Although we have presented descriptive statistics from both data sets, we rely on the BGH data to estimate the model we present next, since their fairly long panel covering eight years of observations enables us to better examine life-cycle patterns. The BGH data also represent a touchstone in the personnel literature and so allow us to connect our findings to extant papers, including some of our own, such as Kahn and Lange [2014] and Pastorino [forthcoming].

3 Model

In this section, we describe the environment, define the equilibrium, and discuss our main assumptions.

3.1 Environment

We consider a labor market populated by heterogeneous risk-averse workers and identical risk-neutral firms. Time is discrete, ranges from 0 to T, and is denoted by t. Workers, denoted by i, differ in ability θ_{it}, which is subject to persistent shocks. When employed, workers exert effort e_{it}, which augments output and modifies human capital as specified below. Ability θ_{it} is unobserved to all market participants, including workers. Workers, unlike firms, observe their effort and human capital. Finally, all firms observe output y_{it} as well as wage contracts. Note that since ability is unknown to all, whereas output is observable to all, ours is a model of symmetric learning about ability.

Production. The output technology is common to all firms, and entry in this market is free.[16] Worker i's output in t is

$$y_{it} = \theta_{it} + k_{it} + e_{it} + \varepsilon_{it}, \tag{1}$$

[16]This market can be one of many segmented by location, occupation, or industry and subject to informational frictions. Each labor market is defined by the distribution of a single index of unknown worker productivity, θ_{it}, as well as common learning, human capital, and output technologies across firms. What is important for our results is that these markets are sufficiently separate that employment opportunities in other markets are irrelevant for workers in a given market. In our empirical application, we focus on the market for managers in a service industry.

where the shock to output ε_{it} captures both variability in the worker's output and noise in its measurement. Worker i's ability evolves over time according to the process $\theta_{it+1} = \theta_{it} + \zeta_{it}$, where ζ_{it} is an unobserved shock realized at the end of period t. A worker's initial ability is normally distributed with mean m_θ and variance σ_θ^2. Similarly, shocks to output and ability are normally distributed with mean zero and variances σ_ε^2 and σ_ζ^2, respectively. When $\sigma_\zeta^2 = 0$, ability is constant over time. Allowing for ability shocks implies that uncertainty about ability need not decrease over time, and so implicit incentives from career concerns do not necessarily decline with experience. As is standard, higher effort e_{it} increases output y_{it} in a first-order stochastic sense and so raises expected output.[17]

Human Capital. The human capital of any worker i evolves over time according to

$$k_{it+1} = \lambda k_{it} + \gamma_t e_{it} + \beta_t, \tag{2}$$

where $(1 - \lambda) \in [0, 1]$ is the depreciation rate, $\gamma_t \in \mathbb{R}$ is the rate at which effort to produce output in period t changes the stock of human capital in period $t + 1$, and β_t is a deterministic term common to all workers. Workers have a common stock of human capital at entry in the labor market, k_0. By absorbing k_0 into m_θ, we let $k_0 = 0$ without loss. This formulation of the human capital process encompasses both the case in which the effort to acquire human capital complements the effort to produce output ($\gamma_t > 0$), as in standard learning-by-doing models, and the case in which the effort to acquire human capital substitutes for the effort to produce output ($\gamma_t < 0$), as in models à la Ben-Porath [1967]. In the first case, the investment in human capital in period t is e_t and the human capital accumulation rate is γ_t. In the second case, the investment in human capital in period t is $\overline{e}_t - e_t$, where $\overline{e}_t$ denotes a worker's endowment of time or efficiency units in t, the human capital accumulation rate is $|\gamma_t|$, and the term $-\gamma_t \overline{e}_t = |\gamma_t| \overline{e}_t$ is absorbed in β_t. We refer to γ_t throughout as the accumulation rate of human capital in period t.[18] Our specification in (2) extends that in Bagger et al. [2014], in which output is $\theta_i + k_{it} + \varepsilon_{it}$, if we interpret β_t as inclusive of firm productivity; see the discussion of firm heterogeneity in Section 9. Unlike these authors, we allow (*i*) the individual heterogeneity parameter θ_{it} to be unknown to workers and firms and vary over time; (*ii*) human capital k_{it} to evolve endogenously as a function of a worker's past effort; and (*iii*) a worker's effort in a period to affect the amount of efficient labor provided. In Section 6, we consider more general formulations of the human capital process, including the case in which this process differs unobservably across workers or depends nonparametrically on effort.

Worker Preferences. In period t, the lifetime utility of a worker who receives wage $w_{t+\tau}$ and exerts effort $e_{t+\tau}$ in period $t+\tau$ for each $0 \leq \tau \leq T-t$ is given by $-\exp\{-r[\sum_{\tau=0}^{T-t} \delta^\tau (w_{t+\tau} - e_{t+\tau}^2/2)]\}$, where $r > 0$ is the coefficient

[17] Like Gibbons and Murphy [1992, p. 476], we allow effort to be negative, as positive effort might not be optimal for a worker. We can then use first-order conditions to characterize the solution to a worker's problem. We later show that effort is positive if piece rates lie in the unit interval and the effort to produce output complements the effort to acquire human capital, which is the empirically relevant range, and derive conditions for equilibrium piece rates to satisfy this restriction in Appendix A.

[18] Note that β_t can also capture additions to human capital from observable investment activities such as formal training. Our production function of skills, $h_{it+1} \equiv \theta_{it+1} + k_{it+1}$, can be interpreted as a log form of the function $\tilde{h}_{it+1} = \tilde{a}_{t+1}(\theta_{it}, \theta_{it+1}) \tilde{\beta}_t \tilde{h}_{it}^\lambda \tilde{e}_{it}^{\gamma_t}$ with $h_{it} = \ln(\tilde{h}_{it})$, $a_{t+1}(\theta_{it}, \theta_{it+1}) = \ln(\tilde{a}_{t+1}(\theta_{it}, \theta_{it+1})) = \theta_{it+1} - \lambda\theta_{it}$, $a_0(\theta_{i-1}, \theta_{i0}) = \theta_{i0}$, $\beta_t = \ln(\tilde{\beta}_t)$, $e_{it} = \ln(\tilde{e}_{it})$, and $h_{i0} = \theta_{i0}$. We can allow for heterogeneous initial stocks of human capital by assuming that k_{i0} is known but random and for β_t to evolve stochastically over time.

of absolute risk aversion, $\delta \in (0, 1)$ is the discount factor, and $e^2/2$ is the monetary cost of effort e. See Gibbons and Murphy [1992] for a virtually identical specification of preferences.[19]

Contracts. In every period t, firms offer workers one-period contracts specifying their wage in t as a function of their output in the period. Following Gibbons and Murphy [1992], we consider linear contracts so that worker i's wage in period t is $w_{it} = a_{it} + b_{it}y_{it}$, where a_{it} is fixed pay, $b_{it}y_{it}$ is variable pay, and b_{it} is the contract piece rate. Restricting attention to one-period contracts is equivalent to requiring long-term contracts to be renegotiation-proof. (See Gibbons and Murphy [1992] for a proof of this result. Their proof immediately extends to our environment.) We focus on linear contracts for three reasons. First, this assumption is standard, so it makes our framework comparable to those commonly studied. Second, incentive contracts are often linear in output, or approximately so, in the data. Third, from a theoretical point of view, linear contracts allow us to summarize the strength of contractual incentives, which is a key feature of interest in our analysis, through a single one-dimensional measure, the piece rate b_{it}.

Wages. Competition among firms implies that expected wages in every period equal expected output. Hence,

$$w_{it} = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it}y_{it} \tag{3}$$

is worker i's wage in period t, where $\mathbb{E}[y_{it}|I_{it}]$ is worker i's expected output in t conditional on the public information I_{it} available about the worker in t. This information consists of the worker's output realizations before t. The term $(1 - b_{it})\mathbb{E}[y_{it}|I_{it}]$ is the fixed component of the wage in period t. This component depends on a worker's conditional expected output in t, which in turn is a function of a worker's conditional expected ability in t.

Strategies and Equilibrium. A worker's history in period t consists of the sequence of the worker's private effort choices and public output realizations up to period $t-1$. A strategy for a firm specifies contract offers to workers conditional on the public portion of workers' histories. A strategy for a worker specifies an effort choice after each history and contract offers by firms. We consider pure-strategy sequential equilibria. An equilibrium specifies strategies for firms and workers such that for each worker (*i*) after any public history for the worker, firms offer a linear contract satisfying (3) that maximizes the worker's expected lifetime utility given the firms' and the worker's future behavior; and (*ii*) the worker's choice of effort in each period is optimal given the worker's history, the contracts firms offer to the worker, and the firms' and the worker's future behavior. Condition (*i*) follows from the assumption of free entry of firms. Condition (*ii*) is sequential rationality. We assume that when indifferent between accepting two or more firms' offers, a worker with given observable characteristics separates from the current employer with probability equal to the corresponding empirical probability of separation from the firm in our data for workers with the same characteristics.

[19] It is straightforward to extend our equilibrium characterization to the case in which the cost of exerting effort e is $g(e)$, where g is twice continuously differentiable and strictly convex. By assuming that workers have constant absolute risk aversion (CARA) preferences, we abstract from wealth effects. Because of its tractability, this assumption is ubiquitous in models of dynamic moral hazard.

Hence, the model is consistent with worker turnover in equilibrium.[20]

3.2 Remarks

Our model nests well-known models of learning about ability, human capital accumulation, and performance incentives. For instance, when ability is known and effort is public, our model reduces to one of human capital accumulation through effort investments that can complement or substitute for the effort expended to produce output. In this case, when also effort is not a choice variable, the model further specializes to one of "passive" human capital acquisition with experience. If, instead, effort does not contribute to human capital and ability is not subject to shocks ($\gamma_t = 0$ at each t and $\sigma_\zeta^2 = 0$), then our model simplifies to the career-concerns model with explicit incentives in Gibbons and Murphy [1992]. When contracts are restricted to fixed pay, the model further reduces to a finite-horizon version of the standard career-concerns model of Hölmstrom [1999]. If, in addition, effort is fixed, our model is an instance of a typical symmetric learning model with ability general across firms like the one in Farber and Gibbons [1996].

Our functional-form assumptions are common in the literature and allow us to completely characterize equilibrium. Since output is linear in its components, contracts are linear in output, shocks to ability are additive, and initial ability, ability shocks, and output shocks are normally distributed, our model with CARA preferences admits a mean-variance representation as in Gibbons and Murphy [1992]. This feature implies that a worker's trade-off between consumption or wages and leisure does not depend on a worker's history, which leads equilibrium to be unique and symmetric with piece rates and effort dependent only on time. In Section 6 and the appendices, we consider more general versions of the model that result from relaxing some of these assumptions, whose equilibria we characterize and which we prove are identified by simple extensions of the arguments presented below.

4 Informal Equilibrium Derivation

In this section, we informally discuss equilibrium and its properties. We first describe the process of learning about ability. We then discuss how career concerns and human capital acquisition affect a worker's incentives to exert effort for a *given* life-cycle profile of piece rates. We conclude by deriving equilibrium piece rates. A more formal characterization of equilibrium follows in Section 5.

4.1 Learning about Ability

Firms and workers learn about a worker's ability over time by observing a worker's output. Consider worker i in period t, whose equilibrium effort and human capital in t are e_t^* and k_t^*, respectively. Denote by $z_{it} = y_{it} - k_t^* - e_t^*$

[20]As Baker et al. [1994a,b] remark, turnover in their data is largely independent of performance and approximately constant with tenure. Specifically, they find no evidence that separations mask a tendency for managers to be laid off or move to other firms in response to poor performance. Hence, we find our approximation that turnover is random from the point of view of the mechanisms of our model not implausible.

the portion of the worker's output in t in excess of the worker's effort and human capital. Then,

$$z_{it} = \theta_{it} + \varepsilon_{it} \tag{4}$$

is the signal about the worker's ability in t that firms and workers extract from the worker's output. Since initial ability and shocks to ability and output are normally distributed, (4) implies that in equilibrium, posterior beliefs about a worker's ability in any period are normally distributed and fully described by their mean $m_{it} = \mathbb{E}[\theta_{it}|I_{it}]$ and variance $\sigma_{it}^2 = \mathrm{Var}[\theta_{it}|I_{it}]$, with $m_{i0} = m_\theta$ and $\sigma_{i0}^2 = \sigma_\theta^2$. We refer to m_{it} as worker i's *reputation* in t. By standard results,

$$m_{it+1} = \frac{\sigma_\varepsilon^2}{\sigma_{it}^2 + \sigma_\varepsilon^2} m_{it} + \frac{\sigma_{it}^2}{\sigma_{it}^2 + \sigma_\varepsilon^2} z_{it} \quad \text{and} \quad \sigma_{it+1}^2 = \frac{\sigma_{it}^2 \sigma_\varepsilon^2}{\sigma_{it}^2 + \sigma_\varepsilon^2} + \sigma_\zeta^2. \tag{5}$$

The recursions for m_{it} and σ_{it}^2 in (5) respectively describe how a worker's reputation and the variance of posterior beliefs about a worker's ability change over time. These expressions are valid even when workers' effort choices deviate from the equilibrium path, since a worker's effort is private and every output realization is possible for any choice of effort. Observe that the variance σ_{it}^2 evolves independently of the realization of z_{it} and so is common to all workers in t. Thus, we can suppress the subscript i and simply denote this variance by σ_t^2. Since signals do not perfectly reveal ability and ability is subject to shocks, uncertainty about ability persists throughout a worker's career and converges to a non-negative fixed point $\sigma_\infty^2 = [\sigma_\zeta^2 + (\sigma_\zeta^4 + 4\sigma_\zeta^2\sigma_\varepsilon^2)^{1/2}]/2$ by (5). In particular, the variance σ_t^2 monotonically decreases to σ_∞^2 if $\sigma_\theta^2 > \sigma_\infty^2$ and monotonically increases to σ_∞^2 if $\sigma_\theta^2 < \sigma_\infty^2$. Note that $\sigma_\infty^2 > 0$ if $\sigma_\zeta^2 > 0$, in which case ability is never fully learned.[21] By iterating on the law of motion for m_{it} in (5), we can trace out the evolution of a worker's reputation as output signals about ability accumulate. With $\mu_t \equiv \sigma_\varepsilon^2/(\sigma_t^2 + \sigma_\varepsilon^2)$ and the convention that $\prod_{k=1}^0 a_k = 1$ for any numeric sequence $\{a_k\}$, we then have the following result.

Lemma 1. *For each worker i and period t, the worker's reputation in period $t+\tau$ with $1 \le \tau \le T-t$ is*

$$m_{it+\tau} = \left(\prod_{k=0}^{\tau-1} \mu_{t+k}\right) m_{it} + \sum_{s=0}^{\tau-1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1-\mu_{t+s}) z_{it+s}.$$

4.2 Dynamic Returns to Effort

We now discuss how the returns to effort depend on workers' incentives to increase their reputation—the market expectation of their ability—and acquire human capital. The results derived here apply to any equilibrium that has the property that piece rates and effort depend only on time. As we show in Section 5, the unique equilibrium of our model has this property. Consider then a sequence of piece rates $\{b_t\}_{t=0}^T$ dependent only on time so that we can suppress a worker's index i to simplify notation. In what follows, we first present a worker's problem and derive the first-order conditions determining a worker's choice of effort in each period when a worker's future effort choices depend only

[21] See Hölmstrom [1999] for a proof of these properties. Kahn and Lange [2014] refer to the case in which $\sigma_\zeta^2 > 0$ as the case of "learning about a moving target" and find evidence for it from the correlation between performance ratings and pay in the BGH data.

on time. We then decompose the returns to effort into terms that capture incentives arising from workers' desire to affect their reputation and human capital.

Worker Problem. Consider worker i's choice of effort in period t. Let $w_{it+\tau}$ be the worker's wage in period $t+\tau$ with $0 \leq \tau \leq T-t$, and let $W_{it} = \sum_{\tau=0}^{T-t} \delta^\tau w_{it+\tau}$ be the present-discounted value of the worker's wages from period t on. The worker chooses effort e_t to maximize the utility $U_{it}(e_t) = \mathbb{E}[-\exp\{-r\left[W_{it} - e_t^2/2\right]\}|h_i^t]$, where we omit the dependence of e_t on i for ease of notation. Note that the expectation in $U_{it}(e_t)$ is conditional on worker i's period-t history h_i^t. Yet, as we will show, the choice of e_t that maximizes $U_{it}(e_t)$ is independent of h_i^t. Since output signals about ability are normally distributed, it follows from (3) and Lemma 1 that the wages $\{w_{it+\tau}\}_{\tau=0}^{T-t}$ are normally distributed, and so is the present-discounted value W_{it}. Recall that if X is normally distributed with mean μ and variance σ^2, then $\mathbb{E}[\exp\{rX\}] = \exp\{r\mu - r^2\sigma^2/2\}$. Thus, e_t maximizes $U_{it}(e_t)$ if, and only if, it maximizes

$$\mathbb{E}[W_{it}|h_i^t] - r\mathrm{Var}[W_{it}|h_i^t]/2 - e_t^2/2 = \sum_{\tau=0}^{T-t} \delta^\tau \mathbb{E}[w_{it+\tau}|h_i^t] - r\mathrm{Var}[W_{it}|h_i^t]/2 - e_t^2/2. \tag{6}$$

First-Order Conditions for Effort. The wage contract in (3) implies that $\partial\mathbb{E}[w_{it}|h_i^t]/\partial e_t = b_t$. Note that worker i's effort in t also influences wages in $t+\tau$ through its effect on the worker's future reputation $m_{it+\tau}$, which impacts the *fixed* component of future pay, and its effect on the worker's future human capital $k_{it+\tau}$, which impacts both the *fixed* and *variable* components of future pay. The first-order condition for worker i's effort in any period t is then[22]

$$e_t = b_t + \sum_{\tau=1}^{T-t} \delta^\tau \frac{\partial\mathbb{E}[w_{it+\tau}|h_i^t]}{\partial e_t}. \tag{7}$$

The right side of (7), which describes the marginal benefit of effort in t, consists of two terms. The first term captures the *static* marginal benefit of effort and is given by the piece rate b_t. The second term captures the *dynamic* marginal benefit of effort, which corresponds to the effect of effort on the present-discounted value of the worker's expected future wages and is different from zero as long as $t < T$.

Marginal Benefit of Effort. In Appendix A, we show that we can express the first-order condition in (7) as

$$e_t = b_t + R_{CC,t} + R_{HK,t}, \tag{8}$$

where[23]

$$R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (1-b_{t+\tau}) \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t) \text{ and } R_{HK,t} = \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\left(b_{t+\tau} + R_{CC,t+\tau}\right). \tag{9}$$

[22] Note that effort does not affect the variance of future wages. To see why, recall that piece rates are taken as given by a worker. As the variance of the signals about ability does not depend on effort, Lemma 1 implies that a worker's effort in period t does not affect the variance of a worker's future reputation. Similarly, a worker's stock of human capital has no impact on the variance of output or wages.

[23] The marginal benefit of effort in t does not vary with e_t, since $R_{CC,t}$ and $R_{HK,t}$ do not depend on e_t. So, (8) is sufficient for optimality. Also, note that (8) implies that effort choices depend only on time and are identical across workers if piece rates satisfy the same properties. Hence, individuals facing the same current and future piece rates, and choosing the same efforts in the future, also behave identically in t. This feature is key to establishing that the unique equilibrium is symmetric and that effort choices and piece rates depend only on time.

The terms $R_{CC,t}$ and $R_{HK,t}$ describe a worker's dynamic marginal benefit of effort that arises from the impact of effort on a worker's reputation and human capital, respectively.[24]

To understand the term $R_{CC,t}$, observe that higher effort in period t on average increases the period-t signal about worker i's ability, which raises the worker's reputation in all future periods. Formally, Lemma 1 implies that

$$\frac{\partial \mathbb{E}[m_{it+\tau}|h_i^t]}{\partial e_t} = \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t)\frac{\partial \mathbb{E}[z_{it}|h_i^t]}{\partial e_t} = \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t)$$

for all $1 \leq \tau \leq T-t$. Since the fixed component of worker i's wage in period $t+\tau$ is $(1-b_{t+\tau})\mathbb{E}[y_{it+\tau}|I_{it+\tau}]$ and $\mathbb{E}[y_{it+\tau}|I_{it+\tau}]$ changes one-for-one with the worker's reputation in period $t+\tau$, a marginal increase in worker i's effort in period t increases the fixed component of the worker's wage in period $t+\tau$ by $(1-b_{t+\tau})(\prod_{k=1}^{\tau-1}\mu_{t+\tau-k})(1-\mu_t)$. The term $R_{CC,t}$ is the present-discounted value of all these marginal increases and captures standard career-concerns incentives (Hölmstrom [1999]): even in the absence of any explicit link between pay and performance, workers have a desire to exert effort to improve their performance in order to influence the market perception of their ability.

To understand the term $R_{HK,t}$, observe that worker i's choice of effort in period t directly affects the *variable* component of the worker's wage in all subsequent periods by affecting the worker's stock of human capital and thus output in each such period. By changing the worker's stock of human capital, effort in period t additionally affects future output signals about the worker's ability, and so the worker's future reputation and *fixed* pay. To elaborate, note that a marginal increase in effort e_t in period t leads to a change in worker i's undepreciated stock of human capital and output in period $t+\tau$ by $\gamma_t\lambda^{\tau-1}$. This change in output affects both the variable component of the worker's wage in period $t+\tau$ by the amount $b_{t+\tau}\gamma_t\lambda^{\tau-1}$ and the magnitude of the signal $z_{it+\tau}$ about the worker's ability observed at the end of $t+\tau$ by the amount $\gamma_t\lambda^{\tau-1}$. By the same argument as that used in the derivation of the term $R_{CC,t}$, the latter change affects worker i's expected future reputation, as larger (respectively, smaller) signals induce the market to infer that a worker is of higher (respectively, lower) ability. As a result, the fixed component of the worker's wage from period $t+\tau$ on changes by $\gamma_t\lambda^{\tau-1}R_{CC,t+\tau}$. Thus, the total change in worker i's expected value of wages in period $t+\tau$ resulting from the impact on human capital of a marginal increase in effort in period t is $\gamma_t\lambda^{\tau-1}(b_{t+\tau}+R_{CC,t+\tau})$. The term $R_{HK,t}$ is the present-discounted value of all these marginal changes.

4.3 Equilibrium Piece Rates

The first-order condition in (8) determines a worker's effort in any period t taking as given current and future piece rates, under the assumption that a worker's future effort choices depend only on time. We now solve for the last-period piece rate and then proceed backward to determine piece rates in the remaining periods. With this characterization of equilibrium piece rates at hand, we rely on the results from the previous subsection to derive equilibrium effort choices, provided that equilibrium efforts and piece rates depend only on time, which will be the case.

[24] It follows from (8) and (9) that effort is positive if piece rates are in the unit interval and $\gamma_t > 0$, which is the empirically relevant case. Indeed, for $1 \leq \tau \leq T-t$, $R_{CC,t+\tau} \geq 0$ if $b_{t+\tau+s} \leq 1$ for all $1 \leq s \leq T-t-\tau$ so that $b_{t+\tau}+R_{CC,t+\tau} \geq 0$ if $b_{t+\tau} \geq 0$.

Last-Period Piece Rates. It is easy to show that equilibrium piece rates in period T are the same for all workers and given by $b_T^* = 1/[1 + r(\sigma_T^2 + \sigma_\varepsilon^2)]$. Intuitively, since no dynamic considerations affect effort decisions in the last period, uncertainty about ability plays the same role as output shocks. Thus, when $t = T$, our model is equivalent to a canonical static linear-normal model of incentives with quadratic effort cost, in which output shocks are normally distributed with variance $\sigma_T^2 + \sigma_\varepsilon^2$. Given that last-period piece rates are the same for all workers independently of their output histories, (8) implies that the last-period equilibrium effort choices are the same for all workers as well.

Piece Rates in Previous Periods. To determine equilibrium piece rates in period $t < T$, suppose that equilibrium efforts and piece rates from period $t + 1$ on depend only on time; we have showed that this property holds when $t = T - 1$. For each $0 \le \tau \le T - t$, let $b_{t+\tau}^*$ be the equilibrium piece rate in $t + \tau$, and define $R_{CC,t}^*$ and $R_{HK,t}^*$ as in (9) with $b_{t+\tau} = b_{t+\tau}^*$ for each τ. Then, a worker's effort in t when the worker's contract piece rate in t is b satisfies

$$e_t = e_t(b) = b + R_{CC,t}^* + R_{HK,t}^*. \tag{10}$$

Let $w_{t+\tau}^* = w_{t+\tau}^*(b)$ and $W_t^* = W_t^*(b)$ respectively be a worker's wage in period $t + \tau$ with $0 \le \tau \le T - t$ and the present-discounted value of the wages from period t on as functions of b. Note that W_t^* depends on b *directly* through the effect of b on the worker's variable pay in t and *indirectly* through the effect of b on the worker's effort in t. Observe also that competition among firms leads firms to offer a piece rate that maximizes a worker's expected lifetime payoff, conditional on the information that firms have about the worker. Then, by the mean-variance representation of worker preferences in (6), a worker's equilibrium piece rate b in t maximizes

$$\mathbb{E}[W_t^*|I_t] - r\text{Var}[W_t^*|I_t]/2 - e_t^2/2, \tag{11}$$

where I_t is the public information about the worker in period t—that is, the worker's output history up to t. We next show that the piece rate maximizing (11) is unique and does not depend on I_t, so it is the same for all workers in t.

First, consider the impact of a marginal change in b on the expected present-discounted value of wages from t on:

$$\frac{\partial \mathbb{E}[W_t^*|I_t]}{\partial b} = \sum_{\tau=0}^{T-t} \delta^\tau \frac{\partial \mathbb{E}[w_{t+\tau}^*|I_t]}{\partial b} = 1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}. \tag{12}$$

The first term on the right side of (12) is $\partial\mathbb{E}[w_t^*|I_t]/\partial b = \partial e_t/\partial b = 1$. The equality $\partial\mathbb{E}[w_t^*|I_t]/\partial b = \partial e_t/\partial b$ follows from the competition among firms, which implies that expected wages equal expected output, and the fact that only effort in period t depends on b. That $\partial e_t/\partial b = 1$ follows from (10). As for the second term on the right side of (12), note that by increasing effort in period t by one unit, a worker not only increases expected output in t by one unit but also changes expected output in period $t + \tau$ with $1 \le \tau \le T - t$ by $\gamma_t \lambda^{\tau-1}$ units, which amounts to the change in the worker's stock of human capital in $t + \tau$. The second term on the right side of (12) is then the present-discounted value of these expected output changes that the worker fully captures.

Consider now the impact of a marginal change in b on the variance of the present-discounted value of wages from period t on. Since, as discussed, effort in period t does not affect the variance of wages, $\text{Var}[W_t^*|I_t]$ depends on b only through its direct effect. As b affects the variance of $w_{t+\tau}^*$ only when $\tau = 0$, we can write

$$\text{Var}[W_t^*|I_t] = \text{Var}[w_t^*|I_t] + 2\sum_{\tau=1}^{T-t} \delta^\tau \text{Cov}[w_t^*, w_{t+\tau}^*|I_t] + \text{Var}_0,$$

where the last term is independent of b. As we show in Appendix A, given that $\text{Var}[w_t^*|I_t] = b^2(\sigma_t^2 + \sigma_\varepsilon^2)$ and the linearity of w_t^* in b implies that $\text{Cov}[w_t^*, w_{t+\tau}^*|I_t]$ is linear in b for all $1 \leq \tau \leq T-t$, it follows that

$$\frac{\partial \text{Var}[W_t^*|I_t]}{\partial b} = 2b(\sigma_t^2 + \sigma_\varepsilon^2) + 2\sum_{\tau=1}^{T-t} \delta^\tau \frac{\partial}{\partial b} \text{Cov}[w_t^*, w_{t+\tau}^*|I_t] = 2b(\sigma_t^2 + \sigma_\varepsilon^2) + 2H_t^*, \tag{13}$$

where $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$.This term reflects the fact that output in t is correlated with future output through the worker's ability. Thus, by increasing b and so the correlation between a worker's wage and ability in t, firms also increase the correlation between a worker's wage in t and wages in future periods, thereby increasing the variance of W_t^*.

Hence, if we again use the fact that $\partial e_t/\partial b = 1$, then the first-order condition for the problem of maximizing (11) is given by $1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - rb(\sigma_t^2 + \sigma_\varepsilon^2) - rH_t^* - e_t = 0$. This equation admits the unique solution

$$b_t^* = \frac{1}{1 + r(\sigma_t^2 + \sigma_\varepsilon^2)} \left(1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^* - R_{CC,t}^* - rH_t^*\right) \tag{14}$$

by (10).[25] See Appendix A for details. Expression (14) is the equilibrium piece rate in period t, which is independent of I_t and so is the same across workers. This independence simplifies the characterization of equilibrium and allows us to make progress in determining how explicit and implicit incentives interact over the life cycle, as we show next.[26]

5 Equilibrium Characterization and Properties

In this section, we characterize the equilibrium and examine the pattern of equilibrium piece rates over the life cycle.

5.1 Recursive Formulation of Equilibrium

We first state and discuss our key characterization result. In order to do so, let $\{\sigma_t^2\}_{t=0}^T$ satisfy the difference equation

$$\sigma_{t+1}^2 = \frac{\sigma_t^2 \sigma_\varepsilon^2}{\sigma_t^2 + \sigma_\varepsilon^2} + \sigma_\zeta^2 \tag{15}$$

with initial condition $\sigma_0^2 = \sigma_\theta^2$. Recall our convention that $\prod_{k=1}^0 a_k = 1$ for any numeric sequence $\{a_k\}$ and that

$$\mu_t = \frac{\sigma_\varepsilon^2}{\sigma_t^2 + \sigma_\varepsilon^2}. \tag{16}$$

[25] That b_t^* maximizes (11) follows from the fact that $\partial \text{E}[W_t^*|I_t]/\partial b$, the marginal benefit to the worker of an increase in b, is constant, whereas $(r/2)\partial \text{Var}[W_t^*|I_t]/\partial b + e_t$, the marginal cost to the worker of an increase in b, increases with b.

[26] Since the equilibrium piece rates in period t are the same for all workers, so are the equilibrium efforts in period t by (8). Thus, if equilibrium efforts and piece rates are symmetric and depend only on time from period $t+1$ on, then they satisfy the same properties from period t on. So, by induction, equilibrium efforts and piece rates are symmetric and depend only on time.

Proposition 1. *The equilibrium is unique and such that piece rates and effort choices are the same for all workers and depend only on time. Let b_t^* and e_t^* be, respectively, the equilibrium piece rate and effort choice in period t. For each t, define b_t^0, $R_{CC,t}^*$, $R_{HK,t}^*$, and H_t^* as*

$$b_t^0 = \frac{1}{1+r(\sigma_t^2+\sigma_\varepsilon^2)}; \tag{17}$$

$$R_{CC,t}^* = \sum\nolimits_{\tau=1}^{T-t} \delta^\tau (1-b_{t+\tau}^*) \left(\prod\nolimits_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t); \tag{18}$$

$$R_{HK,t}^* = \gamma_t \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b_{t+\tau}^* + R_{CC,t+\tau}^*\big); \tag{19}$$

$$H_t^* = \sigma_t^2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau. \tag{20}$$

Then, b_t^* and e_t^* are given recursively by

$$b_t^* = b_t^0 \left(1+\gamma_t \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^* - R_{CC,t}^* - rH_t^*\right) \tag{21}$$

and

$$e_t^* = b_t^* + R_{CC,t}^* + R_{HK,t}^*. \tag{22}$$

In Appendix A, we state and prove analogous results for the more general case in which the law of motion of human capital is $k_{it+1} = \lambda k_{it} + F_t(e_{it})$ with F_t concave; we also provide simple conditions for equilibrium piece rates to lie in the unit interval. Note that the expression in (21) decomposes equilibrium piece rates into five terms. The first term, b_t^0, is the equilibrium piece rate in the static linear-normal model of incentives with exponential utility and quadratic cost of effort when the variance of output is $\sigma_t^2 + \sigma_\varepsilon^2$. The second term, $1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, is the social marginal return to effort in period t, which corresponds to the change in the expected present-discounted value of a worker's lifetime output resulting from a marginal increase in effort in t. As discussed, the third and fourth terms, $R_{CC,t}^*$ and $R_{HK,t}^*$, capture the dynamic private marginal benefit of effort in t arising from career concerns and human capital considerations. The fifth term, rH_t^*, reflects the increase in the variance of the present-discounted value of lifetime wages resulting from a marginal increase in the piece rate in t, scaled by the coefficient of risk aversion.

