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

Braz Camargo† Fabian Lange‡ Elena Pastorino§

July 2021

**Abstract**

Performance pay for most workers makes up only a small fraction of total pay. In this paper, we show that performance pay is nevertheless important for the dynamics of wages over the life cycle because of the incentives it provides for human capital acquisition. We argue so within a model that combines three key mechanisms for wage growth and dispersion, namely, human capital accumulation on the job, employer learning about workers' ability, 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. Our model admits a decomposition of performance pay over the life cycle into four terms that capture: *i*) the trade-off between risk and incentives characteristic of moral-hazard situations; *ii*) the insurance that firms provide against uncertainty about ability; *iii*) incentives for effort due to this uncertainty (career concerns); and *iv*) incentives for effort from human capital acquisition. Despite its parsimony, the model fits the data very well, including the observation that performance pay as a share of total pay, which measures the sensitivity of pay to performance, first increases and then declines with experience after peaking at around 20 years, contrary to the prediction of career-concerns models that this ratio should be increasing especially at the end of the life cycle. Our estimates imply that human capital acquisition and insurance against uncertainty about ability are quantitatively the most important determinants of the sensitivity of pay to performance. Importantly, we also find that through the cumulative impact of effort on human capital acquisition, incentives for performance are a critical source of wage growth and dispersion over the life cycle.

Keywords: Uncertainty and Learning, Human Capital Acquisition, Dynamic Moral Hazard, Performance Incentives, Wage Growth, Inequality, Identification, Structural Estimation

*We thank Joe Altonji, Peter Arcidiacono, Flavio Cunha, Christian Dustmann, Jeremy Fox, George-Levi Gayle, Limor Golan, Hyejin Ku, Attila Lindner, Bob Miller, 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). This paper is dedicated to the memory of Eddie Lazear, whose encouragement, mentorship, and generosity we will never forget.

†Sao Paulo School of Economics. E-mail: braz.camargo@fgv.br.

‡McGill University. Address: Department of Economics, Leacock Building, 855 Sherbrooke Street West, Montreal Quebec H3A2T7, Canada. E-mail: fabolange@gmail.com.

§Hoover Institution, Stanford University, SIEPR, and Federal Reserve Bank of Minneapolis. 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 considered 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 how wages and their dispersion grow with experience as firms and workers learn about differences in ability among workers (Farber and Gibbons [1996]).

Another potential source of 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]). 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 less attention in the study of the dynamics of wages (Rubinstein and Weiss [2006]). However, acquiring human capital normally requires effort and investments in it can either substitute for producing output, as in models of on-the-job training following Ben-Porath [1967], or be complementary to it, as in learning-by-doing models. By influencing workers' effort on the job, performance pay, although small, may then affect how rapidly wages grow with experience by affecting how much human capital workers acquire.

In this paper, we argue that performance incentives can be important for the growth and dispersion of wages over the life cycle because they support the acquisition of human capital. To this purpose, we propose a tractable model of the labor market that combines human capital accumulation, uncertainty and employer learning about workers' ability, and incentives for performance. By so doing, 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 profile of wages and their fixed and variable components over time. 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. Second, based on this decomposition, we show that variable pay provides a rich source of information that can be used to identify our model, which nests common models of human capital investment, learning, and performance incentives. Third, our model resolves an empirical failure that we document for existing "career-concerns" models of learning and performance incentives over the life cycle: such models imply that relative to total pay, performance pay *increases* with labor market experience whereas the data strongly suggest that it eventually *decreases*. Finally, using our estimated model, we demonstrate that 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 over the life cycle when firms are uncertain about

workers' ability and use performance pay to incentivize workers to expend effort on the job. The idea of these so-called career-concerns models is simple. When ability is uncertain and firms gradually learn about workers' unobserved ability based on their output, workers anticipate that good performance favorably influences potential employers' perceptions about their ability and so positively affects their future wages. Accordingly, concerns about the market expectation of their abilities—"career concerns"—stimulate workers to exert effort and so can substitute for explicit performance pay. We add to this framework another dimension of career concerns: by exerting effort on the job, workers not only affect their output in a period but also their 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 their human capital. These implicit incentives interact with explicit incentives from performance pay to determine workers' effort and thus the human capital that workers acquire with experience. Through this mechanism, performance incentives affect both the growth and dispersion of wages.