One way to understand (21) is to compare it with the piece rate that would induce the first-best (efficient) level of effort. This piece rate equates the marginal cost of effort, e_t^*, to its social marginal return, $1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$. From (22), it is immediate that the first-best piece rate is $\chi_t = 1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{CC,t}^* - R_{HK,t}^*$.[27]

The equilibrium piece rate differs from χ_t in two ways. First, it subtracts from χ_t the term rH_t^*, which is positive if $t < T$. Intuitively, any variation in output in period $t < T$ leads to variation not only in wages in t but also in future wages, since wages depend on a worker's reputation, which changes with realized output. Firms partially insure workers against this life-cycle wage risk by means of lower piece rates through rH_t^*, which reduces the correlation between a worker's performance and pay. As long as σ_t^2 declines or does not increase too fast with t, the term rH_t^*

[27] This piece rate reduces to $\chi_t = 1$ only when $t = T$ and no dynamic considerations influence a worker's choice of effort.

also declines with t: as the uncertainty about ability decreases over time, so does the need to insure workers against the resulting variability in wages. Observe also that the magnitude of rH_t^* decreases with r. Naturally, the less risk-averse workers are, the lower the degree of insurance they desire against the risk induced by the uncertainty about ability.

Second, the equilibrium piece rate scales the difference $\chi_t - rH_t^*$ by the factor $b_t^0 < 1$, which also adjusts the piece rate to account for the trade-off between risk and incentives familiar from static models of moral hazard. Specifically, the equilibrium contract weighs the output gain from larger piece rates, which induce higher effort, against the cost of increasing variability in the compensation of a risk-averse worker. This scaling-down effect increases with a worker's risk aversion, r, and effective output risk, $\sigma_t^2 + \sigma_\varepsilon^2$, arising from the uncertainty about ability and the shocks to output.

As in Gibbons and Murphy [1992], the inefficiency in the provision of incentives relative to the first best is due solely to risk aversion. Despite the uncertainty about ability, if workers were risk neutral, then piece rates would be equal to one, and workers' choices of effort would equate the marginal cost of effort to its social marginal return in each period, regardless of the presence or degree of human capital acquisition.[28]

Equation (21) also suggests an alternative decomposition of b_t^* as

$$b_t^* = b_t^0 - b_t^0 R_{CC,t}^* - b_t^0 r H_t^* + b_t^0 \left(\gamma_t \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^* \right). \tag{23}$$

This decomposition isolates the portion of the change in human capital resulting from a marginal increase in effort e_t in t that does not accrue to the worker, $\gamma_t \sum_{\tau=1}^{T-1} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^*$, and is useful for three reasons. First, as we will argue, it helps illustrate how the economic forces at play in our model shape the provision of explicit incentives for effort over time. Second, using this decomposition, we can analytically determine the conditions under which these forces give rise to alternative life-cycle profiles of piece rates, as we discuss in the next subsection. Finally, this decomposition will prove useful for our identification arguments.

The first term in (23) is the equilibrium piece rate b_t^0 in the static linear-normal model of incentives, as discussed. Without dynamic considerations, firms would offer b_t^0 in each period. The second and third terms in (23) capture the contribution of uncertainty and learning about ability to the piece rate and are familiar from the work of Gibbons and Murphy [1992]. In particular, the second term adjusts the explicit incentives provided by piece rates to account for the career-concerns incentives that arise from the presence of uncertainty about ability whenever $t < T$. These implicit incentives induce workers to exert effort even in the absence of any explicit link between pay and performance: by partially substituting for explicit incentives, they lead to lower piece rates.[29] The third term in (23) discounts piece rates so as to provide workers with insurance against the risk in life-cycle wages due to the uncertainty about ability.

The last term in (23), which is novel, captures the contribution of human capital acquisition to the explicit in-

[28] Indeed, if $r = 0$, then $b_T^* = 1$. This, in turn, implies that $R_{CC,T-1}^* = 0$ and $R_{HK,T-1}^* = \gamma_{T-1}\delta$, so that $b_{T-1}^* = 1$. It follows by induction that $b_t^* = 1$, $R_{CC,t}^* = 0$, and $R_{HK,t}^* = \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ for all t. In particular, with risk-neutral workers, implicit incentives for effort would arise only from human capital considerations.

[29] Here, we assume that $R_{CC,t}^*$ is positive when $t < T$. This result holds when piece rates belong to the unit interval.

centives for effort and consists of two parts. The first part is proportional to $\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, which is the present-discounted change in lifetime output that results from the change in a worker's human capital following a marginal increase in effort (to produce output) in period t. The second part, which is proportional to $R^*_{HK,t}$, reflects the implicit incentives for effort arising from the prospect of human capital acquisition, which substitute for explicit incentives and so decrease piece rates when γ_t is greater than zero, and complement explicit incentives and so increase piece rates when γ_t is smaller than zero. Intuitively, in this latter case, e_t negatively contributes to human capital acquisition, which discourages workers from exerting effort to produce output. Thus, higher piece rates help support the incentives to produce output. Overall, this last term adjusts piece rates to better align the private marginal returns to effort with the corresponding social marginal returns, which vary over the life cycle because of discounting and the variation in the rates of human capital accumulation $\{\gamma_t\}_{t=0}^T$. When γ_t is positive or not too negative, this last term, unlike the previous two, contributes positively to equilibrium piece rates when future piece rates are in the unit interval—the empirically relevant case—as our estimates will confirm.

5.2 Piece Rates over the Life Cycle

We now discuss how learning about ability and human capital acquisition affect the life-cycle profile of piece rates. We first consider the cases in which either only learning about ability or only human capital acquisition is present to show how these two forces can lead to opposite patterns for piece rates. This discussion sets the stage for the general case that follows. All proofs are in Appendix B.

Pure Learning-About-Ability Case. If we mute human capital acquisition, setting $\gamma_t = 0$ for all t, then (21) becomes

$$b_t^* = b_t^0(1 - R^*_{CC,t} - rH_t^*).$$

Lemma 2 describes how piece rates evolve in this case.

Lemma 2. *Let $\gamma_t = 0$ for all t. For all σ_θ^2, there exists $T_0 \geq 0$ such that if $T > T_0$, then b_t^* is strictly increasing with t for all $T_0 \leq t \leq T$. Moreover, b_t^* is strictly increasing with t if $\sigma_\theta^2 > \sigma_\infty^2$.*

In our model, ability is subject to persistent shocks. When ability is constant over time, our model specializes to that in Gibbons and Murphy [1992]—under the assumption of a quadratic cost of effort. Lemma 2 thus extends the characterization of the life-cycle profile of piece rates in Gibbons and Murphy [1992] to the case in which ability is subject to shocks. For the first part of the lemma, note that since the degree of uncertainty about ability σ_t^2 eventually converges to σ_∞^2, at some point the only force governing how piece rates evolve over time is the decrease in an individual's working horizon as experience accumulates. Naturally, as the time remaining in a worker's career shortens, the implicit incentives for effort provided by career concerns weaken. Firms compensate for this decline in implicit incentives by increasing explicit incentives through higher piece rates. When $\sigma_\theta^2 > \sigma_\infty^2$, uncertainty about ability decreases monotonically over time, as in Gibbons and Murphy [1992]. In this case, the two forces shaping the provision of explicit incentives—uncertainty about ability and the length of the working horizon—work in the same direction, leading piece rates to strictly increase with experience.[30]

Pure Human Capital Acquisition Case. By setting $\sigma_\theta^2 = \sigma_\zeta^2 = 0$, we eliminate uncertainty and learning about ability from the model. When this is the case, letting $b^0 = 1/(1 + r\sigma_\varepsilon^2)$, (21) becomes

$$b_t^* = b^0 \left[1 + \gamma_t \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} (1 - b_{t+\tau}^*) \right], \tag{24}$$

so piece rates vary over time only because of the incentives to accumulate human capital. The dynamic social marginal return to the effort to produce output in terms of additional future human capital, $\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, differs from the corresponding private one, $\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} b_{t+\tau}^*$. It is intuitive that human capital motives contribute positively to piece rates when $\gamma_t > 0$ and piece rates are smaller than one. Indeed, if efforts to produce output and to acquire human capital are complements and piece rates are smaller than one, then workers do not fully capture the returns to their investments in human capital, and so their willingness to exert effort is reduced. Current piece rates help offset this undersupply of effort, but they do so imperfectly, because of the risk-incentives trade-off. More generally, one can show that piece rates are non-negative as long as $\gamma_t > \lambda - 1/\delta$ for all t. That is, even when the efforts to produce output and to acquire human capital are rival, it is optimal to induce workers to exert more effort out of human capital considerations, provided that the trade-off between output and human capital production is not too severe.

The variation of the rate γ_t of human capital accumulation over the life cycle clearly affects the experience profile of piece rates. We now state and discuss three results that illustrate how for alternative life-cycle patterns of the rates $\{\gamma_t\}_{t=0}^T$, the model can generate increasing or decreasing life-cycle profiles of piece-rates in the pure human capital acquisition case. The first result shows that piece rates are strictly decreasing over time if the rates of human capital accumulation are positive, constant, and not too large. More generally, piece rates are eventually strictly decreasing over time if the rates of human capital accumulation are eventually positive, nonincreasing and not too large.[31]

Lemma 3. *Suppose γ_t is positive and nonincreasing with t for all $T_0 \leq t < T$ for some $0 \leq T_0 < T$. Then, b_t^* is smaller than one and strictly decreasing with t for all $T_0 \leq t \leq T$ if $0 < \gamma_{T_0} < (1 - \delta\lambda)(1 + r\sigma_\varepsilon^2)/\delta[1 - (\delta\lambda)^{T-T_0}]$. In particular, b_t^* is strictly decreasing with t for all t if $\gamma_t \equiv \gamma$ with $0 < \gamma < (1 - \delta\lambda)(1 + r\sigma_\varepsilon^2)/\delta[1 - (\delta\lambda)^T]$.*

Piece rates can also be initially increasing. This is the case, for instance, when the rates of human capital accumulation are initially positive and small but increase rapidly over time, as we prove next.

[30] Piece rates can be initially decreasing when $\sigma_\theta^2 < \sigma_\infty^2$. Indeed, a straightforward backward induction argument shows that piece rates in the pure learning case are always smaller than one, so that $R_{CC,t}^*$ is positive if $\sigma_t^2 > 0$. Hence, if $\sigma_\theta^2 = 0$ and $\sigma_\zeta^2 > 0$, then $b_0^* = b_0^0 > b_0^1 > b_1^*$. By continuity, the same result holds when σ_θ^2 is positive but small.

[31] To understand why the rates of human capital accumulation cannot be too positive for Lemma 3 to hold, first note that since b_T^* is smaller than one, $b_{T-1}^* = b^0[1 + \gamma_{T-1}\delta(1 - b_T^*)]$ is greater than b_T^*. But since b_{T-1}^* is linearly increasing with γ_{T-1}, b_{T-2}^* is *smaller* than b_{T-1}^* when γ_{T-1} and thus b_{T-1}^* are sufficiently large. More generally, if b_{t+1}^* to b_T^* are smaller than one, then b_t^* is strictly increasing with γ_t^* and unbounded above by (24), in which case b_{t-1}^* can be smaller than b_t^* by the same equation.

Lemma 4. *Suppose there exists $0 < T_0 < T$ such that $\gamma_{T_0} > 0$ and $b_t^* < 1$ for all $T_0 \le t \le T$. There exists $\xi \ge 1$ such that if $\gamma_t > 0$ and $\gamma_t > \xi\gamma_{t-1}$ for all $0 < t \le T_0$, then b_t^* is strictly increasing with t for all $0 \le t \le T_0$.*

By Lemma 3, we can ensure that piece rates from period T_0 on are bounded above by one and strictly decreasing over time. So, combining Lemmas 3 and 4, we obtain that piece rates in the pure human capital acquisition case can be hump-shaped if the (positive) rates of human capital accumulation first increase and then (weakly) decrease.

Corollary 1. *Piece rates can be hump-shaped if the rates of human capital accumulation are positive, initially increasing, and then nonincreasing with experience.*

General Case. When learning about ability and human capital acquisition are both present, the stronger force naturally shapes the experience profile of piece rates. For instance, when σ_ζ^2 is small so that ability is effectively known in the long run, human capital acquisition eventually governs the behavior of piece rates if workers are sufficiently long-lived. Intuitively, at some point, the residual uncertainty about ability becomes small enough that learning about it no longer matters for the evolution of piece rates. As a result, piece rates are strictly decreasing over time in the long run if the conditions of Lemma 3 hold. In contrast, when the importance of human capital acquisition declines sufficiently fast over time, learning about ability determines the profile of piece rates in the long run; thus, piece rates eventually become strictly increasing with experience. The next result confirms these intuitions.

Proposition 2. *Suppose that σ_ζ^2 is small. There exists $T_0 \ge 0$ such that if $T > T_0$, γ_t is positive and nonincreasing with t for all $T_0 \le t \le T-1$, and $0 < \gamma_{T-1} \le \gamma_{T_0} < (1-\delta\lambda)(1+r\sigma_\varepsilon^2)/\delta[1-(\delta\lambda)^{T-T_0}]$, then b_t^* is strictly decreasing with t for all $T_0 \le t \le T$. On the other hand, there exists $T_0 \ge 0$ and $\underline{\gamma} > 0$ such that if $T > T_0$ and $|\gamma_t| < \underline{\gamma}$ for all $T_0 \le t \le T-1$, then b_t^* is strictly increasing with t for all $T_0 \le t \le T$.*

As discussed, piece rates can be hump-shaped in the pure human capital acquisition case if the rates of human capital accumulation are positive and initially increasing and then decreasing. By continuity, the same result holds if σ_θ^2 and σ_ζ^2 are small so that uncertainty about ability is initially low and remains so throughout the life cycle. Piece rates can also be hump-shaped when the rates of human capital accumulation are positive and constant over time if σ_θ^2 is large and σ_ε^2 and σ_ζ^2 are small, so that uncertainty about ability is initially large but learning about it occurs rapidly over time. However, Proposition 2 shows that in the presence of learning about ability, if the rates of human capital accumulation become small in absolute value rapidly enough, then piece rates eventually became *strictly increasing* with experience. This result suggests that piece rates can be U-shaped if human capital accumulation is important early on but its importance decreases over time sufficiently fast. We establish this result next.

Proposition 3. *Piece rates can be hump-shaped if one of two conditions hold: (i) σ_θ^2 and σ_ζ^2 are small, and the rates of human capital accumulation are positive and initially increasing and then decreasing; (ii) the rates of human capital* *accumulation are positive and constant over time, σ_θ^2 is large, and σ_ε^2 and σ_ζ^2 are small. Piece rates can be U-shaped if the rates of human capital accumulation are initially positive and large but decrease rapidly over time.*

The last two propositions show that the interplay between learning about ability and human capital acquisition gives rise to complex patterns of explicit incentives, which can lead to opposite experience profiles for piece rates over the life cycle. Yet, Proposition 2 implies that the profile of piece rates at the end of the life cycle transparently reflects the relative importance of learning about ability and human capital acquisition at those levels of experience. This result thus suggests that the profile of piece rates towards the end of the life cycle is especially informative about the primitives of our model. The picture is more nuanced earlier in the life cycle, as the same pattern of piece rates is consistent with varying degrees of importance of learning about ability. Indeed, as Proposition 3 shows, piece rates can be initially increasing both when learning about ability is unimportant throughout the life cycle and when learning about ability matters early on instead. However, a consequence of Lemma 3 and the logic of Proposition 2 is that if the rates of human capital accumulation are constant (and not too large), then piece rates are hump-shaped *only* if learning about ability is important early on. More generally, once the process of learning about ability is pinned down, the observed pattern of piece rates just reflects the process of human capital acquisition, as we show in the next section.

Since, as we also prove in the next section, our model has differing implications for the experience profile of the second moments of the distributions of wages depending on the characteristics of the process of learning about ability, we can infer the relative importance of learning about ability and human capital acquisition at different stages of workers' careers by combining information on the life-cycle profile of piece rates and the covariance structure of wages. We formalize this point in the next section.

6 Identification

In this section, we discuss the identification of the model based on panel data on wages and their fixed or variable components. We start in Section 6.1 by showing that the model is identified from the first and second moments of the distributions of wages and the ratio of variable pay to total pay over the life cycle, up to a level normalization. Specifically, this ratio identifies piece rates in each year of experience. With piece rates known, the second moments of the distributions of wages identify the distributions of workers' initial ability and of the shocks to ability and output, which completely determine the process of learning about ability. Differences in mean wages over time pin down the degree of human capital depreciation. Once these primitives are recovered, by exploiting our characterization of piece rates, we prove that piece rates identify workers' risk preferences and the process of human capital accumulation.[32]

We proceed in Section 6.2 to show that analogous results hold in the presence of unobserved worker heterogeneity,

[32] Our identification arguments do not require any exogenous variation external to the model. Intuitively, the variation of piece rates with experience is informative about the human capital process once the parameters of the learning process are identified from the covariance structure of wages. The model also provides a natural exclusionary restriction in that the dynamic and static piece rates coincide in period T. Hence, the coefficient of risk aversion is pinned down by b_T^*, once the learning parameters are known.

even when wages are measured with error. We also establish that our identification results extend to the case in which human capital evolves nonparametrically with effort according to the law of motion $k_{it+1} = \lambda k_{it} + F_t(e_{it})$ if information about worker performance is available, as is the case for many firm-level data (Frederiksen et al. [2017]).

In our identification arguments, we treat the discount factor δ as known and normalize the quadratic effort cost function so that effort is measured using a money metric. To simplify the exposition, we set the terms $\{\beta_t\}_{t=0}^{T-1}$ in (2) to zero. We discuss the identification of $\{\beta_t\}_{t=0}^{T-1}$ in Appendix C.[33]

6.1 Identification of the Baseline Model

We show that piece rates $\{b_t\}_{t=0}^{T}$; the variances of the distributions of initial ability σ_θ^2, of output shocks σ_ε^2, and of ability shocks σ_ζ^2; the risk aversion coefficient r; and the rates of human capital depreciation $1-\lambda$ and accumulation $\{\gamma_t\}_{t=0}^{T-1}$ are identified from a panel of wages and their variable components, up to the mean initial ability m_θ.

Proposition 4. *The piece rates $\{b_t^*\}_{t=0}^{T}$ and the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified from a panel of wages and their variable components. Once piece rates and the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified, the risk aversion parameter r is identified from the piece rate b_T^*, and the human capital depreciation rate $1-\lambda$ and accumulation rates $\{\gamma_t\}_{t=0}^{T-1}$ are identified from the piece rates $\{b_t^*\}_{t=0}^{T}$ and average wages in $T-1$ and T up to mean initial ability m_θ.*

We divide the proof of Proposition 4 into two parts. First, we show how piece rates and the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified. Then, we show how the risk aversion and human capital parameters are recovered. The logic of the argument is simple. Piece rates in each period are identified by the ratio of average variable pay to average total pay. The variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified from the second moments of the distribution of wages in the first two years of experience. Given these, we can identify the coefficient of risk aversion r from the piece rate in the last period. The depreciation rate of human capital is identified from the difference in average wages between experiences $T-1$ and T and piece rates in these two experience years. The rest of the human capital parameters are identified from the time profile of piece rates in the remaining years of experience. We now provide more details about the argument.

Piece Rates and Variances. The wage of worker i in period t can be expressed as $w_{it} = f_{it} + v_{it}$, where f_{it} and v_{it} are its fixed and variable components, respectively. Since contracts are linear in output, variable pay is given by $v_{it} = b_t^* y_{it}$, and so $\mathbb{E}[w_{it}] = (1-b_t^*)\mathbb{E}[\mathbb{E}[y_{it}|I_{it}]] + b_t^*\mathbb{E}[y_{it}] = \mathbb{E}[y_{it}]$ by (3). Thus, the period-t piece rate is identified as $b_t^* = \mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$. Once piece rates are recovered, the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified as follows. We show in Appendix C that $\mathrm{Var}[w_{i0}] = (b_0^*)^2(\sigma_\theta^2 + \sigma_\varepsilon^2)$, $\mathrm{Cov}[w_{i0}, w_{i1}] = b_0^*\sigma_\theta^2$, and $\mathrm{Var}[w_{i1}] = \sigma_\theta^2 + \sigma_\zeta^2 - \sigma_1^2 + (b_1^*)^2(\sigma_1^2 + \sigma_\varepsilon^2)$. Hence, σ_θ^2 and σ_ε^2 are identified from the variance of wages in the first year and the covariance between first-period

[33] In the more general case in which the derivative of the effort cost function is c, one can show that c and the risk aversion parameter r are separately identified if the rate of human capital depreciation is known and the rates of human capital accumulation in two different periods are equal. Alternatively, one can show that r and c are separately identified if γ_{T-1} is known. Both arguments rely on the fact that whereas the product rc appears at the denominator of the static piece rate b_t^0, only r multiplies the term H_t^* in the expression of the equilibrium piece rate. More generally, our model is identified up to the first derivative of the effort cost function.

and second-period wages. Then, σ_ζ^2 is identified from the variance of wages in the second year, since $\sigma_1^2 = \sigma_\zeta^2 + \sigma_\theta^2\sigma_\varepsilon^2/(\sigma_\theta^2 + \sigma_\varepsilon^2)$. Note that as the parameters of the learning process are identified independently of the parameters of the human capital process, the former can be identified—and estimators of them can be constructed—on the basis of these identification arguments, independently of the specification of the human capital process.

Risk Aversion and Human Capital Parameters. We now establish that the parameters r, λ, and γ_0 to γ_{T-1} are identified up to m_θ from average wages and the identified vector $(\{b_t^*\}_{t=0}^T, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$. First, observe that if this vector is identified, so are the terms σ_t^2, $R_{CC,t}^*$, and H_t^* for all t by (15) to (18) and (20) in Proposition 1. Therefore, r is identified from b_T^*, σ_T^2, and σ_ε^2, given that $b_T^* = 1/[1 + r(\sigma_T^2 + \sigma_\varepsilon^2)]$, and so b_{T-1}^0 is identified from r, σ_{T-1}^2, and σ_ε^2 by (17). Likewise, γ_{T-1} is identified from b_{T-1}^*, b_{T-1}^0, b_T^*, $R_{CC,T-1}^*$, r, and H_{T-1}^*, since $b_{T-1}^* = b_{T-1}^0[1 + \gamma_{T-1}\delta(1 - b_T^*) - R_{CC,T-1}^* - rH_{T-1}^*]$. As for λ, we know from Proposition 1 that e_T^* equals b_T^* and e_{T-1}^* equals $b_{T-1}^* + R_{CC,T-1}^* + \gamma_{T-1}\delta b_T^*$. Thus, effort choices in the last two periods are known from b_{T-1}^*, $R_{CC,T-1}^*$, γ_{T-1}, and b_T^*. Since $\mathbb{E}[w_{it}] = m_\theta + k_t^* + e_t^*$ and the law of motion of human capital is $k_t^* = \lambda k_{t-1}^* + \gamma_{t-1}e_{t-1}^*$ for all $t \geq 1$, it follows that $\mathbb{E}[w_{iT}] - \lambda\mathbb{E}[w_{iT-1}] = e_T^* + (\gamma_{T-1} - \lambda)e_{T-1}^* + (1-\lambda)m_\theta$ when β_{T-1} is zero. Hence, λ is identified from average wages in the last two years, e_T^*, γ_{T-1}, and e_{T-1}^* up to m_θ.[34]

We conclude by showing that the rates γ_0 to γ_{T-2} are identified from $(\{b_t^*\}_{t=0}^T, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r, \lambda)$. Note that

$$b_t^* - b_t^0\big(1 - R_{CC,t}^* - rH_t^*\big) = b_t^0\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(1 - b_{t+\tau}^* - R_{CC,t+\tau}^*\big) \tag{25}$$

for all $t \leq T-2$ by (19) and (21). Since all the terms in (25) except for γ_t are known from $(\{b_t^*\}_{t=0}^T, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r, \lambda)$ in each $t \leq T-2$, the rates γ_0 to γ_{T-2} are identified from this vector by (25). Intuitively, the right side of (25) is the portion of piece rates that cannot be explained by the process of learning about ability alone and so is informative about the degree of human capital accumulation over a worker's career.[35]

6.2 Identification of the Augmented Model

We now extend our identification argument to the case in which there exists unobserved heterogeneity among workers in any of the primitive parameters of the model except for δ or m_θ—recall that these parameters are fixed in the argument in the previous subsection. We show that the model is identified even when wages are measured with error and when the law of motion of human capital depends nonparametrically on effort.

[34] Alternative normalizations are possible. For instance, the parameters λ, γ_0 to γ_{T-1}, and m_θ are all identified from the piece rates b_0^* to b_T^* and the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ if $\gamma_{T-2} = \gamma_{T-1}$. We show that the parameters $\{\beta_t\}_{t=0}^{T-1}$ are identified up to $\beta_{T-1} = 0$.

[35] In these arguments, we have assumed that the coefficient of risk aversion r is constant, whereas the rates of human capital accumulation vary over time. Alternatively, we could have specified $\gamma_t \equiv \gamma$ and allowed r to vary with experience. In this latter case, the identification argument would be virtually identical, provided, say, that the risk aversion parameters satisfy $r_{T-1} = r_T$. The parameter r_T would then be identified from the last-period piece rate. Once γ is identified from the piece rate in $T-1$ and λ is recovered as argued above, piece rates in previous periods would be sufficient to identify the coefficients of risk aversion in periods 0 to $T-2$ by (25). When the rates of human capital acquisition vary over time, we do not need to impose any additional condition such as $r_{T-1} = r_T$, since a natural exclusionary restriction arises from T being the last experience year, so the rate γ_T does not appear in any expression.

Unobserved Heterogeneity and Measurement Error. Suppose there exist J types of workers who differ in their distributions of initial ability, shocks to output, and shocks to ability; their degree of risk aversion; and their human capital process. Mirroring the identification argument of the baseline model, we assume that the terms β_0 to β_{T-1} in the human capital process are zero for all J groups—this assumption can be easily relaxed. Each group j is observable to model agents but not to the econometrician. Denote the probability that a worker is of type j by π_j, and let $\sigma^2_{\theta j}$, $\sigma^2_{\varepsilon j}$, $\sigma^2_{\zeta j}$, r_j, $1-\lambda_j$, and γ_{jt} be, respectively, the variance of the initial distribution of ability, the variance of the output shocks, the variance of the ability shocks, the risk aversion parameter, the depreciation rate of human capital, and the period-t rate of accumulation of human capital for type-j workers. The equilibrium characterization in Proposition 1 holds for each type. Let e^*_{jt}, k^*_{jt}, and b^*_{jt} then be, respectively, the equilibrium effort, stock of human capital, and piece rate in period t for type-j workers. By (3), the wage of worker i of type j with ability θ_{ijt} in t is

$$w_{ijt} = (1-b^*_{jt})\mathbb{E}[\theta_{ijt} + k^*_{jt} + e^*_{jt}|I_{it}] + b^*_{jt}(\theta_{ijt} + k^*_{jt} + e^*_{jt} + \varepsilon_{ijt}),$$

which is normally distributed. Thus, the distribution of wages in each period is a finite mixture of normal distributions. Since finite mixtures of normal distributions are identifiable (Teicher [1963]), both the mixture weights $\{\pi_j\}_{j\in J}$ and the component distributions are identified in each period, and so are their component means $\{\mathbb{E}_j[w_{ijt}]\}_{j\in J}$. Likewise, the variable component of the wage of worker i of type j with ability θ_{ijt} in t, $v_{ijt} = b^*_{jt}(\theta_{ijt} + k^*_{jt} + e^*_{jt} + \varepsilon_{ijt})$, is normally distributed. Then, the distribution of the variable component of wages in each period is also a finite mixture of normal distributions with the same component weights as the corresponding finite mixture distribution of wages. Thus, for each worker type j and period t, mean variable wages $\mathbb{E}_j[v_{ijt}]$ are identified so that, as in the baseline model, the piece rate of type-j workers in t is identified as $b^*_{jt} = \mathbb{E}_j[v_{ijt}]/\mathbb{E}_j[w_{ijt}]$.[36] The rest of the argument proceeds as in the baseline case. First, for each type j, the variances $(\sigma^2_{\theta j}, \sigma^2_{\varepsilon j}, \sigma^2_{\zeta j})$ are identified from the piece rates $\{b^*_{jt}\}^T_{t=0}$ and the second moments of the distributions of wages in the first two experience years. Next, for each type j, the preference and human capital parameters $(r_j, \lambda_j, \{\gamma_{jt}\}^T_{t=0})$ are identified from $(\{b^*_{jt}\}^T_{t=0}, \sigma^2_{\theta j}, \sigma^2_{\varepsilon j}, \sigma^2_{\zeta j})$ up to $m_{\theta j}$.

Proposition 5. *Suppose that each worker is one of J types. For each type j, the piece rates $\{b^*_{jt}\}^T_{t=0}$ and the variances $(\sigma^2_{\theta j}, \sigma^2_{\varepsilon j}, \sigma^2_{\zeta j})$ are identified from a panel of wages and their variable components. Once piece rates and the variances $(\sigma^2_{\theta j}, \sigma^2_{\varepsilon j}, \sigma^2_{\zeta j})$ are identified, the risk aversion parameter r_j is identified from the piece rate b^*_{jT}, and the human capital depreciation rate $1-\lambda_j$ and accumulation rates $\{\gamma_{jt}\}^{T-1}_{t=0}$ are identified from the piece rates $\{b^*_{jt}\}^T_{t=0}$ and average type-specific wages in $T-1$ and T up to mean initial ability $m_{\theta j}$ for each j.*

Proposition 5 immediately extends to the case in which wages and their fixed and variable components are measured with error, provided this error is additive and normally distributed; see Appendix D for the case in which

[36] The correct pairing of the components of the mixtures of total and variable wages in each t is possible by their mixing weights, since the weights of these mixtures are identical type by type. Then, simply imposing the constraint that types be ordered—say, by the size of their mixing weights—not only resolves the usual label ambiguity of finite mixture models but also allows for such pairings.

measurement error is correlated over time. Note that through this latent-type formulation in which workers differ in their ability distribution and human capital process in an unrestricted way, the model accommodates alternative settings in which workers of higher ability may be more or less efficient at acquiring new skills. This more general setup thus relaxes the impact of our functional-form assumptions by leading to a flexible dependence of wages on ability, uncertainty about it, human capital, risk, and workers' risk attitudes.[37]

More General Human Capital Process. Consider now the case in which the law of motion of human capital is $k_{it+1} = \lambda k_{it} + F_t(e_{it})$. We show that this version of the model is also identified if information on workers' performance is available in addition to information on wages. This information is present in the firm-level data sets we examine and in many other commonly used ones (Frederiksen et al. [2017]). For ease of exposition, we start by assuming that the available performance measure is a noisy measure of a worker's effort and later discuss the case in which it provides a noisy signal of both a worker's effort and human capital. Let $p_{it} = e_{it} + \eta_{it}$ then be the performance measure of worker i in period t observed by the econometrician, where η_{it} is a continuously distributed noise term independent across workers and over time with cumulative distribution function G with known mean.[38]

Suppose the equilibrium is such that effort choices and piece rates are the same for all workers and depend only on time, and let e_t^* and k_t^* be, respectively, a worker's equilibrium effort and stock of human capital in period t; we present conditions under which this is the case in Appendix A. It follows from (3) that $\mathbb{E}[w_{it}] = m_\theta + k_t^* + e_t^*$. Since $\mathbb{E}[p_{it}] = e_t^* + \mathbb{E}[\eta_{it}]$ and $\mathbb{E}[\eta_{it}]$ is known, both e_t^* and k_t^* in each t are identified from average wages and average performance in t up to m_θ. Observe next that $\mathbb{E}[w_{iT}] - \lambda\mathbb{E}[w_{iT-1}] = k_T^* + e_T^* - \lambda(k_{T-1}^* + e_{T-1}^*) + (1-\lambda)m_\theta$. Hence, λ is identified from the vector $(k_{T-1}^*, e_{T-1}^*, k_T^*, e_T^*)$ up to m_θ. Now, since $k_{t+1}^* = \lambda k_t^* + F_t(e_t^*)$, we can identify $(F_0(e_0^*), \ldots, F_{T-1}(e_{T-1}^*))$ from λ and the sequence of equilibrium efforts and human capital from 0 to T. Thus, if the functions F_t do not depend on experience or, alternatively, if they do and any of the parameters σ_θ^2, σ_ε^2, σ_ζ^2, r, λ, and γ_t vary across observable groups of workers so that different choices of effort are induced among different groups in each t, then these functions are identified from $(F_0(e_0^*), \ldots, F_{T-1}(e_{T-1}^*))$ and $(e_0^*, \ldots, e_{T-1}^*)$.[39] The identification of piece rates and the remaining parameters follows by the same argument as the proof of Proposition 4. An analogous argument applies when the econometrician observes only a discrete version of p_{it} if G is known; see Appendix C.

The argument so far has relied on a specific functional form for the performance measure p_{it}. In Appendix C, we

[37] We do not estimate this more general version of the model, since our baseline model already fits the data quite well. See Section 7.

[38] That this additional outcome measure is informative about effort (or human capital) is a key step to separately recover the experience profile of effort and human capital in this more general case. As in a standard factor model, the paths of effort and human capital can be identified if the signals about effort and human capital observed by the econometrician—wages and performance, in our case—are common to multiple measurements but the noise in these measurements is not (see, for instance, Cunha et al. [2010]). These conditions are satisfied in our case, since observed wages and performance depend on effort and human capital up to independent measurement errors, ε_{it} and η_{it}.

[39] Our identification argument holds regardless of the length of the time interval between two consecutive periods. So, when the functions F_t are independent of t, an increase in the frequency of the data allows us to identify the common function F at a greater number of points in its support. When the functions F_t depend on t, it is easy to see from the first-order conditions for effort that variation in σ_θ^2, σ_ε^2, σ_ζ^2, r, λ, or γ_t among workers, say, with different age at entry or who entered in the firm in different years, would induce variation in effort in each t that would allow us to identify F_t at every possible equilibrium choice of effort in t.

show that we can extend this argument to the more general case in which $p_{it} = f_t(e_{it}, k_{it}) + \eta_{it}$, where $f_t : \mathbb{R}^2 \mapsto \mathbb{R}$ is a known differentiable function nondecreasing in each of its arguments such that $f_t(\cdot, k_{it})$ is surjective for each $k_{it} \in \mathbb{R}$ and $\partial f_t(e_{it}, k_{it})/\partial e_{it} \neq \partial f_t(e_{it}, k_{it})/\partial k_{it}$ for all $(e_{it}, k_{it}) \in \mathbb{R}^2$. These assumptions, which are trivially satisfied in the case just discussed, imply that on average, higher effort or human capital cannot lead to lower performance, any performance measure is possible for any value of a worker's human capital, and the performance measure is more or less sensitive to changes in effort than to changes in the stock of human capital.[40]

The availability of the performance measure discussed prompts the question of why firms would not offer contracts in which they condition wages not only on output but also on this measure. As argued by Hölmstrom [1979], firms should do so as long as a worker's output is not a sufficient statistic for this additional performance measure. A sizable literature, though, has documented that firms tend to have more information about workers' performance than the information contracts are conditioned on; see the discussion and references in Baker [1992], for instance. A common explanation for this feature of contracts is that although they are observable, performance measures often are not verifiable or are manipulable by workers. When workers' ability is uncertain, although firms cannot or may not want to explicitly link wages to all performance measures, they can still use them to form expectations about workers' ability, which influence offered contracts, even if contracts do not explicitly depend on all these measures. In Appendix D, we account for this effect of additional performance measures on the inference process about ability and show that our characterization and identification results extend to this case as well.

7 Estimation

In this section, we describe the estimation of the model, discuss the parameter estimates, examine the fit of the model to the data, and compare our parameter estimates with analogous ones in the literature.

Estimation Sample. We estimate the model using the well-known BGH data presented in Section 2. This firm-level data set has been extensively studied in the literature and therefore provides a natural starting point for investigating how wages and performance pay vary over the life cycle. The BGH data are also administrative and of high quality (Baker et al. [1994a]) and so less likely to be contaminated by measurement error than commonly used survey data such as the PSID. Crucially, by providing a long panel covering workers of all experience levels, the data permit a meaningful life-cycle analysis. Sample size, however, declines rapidly after 40 years of experience—the maximum level of experience is 47 years—so we exclude observations above this 40-year cutoff. The resulting sample consists of more than 22,000 person-year observations on male managers whose average age is 40 years, with a standard

[40] We can extend the analysis to the case in which the performance measure depends on worker ability by noting that if $p_{it} = f_t(e_{it}, k_{it}, \theta_{it}) + \eta_{it}$, then $\widehat{p}_{it} = \mathbb{E}_\theta[f_t(e_{it}, k_{it}, \theta_{it})] + \eta_{it}$, where $\mathbb{E}_\theta[f_t(e, k, \theta)]$ is the expectation of $f_t(e, k, \theta)$ with respect to θ, plays the role of the performance measure considered so far. Indeed, since we can identify the distribution of workers' abilities in any period t from observed wages and their variable component up to m_θ, we can treat $\widehat{f}_t(e_{it}, k_{it}) = \mathbb{E}_\theta[f_t(e_{it}, k_{it}, \theta_{it})]$ as a known function. It is easy to provide conditions on the functions f_t under which the functions $\hat{f}_t$ satisfy the conditions for identification discussed.

deviation of 9 years.[41] The modal employee in our data holds a college degree. At entry in the firm, on average, managers are 33 years of age, with a standard deviation of 7 years, and have 11 years of labor market experience, with a standard deviation of 8 years. Wage profiles in our data are comparable with those documented in the literature. For instance, the log wages of male college-educated workers increase by 0.67 log points during the first 30 years of labor market experience in our data. This estimate is consistent with a wage growth of about 1 log point documented by Elsby and Shapiro [2012] on the basis of cross-sectional census data between 1960 and 2000. Rubinstein and Weiss [2006] find similar estimates using the PSID and the National Longitudinal Survey of Youth.

Parameterization. In estimation, we fix the discount factor δ at 0.95 and let t range from 1 to 40. Recall that we have assumed the effort cost function $g(e) = e^2/2$. To keep our specification parsimonious, we specify the rates of human capital accumulation according to a polynomial of degree two in experience, $\gamma_t = \psi_0 + \psi_1(t-1) + \psi_2(t-1)^2$. In this baseline exercise, we also assume that the term β_t is zero in each t.[42] As a result, we estimate eight parameters: the parameters σ_θ^2, σ_ε^2, and σ_ζ^2 of the learning process about ability, those governing the human capital acquisition process—namely, ψ_0, ψ_1, and ψ_2 and the human capital depreciation rate $1-\lambda$—and the coefficient of absolute risk aversion r. We estimate these parameters by equally weighted minimum distance targeting 120 moments: the piece rate of the wage contract measured by the ratio of average variable pay to average total pay, the variance of wages, and cumulative wage growth, measured by the difference $\mathbb{E}[w_{it}] - \mathbb{E}[w_{i1}]$ in average wages between experience t and experience 1, for each of the first 40 years of labor market experience.[43]

Table 1: Estimates of Model Parameters

Parameters	Estimates	Standard Errors
σ_θ^2, variance of initial ability	2,024.099	0.0009736
σ_ε^2, variance of shock to output	267,019.845	0.0638429
σ_ζ^2, variance of shock to ability	29.458	0.0000632
ψ_0, coefficient of degree 0 of γ_t	0.892	1.68E-07
ψ_1, coefficient of degree 1 of γ_t	0.035	6.10E-08
ψ_2, coefficient of degree 2 of γ_t	-0.001	1.25E-09
λ, fraction of undepreciated human capital	0.955	3.48E-08
r, coefficient of relative risk aversion	0.0002	9.75E-13

[41] Frederiksen et al. [2017] report several regularities in terms of the distribution of wages and performance management systems across the BGH and five other firm-level data sets. These regularities suggest that the patterns in the BGH data are representative of common compensation and performance management practices. Indeed, many features of the BGH data have been replicated by other studies and are now considered stylized facts about careers in firms. See, for instance, Waldman [2012] on this point. We focus on males because they make up a large majority of the data and the careers of female workers in the 1980s were significantly more likely to be interrupted.

[42] See below the discussion of the estimates for the case in which $g(e) = c^2/2$ with c a free parameter. In the Supplementary Appendix, we provide the estimates of a more general version of the model, in which the law of motion of human capital is $k_{it+1} = \lambda k_{it} + \gamma_t e_{it} + \beta \iota_{it}$, where ι_{it} represents a standard learning-by-doing investment in human capital that accrues for any period of employment and so equals 1 in t if worker i is employed and 0 otherwise. We find estimates very close to those reported here for the parameters that are common across these two versions of the model and model fit only slightly improved compared with that of the baseline model.

[43] We compute these statistics after winsorizing the top and bottom 1% of the distribution of wages at each level of experience and controlling for year, education, race, and individual-specific unobserved effects. All targeted moments are scaled in estimation to be of comparable magnitude. For the properties of the minimum distance estimator, see Newey and McFadden [1994].

Estimates. Table 1 reports the estimates of the model parameters together with their asymptotic standard errors. The large sample size, our parsimonious theoretical framework, and the fact that the model—with eight parameters estimated by targeting 120 moments—is highly overidentified all contribute to the small standard errors. For a sense of magnitudes, recall that wages are measured in thousands of 1988 dollars. The estimates reveal key properties of the process of learning about ability and human capital acquisition at the firm. Consider first the parameters of uncertainty and learning. We estimate that the standard deviation of the distribution of initial ability σ_θ is 44.99 thousand dollars and the standard deviation of the distribution of shocks to ability σ_ζ is 5.43 thousand dollars. Together, these estimates imply that after 40 years of labor market experience, the standard deviation of ability is 56.33 thousand dollars and so about 25% *larger* than when workers enter the labor market. The estimate of the standard deviation of the distribution of output shocks σ_ε, 516.74 thousand dollars, is an order of magnitude larger than the estimate of σ_θ. Thus, learning would occur very slowly even in the absence of ability shocks.[44] In particular, without shocks to ability ($\sigma_\zeta^2 = 0$), uncertainty about it, as measured by the variance of posterior beliefs σ_t^2, would monotonically decline over time but decrease only by 23% over 40 years of experience. Our estimates, however, imply that uncertainty about ability increases with labor market experience because of the shocks to ability and the slow speed of learning. Indeed, we estimate σ_θ^2 to be much smaller than the limiting value of σ_t^2 given by $\sigma_\infty^2 = [\sigma_\zeta^2 + (\sigma_\zeta^4 + 4\sigma_\zeta^2\sigma_\varepsilon^2)^{1/2}]/2 =$ 2,819.38. After 40 years of experience, the variance of posterior beliefs is already more than 20% higher than when workers enter the labor market. Eventually, uncertainty about ability becomes 1.4 times larger than at entry in the labor market.

Consider next the parameters that govern the human capital process—that is, the parameters ψ_0, ψ_1, and ψ_2 for the rate γ_t of human capital accumulation and the parameter $1 - \lambda$ for the rate of human capital depreciation. Since the positive estimates of ψ_0 and ψ_1 and the negative estimate of ψ_2 overall imply positive values for γ_t at each experience, human capital acquisition occurs through learning-by-doing: efforts to produce output and human capital are complements in that a higher level of e_t in a period raises both output in the period and the future stock of human capital. As the estimates of the accumulation rates are also sizable, the returns to effort in terms of additional human capital are fairly large and imply an accumulation path that is concave in experience, as is apparent from the solid blue line in panel (f) of Figure 5. At the margin, an increase in effort that increases current output by 1 dollar raises the stock of human capital by 89 cents at experience 1, 1.12 dollars at experience 10, 1.17 dollars at experience 20, 1.01 dollars at experience 30, and 63 cents at experience 40. At all levels of experience, the contribution of effort to human capital acquisition is therefore substantial: it increases with experience for younger workers but decreases with experience for older workers after peaking at a marginal return of 1.18 dollars at experience 17.[45] The estimate of the

[44] To see how the learning parameters are pinned down by the experience profile of the variance of wages, note that for T large enough, $\mathrm{Var}[w_{iT}] \approx (\sigma_\theta^2 + T\sigma_\zeta^2) - [1 - (b_T^*)^2]\sigma_\infty^2 + (b_T^*)^2\sigma_\varepsilon^2$. Since $\Delta_{iT} - \Delta_{iT-1} \approx [(b_T^*)^2 - 2(b_{T-1}^*)^2 + (b_{T-2}^*)^2](\sigma_\infty^2 + \sigma_\varepsilon^2)$ with $\Delta_{it} \equiv \mathrm{Var}[w_{it}] - \mathrm{Var}[w_{it-1}]$ and σ_∞^2 depends only on σ_ε^2 and σ_ζ^2, changes in the variance of wages late in the life cycle are informative about σ_ε^2 and σ_ζ^2. By contrast, changes in the variance of wages early in the life cycle are informative about σ_θ^2 and σ_ζ^2. Our estimates suggest that learning occurs more slowly than reported in Lange [2007] or Kahn and Lange [2014] but are comparable to those in Pastorino [forthcoming].

[45] These calculations rely on marginal increases in effort. Based on equilibrium effort levels, we find that human capital increases by nearly

depreciation rate $1-\lambda$, 4.5%, implies that it takes about 15 years for a unit of human capital to depreciate by half.

Our estimate of the coefficient of absolute risk aversion r of 2×10^{-4} is consistent with Handel's [2013] estimates from data on health insurance and medical utilization choices of the coefficient of absolute risk aversion in the interval $[1.9, 3.25] \times 10^{-4}$, as well as the estimates in Barseghyan et al. [2016] from data on home and automobile choices. Since estimates of risk aversion may be difficult to compare across different settings, as preferences and choice problems may vary, we follow Cohen and Einav [2007] and assess the degree of risk aversion by calculating the hypothetical amount X that would make an individual indifferent between accepting or rejecting a lottery with a 50% chance of gaining 100 dollars and a 50% chance of losing X dollars. For a risk-neutral individual, X is 100 dollars, whereas for an infinitely risk-averse individual, X is zero. According to our estimate of r, X is 49 dollars, so we estimate an intermediate degree of risk aversion. Another way to interpret our estimate of r is to convert it to an estimate of *relative* risk aversion (RRA), since the coefficient of absolute risk aversion $A(w)$ evaluated at the wage w is related to the coefficient of relative risk aversion $R(w)$ evaluated at w by $R(w) = wA(w)$. Our estimate of r corresponds to an RRA coefficient of approximately 0.5 at the present-discounted value of average yearly earnings in our sample (in thousands of dollars) over the 40 years of experience we consider. This estimate is in line with the range of estimates in the literature; see, for instance, Chetty [2006], who documents an upper bound of 2.[46]

Decomposing Estimated Piece Rates. Expression (23) in Section 5 decomposes piece rates in each period t into five terms, each of which captures a distinct economic force that determines how performance pay evolves relative to total pay over the life cycle. This decomposition, shown in panel (a) of Figure 5 at the estimated parameter values, provides a useful lens through which to interpret our estimates. Consider each component, starting with the static piece rate $b_t^0 = 1/[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)]$ given by the dashed red line in the panel. This term is small because we estimate the variance of the shocks to output σ_ε^2 to be large, and it slightly decreases over the life cycle, since the variance of the posterior beliefs about ability σ_t^2 increases with experience as shocks to ability accumulate.

The second term of the decomposition, $-b_t^0 R^*_{CC,t}$, corresponds to the dashed green line in panel (a) of Figure 5. Recall that $R^*_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b^*_{t+\tau})(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k})(1-\mu_t)$ is the dynamic marginal benefit of effort due to career concerns. Like the term b_t^0, the term $R^*_{CC,t}$ is small because of the large estimated variance of the shocks to output σ_ε^2. With noisy output signals, the process of learning about ability proceeds slowly: the weights $1 - \mu_t = \sigma_t^2/(\sigma_t^2 + \sigma_\varepsilon^2)$ on output signals in the belief updating equation in (5) are small, and thus the learning process is not very sensitive to new observations about a worker's output. As a result, effort has a small effect on beliefs about ability, implying that career-concerns incentives are weak and so have a limited impact on piece rates.

the same amount as output, for instance, roughly by more than 20 thousand dollars by experience 20.

[46] We also estimated a version of the model with r fixed at 0.00085, which corresponds to an RRA coefficient of 2 at the present-discounted value of average yearly earnings in our sample, but allowing the effort cost function to be $ce^2/2$. Since in this exercise risk aversion is set to be four times larger than the level estimated for our baseline model, this model intuitively implies a lower degree of uncertainty about ability. However, the estimated human capital parameters and insurance component of piece rates are very similar to those for the baseline model.

The fifth term of the decomposition, $-b_t^0 R_{HK,t}^*$, is given by the teal dot-dashed line in panel (a) of Figure 5 and is also quantitatively small. The component $R_{HK,t}^* = \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(b_{t+\tau}^* + R_{CC,t+\tau}^*)$, as discussed, is the dynamic marginal benefit of effort (to produce output) arising from human capital acquisition and is small because both the estimated piece rates and, as just noted, the career-concerns terms are small. Intuitively, the private returns to accumulating human capital are modest because the large estimated noise in output and the corresponding slow speed of learning about ability imply that neither the explicit link between output and wages due to piece rates nor the implicit one due to career concerns is strong. Hence, newly acquired human capital at the margin is only gradually reflected in variable pay, which limits the impact of human capital on incentives for effort.

By contrast, the remaining two terms of the decomposition are quantitatively important. The third term, $-b_t^0 r H_t^*$, given by the dashed orange line in panel (a) of Figure 5, is negative and represents the degree of insurance that the wage contract provides against the uncertainty about ability, as measured by the dispersion in posterior beliefs σ_t^2. To elaborate, recall that $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$ is the increase in the variance of the present-discounted value of lifetime wages resulting from a marginal increase in the piece rate in t. The larger the uncertainty about ability, the greater this variance, since variability in beliefs leads to variability in wages. Given that we estimate σ_t^2 to be substantial, this third term largely explains the relatively low level of piece rates throughout the life cycle. Intuitively, workers face a large degree of wage risk induced by the variation in beliefs about their ability as learning takes place. In order to reduce this lifetime wage risk, firms lower piece rates for insurance reasons by an amount proportional to H_t^*—the more so, the more risk-averse workers are. As the working horizon shortens over time, this insurance motive weakens, despite the increase in uncertainty about ability. These observations explain why the dashed orange line in panel (a) of Figure 5 eventually declines in absolute value with experience.

Another way to understand why the insurance component of piece rates is large is by analogy to basic ideas in asset pricing. A central insight of this literature is that risk-averse investors expect to be rewarded for holding assets whose payouts are high when investors value consumption relatively less and low when investors value consumption relatively more—that is, assets whose payouts comove with the market portfolio are less desirable. In our framework, performance pay in any period is positively correlated with output in the period and, through the process of learning about ability, with future output as well. As a result, contracts that specify wages highly sensitive to performance are less attractive to workers, as such contracts increase the correlation between current output, current wages, and future wages, thus amplifying wage risk. By this logic, then, wage contracts tend to feature relatively low performance pay. Empirically, it turns out that this effect is strong, which confirms the intuition in Harris and Hölmstrom [1982] on the importance of the dynamic insurance against ability risk provided by wage contracts for the evolution of wages with experience, and explains why performance pay is small in the data. But, as we discuss in the next section, this result does *not* imply that performance incentives have a small impact on wages.

The fourth term of the decomposition, $b_t^0 \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, is given by the dashed lavender line in panel (a) of

Figure 5. The term $\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ when γ_t is positive is the present-discounted increase in output and human capital resulting from a marginal increase in effort (to produce output), and it is relatively large for $t < T$ because the estimated rate of human capital accumulation γ_t is substantial, whereas the depreciation rate $1 - \lambda$ is small, as noted. Its large and positive contribution to piece rates outweighs the negative insurance term discussed above but declines in magnitude as experience accumulates, thus imparting to piece rates their characteristic hump shape.

Our estimates of the components of piece rates are consistent with the characterization of their life-cycle profile in Section 5.2. Indeed, according to Proposition 3, when initial uncertainty about ability is small and the variance of shocks to ability is not too large, piece rates are hump-shaped if the rates of human capital accumulation are initially increasing and then decreasing. Our estimates satisfy these conditions, thus confirming those intuitions.

Model Fit: Wages. Figure 4 shows how our estimated model fits the data. The model successfully reproduces the experience profile of the ratio of performance pay to total pay (left panel), the variance of wages (middle panel), and cumulative wage growth (right panel) at each level of experience. Having discussed the pattern of piece rates implied by our estimates, we now turn to their implications for the profiles of the variance of wages and wage growth.

As shown in Appendix C, the variance of wages in our model, $\mathrm{Var}[w_{it}] = (\sigma_\theta^2 + t\sigma_\zeta^2) - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$, is the sum of three terms. The first term, $\sigma_\theta^2 + t\sigma_\zeta^2$, is the variance of ability, which increases over the life cycle because of the accumulating shocks to ability. The second term, $-\sigma_t^2$, is negative and accounts for the fact that learning about ability takes place gradually, typically reducing the uncertainty about it and correspondingly raising the variance of wages, as the dispersion in wages increasingly reflects the dispersion in workers' abilities. Indeed, in the absence of shocks to ability ($\sigma_\zeta^2 = 0$), the first two terms would equal $\sigma_\theta^2 - \sigma_t^2$, which eventually increases to σ_θ^2. As is consistent with our estimates, in the presence of shocks to ability ($\sigma_\zeta^2 > 0$), the sum of the first two terms equals $\sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2$, which *increases* with experience because ability shocks progressively increase the dispersion in workers' abilities. The final term, $(b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$, captures how the dispersion in workers' abilities and output shocks contribute to the variance of wages through the variance of variable pay. Since uncertainty about ability σ_t^2 eventually becomes constant, the large estimated value of σ_ε^2 and the declining portion of the profile of piece rates towards the end of the life cycle imply that this third term is eventually *decreasing*. The combination of these terms allows the model to reproduce the first increasing and then decreasing pattern of the variance of wages in the data, as shown in the middle panel of Figure 4.

Consider now wage growth. Since in each period, firms make on average zero profits, wages on average equal output. As average output varies over time because effort and human capital vary—see the next section—the growth in average wages stems from the growth in effort and human capital. Recall that the estimated rate of human capital accumulation γ_t peaks in the middle of the life cycle and subsequently declines, whereas effort is relatively more constant. Now, early in a worker's career, the accumulation of human capital is rapid. However, as the accumulation rate γ_t declines over the second half of the life cycle and the stock of human capital depreciates, the growth in human capital eventually slows. The result is a concave experience profile for (human capital and) mean wages, which eventually plateaus, as shown in the right panel of Figure 4.

8 Performance Incentives and Wages

Our estimated model provides a laboratory that we can use to explore how performance pay affects wages over the life cycle. These exercises illustrate the importance of performance incentives for human capital acquisition and for the dynamics of wages and their fixed and variable components.

8.1 Impact of Performance Incentives on Wage Growth

In our model, average wages change over time as effort and human capital evolve with experience, since $\mathbb{E}[w_t] - \mathbb{E}[w_1] = (e_t^* - e_1^*) + (k_t^* - k_1^*)$. In panel (b) of Figure 5, we isolate how changes in effort measured by $e_t^* - e_1^*$ and changes in human capital measured by $k_t^* - k_1^*$ contribute to cumulative wage growth over the life cycle. Clearly, the accumulation of human capital accounts for most of the life-cycle increase in wages. By contrast, changes in effort do not appear to substantially contribute to wage growth. Such a decomposition, however, measures only the *direct* effect of effort on wages; it does not not account for its *indirect* effect due to the impact of effort on the process of human capital acquisition. Panel (c) of Figure 5 demonstrates the importance of this indirect effect by comparing the estimated wage growth implied by our model with the counterfactual wage growth that would result if we held effort constant over the life cycle at the average level implied by our estimates. The figure shows that wages would grow much less during the first 30 years of labor market experience in this case. Intuitively, in the baseline model, workers exert greater effort when young, since both career concerns and the returns to human capital acquisition are larger then. Later in life, effort declines, together with career concerns, the returns to human capital acquisition, and piece rates. Panel (c) illustrates that this variation in incentives and effort over the life cycle is central to wage growth.[47]

Another way to assess the contribution of effort to wages is to explore how performance pay affects effort and wages. To do so, suppose that we restrict firms to offering contracts without variable pay ($b_t^* \equiv 0$) at each experience t, as in the original career-concerns model of Hölmstrom [1999]. In this case, firms would lack an important instrument for rewarding performance and thereby encouraging effort and the acquisition of human capital. Panels (e) and (f) of Figure 5 show that the resulting equilibrium profiles of effort and human capital (the dashed red lines in both panels) would be much lower relative to their profiles in the baseline model (the solid blue lines in both panels), especially early in workers' careers. In turn, lower effort and human capital would imply a lower growth in wages over the life cycle, as shown in panel (d) of Figure 5 (the dashed red line): by the 20th year of labor market experience, wage growth would be 30% lower. Hence, although performance pay is small relative to total pay, it has a substantial impact

[47] This exercise implies that models of "passive" human capital acquisition, which specify that human capital is just a function of experience, may risk conflating variation in investment, e_t, with variation in its marginal product, γ_t. Panel (c) shows instead that even when γ_t varies over time, it is important to account for variation in investments to correctly infer the impact of human capital acquisition on wage growth.

on wage growth through its indirect effect on workers' effort because effort augments human capital.

8.2 Impact of Performance Incentives on Wage Inequality

To explore the impact of performance incentives on wage dispersion, we start by decomposing the variance of wages into the variance of fixed and variable pay at the estimated parameter values. When doing so, we find that the variance of variable pay accounts for 44% to 100% of the variance of wages over the first 30 years of labor market experience; see panel (a) of Figure 6.[48] Thus, although performance pay represents only a small fraction of total pay at any experience, it is highly variable and so responsible for a large portion of the variability of wages over the life cycle. Panel (b) of Figure 6 further shows that uncertainty about ability is a major source of this variability. In this panel, we compare the variance of wages implied by the model (the solid blue line) with the counterfactual variance of wages that would result *at the estimated piece rates* if ability was muted altogether across workers and over time (the dashed lavender line)—that is, when $\sigma_\theta^2 = \sigma_\zeta^2 = 0$. The difference between the two lines significantly increases over the life cycle as shocks to ability accumulate, thereby contributing to the increase in the variance of wages over time.

This decomposition, however, ignores how wage contracts may differ in the absence of heterogeneity in ability among workers. To measure the variance of wages that would result if workers were homogeneous in their ability, we need to take into account how wage contracts would change in response to the new distribution of abilities. Intuitively, if workers experienced neither uncertainty about ability nor shocks to it, they would face much less risk, so wage contracts would naturally feature higher-powered incentives in the form of higher piece rates, as the trade-off between risk and incentives would be less severe. Higher piece rates could then lead to an overall increase, rather than a decrease, in the variance of wages. More generally, a tension exists between ex-ante wage risk, arising from the initial dispersion in ability among workers, and ex-post wage risk, arising from the variability of fixed and variable pay.

This is precisely what we find when we set $\sigma_\theta^2 = \sigma_\zeta^2 = 0$ and take into account firms' incentives to offer higher piece rates in reponse to the lower uncertainty. The resulting variance of wages is shown by the dashed lavender line in panel (c) of Figure 6, and is more than six times larger than that in the baseline model, represented by the solid blue line. Panel (d) of Figure 6 reports the profile of piece rates in the absence of uncertainty about ability, which are up to three times as large as those in the baseline. These much higher piece rates, in turn, amplify any residual productivity risk faced by workers, leading to a much larger wage dispersion. Hence, compressing the dispersion in ability among workers ex ante induces firms to offer contracts with a higher sensitivity of pay to performance, which more than compensates for the lower variance in ability, giving rise, on balance, to a higher life-cycle variability of wages.

Although stylized, this exercise illustrates the importance of accounting for the endogeneity of the wage structure, as defined by the composition of wages in terms of fixed and variable pay, when assessing the role of alternative

[48] See the related findings by Lemieux et al. [2009] on the importance of the incidence of performance pay for wage inequality. Using PSID data, these authors estimate that the increased prevalence of performance pay between the late 1970s and the early 1990s accounts for about 21% of the increase in the variance of (log) wages over this period.

sources of wage dispersion, in particular heterogeneity in workers' ability. Specifically, this exercise implies that popular reduced-form decompositions of the variance of wages can be misleading, as they implicitly assume that the degree to which firms' attributes, including firm-level or "output" shocks, are reflected in wages does not vary with the degree of heterogeneity in workers' ability or, more generally, the level of uncertainty and risk in the labor market (see Abowd et al. [1999], Card et al. [2013], and, for the importance of the pass-through of firm-level shocks to wages, Guiso et al. [2005]). Here, we have shown that this premise may not always be warranted, since the wage structure is a key endogenous dimension through which firms' characteristics and shocks are transmitted to wages, which depends on the distribution of workers' abilities. In particular, measuring the contribution to wage inequality of "worker" and "firm" heterogeneity as (linearly) separate primitive components affecting wages may be inaccurate when performance incentives matter. In fact, our analysis implies that these sources of wage inequality are interdependent and their impact on wages is mediated by their dispersion for reasons of incentive provision. As a result, once firms' incentives to offer contracts with different sensitivities of wages to performance in response to different levels of uncertainty or output risk are considered, *lower* dispersion in ability (or output risk) can be associated with *greater* wage dispersion, although small decreases in ability dispersion (or output risk)—for given piece rates—lead to lower wage dispersion.[49]

9 Extensions

Here, we explore three dimensions along which our framework can be augmented: wage bargaining, differences in productivity among firms, and multi-job firms. See Appendix D for omitted details.

Bargaining. Although we have assumed that wages are competitively determined, our analysis, starting from the characterization of equilibrium to the identification of the model, would apply essentially unaltered if we allowed firms and workers to bargain over wages. To elaborate, suppose that in each period, workers capture a fraction $\alpha \in (0, 1]$ of the expected value of their match with a firm, and firms absorbed the residual fraction. Then, workers' effort would satisfy the same first-order condition as in (8), but the equilibrium piece rate would now be given by $b_t^* = b_t^0[\alpha(1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}) - R_{CC,t}^* - R_{HK,t}^* - rH_t^*]$. The only difference is that the term $R_{CC,t}$ would be $R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (\alpha - b_{t+\tau})(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k})(1 - \mu_t)$ instead of (9). Identification would proceed as in Section 6 up to α, which can be simply identified by the ratio of total wages paid to firm revenues (see, for instance, Flinn [2006]).

Firm Heterogeneity. By suitably reinterpreting the term β_t in the law of motion for human capital in (2), our model is consistent with settings in which firms differ in their productivity. For instance, suppose that firms are characterized by the productivity parameter p so that the output of worker i when employed by a firm of productivity $p_{it} = p$ in t

[49] Since $\text{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$, it is easy to show that this result holds at any level of experience for a given dispersion in initial ability σ_θ^2, variance of ability shocks σ_ζ^2, and degree of workers' risk aversion r, if the variance of output shocks σ_ε^2 is large enough. Namely, when σ_θ^2 and σ_ζ^2 are decreased to zero, the variance of wages becomes $-\sigma_{t,n}^2 + (b_{t,n}^*)^2(\sigma_{t,n}^2 + \sigma_\varepsilon^2)$, where the subscript n stands for "no uncertainty," $b_{t,n}^* > b_t^*$, and $\sigma_{t,n}^2 < \sigma_t^2$. Then, a sufficient condition for the variance of wages in t to be higher in the absence of uncertainty about ability is that $[(b_{t,n}^*)^2 - (b_t^*)^2]\sigma_\varepsilon^2$ exceeds $\sigma_\theta^2 + t\sigma_\zeta^2$, which can be guaranteed when σ_ε^2 is large enough.

is $y_{it} = p_{it} + \theta_{it} + k_{it} + e_{it} + \varepsilon_{it}$, omitting constants—in our baseline model, which we apply to workers employed at one firm, p is absorbed in β_t. Assume that in each period, workers are matched with a set of firms that are possibly heterogeneous in their productivity and Bertrand-compete for workers in wages. Let there exist at least two firms with each productivity level, so that firms make zero expected profits in equilibrium and workers are employed by the most productive firms—this setup can be relaxed, as in Pastorino [forthcoming]. In this case, if the firm with the highest productivity changes over time—owing to productivity shocks, for instance—then the wage equation implied by our model would be analogous to that in Bagger et al. [2014], where $w_{it} = p_{it} + \theta_{it} + k_{it} + \varepsilon_{it}$, p_{it} denotes a firm effect (up to a discount term that depends on a worker's wage bargaining history) and k_{it} denotes the human capital that a worker (exogenously) accumulates with experience. In addition, it is easy to show that when p_{it} is stochastic or workers experience each period a shock to their productivity that affects their output at any potential firm before wages are determined, such a version of our model can generate very flexible patterns of worker turnover across firms.

Multi-Job Firms. Our analysis so far has abstracted from the possibility of workers' job mobility within firms. We now discuss how our framework can be extended to account for it within a multi-tasking setup in which we interpret different jobs as placing different weights on different tasks. For instance, in the BGH firm, higher-level jobs of the firm's hierarchy consist primarily of general management and strategic planning tasks, whereas for the most part, lower-level jobs entail tasks that involve creating or selling products. Indeed, as a worker progresses through the job hierarchy, the fraction of the former group of tasks increases, whereas the fraction of the latter group of tasks decreases.[50] Consider then a simple multi-tasking extension of our model in which the output of a worker at task $\ell \in \{1, 2\}$ in period t, omitting the worker subscript, is $y_{\ell t} = \beta_\ell \theta + \alpha_\ell k_t + e_{\ell t} + \varepsilon_{\ell t}$, where $e_{\ell t}$ is the worker's effort at task ℓ, θ is the worker's ability, which we assume is time invariant for simplicity, k_t is the worker's human capital, $\varepsilon_{\ell t}$ is the shock to output (or output noise) at task ℓ, and α_ℓ and β_ℓ are non-negative constants. The ratio $e_{\ell t}/(e_{1t} + e_{2t})$ is the weight that the worker places on task ℓ in t. Output shocks are independent across time, workers, and tasks with $\mathrm{Var}[\varepsilon_{\ell t}] = \sigma^2_{\varepsilon,\ell}$. A worker's stock of human capital evolves over time according to $k_{t+1} = \lambda k_t + \gamma_{1t} e_{1t} + \gamma_{2t} e_{2t}$. The (monetary) cost of the effort pair (e_1, e_2) is $c(e_1, e_2) = e_1^2/2 + e_2^2/2 + \nu(e_1 - \underline{e})(e_2 - \underline{e})$, where $|\nu| < 1$ and $\underline{e}$ is low enough for equilibrium effort to be greater than $\underline{e}$ in every period. Note that when $\nu > 0$, tasks are substitutes in the sense that more effort at one task increases the marginal cost of effort at the other task. When $\nu < 0$ instead, tasks are complements in the sense that more effort at one task decreases the marginal cost of effort at the other task.[51] The wage in period t is now $w_t = a_t + b_{1t} y_{1t} + b_{2t} y_{2t}$, where $b_{\ell t}$ is the piece rate associated with the output at task ℓ. Our

[50]Specifically, as described by BGH, at Levels 1 to 4, about 60% of the jobs relate to specific "line" (revenue-generating) business units and correspond to positions that involve direct contact with customers or creating and selling products. Approximately 35% are staff or overhead positions in areas such as accounting, finance, or human resources. At Levels 5 and 6, these two percentages decrease to 45% and 25%, respectively, whereas general management descriptions such as general administration or planning increase to about 30%. At Levels 7 and 8—Level 8 is the highest job level of chairperson-CEO—all jobs are of this latter type.

[51]This framework can be straightforwardly extended to the case of (*i*) task-specific ability and human capital; (*ii*) output shocks correlated across tasks in that $\mathbb{E}[\varepsilon_{1t}\varepsilon_{2t}] = \rho \neq 0$; and (*iii*) more general effort cost functions.

equilibrium characterization easily extends to this multi-tasking case. Since a positive piece rate at one task increases the wage risk associated with the other task, multitasking provides a further rationale for why piece rates are small in the data (Hölmstrom and Milgrom [1991]).

This extension can naturally generate a profile of task assignment with experience such that if we interpret task 1 as "creating or selling products" and task 2 as "management and strategic planning," workers are progressively assigned from jobs in which task 1 is relatively more important to jobs in which task 2 is relatively more important, as we observe in our data.[52] Moreover, if we incorporate heterogeneity in the parameters β_ℓ, α_ℓ, and $\sigma^2_{\varepsilon,\ell}$, then the model can lead to heterogeneous career paths such that some workers progress faster to higher-level jobs than others.

10 Conclusion

The human capital model views workers as engaged in a lifelong process of acquiring new skills. Once formal schooling ends, this process takes place in the workplace, where the effort workers expend at their jobs, by either substituting for the effort to acquire human capital (on-the-job training) or contributing to it (learning-by-doing), becomes central to the accumulation of human capital. This simple insight is the starting point for our exploration of how performance incentives influence the dynamics of wages with labor market experience. Our goal is to examine theoretically and empirically how incentives for effort on the job are affected by human capital considerations, especially when workers' ability is uncertain, and how they in turn shape the structure and evolution of wages over the life cycle.

To this end, we develop and estimate a tractable model of the labor market that integrates three key sources of the dynamics of wages: uncertainty and learning about ability, human capital acquisition, and performance incentives. We use this model to account for the life-cycle profile of wages, their dispersion across individuals, and their composition in terms of fixed and variable pay. This framework nests several known models, including models of learning and matching, standard models of investment in human capital, models of dynamic moral hazard, and "career-concerns" models of learning about ability and performance incentives. We characterize the equilibrium wage contract in this framework and analytically decompose the implied sensitivity of pay to performance into the relative contributions of the basic forces we integrate: the trade-off between output risk and effort incentives characteristic of moral hazard, the insurance against output risk and uncertainty about ability that firms provide to workers through wage contracts, and the implicit incentives for effort arising from workers' desire to affect the market assessment of their ability and accumulate human capital. We prove that the model is identified just from panel data on wages and their fixed or variable components under common assumptions, and obtain simple estimators of the model primitives.

Although variable pay makes up only a small fraction of total pay, our estimates illustrate the centrality of variable

[52]For instance, this occurs if $\beta_1 < \beta_2$ and ability matters more for task 2 than 1. Intuitively, in this case, workers initially demand greater insurance against ability risk at task 2, which translates into a smaller piece rate, and thus less effort, at this task. Over time, as this insurance motive declines, the allocation of workers across tasks increasingly reflects the greater importance of ability at task 2. Alternatively, such job transitions can happen if the rate of human capital accumulation is initially higher at task 1 but eventually becomes higher at task 2.

pay to the dynamics of wages and their components. In particular, we find that performance incentives are a critical source of wage growth and dispersion over the life cycle, owing to the cumulative impact of effort on human capital acquisition and the contribution of variable pay to the variance of wages. We also show the importance of the wage structure as an endogenous mechanism for the transmission of output and ability risk to wages. As a result, although small decreases in the dispersion of ability among workers decrease wage dispersion for a *given* combination of fixed and variable pay, we find that large decreases in the dispersion of ability, by reducing workers' demand for insurance, lead firms to offer contracts providing stronger incentives for effort in the form of a higher fraction of variable pay to total pay. These contracts in turn give rise to a much greater variability of wages across individuals and over time.

Our model rationalizes a novel finding about the life-cycle profile of variable pay that we document: the ratio of variable pay to total pay tends to decline over the second half of workers' careers, precisely when standard models of career concerns predict that variable pay should instead become more and more important. Our estimates suggest that two motives—namely, workers' demand for insurance against the wage risk due to the uncertainty about ability and human capital acquisition—primarily govern variable pay and have effects of opposite sign on the experience profile of variable pay relative to total pay, which explain its low level and peculiar hump shape.

Our estimates assign a crucial role to this first insurance motive. Specifically, since workers face substantial uncertainty about their ability, and, by rewarding high performance, variable pay is positively correlated with ability, wage contracts that specify large variable-pay components are unattractive to workers because they make pay comove with lifetime income. Such contracts thereby compound the wage risk due to the variability of fixed pay over time as beliefs about ability evolve and wage contracts are renegotiated accordingly. We find that this insurance component of wage contracts is quantitatively large and depresses variable pay relative to total pay, especially early in workers' careers. We believe this to be a key rationale for variable pay being small for most workers. Compared with insurance and human capital motives, career-concerns incentives and the contemporaneous trade-off between risk and incentives—the primary component of variable pay in static moral-hazard models—are empirically much less important.

Our analysis has sidestepped questions related to how individuals sort, say, into distinct markets so as to transparently integrate the competing mechanisms of wage growth and dispersion we study within a framework that can be analytically characterized and has empirical content. Such an approach naturally suggests avenues to enrich our analysis and obtain a more complete picture of the forces shaping the structure and dynamics of wages. We hope nonetheless that our results offer a promising first step toward richer models of incentives that can help interpret the sources of the variability of wages across individuals and over time.

References

J. M. Abowd, F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2): 251–333, 1999.

D. A. Ackerberg and M. Botticini. Endogenous matching and the empirical determinants of contract form. *Journal of Political Economy*, 110(3):564–591, 2002.

J. Adda and C. Dustmann. Sources of wage growth. Bocconi University, Mimeo, 2020.

J. Bagger, F. Fontaine, F. Postel-Vinay, and J.-M. Robin. Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, 104(6):1551–1596, 2014.

G. P. Baker. Incentive contracts and performance measurement. *Journal of Political Economy*, 100(3):598–614, 1992.

G. P. Baker, M. Gibbs, and B. Hölmstrom. The internal economics of the firm: Evidence from personnel data. *Quarterly Journal of Economics*, 109(4):881–919, 1994a.

G. P. Baker, M. Gibbs, and B. Hölmstrom. The wage policy of a firm. *Quarterly Journal of Economics*, 109(4): 921–955, 1994b.

L. Barseghyan, F. Molinari, and J. C. Teitelbaum. Inference under stability of risk preferences. *Quantitative Economics*, 7(2):367–409, 2016.

G. S. Becker. Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5, Part 2):9–49, 1962.

Y. Ben-Porath. The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4, Part 1):352–365, 1967.

N. Bloom and J. Van Reenen. Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, 24(1):203–224, 2010.

A. Bowlus and H. Liu. The contributions of search and human capital to earnings growth over the life cycle. *European Economic Review*, 64(1):305–331, 2013.

D. Card, J. Heining, and P. Kline. Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3):967–1015, 2013.

R. Chetty. A new method of estimating risk aversion. *American Economic Review*, 96(5):1821–1834, 2006.

A. Cohen and L. Einav. Estimating risk preferences from deductible choice. *American Economic Review*, 97(3): 745–788, 2007.

F. Cunha. Recent developments in the estimation of production functions of skills. *Fiscal Studies*, 32(2):297–316, 2011.

F. Cunha and J. J. Heckman. Formulating, identifying, and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4):738–782, 2008.

F. Cunha and J. J. Heckman. Decomposing trends in inequality in earnings into forecastable and uncertain components. *Journal of Labor Economics*, 34(S2):S31–S65, 2016.

F. Cunha, J. J. Heckman, and S. Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, 57(2):191–261, 2005.

F. Cunha, J. J. Heckman, and S. M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

C. Dustmann and C. Meghir. Wages, experience and seniority. *Review of Economic Studies*, 72(1):77–108, 2005.

M. W. Elsby and M. D. Shapiro. Why does trend growth affect equilibrium employment? A new explanation of an old puzzle. *American Economic Review*, 102(4):1378–1413, 2012.

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

C. J. Flinn. Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, 74(4):1013–1062, 2006.

J. T. Fox. Estimating the employer switching costs and wage responses of forward-looking engineers. *Journal of Labor Economics*, 28(2):357–412, 2010.

A. Frederiksen, F. Lange, and B. Kriechel. Subjective performance evaluations and employee careers. *Journal of Economic Behavior & Organization*, 134(C):408–429, 2017.

G.-L. Gayle and R. A. Miller. Has moral hazard become a more important factor in managerial compensation? *American Economic Review*, 99(5):1740–1769, 2009.

G.-L. Gayle and R. A. Miller. Identifying and testing models of managerial compensation. *Review of Economic Studies*, 82(3):1074–1118, 2015.

J. Geweke and M. Keane. An empirical analysis of earnings dynamics among men in the PSID: 1968-1989. *Journal of Econometrics*, 96(2):293–356, 2000.

R. Gibbons and K. J. Murphy. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, 100(3):468–505, 1992.

M. Gibbs and W. Hendricks. Do formal salary systems really matter? *ILR Review*, 58(1):71–93, 2004.

T. Gladden and C. Taber. The relationship between wage growth and wage levels. *Journal of Applied Econometrics*, 24(6):914–932, 2009.

L. Golan, G.-L. Gayle, and R. A. Miller. Promotion, turnover and compensation in the executive labor market. *Econometrica*, 83(6):2293–2369, 2015.

W. B. Gordon. On the diffeomorphisms of Euclidean space. *American Mathematical Monthly*, 79(7):755–759, 1972.

L. Guiso, L. Pistaferri, and F. Schivardi. Insurance within the firm. *Journal of Political Economy*, 113(5):1054–1087, 2005.

M. Harris and B. Hölmstrom. A theory of wage dynamics. *Review of Economic Studies*, 49(3):315–333, 1982.

J. J. Heckman, L. Lochner, and C. Taber. Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics*, 1(1):1–58, 1998.

J. J. Heckman, L. Lochner, and R. Cossa. Learning-by-doing versus on-the-job training: Using variation induced by the EITC to distinguish between models of skill formation. In *Designing Inclusion: Tools to Raise Low-End Pay and Employment in Private Enterprise*, Edmund Phelps, Ed., Cambridge:Cambridge University Press, 74–130, 2003.

B. Hölmstrom. Moral hazard and observability. *Bell Journal of Economics*, 10(1):74–91, 1979.

B. Hölmstrom. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182, 1999.

B. Hölmstrom and P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7:24–52, 1991.

L. B. Kahn and F. Lange. Employer learning, productivity, and the earnings distribution: Evidence from performance measures. *Review of Economic Studies*, 81(4):1575–1613, 2014.

F. Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35, 2007.

E. P. Lazear and K. L. Shaw. Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, 21(4):91–114, 2007.

E. P. Lazear and K. L. Shaw. A personnel economics approach to productivity enhancement. *Nordic Economic Policy Review*, 2:209–251, 2011.

E. P. Lazear and K. L. Shaw. Introduction: Firms and the distribution of income: The roles of productivity and luck. *Journal of Labor Economics*, 36(S1):S1–S12, 2018.

T. Lemieux, W. B. MacLeod, and D. Parent. Performance pay and wage inequality. *Quarterly Journal of Economics*, 124(1):1–49, 2009.

L. Lochner and Y. Shin. Understanding earnings dynamics: Identifying and estimating the changing roles of unobserved ability, permanent and transitory shocks. NBER Working Paper 20068, 2014.

L. Lochner, Y. Park, and Y. Shin. Wage dynamics and returns to unobserved skill. NBER Working Paper 24220, 2018.

M. Margiotta and R. A. Miller. Managerial compensation and the cost of moral hazard. *International Economic Review*, 41(3):669–719, 2000.

C. Meghir and L. Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.

R. A. Miller. Job matching and occupational choice. *Journal of Political Economy*, 92(6):1086–1120, 1984.

J. A. Mincer. Schooling, experience, and earnings. *NBER Books*, 1974.

W. K. Newey and D. L. McFadden. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Volume IV, Robert F. Engle and Daniel L. McFadden, Eds., Amsterdam:North–Holland, 2111–2245, 1994.

E. Pastorino. Careers in firms: The role of learning and human capital. *Journal of Political Economy*, forthcoming.

I. Perrigne and Q. Vuong. Nonparametric identification of a contract model with adverse selection and moral hazard. *Econometrica*, 79(5):1499–1539, 2011.

Y. Rubinstein and Y. Weiss. Post schooling wage growth: Investment, search and learning. In *Handbook of the Economics of Education*, Volume I, Eric Hanushek and Finis Welch, Eds., Amsterdam:North–Holland, 1–67, 2006.

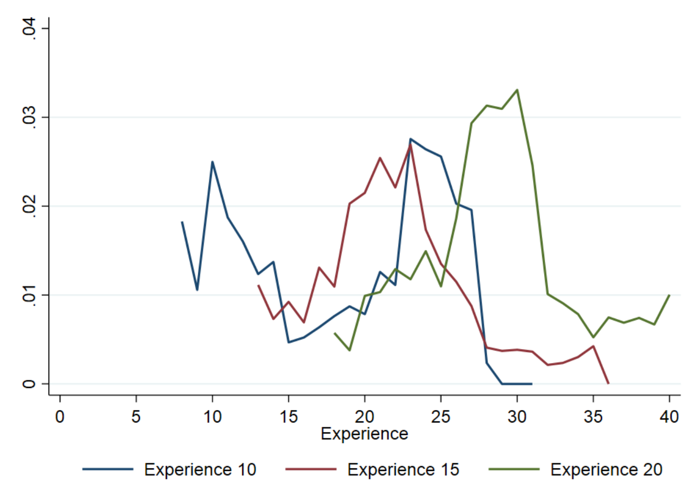
C. Sanders and C. Taber. Life-cycle wage growth and heterogeneous human capital. *Annual Review of Economics*, 4, 2012.

H. Teicher. Identifiability of finite mixtures. *Annals of Mathematical Statistics*, 34(4):1265–1269, 1963.

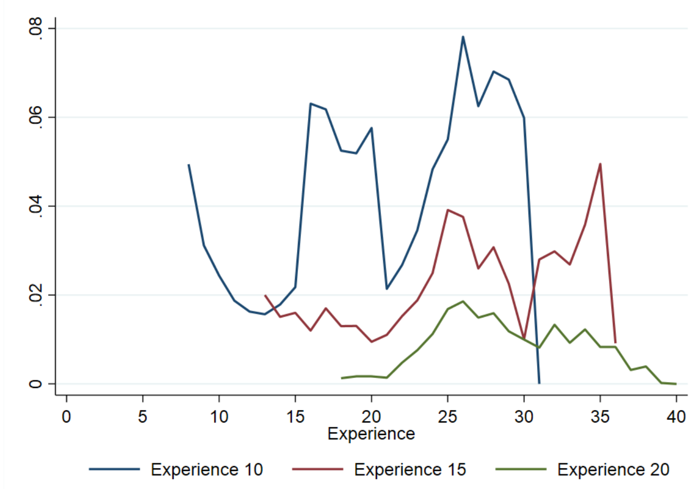
M. Waldman. Theory and evidence in internal labor markets. In *Handbook of Organizational Economics*, Robert Gibbons and John Roberts, Eds., Princeton, NJ:Princeton University Press, 520–572, 2012.

Figure 1: Ratio of Performance Pay to Total Pay by Industry in PSID

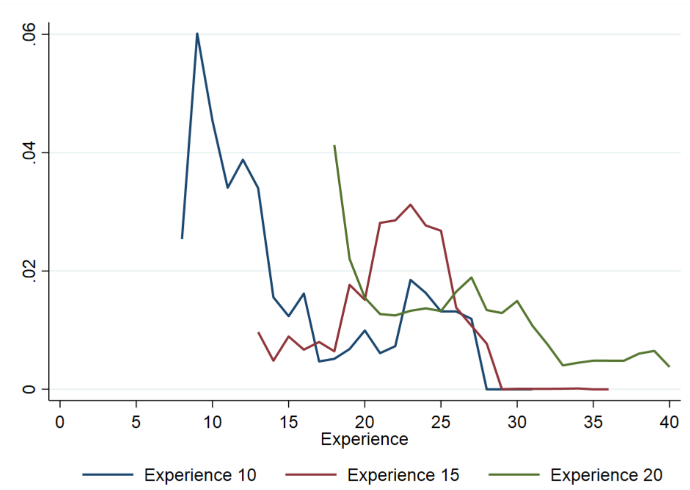
(a) Manufacturing



(b) Transport



(c) Services



(d) FIRE

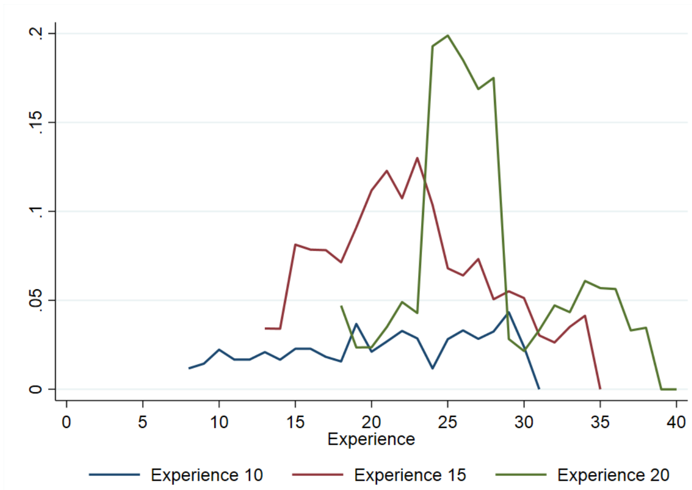


Figure 2: Ratio of Performance Pay to Total Pay in BGH Data

(a) All (b) College (c) No College

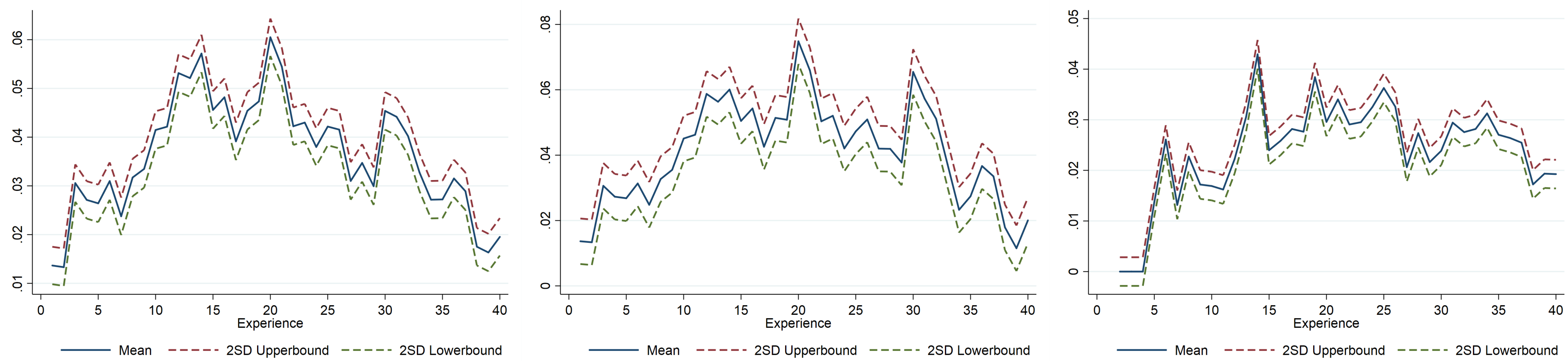


Figure 3: Ratio of Performance Pay to Total Pay in GH Data

(a) All (b) College (c) No College

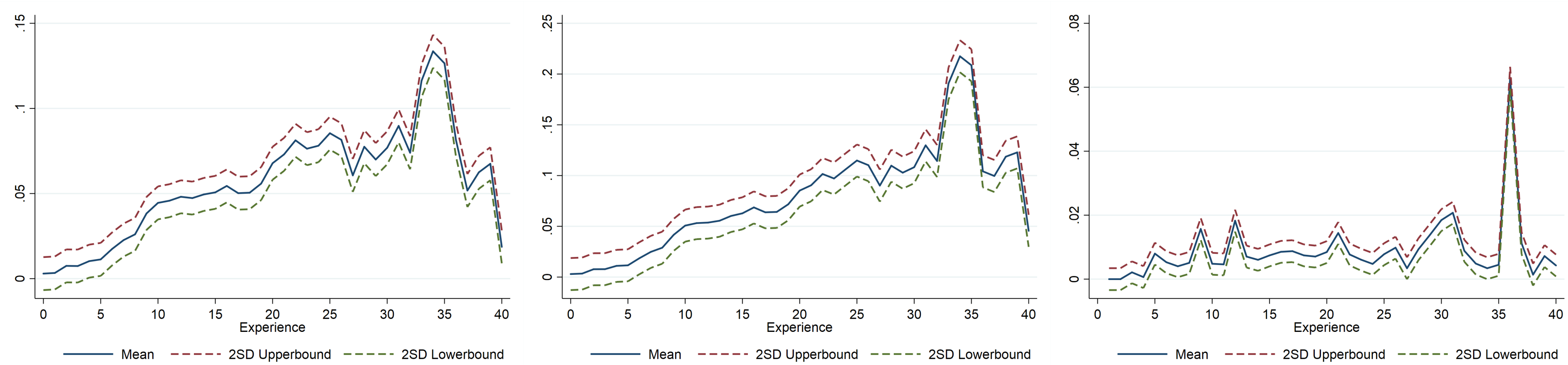


Figure 4: Model Fit to Targeted Moments from BGH Data

(a) Ratio of Performance Pay to Total Pay

(b) Wage Variance

(c) Wage Growth Relative to Initial Wage

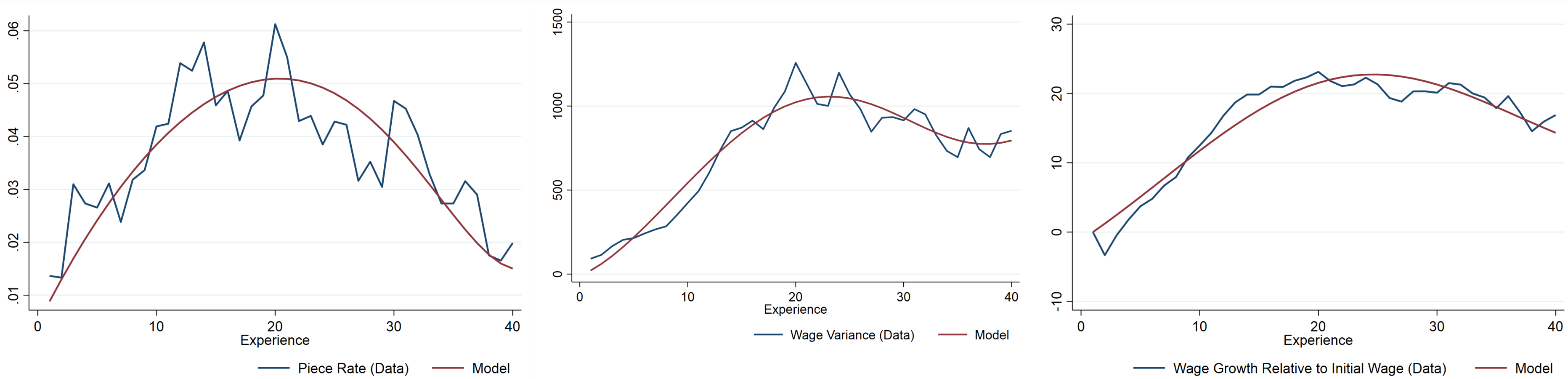


Figure 5: Piece Rates, Wage Growth, Effort, and Human Capital

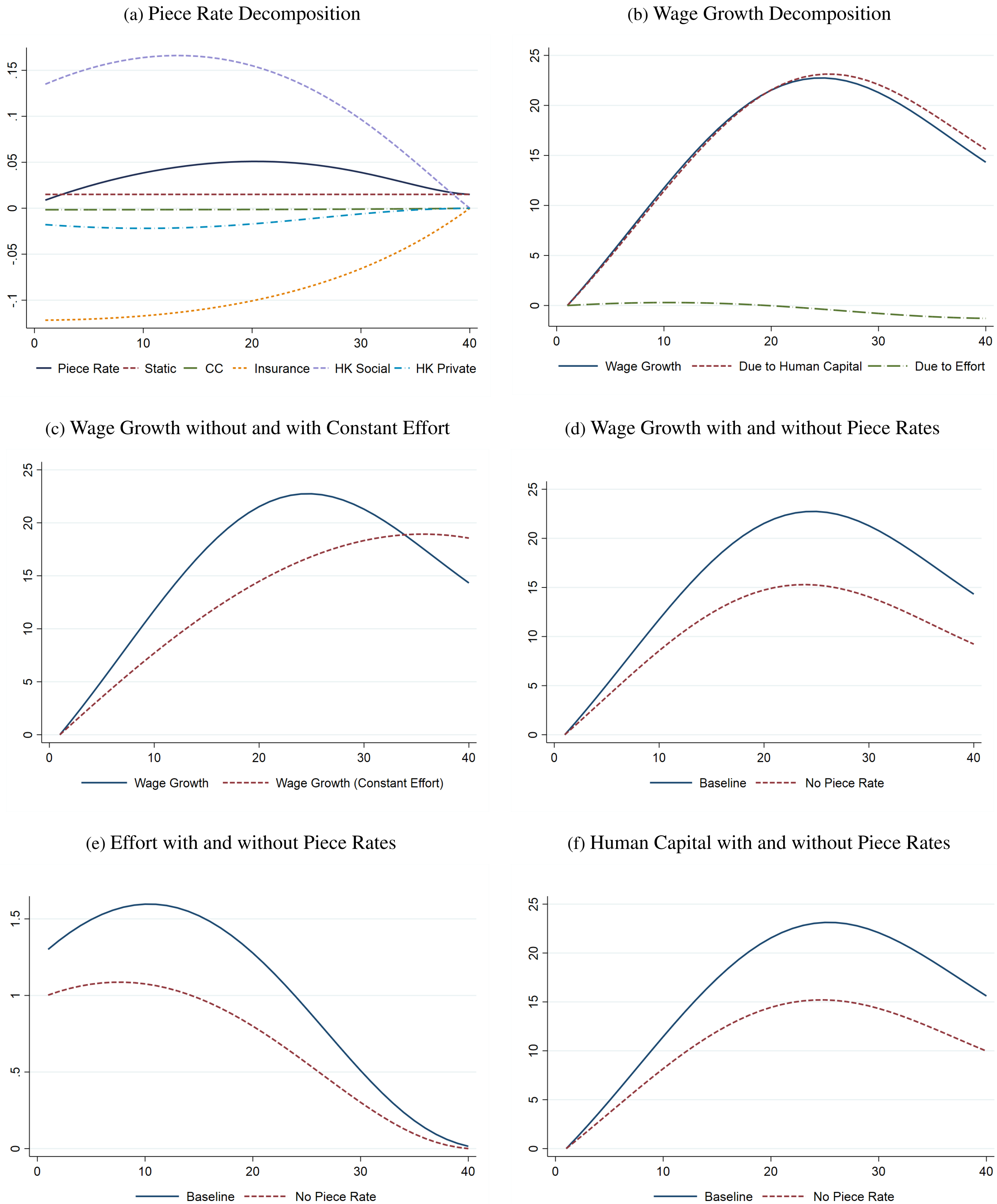
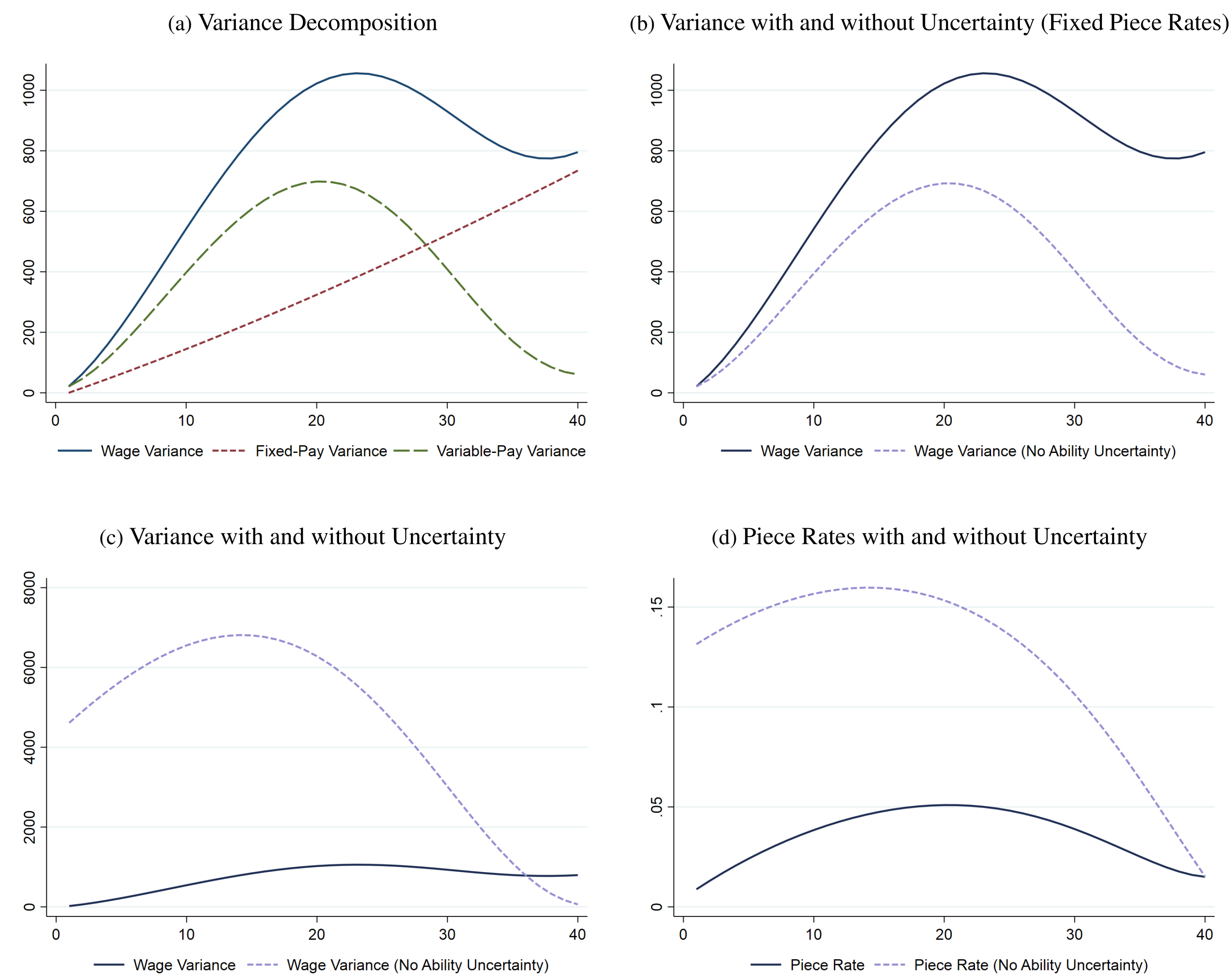


Figure 6: Role of Ability Uncertainty for the Variance of Wages and Piece Rates



A Appendix: Equilibrium Derivation

In this appendix, we work with the more general case in which the law of motion for human capital is

$$k_{it+1} = \lambda k_{it} + F_t(e_{it}), \tag{A1}$$

where $F_t : \mathbb{R} \to \mathbb{R}$ is thrice differentiable and weakly concave with $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$, F_t''' non-positive and non-decreasing, and $\inf_{e\in\mathbb{R}} F_t''(e) > -\infty$. This case reduces to the case of the main text when $F_t(e) = \gamma_t e$ for all t. We first derive the first-order conditions for the optimal choices of effort for a worker when piece rates and the worker's future effort choices depend only on time. We then determine the equilibrium piece rates and show that they are the same for all workers and depend only on time. We conclude by presenting our equilibrium characterization and discussing conditions under which equilibrium piece rates are in the unit interval. Our equilibrium characterization includes Proposition 1 as a special case.

A.1 First-Order Conditions for Effort

We first show that if piece rates for a worker are $\{b_t\}_{t=0}^{T}$ and thus depend only on time, then the first-order condition for the worker's optimal choice of effort in period t when the worker's future behavior depends only on time is

$$e_t = b_t + R_{CC,t} + R_{HK,t}(e_t), \tag{A2}$$

where

$$R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau}) \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k} \right) (1 - \mu_t) \text{ and } R_{HK,t}(e) = F_t'(e) \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b_{t+\tau} + R_{CC,t+\tau}\big). \tag{A3}$$

Note that (A2) reduces to (8) when $F_t(e) = \gamma_t e$. The assumption that $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$ ensures that (A2) always has a solution. This solution need not be an optimal choice of effort for the worker, though. Additional assumptions, which we will discuss, are necessary for this to be the case. We start with the following auxiliary result.

Lemma 5. *Fix $\{\xi_t\}_{t=1}^{T}$. For each period $t \le T - 1$,*

$$\sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau}) \sum_{s=1}^{\tau-1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k} \right) (1 - \mu_{t+s}) \xi_s = \sum_{\tau=1}^{T-t} \delta^\tau \xi_\tau R_{CC,t+\tau}.$$

Proof. The result is trivially true when $t = T - 1$, since $R_{CC,T} = 0$. Fix $t \le T - 2$, and let $u, v \in \mathbb{R}^{T-t-1}$ be such that $u = (\xi_t, \dots, \xi_{T-t-1})$ and $v = \big(\delta^2(1 - b_{t+2}), \dots, \delta^{T-t}(1 - b_T)\big)$. Moreover, let A be the square matrix of order $T - t - 1$ such that $A_{ij} = 0$ if $i < j$ and $A_{ij} = \big(\prod_{k=1}^{i-j} \mu_{t+i+1-k}\big)(1 - \mu_{t+j})$ if $i \ge j$. If we let $\langle v, Au\rangle$ denote the scalar product of the vectors v and Au, then

$$\begin{aligned}
\langle v, Au\rangle &= \sum_{i=1}^{T-t-1} \delta^{i+1}(1 - b_{t+1+i}) \sum_{j=1}^{i} \left(\prod_{k=1}^{i-j} \mu_{t+i+1-k} \right)(1 - \mu_{t+j})\xi_j \\
&= \sum_{i=1}^{T-t} \delta^{i}(1 - b_{t+i}) \sum_{j=1}^{i-1} \left(\prod_{k=1}^{i-1-j} \mu_{t+i-k} \right)(1 - \mu_{t+j})\xi_j,
\end{aligned}$$

where the second equality follows from the change of variable $i \mapsto i - 1$ and the fact that the term $i = 1$ in the sum is zero. Now let D be the diagonal matrix of order $T - t - 1$ such that $D_{ii} = \delta^i$ and denote the transpose of a matrix M by M'. Then, since $\langle v, Au\rangle = v'Au = \langle A'v, u\rangle$, it follows that

$$\langle v, Au\rangle = \langle v, AD^{-1}Du\rangle = \langle (AD^{-1})'v, Du\rangle = \langle (D^{-1})'A'v, Du\rangle = \langle D^{-1}A'v, Du\rangle. \tag{A4}$$

On the other hand, since $(D^{-1}A'v)_i = \delta^{-i}(A'v)_i$, it follows that

$$
\begin{aligned}
(D^{-1}A'v)_i &= \delta^{-i}\sum_{j=1}^{T-t-1} A_{ji}v_j = \delta^{-i}\sum_{j=i}^{T-t-1}\left(\prod_{k=1}^{j-i}\mu_{t+j+1-k}\right)(1-\mu_{t+i})\delta^{j+1}(1-b_{t+1+j}) \\
&= \sum_{j=1}^{T-t-i}\left(\prod_{k=1}^{j-1}\mu_{t+i-k}\right)(1-\mu_{t+i})\delta^{j}(1-b_{t+i+j}) = R_{CC,t+i}
\end{aligned}
$$

for each $1 \leq i \leq T-t-1$—note the change of variables $j \mapsto j+i-1$ in the last equality. So, (A4) implies that

$$
\langle v, Au\rangle = \sum_{i=1}^{T-t-1}\delta^i \xi_i R_{CC,t+i} = \sum_{i=1}^{T-t}\delta^i \xi_i R_{CC,t+i},
$$

where we used the fact that $R_{CC,T} = 0$ a second time. This establishes the desired result. $\square$

Suppose that piece rates are $\{b_t\}_{t=0}^T$, and consider worker i's choice of effort in period t when the worker's future effort choices depend only on time. The argument in the main text—whether the functions $\{F_t\}_{t=0}^T$ are linear or not does not matter—shows that the first-order condition for the worker's choice of effort is

$$
e_t = b_t + \sum_{\tau=1}^{T-t}\delta^\tau \frac{\partial \mathbb{E}[w_{it+\tau}|h_i^t]}{\partial e_t}, \tag{A5}
$$

where $w_{t+\tau}$ and h^t are, respectively, the worker's wage in period $t+\tau$ and history in period t. In what follows, we show that (A5) reduces to (A2). In particular, the worker's optimal choice of effort does not depend on h_i^t.

First, recall from (3) that $w_{it+\tau} = (1-b_{t+\tau})\mathbb{E}[y_{it+\tau}|I_{it+\tau}] + b_{t+\tau}y_{it+\tau}$ for all $1 \leq \tau \leq T-t$, where $y_{it+\tau}$ is the worker's output in period $t+\tau$ and $I_{it+\tau}$ is the public information about the worker that is available in the same period (which depends on $h_i^{t+\tau}$). Let $m_{it+\tau}$ be the worker's reputation in period $t+\tau$; note that $m_{it+\tau}$ depends on $I_{it+\tau}$. Since for each $1 \leq \tau \leq T-t$, the worker's choice of effort in period t affects $\mathbb{E}[y_{it+\tau}|I_{it+\tau}]$ only through its impact on $m_{it+\tau}$, as the other terms in the conditional expectation depend on the worker's *conjectured* effort and stock human capital in period $t+\tau$ and the worker's future effort choices depend only on time, it follows that

$$
\frac{\partial \mathbb{E}[w_{it+\tau}|h_i^t]}{\partial e_t} = (1-b_{t+\tau})\frac{\partial \mathbb{E}[m_{it+\tau}|h_i^t]}{\partial e_t} + b_{t+\tau}\frac{\partial \mathbb{E}[y_{it+\tau}|h_i^t]}{\partial e_t}
$$

for all $1 \leq \tau \leq T-t$. By (A1) and the fact that behavior from period $t+1$ on depends only on time,

$$
\frac{\partial \mathbb{E}[y_{it+\tau}|h_i^t]}{\partial e_t} = \lambda^{\tau-1}F_t'(e_t)
$$

for all $1 \leq \tau \leq T-t$. Finally, note from Lemma 1 that

$$
\begin{aligned}
\frac{\partial \mathbb{E}[m_{it+\tau}|h_i^t]}{\partial e_t} &= \sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})\frac{\partial \mathbb{E}[z_{it+s}|h_i^t]}{\partial e_t} \\
&= \left(\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}\right)(1-\mu_t)\frac{\partial \mathbb{E}[z_{it}|h_i^t]}{\partial e_t} + \sum_{s=1}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})\frac{\partial \mathbb{E}[z_{it+s}|h_i^t]}{\partial e_t},
\end{aligned}
$$

where z_{it+s} is the signal about the worker's ability in period $t+s$. Given that $\partial \mathbb{E}[z_{it}|h_i^t]/\partial e_t = 1$ and

$$
\frac{\partial \mathbb{E}[z_{it+s}|h_i^t]}{\partial e_t} = \frac{\partial \mathbb{E}[y_{it+s}|h_i^t]}{\partial e_t} = \lambda^{s-1}F_t'(e_t)
$$

for all $1 \leq s \leq T-t$, we can rewrite (A5) as

$$\begin{aligned} e_t &= b_t + F_t'(e_t) \sum_{\tau=1}^{T-t} \delta^\tau \left\{ (1-b_{t+\tau}) \sum_{s=1}^{\tau-1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k} \right) (1-\mu_{t+s}) \lambda^{s-1} + b_{t+\tau} \lambda^{\tau-1} \right\} \\ &\quad + \sum_{\tau=1}^{T-t} \delta^\tau (1-b_{t+\tau}) \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k} \right) (1-\mu_t). \end{aligned}$$

The desired result follows from Lemma 5 with $\xi_\tau = \lambda^{\tau-1}$.

The first-order condition (A2) is necessary for optimality. In the benchmark case in which the functions $\{F_t\}_{t=0}^T$ are linear, this condition is also sufficient for optimality. Indeed, the marginal benefit of effort to the worker—the right side of (A2)—is independent of the worker's effort, whereas the marginal cost of effort to the worker—the left side of (A2)—is increasing with effort. When the functions $\{F_t\}_{t=0}^T$ are nonlinear, (A2) need not be sufficient for optimality, though. Since F_t is concave, a sufficient condition for (A2) to be sufficient for optimality is

$$\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b_{t+\tau} + R_{CC,t+\tau} \big) \geq 0. \tag{A6}$$

Indeed, $R_{HK,t}(e)$ is nonincreasing with e if (A6) holds, in which case the marginal benefit of effort is nonincreasing with effort. Condition (A6) holds if piece rates are in the unit interval.

A.2 Equilibrium Piece Rates

We now derive the equilibrium piece rates. We consider the linear and nonlinear cases separately. Whereas the derivation of the equilibrium piece rates in the linear case is valid in general, the derivation of the equilibrium piece rates in the nonlinear case holds under some restrictions that we discuss.

A.2.1 Linear Case

Suppose that $F_t(e) = \gamma_t e$ with $\gamma_t \in \mathbb{R}$ for all t. We first derive the last-period equilibrium piece rates and then derive the equilibrium piece rates in previous periods when future equilibrium piece rates and effort choices are the same for all workers and depend only on time. We then prove that equilibrium piece rates and effort choices are the same for all workers and depend only on time, and we derive a recursive characterization for the former.

Last-Period Piece Rates. A standard argument shows that the period-T equilibrium piece rates are the same for all workers and equal to

$$b_T^* = \frac{1}{1 + r(\sigma_T^2 + \sigma_\varepsilon^2)}.$$

Since it follows from (A2) and (A3) that the period-T effort choice of a worker with piece rate b is $e_T = b$ regardless of the worker's history, the period-T equilibrium effort choices are also the same for all workers. Clearly, last-period equilibrium piece rates and effort choices depend only on time.

Piece Rates in Previous Periods. Let $t < T$ and suppose that the equilibrium piece rates and effort choices from period $t+1$ on are the same for all workers and depend only on time; this is true for $t = T-1$. We show that the equilibrium piece rates and effort choices from period t on are the same for all workers and depend only on time, and we derive an expression for the equilibrium piece rate in period t. In what follows, let $b_{t+\tau}^*$ be the equilibrium piece rate in period $t+\tau$ with $1 \leq \tau \leq T-t$, and define $R_{CC,t}^*$ and $R_{HK,t}^*(e)$ to be given by (A3) with $b_{t+\tau} = b_{t+\tau}^*$ for all τ. Note that $R_{HK,t}^*(e)$ is independent of e; that is, $R_{HK,t}^*(e) \equiv R_{HK,t}^*$.

Consider first a worker's optimal choice of effort in period t. We know from above that if the worker's piece rate is b, then the worker's optimal choice of effort is

$$e_t = e_t(b) = b + R_{CC,t}^* + R_{HK,t}^*. \tag{A7}$$

In particular, since (A7) does not depend on a worker's history, the workers' equilibrium choices of effort in t do not depend on their histories and so are the same for all workers if the same is true for equilibrium piece rates in t. It follows immediately from (A7) that e_t is strictly increasing with b and such that $\partial e_t/\partial b = 1$.

Now let $w^*_{t+\tau} = w^*_{t+\tau}(b)$ be a worker's wage in period $t+\tau$ with $0 \leq \tau \leq T-t$ when the worker's piece rate in period t is b, and define $W^*_t = W^*_t(b)$ to be such that $W^*_t = \sum_{\tau=0}^{T-t} \delta^\tau w^*_{t+\tau}$. The argument in the main text shows that a worker's equilibrium piece rate in period t is the choice of b that maximizes

$$\mathbb{E}[W^*_t|I_t] - r\mathrm{Var}[W^*_t|I_t]/2 - e_t^2/2, \tag{A8}$$

where I_t is the public information about the worker in t. In what follows, we show that (A8) has a unique maximizer and that this maximizer is independent of I_t. Thus, equilibrium piece rates in period t are the same for all workers.

Let $y^*_{t+\tau} = y^*_{t+\tau}(b)$ be the worker's output in period $t+\tau$ with $0 \leq \tau \leq T-t$ as a function of b. Note that $y^*_{t+\tau}$ depends on b only through the impact of b on e_t. It follows from (3) that $\mathbb{E}[w^*_{t+\tau}|I_t] = \mathbb{E}[y^*_{t+\tau}|I_t]$ for all τ. Since $\partial\mathbb{E}[y^*_t|I_t]/\partial b = \partial e_t/\partial b$, $\partial\mathbb{E}[y^*_{t+\tau}|I_t]/\partial b = F'_t(e_t)\lambda^{\tau-1}\partial e_t/\partial b$ for all $1 \leq \tau \leq T-t$, and $\partial e_t/\partial b = 1$, we have that

$$\frac{\partial\mathbb{E}[W^*_t|I_t]}{\partial b} = \left[1 + F'_t(e_t)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\right]\frac{\partial e_t}{\partial b} = 1 + \gamma_t\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}. \tag{A9}$$

We show below that

$$\frac{\partial\mathrm{Var}[W^*_t|I_t]}{\partial b} = 2\big[b(\sigma_t^2 + \sigma_\varepsilon^2) + H^*_t\big], \tag{A10}$$

where $H^*_t = \sigma_t^2\sum_{\tau=1}^{T-t}\delta^\tau$. From (A7), the first-order condition for the problem of maximizing (A8) is then

$$1 + \gamma_t\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - R^*_{HK,t} - R^*_{CC,t} - rH^*_t - b\big[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)\big] = 0. \tag{A11}$$

We now establish (A10). We know from the main text that

$$\mathrm{Var}[W^*_t|I_t] = b^2(\sigma_t^2 + \sigma_\varepsilon^2) + 2\sum_{\tau=1}^{T-t}\delta^\tau\mathrm{Cov}[w^*_t, w^*_{t+\tau}|I_t] + \mathrm{Var}_0,$$

where Var_0 is independent of b. We claim that $\mathrm{Cov}[w^*_t, w^*_{t+\tau}|I_t] = b\sigma_t^2$ for all $1 \leq \tau \leq T-t$, from which (A10) follows. Since the worker's reputation in period t is nonrandom conditional on I_t, it follows from (3) that

$$\mathrm{Cov}[w^*_t, w^*_{t+\tau}|I_t] = b\,\mathrm{Cov}[y^*_t, w^*_{t+\tau}|I_t]$$

for all $1 \leq \tau \leq T-t$. Now observe, once again from (3), that

$$\mathrm{Cov}[y^*_t, w^*_{t+\tau}|I_t] \;=\; b_{t+\tau}\mathrm{Cov}[y^*_t, y^*_{t+\tau}|I_t] + (1-b_{t+\tau})\mathrm{Cov}[y^*_t, m^*_{t+\tau}|I_t]$$

for all $1 \leq \tau \leq T-t$, where $m^*_{t+\tau} = m^*_{t+\tau}(b)$ is a worker's reputation in period $t+\tau$ as a function of the period-t piece rate. Like $y^*_{t+\tau}$, the reputation $m^*_{t+\tau}$ depends on b only through the impact of b on e_t. So, if $z^*_{t+s} = z^*_{t+s}(b)$ with $0 \leq s \leq T-t$ is the signal about ability in period $t+s$ as a function of b, then Lemma 1 implies that

$$\mathrm{Cov}[y^*_t, w^*_{t+\tau}|I_t] = b_{t+\tau}\mathrm{Cov}[y^*_t, y^*_{t+\tau}|I_t] + (1-b_{t+\tau})\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})\mathrm{Cov}[y^*_t, z^*_{t+s}|I_t]$$

for all $1 \leq \tau \leq T-t$. Since $\mathrm{Cov}[y^*_t, y^*_{t+\tau}|I_t] = \sigma_t^2$ for all $1 \leq \tau \leq T-t$ and

$$\mathrm{Cov}[y^*_t, z^*_{t+s}|I_t] = \begin{cases} \sigma_t^2 + \sigma_\varepsilon^2 & \text{if} \quad s = 0 \\ \sigma_t^2 & \text{if} \quad 1 \leq s \leq T-t \end{cases},$$

we then have that

$$\mathrm{Cov}[y_t^*, w_{t+\tau}^*|I_t] = \sigma_t^2\left[(1-b_{t+\tau})\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})+b_{t+\tau}\right]+\sigma_\varepsilon^2(1-b_{t+\tau})\left(\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}\right)(1-\mu_t).$$

To conclude, observe that $\sigma_\varepsilon^2(1-\mu_t) = \sigma_t^2\mu_t$ and $\mu_t\prod_{k=1}^{\tau-1}\mu_{t+\tau-k} = \prod_{k=1}^{\tau}\mu_{t+\tau-k}$ together imply that we can rewrite the above expression for $\mathrm{Cov}[y_t^*, w_{t+\tau}^*|I_t]$ as

$$\mathrm{Cov}[y_t^*, w_{t+\tau}^*|I_t] = \sigma_t^2\left\{(1-b_{t+\tau})\left[\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})+\prod_{k=1}^{\tau}\mu_{t+\tau-k}\right]+b_{t+\tau}\right\}.$$

The desired result follows from the fact that the term in square brackets equals one.

The first-order condition (A11) has a unique solution,

$$b_t^* = \frac{1}{1+r(\sigma_t^2+\sigma_\varepsilon^2)}\left[1+\gamma_t\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - R_{HK,t}^* - R_{CC,t}^* - rH_t^*\right], \tag{A12}$$

which is independent of I_t, and thus the same for all workers. Note that b_T^* given by (A12) reduces to the last-period piece rate derived above. That b_t^* maximizes (A8), and so is the equilibrium piece rate in period t, follows from the fact that $\partial\mathbb{E}[W_t^*|I_t]/\partial b$, the marginal benefit to a worker of an increase in b, is constant in b, while

$$\frac{r}{2}\frac{\partial\mathrm{Var}[W_t^*|I_t]}{\partial b} + e_t = r\left[b(\sigma_t^2+\sigma_\varepsilon^2)\right] + e_t,$$

the marginal cost to a worker of an increase in b, is strictly increasing with b. So, the first-order condition (A11) is necessary and sufficient for optimality.

Recursive Characterization of Equilibrium Piece Rates. The above argument shows that if there exists $t < T$ such that from period $t+1$ on, the equilibrium piece rates and effort choices are the same for all workers and depend only on time, then the equilibrium piece rates and effort choices from period t on have the same properties. Since the last-period equilibrium piece rates and effort choices are the same for all workers and (trivially) depend only on time, it follows by induction that the equilibrium piece rates are the same for all workers and depend only on time. From this, it further follows, once again by the above argument, that the equilibrium piece rate in any period t is determined by the equilibrium piece rates in subsequent periods through equation (A12).

A.2.2 Nonlinear Case

Suppose now that the functions $\{F_t\}_{t=0}^T$ are nonlinear for at least one $t < T$ and such that

$$\frac{\sigma_t^2}{\sigma_\varepsilon^2} < F_t'(e) < \frac{\sigma_t^2}{\sigma_\varepsilon^2}\left[1+r(\sigma_t^2+\sigma_\varepsilon^2)\right] \text{ for all } e\in\mathbb{R} \text{ and } t<T.$$

We first derive the last-period equilibrium piece rates. We then derive a necessary and sufficient condition for the equilibrium piece rates in previous periods when (*i*) future equilibrium piece rates and effort choices are the same for all workers and depend only on time; and (*ii*) future equilibrium piece rates are in the interval $(0,1)$. We next show that the equilibrium piece rates and effort choices are the same for all workers and depend only on time and that the equilibrium piece rates are in the interval $(0,1)$ if r is small enough. We use this fact to derive a recursive characterization of the equilibrium piece rates. For simplicity, we assume that $\lambda = 1$ in the final step of the equilibrium derivation. We conclude the nonlinear case by showing how to extend the argument in the final step to the case in which λ is smaller than one but close to it and discussing the role of the restrictions on the model's parameters in the equilibrium derivation.

Last-Period Piece Rates. Since only static considerations matter when $t = T$, the last-period equilibrium piece rates and effort choices are the same in the nonlinear case as in the linear case. Specifically, they are the same for all workers and (trivially) depend only on time.

Piece Rates in Previous Periods. Let $t < T$, and suppose that the equilibrium piece rates and effort choices from period $t+1$ on are the same for all workers and depend only on time, and piece rates belong to the interval $(0,1)$; this is true when $t = T-1$. We show that the equilibrium piece rates from period t on are the same for all workers and depend only on time, and derive an expression for the equilibrium piece rates in t. Again, $b^*_{t+\tau}$ is the equilibrium piece rate in $t+\tau$ with $1 \le \tau \le T-t$, and $R^*_{CC,t}$ and $R^*_{HK,t}(e)$ are given by (A3) with $b_{t+\tau} = b^*_{t+\tau}$ for all τ.

Consider first a worker's optimal choice of effort in period t. Since $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big) \ge 0$ when $b^*_{t+\tau} \in (0,1)$ for all $1 \le \tau \le T-t$, it follows that if a worker's piece rate in period t is b, then the worker's optimal choice of effort is the unique solution to the necessary and sufficient first-order condition

$$e_t = b + R^*_{CC,t} + R^*_{HK,t}(e_t). \tag{A13}$$

Recall that $R^*_{HK,t}(e)$ is monotone decreasing with e, since $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big) \ge 0$. As in the linear case, given that (A13) does not depend on a worker's history, the workers' equilibrium choices of effort in t are independent of their history, and so the same for all workers, if the period-t piece rates are the same for all workers.

Equation (A13) implicitly defines a worker's optimal choice of effort in period t as a function of the worker's piece rate in period t. In an abuse of notation, denote this function by $e_t = e_t(b)$. The implicit function theorem implies that e_t is continuously differentiable, with

$$\frac{\partial e_t}{\partial b} = \frac{1}{1 - F_t''(e_t)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big)}. \tag{A14}$$

Given that $F_t'''(e) \le 0$ for all $e \in \mathbb{R}$ and $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big) \ge 0$, it follows from (A14) that $\partial e_t/\partial b$ is positive, bounded above by one, and nonincreasing with b.

Once again, let $W^*_t = W^*_t(b)$ be the present-discounted value of the wage payments to the worker from period t on when the worker's piece rate in period t is b. Competition among firms and the mean-variance representation of worker preferences imply that the worker's equilibrium piece rate in period t when the public information about the worker in t is I_t is the choice of b maximizing (A8). In what follows, we first derive the (necessary) first-order condition for this problem. We then show that this first-order condition is sufficient for optimality and has a unique solution that is independent of I_t and so is the same for all workers. We know from (A9) that

$$\frac{\partial \mathbb{E}[W^*_t|I_t]}{\partial b} = \left[1 + F_t'(e_t)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\right]\frac{\partial e_t}{\partial b}.$$

Since the functions $\{F_t\}_{t=0}^T$ do not matter for the derivation of $\mathrm{Var}[W^*_t|I_t]$, it follows $\partial \mathrm{Var}[W^*_t|I_t]/\partial b$ is still given by (A10). So, the first-order condition for the problem of maximizing (A8) is

$$\left[1 + F_t'(e_t)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - e_t\right]\frac{\partial e_t}{\partial b} - r\big[b(\sigma_t^2 + \sigma_\varepsilon^2) + H^*_t\big] = 0 \tag{A15}$$

with $H^*_t = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^{\tau-1}$. Using (A13) and the definition of $R^*_{HK,t}(e)$, we can rewrite (A15) as

$$b = \left[1 + \frac{r(\sigma_t^2 + \sigma_\varepsilon^2)}{\partial e_t/\partial b}\right]^{-1}\left[1 + F_t'(e_t)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(1 - b^*_{t+\tau} - R^*_{CC,t+\tau}) - R^*_{CC,t} - \frac{rH^*_t}{\partial e_t/\partial b}\right]. \tag{A16}$$

Note that the solutions to (A16), if they exist, do not depend on I_t and so are the same for every worker.

In order to establish that (A15) is sufficient for optimality, let

$$MB_t(b) = \left[1 + F_t'(e_t)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\right]\frac{\partial e_t}{\partial b}$$

be the marginal benefit to the worker of an increase in b and

$$MC_t(b) = r\big[b(\sigma_t^2 + \sigma_\varepsilon^2) + H_t^*\big] + e_t\frac{\partial e_t}{\partial b}$$

be the marginal cost to the worker of an increase in b. Given that e_t is nondecreasing with b and $\partial e_t/\partial b$ is nonincreasing with b, it follows that MB_t is nonincreasing with b. Now note that since the fact that F_t''' is non-positive and nondecreasing implies that $F_t'''(e)e \geq F_t''(e)$ for all $e \in \mathbb{R}$, we then have from (A14) that[53]

$$\begin{aligned}\frac{d}{db}\left(e_t\frac{\partial e_t}{\partial b}\right) &= \left(\frac{\partial e_t}{\partial b}\right)^2\left[1 + \frac{e_tF_t'''(e_t)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(b_{t+\tau}^* + R_{CC,t+\tau}^*)}{1 - F_t''(e_t)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(b_{t+\tau}^* + R_{CC,t+\tau}^*)}\right] \\ &\geq \left(\frac{\partial e_t}{\partial b}\right)^2\frac{1}{1 - F_t''(e_t)\sum_{\tau=1}^{T-t}\delta^\tau(b_{t+\tau}^* + R_{CC,t+\tau}^*)} > 0.\end{aligned}$$

Thus, MC_t is strictly increasing with b, which establishes the sufficiency of (A15).

We conclude this step by showing that (A15), and so (A16), has a unique solution b_t^*, which we know does not depend on I_t. First note that MB_t is bounded since $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$ and that $\partial e_t/\partial b$ belongs to the unit interval. On the other hand, given that $e_t\partial e_t/\partial b$ is strictly increasing with b, it follows from the expression for MC_t that $\lim_{b\to-\infty} MC_t(b) = -\infty$ and $\lim_{b\to+\infty} MC_t(b) = +\infty$. So, (A15) has a solution, which is unique given the properties of MB_t and MC_t established above. Note that $b_T^* = 1/[1 + r(\sigma_T^2 + \sigma_\varepsilon^2)]$, since $\partial e_T/\partial b = 1$.

Recursive Characterization of Equilibrium Piece Rates. The above argument shows that if there exists $t < T$ such that from period $t+1$ on, the equilibrium piece rates and effort choices are the same for all workers and depend only on time, and the equilibrium piece rates are in the unit interval, then the equilibrium piece rates and effort choices from period t on are the same for all workers and depend only on time. We now show that if $\lambda = 1$, then the equilibrium piece rates in period t are also in the interval $(0, 1)$ if r is sufficiently small. From this, we are able to show that when $\lambda = 1$, the equilibrium piece rates are the same for all workers, depend only on time, and belong to the interval $(0, 1)$ if r is sufficiently small. We conclude by using this last fact to derive a recursive characterization of the equilibrium piece rates when $\lambda = 1$ and r is small enough.

Suppose that $\lambda = 1$. We first show that $F_t'(e) < (\sigma_t^2/\sigma_\varepsilon^2)[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)]$ for all $e \in \mathbb{R}$ implies that $b_t^* < 1$. Observe from Lemma 5 that

$$\sum_{\tau=1}^{T-t}\delta^\tau(1 - b_{t+\tau}^* - R_{CC,t+\tau}^*) = \sum_{\tau=1}^{T-t}\delta^\tau(1 - b_{t+\tau}^*)\left[1 - \sum_{s=1}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1 - \mu_{t+s})\right].$$

Since

$$\sum_{s=1}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1 - \mu_{t+s}) + \prod_{k=1}^{\tau-1}\mu_{t+\tau-k} = 1, \tag{A17}$$

[53]The desired inequality is immediate if $e \leq 0$. Consider the case in which $e > 0$. Then $F_t''(e) = F_t''(0) + \int_0^e F_t'''(s)ds$ implies that $F_t''(e) \leq \int_0^e F_t'''(s)ds \leq \int_0^e F'''(e)ds = eF_t'''(e)$: the first inequality holds since $F_t''(0) \leq 0$, the second inequality holds since $F_t'''(s) \leq F_t'''(e)$ for all $s \leq e$, and the last equality holds since $e > 0$.

we then have that

$$\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b_{t+\tau}^{*}-R_{CC,t+\tau}^{*})=\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b_{t+\tau}^{*})\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}=\frac{\sigma_{\varepsilon}^{2}}{\sigma_{t}^{2}}R_{CC,t}^{*}, \tag{A18}$$

where the second equality follows from (A3) and the fact that $\mu_t/(1-\mu_t)=\sigma_\varepsilon^2/\sigma_t^2$ by (16). Now observe that the right-hand side of (A16), and so b_t^*, is smaller than one if, and only if,

$$F_t'(e_t)\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b_{t+\tau}^{*}-R_{CC,t+\tau}^{*})-R_{CC,t}^{*}<\frac{r}{\partial e_t/\partial b}\left(\sigma_{\varepsilon}^{2}+\sigma_t^2\sum_{\tau=0}^{T-t}\delta^{\tau}\right), \tag{A19}$$

where $H_t^*=\sigma_t^2\sum_{\tau=1}^{T-t}\delta^\tau$. Since $\partial e_t/\partial b\le 1$, (A18) implies that a sufficient condition for (A19) is

$$R_{CC,t}^{*}\left[\frac{\sigma_{\varepsilon}^{2}}{\sigma_{t}^{2}}F_t'(e_t)-1\right]<r\left(\sigma_{\varepsilon}^{2}+\sigma_t^2\sum_{\tau=0}^{T-t}\delta^{\tau}\right).$$

The above inequality holds as $F_t'(e)<(\sigma_t^2/\sigma_\varepsilon^2)\big[1+r(\sigma_t^2+\sigma_\varepsilon^2)\big]$ for all $e\in\mathbb{R}$ by assumption, and by (A3) and the assumption that $b_{t+\tau}^*$ is in the unit interval for all $1\le\tau\le T-t$,

$$R_{CC,t}^{*}\le(1-\mu_t)\sum_{\tau=1}^{T-t}\delta^{\tau}=\frac{\sigma_t^2}{\sigma_t^2+\sigma_\varepsilon^2}\sum_{\tau=1}^{T-t}\delta^{\tau}<\frac{1}{\sigma_t^2+\sigma_\varepsilon^2}\left(\sigma_{\varepsilon}^{2}+\sigma_t^2\sum_{\tau=0}^{T-t}\delta^{\tau}\right).$$

We now show that $F_t'(e)>\sigma_t^2/\sigma_\varepsilon^2$ for all $e\in\mathbb{R}$ implies that there exists $\overline{r}>0$ such that $b_t^*>0$ for all $r\in(0,\overline{r})$. For this, observe, again using Lemma 5, that

$$\begin{aligned}\sum_{\tau=1}^{T-t}\delta^{\tau}(b_{t+\tau^*}+R_{CC,t+\tau}^{*}) &= \sum_{\tau=1}^{T-t}\delta^{\tau}\left[b_{t+\tau}^{*}+(1-b_{t+\tau}^{*})\sum_{s=1}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})\right]\\ &= \sum_{\tau=1}^{T-t}\delta^{\tau}\left[1-(1-b_{t+\tau}^{*})\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}\right]<\frac{\delta}{1-\delta},\end{aligned}$$

where the second equality follows from (A17) and the assumption that $b_{t+\tau}^*<1$ for all $1\le\tau\le T-t$. Thus,

$$\frac{rH_t^*}{\partial e_t/\partial b}=r\left[1-F_t''(e_t)\sum_{\tau=1}^{T-t}\delta^{\tau}(b_{t+\tau}^{*}+R_{CC,t+\tau}^{*})\right]\sigma_t^2\sum_{\tau=1}^{T-t}\delta^{\tau}<r\sigma_t^2\left[1-F_t''(\infty)\frac{\delta}{1-\delta}\right]\frac{\delta}{1-\delta}. \tag{A20}$$

Now note that $F_t'(e)>\sigma_t^2/\sigma_\varepsilon^2$ for all $e\in\mathbb{R}$ by assumption and the argument leading to (A18) together imply that

$$F_t'(e_t)\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b_{t+\tau}^{*}-R_{CC,t+\tau}^{*})-R_{CC,t}^{*}=R_{CC,t}^{*}\left[\frac{\sigma_{\varepsilon}^{2}}{\sigma_{t}^{2}}F_t'(e_t)-1\right]>0.$$

Since $b_{t+\tau}^*<1$ for all $1\le\tau\le T-t$, we have that $R_{CC,t}^*>0$. Then, by (A20) there exists $\overline{r}>0$ such that

$$1+F_t'(e_t)\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b_{t+\tau}^{*}-R_{CC,t+\tau}^{*})-R_{CC,t}^{*}-\frac{rH_t^*}{\partial e_t/\partial b}>0 \tag{A21}$$

if $r\in(0,\overline{r})$. This, in turn, implies that the right-hand side of (A16) is positive, and so is b_t^*.

Summing up, there exists $\overline{r}>0$ such that $b_t^*\in(0,1)$ provided that $r\in(0,\overline{r})$. Since σ_t^2 is monotonically decreasing if $\sigma_\theta^2>\sigma_\zeta^2$ and monotonically increasing if $\sigma_\theta^2<\sigma_\zeta^2$, it follows that $\sigma_t^2\le\max\{\sigma_\theta^2,\sigma_\zeta^2\}$. Thus, from

(A20), we can take the upper bound $\overline{r}$ on the worker's risk aversion to be independent of t. This fact is useful below.

We established that if there exists $t < T$ such that from period $t+1$ on, the equilibrium piece rates and effort choices are the same for all workers and depend only on time, and the equilibrium piece rates are in the interval $(0,1)$, then there exists $\overline{r} > 0$ independent of t such that the equilibrium piece rates and effort choices from period t on are the same for all workers and depend only on time, provided that $r \in (0, \overline{r})$ and $\lambda = 1$. Since the last-period equilibrium piece rates and effort choices are the same for all workers and (trivially) depend only on time, and the last-period equilibrium piece rate is in the interval $(0,1)$, a straightforward induction argument shows that if $\lambda = 1$ and $r \in (0, \overline{r})$, then the equilibrium piece rates and effort choices are the same for all workers and depend only on time, and the equilibrium piece rates are in the interval $(0,1)$. Moreover, by (A3) and (A14), equations (A13) and (A16) imply that the equilibrium piece rate in t is defined recursively as

$$b_t^* = \frac{1 + F_t'(e_t^*) \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^*(e_t^*) - R_{CC,t}^* - r\left[1 - \dfrac{F_t''(e_t^*)}{F_t'(e_t^*)} R_{HK,t}^*(e_t^*)\right] H_t^*}{1 + r(\sigma_t^2 + \sigma_\varepsilon^2)\left[1 - \dfrac{F_t''(e_t^*)}{F_t'(e_t^*)} R_{HK,t}^*(e_t^*)\right]},$$

where e_t^* is the unique solution to $e_t^* = b_t^* + R_{CC,t}^* + R_{HK,t}^*(e_t^*)$. This concludes the equilibrium derivation.

Discussion. We can relax the assumption that $\lambda = 1$ in the final step of the equilibrium derivation. First, note that (A13), (A14), and (A16) define the equilibrium piece rates continuously as a function of λ.[54] So, for each t, the map $\lambda \mapsto \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(1 - b_{t+1}^* - R_{CC,t+\tau}^*)$ is continuous. From this, it follows that if we take λ sufficiently close to one, then the inequalities (A19) and (A21) will continue to hold when $r \in (0, \overline{r})$, where $\overline{r}$ is the upper bound on r in the case in which $\lambda = 1$. The implied restrictions on the rates of human capital accumulation are natural. The rates of human capital accumulation cannot be too positive; otherwise, piece rates can be greater than one. Likewise, the rates of human capital accumulation have to be sufficiently positive, otherwise the learning-about-ability motive dominates, and we know from Gibbons and Murphy [1992] that it can lead to negative piece rates. In addition, workers cannot be too risk averse; otherwise, the demand for insurance against the lifetime wage risk due to the uncertainty about ability overwhelms all other factors determining equilibrium piece rates. Finally, since human capital depreciation effectively acts to reduce the rates of human capital accumulation, it cannot be too large.

A.3 Equilibrium Characterization

We can now state our main equilibrium characterization result. It includes Proposition 1 as a special case.

Proposition 6. *Suppose that either (i) $F_t(e) = \gamma_t e$ with $\gamma_t \in \mathbb{R}$ or (ii)*

$$\frac{\sigma_t^2}{\sigma_\varepsilon^2} < F_t'(e) < \frac{\sigma_t^2}{\sigma_\varepsilon^2}\left[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)\right] \textit{ for all } e \in \mathbb{R} \textit{ and } t < T.$$

In case (i), the unique equilibrium is such that piece rates and effort choices are the same for all workers and depend only on time. In case (ii), there exist $\underline{\lambda} \in (0,1)$ and $\overline{r} > 0$ such that if $\lambda \in (\underline{\lambda}, 1]$ and $r \in (0, \overline{r})$, then the unique equilibrium is such that piece rates and effort choices are the same for all workers and depend only on time, and piece rates are in the interval $(0,1)$. Let b_t^* and e_t^* be, respectively, the equilibrium piece rate and effort choice in period t. For each t, let $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$ and define $R_{CC,t}^*$ and $R_{HK,t}^*(e)$ to be such that

$$R_{CC,t}^* = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau}^*) \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1 - \mu_t) \textit{ and } R_{HK,t}^*(e_t^*) = F_t'(e_t^*) \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b_{t+\tau}^* + R_{CC,t+\tau}^*\big).$$

[54] The recursive structure of the equilibrium piece rates implies that if future pieces rates depend continuously on λ, then current piece rates are also continuous functions of λ. Since the last-period piece rate is continuous in λ, so are the equilibrium piece rates in all previous periods.

Then, b_t^* and e_t^* are given recursively by

$$b_t^* = \frac{1 + F_t'(e_t^*)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - R_{HK,t}^*(e_t^*) - R_{CC,t}^* - r\left[1 - \dfrac{F_t''(e_t^*)}{F_t'(e_t^*)}R_{HK,t}^*(e_t^*)\right]H_t^*}{1 + r(\sigma_t^2+\sigma_\varepsilon^2)\left[1 - \dfrac{F_t''(e_t^*)}{F_t'(e_t^*)}R_{HK,t}^*(e_t^*)\right]};$$

$$e_t^* = b_t^* + R_{CC,t}^* + R_{HK,t}^*(e_t^*).$$

Note that the expression for b_t^* in Proposition 6 reduces to the expression for b_t^* in Proposition 1 in the linear case. Clearly, the conditions under which equilibrium piece rates are in the unit interval apply in the linear case with γ_t in place of $F_t'(e)$. We conclude the equilibrium characterization by providing an alternative set of conditions under which equilibrium piece rates in the linear case are in the unit interval. They apply even in the absence of learning about ability. The proof of Corollary 2 is straightfoward and thus omitted.

Corollary 2. *Consider the linear case and suppose that*

$$\frac{\sigma_t^2}{\sigma_\varepsilon^2} < \gamma_t < \left(\frac{1-\delta\lambda}{\delta}\right) r(\sigma_t^2+\sigma_\varepsilon^2) \text{ for all } t < T.$$

There exists $\underline{\lambda} \in (0,1)$ and $\overline{r} > 0$ such that if $\lambda \in (\underline{\lambda}, 1]$ and $r \in (0, \overline{r})$, then $b_t^* \in (0,1)$ for all t.

B Appendix: Equilibrium Properties

B.1 Proof of Lemma 2

Consider first the case in which $\sigma_\theta^2 \geq \sigma_\infty^2$, so that σ_t^2 is nonincreasing with t and thus b_t^0 is nondecreasing with t. Note that $H_{T-1}^* > H_T^* = 0$ and, since $b_T^* \in (0,1)$ and $\mu_{T-1} \in (0,1)$, that $R_{CC,T-1}^* = \delta(1-b_T^*)(1-\mu_{T-1}) > R_{CC,T}^* = 0$. Thus,

$$b_{T-1}^* = b_{T-1}^0\big(1 - R_{CC,T-1}^* - rH_{T-1}^*\big) < b_{T-1}^0 \leq b_T^0 = b_T^*.$$

Now suppose, by induction, that there exists $1 \leq t \leq T-1$ such that $R_{CC,t+\tau}^* > R_{CC,t+\tau+1}^*$ and $b_{t+\tau}^* < b_{t+\tau+1}^*$ for all $0 \leq \tau \leq T-t-1$. We are done if we show that $R_{CC,t-1}^* > R_{CC,t}^*$ and $b_{t-1}^* < b_t^*$. Let $s = t-1$. Then,

$$\begin{aligned} R_{CC,s}^* &= \sum_{\tau=1}^{T-s}\delta^\tau(1-b_{s+\tau}^*)\left(\prod_{k=1}^{\tau-1}\mu_{s+\tau-k}\right)(1-\mu_s) > \sum_{\tau=1}^{T-s-1}\delta^\tau(1-b_{s+\tau}^*)\left(\prod_{k=1}^{\tau-1}\mu_{s+\tau-k}\right)(1-\mu_s) \\ &> \sum_{\tau=1}^{T-s-1}\delta^\tau(1-b_{s+1+\tau}^*)\left(\prod_{k=1}^{\tau-1}\mu_{s+\tau-k}\right)(1-\mu_s), \end{aligned}$$

where the first inequality follows from the fact that $b_T^* \in (0,1)$ and $\mu_t \in (0,1)$ for $0 \leq t \leq T$ and the second inequality follows from $b_{s+1+\tau}^* > b_{s+\tau}^*$ for all $1 \leq \tau \leq T-s-1$ by the induction hypothesis. Hölmstrom [1999] shows that $(1-\mu_s)\prod_{k=1}^{\tau-1}\mu_{s+\tau-k}$ is a decreasing function of μ_s (see the argument leading to Proposition 2 at page 174). Since $\mu_{s+1} \geq \mu_s$, we then have that

$$R_{CC,s}^* > \sum_{\tau=1}^{T-s-1}\delta^\tau(1-b_{s+1+\tau}^*)\left(\prod_{k=1}^{\tau-1}\mu_{s+1+\tau-k}\right)(1-\mu_{s+1}) = R_{CC,s+1}^* = R_{CC,t}^*.$$

To conclude the argument, observe that if $b \geq b_t^*$, then

$$1 - R_{CC,t}^* - rH_t^* - b\big[1 + r(\sigma_t^2+\sigma_\varepsilon^2)\big] \leq 0.$$

Thus, given that $R^*_{CC,s} > R^*_{CC,t}$, $H^*_s \geq H^*_t$, and $\sigma^2_s \geq \sigma^2_t$, it follows that $b \geq b^*_t$ implies that

$$1 - R^*_{CC,s} - rH^*_s - b\big[1 + r(\sigma^2_s + \sigma^2_\varepsilon)\big] < 0.$$

We know from the proof of Proposition 6 that the first-order condition (A11) is necessary and sufficient for the equilibrium piece rates. Thus, $b^*_s = b^*_{t-1} < b^*_t$. This concludes the case in which $\sigma^2_\theta \geq \sigma^2_\infty$.

Now consider the case in which $\sigma^2_\theta < \sigma^2_\infty$. Fix $T_0 \geq 0$ and let $T > T_0$; we pin down T_0 below. Moreover, let $\mu_\infty = \sigma^2_\varepsilon/(\sigma^2_\infty + \sigma^2_\varepsilon)$ and consider the difference equation

$$\widehat{b}_t = \frac{1}{1 + r(\sigma^2_\infty + \sigma^2_\varepsilon)}\left[1 - \sum_{\tau=1}^{T-t} \delta^\tau (1 - \widehat{b}_{t+\tau})\mu_\infty^{\tau-1}(1 - \mu_\infty) - r\sigma^2_\infty \sum_{\tau=1}^{T-1} \delta^\tau\right]$$

for $T_0 \leq t \leq T$. By construction, $\widehat{b}_t$ is the equilibrium piece in period t if uncertainty about ability from period T_0 on were constant and equal to σ^2_∞. It follows from the first case that $\widehat{b}_t$ is strictly increasing with t for all $T_0 \leq t \leq T$. We claim that $\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} b^*_t = \widehat{b}_t$ for all $T_0 \leq t \leq T$. First, note that $\sigma^2_{T_0} < \sigma^2_T < \sigma^2_\infty$ implies that $\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} b^*_T = \widehat{b}_T$. Now suppose, by induction, that there exists $T_0 < t \leq T$ such that $\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} b^*_{t+\tau} = \widehat{b}_{t+\tau}$ for all $0 \leq \tau \leq T - t$. Let $s = t - 1$. We obtain the desired result if $\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} b^*_s = \widehat{b}_s$. For this, note that

$$b^*_s = \frac{1}{1 + r(\sigma^2_s + \sigma^2_\varepsilon)}\left[1 - \sum_{\tau=1}^{T-s} \delta^\tau (1 - b^*_{s+\tau})\left(\prod_{k=1}^{\tau-1} \mu_{s+\tau-k}\right)(1 - \mu_s) - r\sigma^2_s \sum_{\tau=1}^{T-1} \delta^\tau\right].$$

Since $\sigma^2_{T_0} \leq \sigma^2_{s+\tau} < \sigma^2_\infty$ for all $0 \leq \tau \leq T - s$, it follows that $\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} \sigma^2_{s+\tau} = \sigma^2_\infty$ for all $0 \leq \tau \leq T - s$, and so $\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} \mu^2_{s+\tau} = \mu^2_\infty$ for all $0 \leq \tau \leq T - s$ as well. This, in turn, implies that

$$\lim_{\sigma^2_{T_0} \to \sigma^2_\infty} b^*_s = \frac{1}{1 + r(\sigma^2_\infty + \sigma^2_\varepsilon)}\left[1 - \sum_{\tau=1}^{T-s} \delta^\tau (1 - \widehat{b}_{s+\tau})\mu_\infty^{\tau-1}(1 - \mu_\infty) - r\sigma^2_\infty \sum_{\tau=1}^{T-1} \delta^\tau\right] = \widehat{b}_s$$

by the induction hypothesis and the fact that b^*_s is jointly continuous in $(b^*_{s+1}, \ldots, b^*_T, \sigma^2_s, \mu_s, \ldots, \mu_T)$. To conclude, given that $\widehat{b}_t$ is strictly increasing with t for all $T_0 \leq t \leq T$, we note that there exists $\eta > 0$ such that if $|b^*_t - \widehat{b}_t| \leq \eta$ for all $T_0 \leq t \leq T$, then b^*_t is strictly increasing with t for all $T_0 \leq t \leq T$ as well. Since $\lim_{T_0 \to \infty} \sigma^2_{T_0} = \sigma^2_\infty$, we then have that there exists $T_0 \geq 0$ such that $|b^*_t - \widehat{b}_t| \leq \eta$, and thus b^*_t is strictly increasing with t for all $T_0 \leq t \leq T$.

B.2 Proof of Lemma 3

Fix $0 \leq T_0 < T$. We first show that $\gamma_t < (1 - \delta\lambda)(1 + r\sigma^2_\varepsilon)/\delta[1 - (\delta\lambda)^{T-T_0}]$ for all $T_0 \leq t < T$ implies that $b^*_t \in [0, 1)$ for $T_0 \leq t \leq T$. Suppose, by induction, that there exists $T_0 + 1 \leq t \leq T$ such that $b^*_{t+\tau} \in [b^0, 1)$ for all $0 \leq \tau \leq T - t$; the induction hypothesis is true for $t = T$. Let $s = t - 1$. The desired result follows if $b^*_s \in [b^0, 1)$. First, note that $b^*_{s+\tau} < 1$ for all $1 \leq \tau \leq T - s$ and (24) imply that $b^*_s \geq b^0$. Now, note that $b^*_{s+\tau} \geq b^0$ for all $1 \leq \tau \leq T - s$ and (24) imply that

$$b^*_s = b^0\left[1 + \gamma_s \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1}(1 - b^*_{s+\tau})\right] \leq b^0\left[1 + \gamma_s \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1}(1 - b^0)\right] = b^0\left(1 + \gamma_s r\sigma^2_\varepsilon b^0 \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1}\right),$$

since $1 - b^0 = r\sigma^2_\varepsilon b^0$. Thus, a sufficient condition for $b^*_s < 1$ is that $\gamma_s r\sigma^2_\varepsilon b^0 \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1} < r\sigma^2_\varepsilon$, which holds if $\gamma_s < (1 - \delta\lambda)(1 + r\sigma^2_\varepsilon)/\delta[1 - (\delta\lambda)^{T-T_0}]$.

We now show that γ_t positive and nonincreasing with t for all $T_0 \leq t < T$ implies that b^*_t is strictly decreasing with t for all $T_0 \leq t \leq T$. We known from the main text that $b^*_{T-1} > b^*_T$. So, assume that $T_0 < T - 1$, and suppose, by induction, that there exists $T_0 + 1 \leq t \leq T - 1$ such that $b^*_{t+\tau} > b^*_{t+1+\tau}$ for all $0 \leq \tau \leq T - t - 1$. Let $s = t - 1$.

Then,

$$b_s^* > b^0\left[1+\gamma_s\sum_{\tau=1}^{T-s-1}\delta^\tau\lambda^{\tau-1}(1-b_{s+\tau}^*)\right] > b^0\left[1+\gamma_s\sum_{\tau=1}^{T-s-1}\delta^\tau\lambda^{\tau-1}(1-b_{s+1+\tau}^*)\right]$$
$$\geq b^0\left[1+\gamma_{s+1}\sum_{\tau=1}^{T-s-1}\delta^\tau\lambda^{\tau-1}(1-b_{s+1+\tau}^*)\right] = b_{s+1}^*,$$

where the first inequality follows since $b_T^* \in (0,1)$, the second inequality follows from the induction hypothesis, and the third inequality follows since piece rates are in $[0,1]$ and $\gamma_s \geq \gamma_{s+1}$. This concludes the proof.

B.3 Proof of Lemma 4

Suppose $0 < T_0 < T$ is such that $\gamma_{T_0} > 0$ and $b_t^* < 1$ for all $T_0 \leq t \leq T$. Since both $\sum_{\tau=1}^{T-T_0}\delta^\tau\lambda^{\tau-1}(1-b_{T_0+\tau}^*)$ and $\sum_{\tau=1}^{T-T_0+1}\delta^\tau\lambda^{\tau-1}(1-b_{T_0-1+\tau}^*)$ are positive by assumption, there exists $\gamma_{T_0-1} > 0$ such that

$$\gamma_{T_0-1}\sum_{\tau=1}^{T-T_0}\delta^\tau\lambda^{\tau-1}(1-b_{T_0+\tau}^*) < \gamma_{T_0}\sum_{\tau=1}^{T-T_0-1}\delta^\tau\lambda^{\tau-1}(1-b_{T_0+1+\tau}^*).$$

By reducing γ_{T_0-1} if necessary, we can ensure that $\gamma_{T_0} > \gamma_{T_0-1}$. From (24), it follows that $b_{T_0-1}^* \in (b^0, b_{T_0}^*)$. Given that $b_{T_0-1}^* < 1$, we can repeat the step for $t = T_0 - 1$ to show that there exists $\gamma_{T_0-2} \in (0, \gamma_{T_0-1})$ such that $b_{T_0-2}^* \in (b^0, b_{T_0-1}^*)$. Continuing backward, we obtain the desired result.

B.4 Proof of Proposition 2

We first show that when σ_ζ^2 is small, there exists $T_0 \geq 0$ such that if $T > T_0$, γ_t is nonincreasing with t for all $T_0 \leq t < T$, and $0 < \gamma_{T-1} \leq \gamma_{T_0} < (1-\delta\lambda)(1+r\sigma_\varepsilon^2)/\delta[1-(\delta\lambda)^{T-T_0}]$, then b_t^* is strictly decreasing with t for all $T_0 \leq t \leq T$. For simplicity, assume that $\sigma_\zeta^2 = 0$. Since the equations for the equilibrium piece rates depend continuously on σ_ζ^2 and $\lim_{t\to\infty}\sigma_t^2 \approx 0$ when $\sigma_\zeta^2 \approx 0$, we can extend the argument to the case in which σ_ζ^2 is positive but small. Fix $T_0 > 0$ and let $T > T_0$; we pin down T_0 below. Now consider the difference equation

$$\widehat{b}_t = \frac{1}{1+r\sigma_\varepsilon^2}\left[1+\gamma_t\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(1-\widehat{b}_{t+1})\right]$$

for $T_0 \leq t \leq T$. By definition, $\widehat{b}_t$ is the piece rate in period $T_0 \leq t \leq T$ if only human capital acquisition is present. The same argument as the one in the proof of Lemma 2 shows that $\lim_{\sigma_{T_0}^2\to 0} b_t^* = \widehat{b}_t$ for all $T_0 \leq t \leq T$. Since, by Lemma 3, $\widehat{b}_t$ is strictly decreasing with t for all $T_0 \leq t \leq T$ and $\lim_{T_0\to\infty}\sigma_{T_0}^2 = 0$, it then follows, also by the same argument as the one in the proof of Lemma 2, that we can choose $T_0 \geq 0$ such that b_t^* is strictly decreasing with t for all $T_0 \leq t \leq T$.

We now show that there exist $T_0 \geq 0$ and $\underline{\gamma} > 0$ such that if $T > T_0$ and $|\gamma_t| < \underline{\gamma}$ for all $T_0 \leq t < T$, then b_t^* is strictly increasing with t for all $T_0 \leq t \leq T$. Fix $T_0 \geq 0$, let $T > T_0$, and assume that $\gamma_t = 0$ for all $T_0 \leq t < T$. Since the equations for the equilibrium piece rates depend continuously on the rates of human capital accumulation, we can extend the argument to the case in which γ_{T_0} to γ_{T-1} are small in absolute value. Given that from period T_0 on, the equilibrium piece rates coincide with the equilibrium piece rates in the pure learning-about-ability case, it follows from Lemma 2 that b_t^* is strictly increasing with t for all $T_0 \leq t \leq T$ if T_0 is sufficiently large.

B.5 Proof of Proposition 3

That piece rates can be hump-shaped if σ_θ^2 and σ_ζ^2 are small and the rates of human capital accumulation are initially increasing and then decreasing follows immediately from Corollary 1 and continuity. We now show that piece rates can also be hump-shaped when the rates of human capital accumulation are positive and constant over time, σ_θ^2 is large, and σ_ε^2 and σ_ζ^2 are small. Suppose that $\gamma_t \equiv \gamma$ with $0 < \gamma < (1-\delta\lambda)(1+r\sigma_\varepsilon^2)/\delta[1-(\delta\lambda)^T]$, and for simplicity, assume that $\lambda = 1$ and $\sigma_\varepsilon^2 = 0$. Given that the equations for the equilibrium piece rates are continuous in λ and σ_ε^2, the results extend to the case in which σ_ε^2 is positive but small and λ is different from, but close to, one. Note that $\sigma_\varepsilon^2 = 0$ implies that $\sigma_t^2 = \sigma_\zeta^2$ for all $1 \leq 1 \leq T$ and, since $\mu_t \equiv 0$, that $R^*_{CC,t} = \delta(1-b^*_{t+1})$ for all $0 \leq t < T$. Therefore,

$$\begin{aligned}
b_t^* &= \frac{1}{1+r\sigma_\zeta^2}\left[1+\gamma\sum_{\tau=1}^{T-t}\delta^\tau(1-b^*_{t+\tau}-R^*_{CC,t+\tau}) - R^*_{CC,t} - r\sigma_\zeta^2\sum_{\tau=1}^{T-1}\delta^\tau\right]\\
&= \frac{1}{1+r\sigma_\zeta^2}\left[1+\gamma\sum_{\tau=1}^{T-t}\delta^\tau(1-b^*_{t+\tau}) - \gamma\sum_{\tau=1}^{T-t-1}\delta^{\tau+1}(1-b^*_{t+1+\tau}) - \delta(1-b^*_{t+1}) - r\sigma_\zeta^2\sum_{\tau=1}^{T-t}\delta^\tau\right]\\
&= \frac{1}{1+r\sigma_\zeta^2}\left[1-\delta+\gamma\delta+(1-\gamma)\delta b^*_{t+1} - r\sigma_\zeta^2\sum_{\tau=1}^{T-t}\delta^\tau\right]
\end{aligned}$$

for all $1 \leq t < T$ with $b_T^* = 1/(1+r\sigma_\zeta^2)$, where the second equality follows from the fact that $R^*_{CC,T} = 0$. We claim that there exists $\eta > 0$ such that $b_t^* > \eta$ for all $1 \leq t \leq T$ if σ_ζ^2 is sufficiently small. Indeed, in the limit as σ_ζ^2 converges to zero, the above equations for b_t^* reduce to $b_t^* = 1-\delta+\gamma\delta+(1-\gamma)\delta b^*_{t+1}$ for all $1 \leq t < T$ with $b_T^* = 1$. In this limiting case, it follows immediately that $b_t^* = 1$ for all $t \geq 1$. The desired result now follows, since the equations for b_t^* depend continuously on σ_ζ^2. Next, note that

$$b_0^* = \frac{1}{1+r\sigma_\theta^2}\left[1-\delta+\gamma\delta+(1-\gamma)\delta b_1^* - r\sigma_\theta^2\sum_{\tau=1}^{T}\delta^\tau\right],$$

and so b_0^* is smaller than η if σ_θ^2 is sufficiently large. Since b_1^* does not depend on σ_θ^2, it then follows that $b_0^* < b_1^*$ if σ_θ^2 is sufficiently large and σ_ζ^2 is sufficiently small. To finish, note, by continuity and Lemma 3, that b_t^* is strictly decreasing with t for $1 \leq t \leq T$, reducing σ_ζ^2 further if necessary.

We now show that equilibrium piece rates can be U-shaped if human capital accumulation is important early on but its importance decreases quickly enough over time. For simplicity, let $\lambda = 1$; as was the case above, the argument extends to the case in which λ is different from but close to one. It follows from the proof of Proposition 6 in Appendix A—and (A18) in particular—that

$$b_t^* = \frac{1}{1+r(\sigma_t^2+\sigma_\varepsilon^2)}\left[1+R^*_{CC,t}\left(\gamma_t\frac{\sigma_\varepsilon^2}{\sigma_t^2}-1\right) - r\sigma_t^2\sum_{\tau=1}^{T-t}\delta^\tau\right] \tag{B22}$$

for all $0 \leq t \leq T$. Fix $0 < T_0 < T$ and suppose that $\gamma_t = 0$ for all $T_0 \leq t \leq T$. Moreover, assume that $\sigma_\theta^2 > \sigma_\infty^2$. We know from the proofs of Lemma 2 and Proposition 2 that both assumptions can be relaxed. Then, b_t^* is strictly increasing with t for all $T_0 \leq t \leq T$. Since it is also the case that $b_t^* < 1$ for all $T_0 \leq t \leq T$, we have that $R^*_{CC,T_0-1} > 0$. By (B22), we can choose $\gamma_{T_0-1} > 0$ so that $b^*_{T_0-1} > b^*_{T_0}$. Since $b^*_{T_0} < 1$, we can ensure that $b^*_{T_0-1} < 1$ as well. Since $b^*_{T_0-1} < 1$, we can repeat the step for $t = T_0 - 1$ to show that there exists a value of γ_{T_0-2} for which $b^*_{T_0-1} < b^*_{T_0-2} < 1$. Continuing backward, we obtain the desired result.

C Appendix: Identification

C.1 Second Moments of Wage Distributions

In this appendix, we calculate the second moments of the wage distribution. From (3), we can express worker i's wage in period t as $w_{it} = \overline{w}_{it} + r_{it}$, where $r_{it} = (1-b_t^*)\mathbb{E}[\theta_{it}|I_{it}] + b_t^*(\theta_{it}+\varepsilon_{it})$ is the random part of w_{it}. Assume without loss that $m_\theta = 0$, in which case $\mathbb{E}[\theta_{it}] \equiv 0$, and so $\mathbb{E}[r_{it}] \equiv 0$. Since the conditional expectation is an orthogonal projection, we have that $\mathbb{E}[\theta_{it}|I_{it}] \perp \theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]$. We use this fact repeatedly in what follows.

Variances of Wage Residuals. We claim that $\text{Var}[r_{it}] = \text{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$. Indeed, since

$$r_{it} = \mathbb{E}[\theta_{it}|I_{it}] + b_t^*\big(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}] + \varepsilon_{it}\big), \tag{C23}$$

we have that

$$\text{Var}[r_{it}] = \text{Var}[\mathbb{E}[\theta_{it}|I_{it}]] + (b_t^*)^2\text{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] + (b_t^*)^2\sigma_\varepsilon^2. \tag{C24}$$

Now note that $\text{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] = \text{Var}[\theta_{it}] - \text{Var}[\mathbb{E}[\theta_{it}|I_{it}]]$.[55] Moreover, given that $\theta_{it}|I_{it}$ is normally distributed with mean $\mathbb{E}[\theta_{it}|I_{it} = \iota_t]$ and variance σ_t^2 when $I_{it} = \iota_t$, the random variable $(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}])|I_{it}$ is normally distributed with mean zero and variance σ_t^2. Thus, $\text{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] = \mathbb{E}[\text{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]|I_{it}] = \sigma_t^2$, and so, since $\text{Var}[\theta_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2$, it follows that $\text{Var}[\mathbb{E}[\theta_{it}|I_{it}]] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2$. The desired result follows from (C24).

Covariances of Wage Residuals. We claim that $\text{Cov}[r_{it}, r_{it+s}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + b_t^*\sigma_t^2$ for all $1 \leq s \leq T - t$. Let $\eta_{it}^s = \mathbb{E}[\theta_{it+s}|I_{t+s}] - \mathbb{E}[\theta_{it}|I_t]$. Since

$$\begin{aligned} r_{it+s} &= \mathbb{E}[\theta_{it+s}|I_{it+s}] + b_{t+s}^*(\theta_{it} + \zeta_{it} + \cdots + \zeta_{it+s-1} - \mathbb{E}[\theta_{it+s}|I_{it+s}] + \varepsilon_{it+s}) \\ &= \mathbb{E}[\theta_{it}|I_{it}] + b_{t+s}^*(\theta_{it} + \zeta_{it} + \cdots + \zeta_{it+s-1} - \mathbb{E}[\theta_{it}|I_{it}] + \varepsilon_{it+s}) + (1 - b_{t+s}^*)\eta_{it}^s, \end{aligned}$$

we then have that

$$\begin{aligned} \mathbb{E}[r_{it}r_{it+s}] &= \text{Var}[\mathbb{E}[\theta_{it}|I_{it}]] + (1 - b_{t+s}^*)\mathbb{E}\big[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s\big] \\ &\quad + b_t^*b_{t+s}^*\text{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] + (1 - b_{t+s}^*)b_t^*\mathbb{E}[(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}] + \varepsilon_{it})\eta_{it}^s] \\ &= \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + b_t^*b_{t+s}^*\sigma_t^2 + (1 - b_{t+s}^*)b_t^*\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta_{it}^s] + (1 - b_t^*)(1 - b_{t+s}^*)\mathbb{E}[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s]. \end{aligned}$$

We now show that $\mathbb{E}[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s] = 0$ and $\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta_{it}^s] = \sigma_t^2$, which implies the desired result. First, note that

$$\eta_{it}^s = \sum_{k=0}^{s-1}\left(\prod_{j=1}^{s-1-k}\mu_{t+s-j}\right)(1 - \mu_{t+k})(\theta_{it+k} + \varepsilon_{it+k} - \mathbb{E}[\theta_{it}|I_{it}])$$

by Lemma 1. Since $\theta_{it+k} = \theta_{it} + \zeta_{it} + \cdots + \zeta_{it+k-1}$, it easily follows that $\mathbb{E}[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s] = 0$. Moreover,

$$\begin{aligned} (\theta_{it} + \varepsilon_{it})\eta_{it}^s &= (\theta_{it} + \varepsilon_{it})\big(\theta_{it} + \varepsilon_{it} - \mathbb{E}[\theta_{it}|I_{it}]\big)\left(\prod_{j=1}^{s-1}\mu_{t+s-j}\right)(1 - \mu_t) \\ &\quad + \theta_{it}\big(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]\big)\sum_{k=1}^{s-1}\left(\prod_{j=1}^{s-1-k}\mu_{t+s-j}\right)(1 - \mu_{t+k}) + \Lambda_t^s, \end{aligned}$$

where Λ_t^s is a random variable with zero mean. Given that $\mathbb{E}[(\theta_{it} + \varepsilon_{it})(\theta_{it} + \varepsilon_{it} - \mathbb{E}[\theta_{it}|I_{it}])] = \sigma_t^2 + \sigma_\varepsilon^2$ and $\mathbb{E}[\theta_{it}(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}])] = \sigma_t^2$, it then follows from (A17) and $(\sigma_t^2 + \sigma_\varepsilon^2)(1 - \mu_t) = \sigma_t^2$ that

$$\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta_{it}^s] = (\sigma_t^2 + \sigma_\varepsilon^2)(1 - \mu_t)\left(\prod_{j=1}^{s-1}\mu_{t+s-j}\right) + \sigma_t^2\sum_{k=1}^{s-1}\left(\prod_{j=1}^{s-1-k}\mu_{t+s-j}\right)(1 - \mu_{t+k}) = \sigma_t^2.$$

The following lemma summarizes these results.

Lemma 6. *For all t and $1 \leq s \leq T - t$, we have that (i)* $\text{Var}[r_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$*; and (ii)* $\text{Cov}[r_{it}, r_{it+s}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + b_t^*\sigma_t^2$.

[55] Indeed, $\text{Var}[A - B] = \text{Var}[A] + \text{Var}[B] - 2\text{Cov}[A, B]$ and $\text{Cov}[\theta_{it}, \mathbb{E}[\theta_{it}|I_{it}]] = \text{Var}[\mathbb{E}[\theta_{it}|I_{it}]]$.

C.2 More General Human Capital Process

We first consider the case in which the econometrician observes a discrete version of the continuous performance measure p_{it} discussed in Section 6 and then consider the case in which the performance measure p_{it} is a general function of a worker's effort and human capital.

Discrete Performance Measure. Consider the case in which the econometrician observes only a discrete version of p_{it} and the cumulative distribution function G for the noise in the performance measure is known. Namely, assume that for each t, there exist thresholds $\overline{p}_{1t} < \ldots < \overline{p}_{Kt}$ and that the econometrician observes p^o_{it} given by

$$p^o_{it} = \begin{cases} 0 & \text{if} \quad p_{it} \leq \overline{p}_{1t} \\ k & \text{if} \quad \overline{p}_{kt} < p_{it} \leq \overline{p}_{k+1t} \quad \text{for } k \in \{1, \ldots, K-1\} \\ K & \text{if} \quad p_{it} > \overline{p}_{Kt} \end{cases} .$$

This is a plausible representation of performance scales in firms; see, for instance, Baker et al. [1994a]. Given that $\mathbb{P}\{p^o_{it} = K\} = 1 - \mathbb{P}\{p_{it} \leq \overline{p}_{Kt}\}$ and $\mathbb{P}\{p^o_{it} = k\} = \mathbb{P}\{p_{it} \leq \overline{p}_{k+1t}\} - \mathbb{P}\{p_{it} \leq \overline{p}_{kt}\}$ for all k by definition of p^o_{it}, it follows immediately that the probabilities $\mathbb{P}\{p_{it} \leq \overline{p}_{1t}\}$ to $\mathbb{P}\{p_{it} \leq \overline{p}_{Kt}\}$ are identified from the probabilities $\mathbb{P}\{p^o_{it} = 1\}$ to $\mathbb{P}\{p^o_{it} = K\}$—that is, from the distribution of the discrete performance measure in period t.

Let k^*_t and e^*_t be, respectively, the workers' equilibrium stock of human capital and effort in period t. Since $\mathbb{E}[w_{it}] = m_\theta + k^*_t + e^*_t$ and $\mathbb{P}\{p_{it} \leq \overline{p}_{kt}\} = \mathbb{P}\{\eta_{it} \leq \overline{p}_{kt} - e^*_t\} = G(\overline{p}_{kt} - e^*_t)$ for each k with G strictly increasing and so invertible, we obtain a linear system of $K+1$ equations

$$\begin{aligned} &k^*_t + e^*_t = \mathbb{E}[w_{it}] - m_\theta \\ &\overline{p}_{1t} - e^*_t = G^{-1}(\mathbb{P}\{p_{it} \leq \overline{p}_{1t}\}) \\ &\vdots \\ &\overline{p}_{Kt} - e^*_t = G^{-1}(\mathbb{P}\{p_{it} \leq \overline{p}_{Kt}\}) \end{aligned}$$

in the $K+3$ unknowns $(e^*_t, k^*_t, m_\theta, \overline{p}_{1t}, \ldots, \overline{p}_{Kt})$ for each t. This system has a unique solution up to m_θ and, say, $\overline{p}_{1t}$. Indeed, given that $\mathbb{P}\{p_{it} \leq \overline{p}_{1t}\}$ is identified from the distribution of the discrete performance measure in t, the sub-system that consists of the first two equations of the system admits a unique solution for e^*_t and k^*_t if m_θ and $\overline{p}_{1t}$ are known. We can then recover $\overline{p}_{kt}$ for each $k \geq 2$ as $\overline{p}_{kt} = e^*_t + G^{-1}(\mathbb{P}\{p_{it} \leq \overline{p}_{kt}\})$, since the probabilities $\mathbb{P}\{p_{it} \leq \overline{p}_k\}$ for $k \geq 2$ are identified as discussed. Hence, the vector $(e^*_t, k^*_t, \overline{p}_{2t}, \ldots, \overline{p}_{Kt})$ is identified from mean wages and the distribution of workers' performance up to m_θ and $\overline{p}_{1t}$ in each period t. The rest of the argument proceeds as in the case of a continuous performance measure in Section 6.

General Performance Function. Consider now the case in which

$$p_{it} = f_t(e_{it}, k_{it}) + \eta_{it},$$

where the noise in performance has the same properties as in the main text and for each t the function $f_t : \mathbb{R}^2 \to \mathbb{R}$ is known and continuously differentiable. We only consider the case in which the econometrician observes p_{it}, as it will be clear that we can extend the analysis to the case in which the econometrician observes the truncated version p^0_{it} by following the approach discussed above. Suppose that the equilibrium is such that effort choices and piece rates are the same for all workers and depend only on time, and let e^*_t and k^*_t be, respectively, workers' equilibrium effort and stock of human capital in period t. For each t, we have the following system of equations:

$$\begin{aligned} e^*_t + k^*_t &= \mathbb{E}[w_{it}] - m_\theta \\ f_t(e^*_t, k^*_t) &= \mathbb{E}[p_{it}] - \mathbb{E}[\eta_{it}] \end{aligned} , \tag{C25}$$

where $\mathbb{E}[w_{it}]$ and $\mathbb{E}[p_{it}]$ are observed by the econometrician and $\mathbb{E}[\eta_{it}]$ is known. We claim that (C25) has a unique solution if $e \mapsto f_t(e, \alpha - e)$ is surjective for all $\alpha \in \mathbb{R}$ and $\partial f_t(e,k)/\partial e \neq \partial f_t(e,k)/\partial k$ for all $(e,k) \in \mathbb{R}^2$. Indeed, using the first equation in (C25) to solve for k_t^*, we can rewrite the second equation in (C25) as

$$f_t(e_t^*, \mathbb{E}[w_{it}] - m_\theta - e_t^*) = \mathbb{E}[p_{it}] - \mathbb{E}[\eta_{it}]. \tag{C26}$$

Since $e \mapsto f_t(e, \alpha - e)$ is surjective for all $\alpha \in \mathbb{R}$, equation (C26) has a solution regardless of m_θ, $\mathbb{E}[w_{it}]$, $\mathbb{E}[p_{it}]$, and $\mathbb{E}[\eta_{it}]$. Now let $h(e) = f_t(e, \mathbb{E}[w_{it}] - m_\theta - e)$. Since $\partial f_t(e,k)/\partial e \neq \partial f_t(e,k)/\partial k$ for all $(e,k) \in \mathbb{R}^2$ implies that $h'(e) \neq 0$ for all $e \in \mathbb{R}$, the solution to (C26) is unique. So, if the functions $\{f_t\}_{t=0}^T$ have the properties described above, then the workers' effort and stock of human capital in each period t are identified from mean wages and mean performance measures in t up to m_θ. The rest of the identification argument proceeds as in Section 6.

We conclude by proving that $\partial f_t(e,k)/\partial e \neq \partial f_t(e,k)/\partial k$ for all $(e,k) \in \mathbb{R}^2$ and t and $e \mapsto f_t(e, \alpha - e)$ surjective for all $\alpha \in \mathbb{R}$ and t are necessary for identification. Fix t and let $G_t : \mathbb{R}^2 \to \mathbb{R}^2$ be such that $G_t(e,k) = (e + k, f_t(e,k))$. A necessary condition for identification is that the implicit equation $G_t(e,k) = v$ has a solution for e and k for any $v \in \mathbb{R}^2$. Given that G_t is continuously differentiable, it follows from Haddamard's global inverse function theorem (see Gordon [1972]) that G_t is a (C^1) diffeomorphism if, and only if, $DG_t(e,k)$, the Jacobian matrix of G_t evaluated at (e,k), has non-zero determinant for all $(e,k) \in \mathbb{R}^2$ and $\lim_{||(e,k)|| \to \infty} ||G_t(e,k)|| = \infty$, where $|| \cdot ||$ is the Euclidian norm.[56] So, a necessary condition for identification is that $\det DG_t(e,k) \neq 0$ for all $(e,k) \in \mathbb{R}^2$ and $\lim_{||(e,k)|| \to \infty} ||G_t(e,k)|| = \infty$. Since $\det DG_t(e,k) = \partial f_t(e,k)/\partial k - \partial f_t(e,k)/\partial e$, it then follows that $\partial f_t(e,k)/\partial e \neq \partial f_t(e,k)/\partial k$ for all $(e,k) \in \mathbb{R}^2$ is necessary for identification. Now observe that f_t continuously differentiable implies that either $\partial f_t(e,k)/\partial k > \partial f_t(e,k)/\partial e$ for all $(e,k) \in \mathbb{R}^2$ or $\partial f_t(e,k)/\partial k < \partial f_t(e,k)/\partial e$ for all $(e,k) \in \mathbb{R}^2$. Assume that the latter condition holds. The same argument applies when the former condition holds. Hence, $f_t(e, \alpha - e)$ is strictly increasing in e for all $\alpha \in \mathbb{R}$. Given that $||G_t(e, \alpha - e)|| = \sqrt{\alpha^2 + f_t(e, \alpha - e)^2}$ and for all $\alpha \in \mathbb{R}$, we have that $||(e, \alpha - e)|| \to \infty$ if, and only if, $|e| \to \infty$, a necessary condition for $\lim_{||(e,k)|| \to \infty} ||G_t(e,k)|| = \infty$ is that $\lim_{|e| \to \infty} |f_t(e, \alpha - e)| = \infty$ for all $\alpha \in \mathbb{R}$. Since $f_t(e, \alpha - e)$ is strictly increasing in e for all $\alpha \in \mathbb{R}$, this last condition is then equivalent to $\lim_{e \to \infty} f_t(e, \alpha - e) = \infty$ and $\lim_{e \to -\infty} f_t(e, \alpha - e) = -\infty$. Thus, $e \mapsto f_t(e, \alpha - e)$ surjective for all $\alpha \in \mathbb{R}$ is also necessary for identification. This concludes the argument.

C.3 Identification of the Remaining Human Capital Parameters in the Baseline Model

Consider the more general version of the model in which the parameters $\{\beta_t\}_{t=0}^{T-1}$ in the law of motion for human capital (2) are unknown. By the same argument as that in Section 6, the piece rates $\{b_t^*\}_{t=0}^T$ and the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified from a panel of wages and their variable components. Likewise, if the vector $(\{b_t^*\}_{t=0}^T, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ is identified, so are the terms σ_t^2, $R_{CC,t}^*$, and H_t^* for all t. In particular, r is still identified from b_T^*, σ_T^* and σ_ε^2, and b_t^0 is identified for all t from r, σ_t^2, and σ_ε^2.

First, consider the case in which the depreciation rate $1 - \lambda$ is known. Since for all $t \leq T - 1$, all terms in

$$b_t^* - b_t^0\big(1 - R_{CC,t}^* - rH_t^*\big) = b_t^0 \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(1 - b_{t+\tau}^* - R_{CC,t+\tau}^*\big)$$

except for γ_t are known from $(\{b_t^*\}_{t=0}^T, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r, \lambda)$, the parameters $\{\gamma_t\}_{t=0}^{T-1}$ are identified from this vector by the above equation, and so are the terms $R_{HK,t}^*$ for all $t \leq T - 1$. Thus, by (22), the effort choices e_t^* are identified for all t from $(\{b_t^*\}_{t=0}^T, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, \{\gamma_t\}_{t=0}^{T-1}, \lambda)$. To conclude, since for all $t \geq 1$, all terms in

$$\mathbb{E}[w_{it+1}] - \lambda \mathbb{E}[w_{it}] = e_t^* + (\gamma_{t-1} - \lambda)e_{t-1}^* + (1 - \lambda)m_\theta + \beta_{t-1} \tag{C27}$$

but β_{t-1} are known from $(\mathbb{E}[w_{it+1}], \mathbb{E}[w_{it}], \lambda, e_t^*, e_{t-1}^*)$ up to m_θ, the parameters $\{\beta_t\}_{t=0}^{T-1}$ are identified from the latter vector by the above equation up to m_θ.

Now consider the case in which $1 - \lambda$ is unknown. As in Section 6, the parameter γ_{T-1} is identified from b_{T-1}^*,

[56] A continuously differentiable function $G : \mathbb{R}^n \to \mathbb{R}^n$ with $n \geq 1$ is a diffeomorphism if G is invertible and both G and G^{-1} are continuously differentiable.

b^0_{T-1}, b^*_T, $R^*_{CC,T-1}$, r, and H^*_{T-1}, and the efforts e^*_T and e^*_{T-1} are identified from b^*_{T-1}, b^*_T, $R^*_{CC,T-1}$, and γ_{T-1}. Thus, if β_{T-1} is known, then λ is identified from (C27) evaluated at $t = T-1$ up to m_θ. Hence, the parameters $\{\gamma_t\}_{t=0}^{T-2}$ are identified by (25) from $(\{b^*_t\}_{t=0}^T, \sigma^2_\theta, \sigma^2_\varepsilon, \sigma^2_\zeta, r, \lambda)$, and so is the term $R^*_{HK,t}$ for all $t \leq T-1$. The same argument as that in the previous paragraph establishes the identification of $\{e^*_t\}_{t=0}^{T-2}$ and so of $\{\beta_t\}_{t=0}^{T-2}$.

We conclude this part by showing that it is possible for the parameter λ to be identified without assuming that β_{T-1} is known. First, note that $\mathbb{E}[w_{i0}] = e^*_0$, since we have normalized k_0 to zero. Then, from the first-order condition for effort in $t = 0$, it follows that

$$A_0 = \mathbb{E}[w_{i0}] - b^*_0 - R^*_{CC,0} = R^*_{HK,0} = \frac{\gamma_0}{\delta} \sum\nolimits_{\tau=1}^{T} (\delta\lambda)^{\tau-1}\big(b^*_\tau + R^*_{CC,\tau}\big). \tag{C28}$$

The expression of the equilibrium piece rate in $t = 0$ implies that

$$B_0 = \frac{b^*_0}{b^0_0} - \big(1 - R^*_{CC,0} - rH^*_0\big) = \frac{\gamma_0}{\delta} \sum\nolimits_{\tau=1}^{T} (\delta\lambda)^{\tau-1}\big(1 - b^*_\tau - R^*_{CC,\tau}\big). \tag{C29}$$

Both A_0 and B_0 are known from $(\{b^*_t\}_{t=0}^T, \sigma^2_\theta, \sigma^2_\varepsilon, \sigma^2_\zeta, r)$. Taking the ratio of (C28) and (C29) yields

$$\frac{\sum_{\tau=1}^{T} (\delta\lambda)^{\tau-1}\big(1 - b^*_\tau - R^*_{CC,\tau}\big)}{\sum_{\tau=1}^{T} (\delta\lambda)^{\tau-1}\big(b^*_\tau + R^*_{CC,\tau}\big)} = \frac{B_0}{A_0},$$

which can be further manipulated to obtain

$$\sum\nolimits_{\tau=1}^{T} (\delta\lambda)^{\tau-1} \left[A_0 - (A_0 + B_0)(b^*_\tau + R^*_{CC,\tau})\right] = 0. \tag{C30}$$

Equation (C30) is a polynomial of degree $T-1$ in $\delta\lambda$ with known coefficients. We can then apply Descartes's rule of signs to determine the number n of positive roots of this polynomial equation by counting the number of sign changes in the coefficients of the polynomial proceeding from lower to higher powers. If the data are such that $n = 1$, then we can determine $\delta\lambda$ and thus λ, since δ is assumed to be known.

D Supplementary Appendix

In this appendix, we provide omitted model, identification, and estimation details.

D.1 Equilibrium Contracts in the Presence of Multiple Performance Measures

In this section, we extend our analysis to the case in which there exists an observable but unverifiable additional performance measure for workers. Since the argument in this case follows many of the steps of the corresponding derivations in the case without the additional performance measure, the exposition will be terse. The environment is the same as in the case with the general human capital process, except that for each worker i and in every period t, firms now observe a noisy measure of workers' performance, p_{it}, in addition to output, y_{it}. Assume that

$$p_{it} = \gamma^e_t e_{it} + \gamma^k_t k_{it} + \theta_{it} + \eta_{it},$$

where γ^e_t and γ^k_t are known constants and η_{it} is an unobserved idiosyncratic shock to worker i's performance measure in t that is normally distributed with mean zero and variance σ^2_η and is independent of all other shocks. For ease of exposition, we assume that $\gamma^e_t \equiv 1$ and $\gamma^k_t \equiv 0$. Our analysis extends to the more general case if, and only if, $\gamma^e_t \neq \gamma^k_t$ for all t. Since the performance measure is unverifiable, firms still offer linear one-period output-contingent contracts to workers. So, worker i's wage in period t is given by $w_{it} = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it}y_{it}$, where b_{it} is the worker's piece rate in period t and I_{it} is the public information about the worker available in t. However, this case differs from the one without the performance measure in that I_{it} contains not only the worker's output realizations before t but also the realizations of the worker's performance measure before t. The definition of equilibrium is the same as before. As we did before, we focus on pure-strategy equilibria.

Learning about Ability. We first discuss how the presence of the performance measure affects learning about workers' ability in equilibrium. Consider worker i in period t, and let e_{it}^* and k_{it}^* be the worker's equilibrium effort and stock of human capital in period t, respectively. As is the case in the main text, e_{it}^* and k_{it}^* can depend on the worker's history in period t. Let $z_{it}^y = y_{it} - e_{it}^* - k_{it}^*$ and $z_{it}^p = p_{it} - e_{it}^*$ be, respectively, the part of worker i's output and performance measure in period t that cannot be explained by the worker's effort and stock of human capital in t. Since in equilibrium, agents correctly anticipate a worker's effort and stock of human capital at any point in time, the same argument as that in the main text shows that posterior beliefs about worker i's ability in period t are normally distributed with mean m_{it} and variance σ_{it}^2. In an abuse of notation, let $\sigma_{it+1/2}^2 = \sigma_{it}^2\sigma_\varepsilon^2/(\sigma_{it}^2 + \sigma_\varepsilon^2)$. By standard results, m_{it} and σ_{it}^2 evolve over time according to

$$m_{it+1} = \frac{\sigma_\eta^2}{\sigma_{it+1/2}^2 + \sigma_\eta^2}\left(\frac{\sigma_\varepsilon^2}{\sigma_t^2 + \sigma_\varepsilon^2} m_{it} + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2} z_{it}^y\right) + \frac{\sigma_{it+1/2}^2}{\sigma_{it+1/2}^2 + \sigma_\eta^2} z_{it}^p \text{ and } \sigma_{it+1}^2 = \frac{\sigma_{it+1/2}^2\sigma_\eta^2}{\sigma_{it+1/2}^2 + \sigma_\eta^2} + \sigma_\zeta^2.$$

The equations for the evolution of m_{it} and σ_{it}^2 follow from a belief-updating process in which in each period, agents first update their beliefs about a worker's ability based on the worker's output and then update their beliefs based on the realization of the worker's performance measure.[57] Now let $\sigma_{\varepsilon\eta}^2 = \sigma_\varepsilon^2\sigma_\eta^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$ and

$$z_{it} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} z_{it}^y + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} z_{it}^p. \tag{D31}$$

Straightforward algebra shows that m_{it} and $\sigma_{it}^2 \equiv \sigma_t^2$ evolve over time according to

$$m_{it+1} = \frac{\sigma_{\varepsilon\eta}^2}{\sigma_t^2 + \sigma_{\varepsilon\eta}^2} m_t + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_{\varepsilon\eta}^2} z_{it} \text{ and } \sigma_{t+1}^2 = \frac{\sigma_t^2\sigma_{\varepsilon\eta}^2}{\sigma_t^2 + \sigma_{\varepsilon\eta}^2} + \sigma_\zeta^2.$$

Thus, the evolution of posterior means and variances follow the *same* laws of motion as those in the case without the additional performance measure, except that $\sigma_{\varepsilon\eta}^2$ plays the role of the variance of the noise in output and z_{it} given by (D31) plays the role of the signal about worker i's ability in period t. When $\sigma_\eta^2 = \infty$ and the performance measure is uninformative, the laws of motion for m_{it} and σ_t^2 reduce to the laws of motion in the absence of the performance measure. If we let $\mu_t = \sigma_{\varepsilon\eta}^2/(\sigma_t^2 + \sigma_{\varepsilon\eta}^2)$, it then follows that the law of motion for a worker's reputation is still given by the expression in Lemma 1.

Dynamic Returns to Effort. We now consider the first-order conditions for worker effort when piece rates and future effort choices depend only on time. Since for any worker i, we have that $\partial\mathbb{E}[z_{it}|h_i^t]/\partial e_t = 1$ for any period t and any period-t private history h_i^t for the worker, it follows that the expressions for $R_{CC,t}$ and $R_{HK,t}(e_t)$ are the same as they are in the case with the general human capital process without the performance measure, and so are the first-order conditions for worker effort when piece rates and future behavior depend only on time.[58]

Equilibrium Piece Rates. Since the first-order conditions for effort when piece rates and future effort choices depend only on time are the same as they are in the case with the general human capital process without the performance measure, the derivation of the equilibrium piece rates follows exactly the same steps as in Appendix A. The only step in which the presence of the performance measure can alter the derivation of equilibrium piece rates is in

[57] The order in which agents use the information about a worker to update their beliefs about the worker is clearly irrelevant.

[58] More generally, $\partial\mathbb{E}[z_{it}|h_i^t]/\partial e_t = (\sigma_\eta^2 + \sigma_\varepsilon^2)^{-1}(\sigma_\eta^2 + \gamma_t^e\sigma_\varepsilon^2)$, in which case

$$R_{CC,t} = \sum_{\tau=1}^{T-t}\delta^\tau(1 - b_{t+\tau})\left(\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}\right)(1-\mu_t)(\sigma_\eta^2 + \sigma_\varepsilon^2)^{-1}(\sigma_\eta^2 + \gamma_t^y\sigma_\varepsilon^2).$$

The expression for $R_{HK,t}(e_t)$ remains the same. Since, as we show below, we can identify the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2)$ from a panel of wages by experience with information on their fixed or variable components and $p_{it} = \widehat{f}(e_{it}, k_{it}) + \eta_{it}$, where $\widehat{f}_t(e,k) = \gamma_t^e e + \gamma_t^k k + m_\theta$ is known up to m_θ and satisfies the conditions for identification for the case with the more general human capital process if, and only if, $\gamma_t^e \neq \gamma_t^k$, we can adapt the identification argument below to this more general case.

the calculation of the derivative $\partial \text{Var}[W_t^*|I_t]/\partial b$, as the presence of the performance measure potentially affects the covariance of wage payments across periods; note that I_t now describes past output and performance-measure realizations. We claim that $\partial \text{Var}[W_t^*|I_t]/\partial b$ has the same expression as in the case without the performance measure, so that the expression for equilibrium piece rates remains unchanged. The only change relative to the case without the performance measure concerns the evolution of the variance of posterior beliefs about ability. It still follows that

$$\text{Var}[W_t^*|I_t] = b^2(\sigma_t^2 + \sigma_\varepsilon^2) + 2\sum_{\tau=1}^{T-t} \delta^\tau \text{Cov}[w_t^*, w_{t+\tau}^*|I_t] + \text{Var}_0,$$

where Var_0 does not depend on b. As is the case in the main text, $w_{t+\tau}^* = w_{t+\tau}^*(b)$ with $0 \leq \tau \leq T-t$ is a worker's wage in period $t+\tau$ as a function of the piece rate in period t. We claim that $\text{Cov}[w_t^*, w_{t+\tau}^*|I_t] = b\sigma_t^2$ for all $\tau \geq 1$, which implies the desired result. As in Appendix A, $\text{Cov}[w_t^*, w_{t+\tau}^*|I_t] = b\,\text{Cov}[y_t^*, w_{t+\tau}^*|I_t]$ and

$$\text{Cov}[y_t^*, w_{t+\tau}^*|I_t] \;=\; b_{t+\tau}^*\text{Cov}[y_t^*, y_{t+\tau}^*|I_t] + (1-b_{t+\tau}^*)\text{Cov}[y_t^*, m_{t+\tau}^*|I_t]$$

for all $\tau \geq 1$, where $y_{t+\tau}^* = y_{t+\tau}^*(b)$ and $m_{t+\tau}^* = m_{t+\tau}^*(b)$ still respectively denote a worker's output and reputation in period $t+\tau$ as a function of the period-t piece rate. Hence, if $z_{t+s}^* = z_{t+s}^*(b)$ with $0 \leq s \leq T-t$ is once again the signal about a worker's ability in period $t+s$ as a function of b, then Lemma 1 implies that for all $\tau \geq 1$,

$$\text{Cov}[y_t^*, w_{t+\tau}^*|I_t] = b_{t+\tau}^*\text{Cov}[y_t^*, y_{t+\tau}^*|I_t] + (1-b_{t+\tau}^*)\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})\text{Cov}[y_t^*, z_{t+s}^*|I_t].$$

The presence of the performance measure does not change the fact that $\text{Cov}[y_t^*, y_{t+\tau}^*|I_t] = \sigma_t^2$ for all $\tau \geq 1$. Now observe that since $z_{t+s}^* = [\sigma_\eta^2/(\sigma_\eta^2+\sigma_\varepsilon^2)]z_{it}^{y*} + [\sigma_\varepsilon^2/(\sigma_\eta^2+\sigma_\varepsilon^2)]z_{it}^{p*}$,

$$\text{Cov}[y_t^*, z_{t+s}^*|I_t] = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\text{Cov}[y_t^*, z_{t+s}^{y*}|I_t] + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\text{Cov}[y_t^*, z_{t+s}^{p*}|I_t].$$

Given that $\text{Cov}[y_t^*, z_{t+s}^{p*}|I_t] \equiv \sigma_t^2$ and

$$\text{Cov}[y_t^*, z_{t+s}^{y*}|I_t] = \begin{cases} \sigma_t^2 + \sigma_\varepsilon^2 & \text{if} \quad s = 0 \\ \sigma_t^2 & \text{if} \quad s \geq 1 \end{cases},$$

we then have that

$$\text{Cov}[y_t^*, w_{t+\tau}^*|I_t] = \sigma_t^2\left[(1-b_{t+\tau}^*)\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s}) + b_{t+\tau}^*\right] + \sigma_{\varepsilon\eta}^2(1-b_{t+\tau})\left(\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}\right)(1-\mu_t).$$

The desired result follows from the fact that $\sigma_{\varepsilon\eta}^2(1-\mu_t) = \sigma_{\varepsilon\eta}^2\sigma_t^2/(\sigma_{\varepsilon\eta}^2+\sigma_t^2) = \sigma_t^2\mu_t$.

Identification. As in Section 6, equilibrium piece rates are identified from a panel of wages and their variable components. Now since $\text{Var}[w_{i0}] = (b_0^*)^2(\sigma_\theta^2+\sigma_\varepsilon^2)$, $\text{Cov}[w_{i0}, w_{i1}] = b_0^*\sigma_\theta^2$, and $\text{Var}[p_{i0}] = \sigma_\theta^2 + \sigma_\eta^2$, the vector $(\sigma_\theta^2, \sigma_\eta^2, \sigma_\varepsilon^2)$ is identified from $\text{Var}[w_{i0}]$, $\text{Cov}[w_{i0}, w_{i1}]$, and $\text{Var}[p_{i0}]$. In particular, we do not need to assume that the distribution of the shock terms η_{it} is known in order to obtain identification. The variance σ_ζ^2 is then identified from $\text{Var}[w_{i1}]$, since $\text{Var}[w_{i1}] = \sigma_\theta^2 + \sigma_\zeta^2 - \sigma_1^2 + (b_1^*)^2(\sigma_1^2+\sigma_\varepsilon^2)$ and σ_1^2 is known from $(\sigma_\theta^2, \sigma_\eta^2, \sigma_\varepsilon^2)$. Finally, given that $p_{it} = \widehat{f}(e_{it}, k_{it}) + \eta_{it}$, where $\widehat{f}_t(e,k) = e + m_\theta$ is known up to m_θ and satisfies the conditions for identification for the case with the more general human capital process, the rest of identification proceeds as in Section 6.

D.2 Additional Estimation and Identification Results

In this section, we discuss two augmented versions of our model: one in which we allow for a more flexible human capital process and another that allows for correlated measurement error in wages.

Augmented Human Capital Function. We report in Table 2 the estimates of the parameters of a more general version of our model in which the law of motion of human capital is $k_{it+1} = \lambda k_{it} + \gamma_t e_{it} + \beta \iota_{it}$, where ι_{it} represents a pure learning-by-doing investment in human capital that accrues for any period a worker spends in the labor market in that ι_{it} equals 1 in t if worker i is employed and equals 0 otherwise. As is apparent from Table 2, the estimates of the parameters of this version of the model are very similar to those of the baseline version—and the fit of the model is virtually unchanged. For instance, the estimated standard deviation of the initial distribution of ability σ_θ, output shocks σ_ε, and ability shocks σ_ζ are, respectively, 44.99, 516.74, and 5.43 for the baseline model and 48.15, 529.50, and 5.77 for the augmented model, whereas the estimates of γ_1, γ_2, and r are virtually identical across the two versions of the model. The estimate of γ_0 for the augmented model, 0.739, is somewhat lower than that for the baseline model, 0.892, much like the estimate of λ, which is 0.932 for the augmented model and 0.955 for the baseline model. Intuitively, since in the augmented model, we allow for an additional channel through which workers acquire human capital—and they do so costlessly—it is not surprising that the marginal product of effort in the production of human capital is lower. Interestingly, though, this reduction in the marginal contribution of effort to human capital is very small, thus confirming qualitatively and quantitatively the implications of the baseline model. That is, although workers now can also acquire new skills simply by working, effort still plays a key role in the human capital accumulation process and so is central to the dynamics of wages.

Table 2: Estimates of Augmented Model Parameters

Parameters	Estimates	Standard Errors
σ^2_θ, variance of initial ability	2,318.081	0.0013288
σ^2_ε, variance of shock to output	280,372.479	0.1114008
σ^2_ζ, variance of shock to ability	33.286	0.0000792
ψ_0, coefficient of degree 0 of γ_t	0.739	0.0000006
ψ_1, coefficient of degree 1 of γ_t	0.035	0.0000001
ψ_2, coefficient of degree 2 of γ_t	-0.001	1.45E-09
λ, fraction of undepreciated human capital	0.932	0.0000001
r, coefficient of relative risk aversion	0.0002	1.52E-10
β, coefficient on experience	0.844	0.0000025

For a sense of magnitudes, at the margin, an increase in effort that increases current output by 1 dollar raises the stock of human capital by 74 cents (89 cents in the baseline) at experience 1, 96 cents (1.12 dollars in the baseline) at experience 10, 1 dollar (1.17 dollars in the baseline) at experience 20, 81 cents (1.01 dollars in the baseline) at experience 30, and 39 cents (63 cents in the baseline) at experience 40. As is the case in the baseline model, the contribution of effort to human capital is sizable in all years, increasing with experience for younger workers, and declining with experience for older ones, after peaking at a marginal return of 1.01 dollars at experience 17.

Correlated Measurement Error in Wages. We now consider a more general version of the model in which we allow for measurement error in wages. Specifically, we assume that wages are observed with additive and orthogonal measurement error that follows an AR(1) process. Using the notation of Appendix C, we express the random component of the wage as r_{it} and assume that the measured random component of the wage is $\tilde{r}_{it} = r_{it} + u_{it}$, where

$$u_{it+1} = \rho u_{it} + \nu_{it+1},\ \nu_{it} \text{ i.i.d. with variance } \sigma^2_\nu, \text{ and } \mathrm{Var}(u_{it}) = \sigma^2_\nu/(1-\rho^2).$$

The identification of this version of the model proceeds as follows. We need to identify the covariance matrix of r_{it} as well as the parameters (ρ, σ^2_ν), in addition to the other parameters of the model. It is easy to verify that the covariance matrix of r_{it} is identified from the covariance matrix of $\tilde{r}_{it}$ once (ρ, σ^2_ν) are identified. Thus, we are left to show how (ρ, σ^2_ν) can be recovered from the covariance matrix of observed wages. Once this is established, the covariance matrix of "true" wages is pinned down, and so we can proceed with the identification argument presented in Section 6. To this end, consider

$$\mathrm{Cov}(\tilde{r}_{it}, \tilde{r}_{it+s}) = \mathrm{Cov}(r_{it}, r_{it+s}) + \mathrm{Cov}(u_{it}, u_{it+s}) = \mathrm{Cov}(r_{it}, r_{it+s}) + \rho^s \mathrm{Var}(u_{it}).$$

Using the fact that $\mathrm{Cov}(r_{it}, r_{it+s}) = \mathrm{Cov}(r_{it}, r_{it+k})$ for all k and s, we obtain that

$$\mathrm{Cov}(\tilde{r}_{it}, \tilde{r}_{it+2} - \tilde{r}_{it+1}) = \rho(\rho-1)\mathrm{Var}(u_{it}) \text{ and } \mathrm{Cov}(\tilde{r}_{it}, \tilde{r}_{it+3} - \tilde{r}_{it+1}) = \rho(\rho^2-1)\mathrm{Var}(u_{it}),$$

which implies that

$$\frac{\mathrm{Cov}(\tilde{r}_{it}, \tilde{r}_{it+3} - \tilde{r}_{it+1})}{\mathrm{Cov}(\tilde{r}_{it}, \tilde{r}_{it+2} - \tilde{r}_{it+1})} = \frac{\rho^2-1}{\rho-1} = 1+\rho,$$

and so ρ is identified. To see how σ_ν^2 can be recovered, note that

$$\mathrm{Cov}(\tilde{r}_{it}, \tilde{r}_{it+2} - \tilde{r}_{it+1}) = -\rho\sigma_\nu^2/(1+\rho).$$

Further details about this version of the model are available upon request.

D.3 Bargaining

In this section, we consider an extension of our baseline model in Section 3 in which in every period workers capture a fraction $\alpha \in (0, 1]$ of the expected value of their match with a firm, which reduces to our baseline model when $\alpha = 1$.[59] We omit most of the details in what follows, since derivations for this more general model follow very closely derivations for the baseline model.

Wages. Consider worker i in period t. The expected value of the match between the worker and a firm is $\mathbb{E}[y_{it}|I_{it}]$. So, if Π_{it} is the expected flow profit of the firm that employs worker i in period t, then $\Pi_{it} = (1-\alpha)\mathbb{E}[y_{it}|I_{it}]$, as the firm captures a fraction $1-\alpha$ of the expected value of the match. On the other hand, since $w_{it} = a_{it} + b_{it}y_{it}$, we have that $\Pi_{it} = \mathbb{E}[y_{it} - w_{it}|I_{it}] = (1-b_{it})\mathbb{E}[y_{it}|I_{it}] - a_{it}$. It follows that $a_{it} = (\alpha - b_{it})\mathbb{E}[y_{it}|I_{it}]$.

Learning about Ability. The process of learning about ability is the same as when $\alpha = 1$. So, posterior beliefs about a worker's ability are normally distributed with mean and variance that evolve according to (5), and the law of motion for workers' reputation can be expressed as in Lemma 1.

Dynamic Returns to Effort. Consider now the first-order condition for effort when piece rates and future effort choices depend only on time. If piece rates are $\{b_t\}_{t=0}^T$, then this necessary and sufficient first-order condition is

$$e_t = b_t + R_{CC,t} + R_{HK,t},$$

where

$$R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (\alpha - b_{t+\tau}) \left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k} \right) (1-\mu_t) \text{ and } R_{HK,t} = \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} (b_{t+\tau} + R_{CC,t+\tau}).$$

The intuition for this result is simple. The derivation of $R_{HK,t}$ does not depend on the surplus-sharing rule, so its expression does not change. The expression for $R_{CC,t}$ follows from the fact that the fixed component of a worker's wage in period $t+\tau$ with $0 \le \tau \le T-t$ is now a fraction $\alpha - b_{t+\tau}$ of the worker's expected output in $t+\tau$.

Equilibrium Piece Rates. The derivation of equilibrium piece rates follows the same steps as those in the main text. The expression for $\mathrm{Var}[W_t^*|I_t]$ is the same as when $\alpha = 1$, since the surplus-sharing rule has no impact on the second moments of wages. Since

$$\frac{\partial \mathbb{E}[W_t^*|I_t]}{\partial b} = \alpha \left(1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \right),$$

[59] The model easily extends to the case in which the fraction α depends on t. Details are available upon request.

as now workers capture only a fraction α of their expected output, it then follows that the equilibrium piece rate b_t^* is

$$b_t^* = b_t^0\left[\alpha\left(1+\gamma_t\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\right) - R_{HK,t}^* - R_{CC,t}^* - rH_t^*\right],$$

where $R_{CC,t}^*$ and $R_{HK,t}^*$ are the expressions $R_{CC,t}$ and $R_{HK,t}$ given above, with b_t^* in place of b_t for each period t, and b_t^0 and H_t^* are the same as in the baseline model.

Identification. Let $v_{it} = b_t^* y_{it}$ be worker i's variable pay in period t. Since $\mathbb{E}[w_{it}] = \mathbb{E}[(\alpha - b_t^*)\mathbb{E}[y_{it}|I_{it}] + b_t^* y_{it}] = \alpha\mathbb{E}[y_{it}]$, it follows that

$$b_t^* = \frac{\mathbb{E}[v_{it}]}{\mathbb{E}[y_{it}]} = \alpha\frac{\mathbb{E}[v_{it}]}{\mathbb{E}[w_{it}]}.$$

So, if we can identify α, then we can identify the piece rates $\{b_t^*\}_{t=0}^T$ from a panel of wages and their variable components. In order to identify the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$, note that the period-$t$ wage residual is

$$r_{it} = (\alpha - b_t^*)\mathbb{E}[\theta_{it}|I_{it}] + b_t^*(\theta_{it} + \varepsilon_{it}).$$

By the same steps as those in the derivation of the variances of the wage residuals in Appendix C (when $\alpha = 1$),

$$\mathrm{Var}[r_{it}] = \alpha^2\mathrm{Var}\big[\mathbb{E}[\theta_{it}|I_{it}]\big] + (b_t^*)^2\mathrm{Var}\big[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]\big] + (b_t^*)^2\sigma_\varepsilon^2 = \alpha^2(\sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2) + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2).$$

Now, using the fact that

$$r_{it+s} = \alpha\mathbb{E}[\theta_{it}|I_{it}] + b_{t+s}^*\big(\theta_{it} + \zeta_{it} + \cdots + \zeta_{it+s-1} - \mathbb{E}[\theta_{it}|I_{it}] + \varepsilon_{it+s}\big) + (\alpha - b_{t+s}^*)\eta_{it}^s,$$

where $\eta_{it}^s = \mathbb{E}[\theta_{it+s}|I_{it+s}] - \mathbb{E}[\theta_{it}|I_{it}]$, one can follow the same steps as those in the derivation of the covariances of wage residuals in Appendix C to show that

$$\begin{aligned}
&\mathrm{Cov}[r_{it}, r_{it+s}] \\
&\quad= \alpha^2\mathrm{Var}\big[\mathbb{E}[\theta_{it}|I_{it}]\big] + b_t^* b_{t+s}^*\sigma_t^2 + (\alpha - b_{t+s}^*)b_t^*\mathbb{E}\big[(\theta_{it} + \varepsilon_{it})\eta_{it}^s\big] + (\alpha - b_t^*)(\alpha - b_{t+s}^*)\mathbb{E}\big[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s\big] \\
&\quad= \alpha^2(\sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2) + \alpha b_t^*\sigma_t^2.
\end{aligned}$$

The rest of the identification argument follows the steps of the identification argument in Section 6.

D.4 Multi-Job Firms

We finally consider a multi-tasking extension of our model. As discussed in Section 9, we can interpret different jobs as placing different weights on different tasks. Thus, by generating a life-cycle profile of task weights, this extension can account for job mobility over the life cycle.

Setup. There are two tasks, denoted by $\ell \in \{1, 2\}$. A worker's output at task ℓ in period t is $y_{\ell t} = \beta_\ell\theta + \alpha_\ell k_t + e_{\ell t} + \varepsilon_{\ell t}$, where $e_{\ell t}$ is the worker's effort at ℓ, θ is the worker's time-invariant ability, k_t is the worker's human capital, $\varepsilon_{\ell t}$ is the shock to output at ℓ, and α_ℓ and β_ℓ are non-negative constants respectively capturing the importance of human capital and ability at ℓ. The noise terms are independent across time, workers, and tasks with $\mathrm{Var}[\varepsilon_{\ell t}] = \sigma_{\varepsilon,\ell}^2$. A worker's stock of human capital evolves over time according to $k_{t+1} = \lambda k_t + \gamma_{1t}e_{1t} + \gamma_{2t}e_{2t}$. The cost of the effort pair (e_1, e_2) is $c(e_1, e_2) = e_1^2/2 + e_2^2/2 + \nu(e_1 - \underline{e})(e_2 - \underline{e})$, where $|\nu| < 1$ and $\underline{e}$ is low enough for the equilibrium choices of effort to be greater than $\underline{e}$ in every period. Wages now are $w_t = a_t + b_{1t}y_{1t} + b_{2t}y_{2t}$, where $b_{\ell t}$ is the piece rate at task ℓ. Competition among firms implies that $a_t = (1 - b_{1t})\mathbb{E}[y_{1t}|I_t] + (1 - b_{2t})\mathbb{E}[y_{2t}|I_t]$, where I_t is the public information about a worker available at t. The definition of equilibrium is the same as the one in the single-task case. As we did before, we focus on pure-strategy equilibria. As in the single-task case, equilibrium is such that effort choices and piece rates are the same for all workers and depend only on time.

For simplicity, we assume that ability is time-invariant. Worker ability is common across tasks but can matter differently for the two tasks. We can extend the model to allow for task-specific abilities—the case we consider corresponds to the case of perfectly correlated abilities. Human capital is also common across tasks but can matter differently at different tasks. A more general model allowing for task-specific human capital is possible. Such an extension is straightforward and does not affect the substance of our analysis. Finally, we can extend the analysis to the case in which output shocks are correlated across tasks. This case is more natural to consider in the presence of task-specific abilities, as correlated abilities naturally lead to correlated signals about ability.

Learning about Ability. Let $z_{\ell t} = (y_{\ell t} - \alpha_\ell k_t^* - e_{\ell t}^*)/\beta_\ell$, where $e_{\ell t}^*$ is the equilibrium effort at task ℓ in period t and k_t^* is the equilibrium human capital in period t, be the signal about ability provided by output at ℓ in t.[60] In equilibrium, $z_{\ell t} = \theta + \widetilde{\varepsilon}_{\ell t}$ with $\widetilde{\varepsilon}_{\ell t} = \varepsilon_{\ell t}/\beta_\ell$. Thus, as in the single-task case, beliefs about a worker's ability conditional on I_t are normally distributed with mean $m_{it} = \mathbb{E}[\theta|I_t]$ and variance $\sigma_t^2 = \mathrm{Var}[\theta|I_t]$ in each t. Let $\widetilde{\sigma}_{\varepsilon,\ell}^2 = \sigma_{\varepsilon,\ell}^2/\beta_\ell$ be the variance of $\widetilde{\varepsilon}_{\ell t}$. Moreover, let

$$\sigma_\varepsilon^2 = \frac{\sigma_{\varepsilon,1}^2\sigma_{\varepsilon,2}^2}{\beta_2^2\sigma_{\varepsilon,1}^2 + \beta_1^2\sigma_{\varepsilon,2}^2},\ \xi_1 = \frac{\beta_1^2\sigma_{\varepsilon,2}^2}{\beta_2^2\sigma_{\varepsilon,1}^2 + \beta_1^2\sigma_{\varepsilon,2}^2},\ \xi_2 = \frac{\beta_2^2\sigma_{\varepsilon,1}^2}{\beta_2^2\sigma_{\varepsilon,1}^2 + \beta_1^2\sigma_{\varepsilon,2}^2},\ \text{ and }\ z_t = \xi_1 z_{1t} + \xi_2 z_{2t}.$$

It is possible to show that the laws of motion for m_t and σ_t^2 are

$$m_{t+1} = \frac{\sigma_\varepsilon^2}{\sigma_t^2 + \sigma_\varepsilon^2} m_t + \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2} z_t \ \text{ and }\ \sigma_{t+1}^2 = \frac{\sigma_t^2\sigma_\varepsilon^2}{\sigma_t^2 + \sigma_\varepsilon^2}.$$

As the noise terms are independent across tasks, we can break the belief-updating process in any period in two parts. First, firms and workers update their beliefs about θ using z_{1t}, then they update their beliefs about θ using z_{2t}. We obtain the above formulas by applying well-known results that we already used in the single-task case. For each t and $0 \le \tau \le T - t$, let $\Sigma_{t+\tau} = \sigma_t^2/(\tau\sigma_t^2 + \sigma_\varepsilon^2)$. It follows from the law of motion for m_t that a worker's reputation in period $t + \tau$ can be expressed as

$$m_{t+\tau} = \frac{\sigma_\varepsilon^2}{\tau\sigma_t^2 + \sigma_\varepsilon^2} m_t + \Sigma_{t+\tau}\sum_{s=0}^{\tau-1} z_{t+s}.$$

Note that if $\beta_1 = 0$ and ability does not matter for task 1, then $\xi_1 = 0$, $\xi_2 = 1$, and $\sigma_\varepsilon^2 = \widetilde{\sigma}_{\varepsilon,2}^2$ and the above formulas reduce to the formulas in the single-task case. This result is expected, as in this case, firms can learn about a worker's ability only by observing the worker's performance at task 2. Similar results hold if $\beta_2 = 0$. Also note that σ_ε^2 is strictly decreasing with both β_1 and β_2. Intuitively, increasing the importance of ability for either task makes workers' performance more informative about it.

Dynamic Returns to Effort. Consider now the first-order conditions for effort at each task when piece rates and future effort choices depend only on time. Suppose the piece rates for task ℓ are $\{b_{\ell t}\}_{t=0}^T$, and let

$$R_{CC,\ell t} = \sum_{\tau=1}^{T-t}\delta^\tau\big[\beta_1(1 - b_{1t+\tau}) + \beta_2(1 - b_{2t+\tau})\big](\xi_\ell/\beta_\ell)\Sigma_{t+\tau};$$
$$R_{HK,\ell t} = \gamma_{\ell t}\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\big[\alpha_1(b_{1t+\tau} + R_{CC,1t+\tau}) + \alpha_2(b_{2t+\tau} + R_{CC,2t+\tau})\big].$$

The necessary and sufficient first-order conditions for effort are

$$e_{1t} + \nu(e_{2t} - \underline{e}) = b_{1t} + R_{CC,1t} + R_{HK,1t}; \tag{D32}$$
$$e_{2t} + \nu(e_{1t} - \underline{e}) = b_{2t} + R_{CC,2t} + R_{HK,2t}. \tag{D33}$$

These equations state that at each task, the marginal cost of effort at the task is equal to its marginal benefit.

To understand the term $R_{CC,\ell t}$, note from the law of motion for a worker's reputation that $\partial m_{t+\tau}/\partial z_{\ell t} = \xi_\ell\Sigma_{t+\tau}$.

[60] Once again, $e_{\ell t}^*$ and k_t^* can depend on a worker's history in t.

Since $\partial\mathbb{E}[z_{\ell t}|h^t]/\partial e_{\ell t} = 1/\beta_\ell$, where h^t is a worker's private history in t, the increase in a worker's reputation in $t+\tau$ following a marginal increase in the effort at task ℓ in period t is then equal to $(\xi_\ell/\beta_\ell)\Sigma_{t+\tau}$. Now note that the signal about ability at one task influences future fixed pay at both tasks and that the importance of this signal for task ℓ is proportional to the importance of ability for performance at ℓ as measured by β_ℓ. To understand the term $R_{HK,\ell t}$, note that effort at task ℓ changes human capital at rate $\gamma_{\ell t}$ and that the importance of human capital for ℓ is proportional to α_ℓ. As in the single-task case, higher human capital affects both the variable component of a worker's future wages and the future signals about the worker's ability.

Solving the system (D32) and (D33) for e_{1t} and e_{2t}, we obtain that

$$e_{1t} = \frac{\nu}{\nu+1}\underline{e} + \frac{1}{1-\nu^2}\left[b_1 + R_{CC,1t} + R_{HK,1t} - \nu\big(b_2 + R_{CC,2t} + R_{HK,2t}\big)\right]; \tag{D34}$$

$$e_{2t} = \frac{\nu}{\nu+1}\underline{e} + \frac{1}{1-\nu^2}\left[b_2 + R_{CC,2t} + R_{HK,2t} - \nu\big(b_1 + R_{CC,1t} + R_{HK,1t}\big)\right]. \tag{D35}$$

Note that

$$\frac{\partial e_{1t}}{\partial b_1} = \frac{\partial e_{2t}}{\partial b_2} = \frac{1}{1-\nu^2} > 0 \text{ and } \frac{\partial e_{1t}}{\partial b_2} = \frac{\partial e_{2t}}{\partial b_1} = -\frac{\nu}{1-\nu^2}.$$

So, an increase in a task's piece rate increases effort at the task. Whether such an increase increases or decreases effort at the other task depends on whether tasks are complements ($\nu < 0$) or substitutes ($\nu > 0$). If tasks are complements, then increasing the piece rate at one task increases effort at the other task. If tasks are instead substitutes, then increasing the piece rate at one task decreases effort at the other task.

Equilibrium Piece Rates. As in the single-task case, (D34) and (D35) imply that if piece rates and effort choices from period $t+1$ on are the same for all workers and depend only on time, then effort choices from period t on are the same for all workers and depend only on time. A straightforward modification of the argument in the single-task case shows that if equilibrium piece rates and effort choices from period $t+1$ on are the same for all workers and depend only on time, then equilibrium piece rates from period t on are the same for all workers and depend only on time. Since in the last period, our multi-tasking extension reduces to the static multi-tasking model of Hölmstrom and Milgrom [1991], last-period equilibrium piece rates and effort choices are the same for all workers and (trivially) depend only on time. So, equilibrium piece rates and effort choices are the same for all workers and depend only on time. We derive the expressions for equilibrium piece rates in what follows. For ease of notation, we use the subscript $-\ell$ to denote the task other than task ℓ. So, for instance, $e_{-\ell}$ is the effort at task 2 if $\ell = 1$ and the effort at task 1 if $\ell = 2$. For each t and $0 \le \tau \le T - t$, let $w^*_{t+\tau} = w^*_{t+\tau}(b_1, b_2)$ and $W^*_t = W^*_t(b_1, b_2)$ be, respectively, a worker's wage in period $t+\tau$ and the present discounted value of a worker's wages from period t on as a function of the piece rates in t assuming that future equilibrium piece rates and effort choices depend only on time. Moreover, let $e_{1t} = e_{1t}(b_1, b_2)$ and $e_{2t} = e_{2t}(b_1, b_2)$ be given respectively by (D34) and (D35) with $R_{CC,\ell t}$ and $R_{HK,\ell t}$ evaluated at the future equilibrium piece rates. Competition among firms implies that the equilibrium piece rates in period t for a worker with public history I_t maximize

$$\mathbb{E}[W^*_t|I_t] - r\mathrm{Var}[W^*_t|I_t]/2 - c(e_{1t}, e_{2t}). \tag{D36}$$

Let us determine the solution to the problem of maximizing (D36). Since workers capture the entire value of their matches with firms, it follows that

$$\frac{\partial\mathbb{E}[w^*_t|I_t]}{\partial b_\ell} = \frac{\partial e_{1t}}{\partial b_\ell} + \frac{\partial e_{2t}}{\partial b_\ell} \text{ and } \frac{\partial\mathbb{E}[w^*_{t+\tau}|I_t]}{\partial b_\ell} = (\alpha_1+\alpha_2)\lambda^{\tau-1}\left(\gamma_{1,t}\frac{\partial e_{1t}}{\partial b_\ell} + \gamma_{2,t}\frac{\partial e_{2t}}{\partial b_\ell}\right) \text{ for } 1 \le \tau \le T-t.$$

Therefore,

$$\frac{\partial\mathbb{E}[W^*_t|I_t]}{\partial b_\ell} = \sum_{i=1,2}\frac{\partial e_{it}}{\partial b_\ell}\left[1 + \gamma_{i,t}(\alpha_1+\alpha_2)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\right].$$

Now note that

$$\mathrm{Var}[W_t^*|I_t] = b_1^2(\beta_1^2\sigma_t^2+\sigma_{\varepsilon,1}^2)+b_2^2(\beta_2^2\sigma_t^2+\sigma_{\varepsilon,2}^2)+2b_1b_2\beta_1\beta_2\sigma_t^2+2\sum_{\tau=1}^{T-1}\delta^\tau \mathrm{Cov}[w_t^*,w_{t+\tau}^*|I_t]+\mathrm{Var}_0,$$

where Var_0 is independent of (b_1,b_2), and as in the single-task case, $\mathrm{Cov}[w_t^*,w_{t+\tau}^*|I_t]$ is linear in b_1 and b_2. So,

$$\frac{\partial \mathrm{Var}[W_t^*|I_t]}{\partial b_\ell} = 2b_\ell(\beta_\ell^2\sigma_t^2+\sigma_{\varepsilon,\ell}^2)+2b_{-\ell}\beta_1\beta_2\sigma_t^2+2H_{\ell t}^*,$$

where

$$H_{\ell t}^* = \sum_{\tau=1}^{T-1}\delta^{\tau-1}\frac{\partial \mathrm{Cov}[w_t^*,w_{t+\tau}^*|I_t]}{\partial b_\ell}$$

is independent of b_1 and b_2. Given that (D32) and (D33) imply that

$$\frac{\partial c(e_{1t},e_{2t})}{\partial b_\ell} \;=\; \left(b_\ell+R_{CC,\ell t}^*+R_{HK,\ell t}^*\right)\frac{\partial e_{\ell t}}{\partial b_\ell}+\left(b_{-\ell}+R_{CC,-\ell t}^*+R_{HK,-\ell t}^*\right)\frac{\partial e_{-\ell t}}{\partial b_\ell},$$

where $R_{CC,\ell t}^*$ and $R_{HK,\ell t}^*$ are, respectively, $R_{CC,\ell t}$ and $R_{HK,\ell t}$ evaluated at the equilibrium piece rates, the necessary and sufficient first-order conditions for the problem of maximizing (D36) are

$$\begin{aligned}
&\sum_{\ell=1,2}\frac{\partial e_{\ell t}}{\partial b_1}\left[1+\gamma_{\ell,t}(\alpha_1+\alpha_2)\sum_{\tau=1}^{T-1}\delta^\tau\lambda^{\tau-1}-R_{HK,\ell t}^*-R_{CC,\ell t}^*\right]\\
&\qquad\qquad -b_1\left[\frac{\partial e_{1t}}{\partial b_1}+r(\beta_1^2\sigma_t^2+\sigma_{\varepsilon,1}^2)\right]-b_2\left(\frac{\partial e_{2t}}{\partial b_1}+r\beta_1^2\beta_2^2\sigma_t^2\right)-rH_{1t}^*=0;\\
&\sum_{\ell=1,2}\frac{\partial e_{\ell t}}{\partial b_2}\left[1+\gamma_{\ell,t}(\alpha_1+\alpha_2)\sum_{\tau=1}^{T-1}\delta^\tau\lambda^{\tau-1}-R_{HK,\ell t}^*-R_{CC,\ell t}^*\right]\\
&\qquad\qquad -b_2\left[\frac{\partial e_{2t}}{\partial b_2}+r(\beta_2^2\sigma_t^2+\sigma_{\varepsilon,2}^2)\right]-b_1\left(\frac{\partial e_{1t}}{\partial b_2}+r\beta_1^2\beta_2^2\sigma_t^2\right)-rH_{2t}^*=0.
\end{aligned}$$

To conclude, let $\mathcal{W}_{\ell t}=1+\gamma_{\ell,t}(\alpha_1+\alpha_2)\sum_{\tau=1}^{T-1}\delta^\tau\lambda^{\tau-1}-R_{HK,\ell t}^*-R_{CC,\ell t}^*$ and

$$b_{\ell t}^0=\left[1+\frac{r(\beta_\ell^2\sigma_t^2+\sigma_{\varepsilon,\ell}^2)}{\partial e_{\ell t}/\partial b_\ell}\right]^{-1}=\frac{1}{1+r(1-\nu^2)(\beta_\ell^2\sigma_t^2+\sigma_{\varepsilon,\ell}^2)}.$$

Since $\partial e_{-\ell t}/\partial b_\ell=-\nu\partial e_{\ell t}/\partial b_\ell$, we can rewrite the above first-order conditions for piece rates as

$$\begin{aligned}
b_{1t}&=b_{1t}^0\big[\mathcal{W}_{1t}-\nu(\mathcal{W}_{2t}-b_{2t})-r(1-\nu^2)(H_{1t}^*+b_{2t}\beta_1\beta_2\sigma_t^2)\big];\\
b_{2t}&=b_{2t}^0\big[\mathcal{W}_{2t}-\nu(\mathcal{W}_{1t}-b_{1t})-r(1-\nu^2)(H_{2t}^*+b_{1t}\beta_1\beta_2\sigma_t^2)\big].
\end{aligned}$$

This last system of equations admits a unique solution (b_{1t}^*,b_{2t}^*), which is independent of I_t and is the pair of equilibrium piece rates in t.

Note that the expression for $H_{\ell t}^*$ does not matter per se for the derivation of equilibrium piece rates, since

$$H_{\ell t}^*=\beta_\ell\sigma_t^2\sum_{\tau=1}^{T-t}\delta^{\tau-1}\left[\beta_1b_{1t+\tau}^*+\beta_2b_{2t+\tau}^*+(1-b_{1t+\tau}^*-b_{2t+\tau}^*)\frac{\tau(\xi_1\beta_1+\xi_2\beta_2)\sigma_t^2+\sigma_\varepsilon^2}{\tau\sigma_t^2+\sigma_\varepsilon^2}\right].$$

In particular, if $\beta_1=\beta_2=1$, then $H_{\ell t}^*=\sigma_t^2\sum_{\tau=1}^{T-t}\delta^{\tau-1}$.

By definition, $\mathcal{W}_{\ell t}$ is the wedge in period t between the social marginal benefit of effort at task ℓ and the private marginal benefit of effort at the task. A piece rate at task ℓ in period t equal to $\mathcal{W}_{\ell t}$ would induce workers to exert the first-best level of effort at task ℓ. As in the single-task case, the piece rate at task ℓ in period t is proportional to $\mathcal{W}_{\ell t}$ minus a term, $r(1-\nu^2)(H^*_{\ell t}+b^*_{-\ell t}\beta_1^2\beta_2^2\sigma_t^2)$, that reflects the insurance workers demand against the uncertainty about their ability. Also as in the single task case, the constant of proportionality captures the standard risk-incentives trade-off. In contrast to the single-task case, the insurance component of the piece rate at task ℓ in t features an additional term that depends on the period-t piece rate at the other task. This result is intuitive. Because ability is common across tasks, uncertainty about ability implies that an increase in the piece rate at a task increases the risk associated with (the contemporaneous) performance at the other task. Another difference from the single-task case is that the piece rate at task ℓ in period t features an additional term proportional to $-\nu(W_{-\ell t} - b_{-\ell t})$. This term captures both the interdependence in the human capital accumulation process across tasks—by exerting effort at one task, workers affect their productivity at both tasks—and the fact that providing incentives for effort at one task affects the incentives for effort at the other task. When tasks are substitutes, this term tends to depress piece rates.