More formally, we model the labor market as consisting of homogeneous risk-neutral firms and heterogeneous risk-averse workers of unknown ability, which is subject to persistent shocks. Employed workers exert effort, which influences both a worker's output and human capital. A worker's effort and acquired human capital are observed only by the worker. Instead, a worker's output in a period or *performance*, 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 it over time. Firms compete for workers by offering short-term employment contracts in which variable pay depends on a worker's output.[1]

We characterize equilibrium wages and, as mentioned, decompose the ratio of performance pay to total pay or "piece rate" of the equilibrium contract, which measures the sensitivity of pay to performance, into four terms that reflect fundamental life-cycle forces, are readily interpretable, and take the form of simple functions of the model primitives.[2] The first term of this decomposition captures the standard trade-off between risk and incentives familiar from static moral-hazard models (Hölmstrom [1979]) and changes in this trade-off over time as uncertainty about a worker's ability varies over the life cycle. The remaining three terms capture how workers' demand for insurance against uncertainty about ability, learning about ability, and human capital acquisition lead to deviations between the statically optimally piece rate and the dynamically optimal one implied by our model.

To elaborate, the second and third terms negatively affect piece rates. The second term captures workers' value for the insurance that a wage contract provides, through lower piece rates, against the uncertainty workers face because of the process of learning about ability and the shocks to ability. Intuitively, lower piece rates partially insure workers against wage risk as they reduce the contemporaneous covariance between pay and performance. But

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

[2] In our model, variable pay is proportional to performance. The factor of proportionality, namely, the contract piece rate, then equals both the ratio of performance pay to total pay and the (marginal) sensitivity of pay to performance.

as workers accumulate experience and face a shorter time horizon, they demand less insurance and so the size of this term decreases in magnitude. This term then contributes to an increase in the 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 become less important over time. 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, which is positive, is proportional to the difference between the social and private marginal returns to human capital acquisition. The private-returns component of this term is reminiscent of the career-concerns term described in the previous paragraph, as it represents the incentive for effort a worker faces out of the desire to invest in human capital. However, because of learning about ability and workers' risk aversion, these private returns tend to be smaller than the social returns at any point in time. When the difference between the social and private returns to human capital is large, firms pay workers larger piece rates to encourage workers to exert more effort. Thus, unlike the previous two terms, this fourth term tends to positively contribute to piece rates. As experience accumulates, though, this term declines in magnitude and adds progressively less to piece rates.

We show that the model is identified from panel data on wages and their fixed or variable components. In particular, 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.[3] The life-cycle profile of piece rates itself can be recovered from the ratio of variable pay to total pay in each year of experience under the assumption of free entry of firms in the labor market. We also show that these arguments extend to the case of observable and unobservable heterogeneity in the model primitives, including in workers' human capital process and degree of risk aversion, and in the process that governs learning about workers' ability. These identification results also apply 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 several firm-level data sets.[4]

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 sector. We use information on wages and their variable component—performance pay—that is available between 1981 and 1988. We document that in the BGH data, performance pay as a fraction of total pay first increases and then declines with experience. We confirm that this same hump-shaped pattern of performance pay is present in other firm-level data as well as in data on variable pay from the Panel Study of Income Dynamics (PSID). These findings

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

[4]For important related work on the identification and estimation of static and dynamic moral hazard models of executive compensation, see Margiotta and Miller [2000], Gayle and Miller [2009, 2015], and Golan et al. [2015]. Differently from these authors, we consider a model that features uncertainty and learning about ability and persistent shocks to ability, and rely only on the experience profile of wages and their variable component for its identification.

directly contradict the prediction of career-concerns models that performance pay relative to total pay increases throughout 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, not only successfully matches the hump-shaped pattern of performance pay relative to total pay but also the increase and curvature of mean wages, as well as the hump-shaped profile of the variance of wages, with experience.

Our estimates suggest that individuals differ in their ability at entry in the labor market and that uncertainty about ability is present throughout the life cycle—in fact, it increases with experience due to the shocks to ability that accumulate over time, despite firms and workers learning about ability. We estimate that the insurance against this uncertainty that a wage contract provides is the major factor depressing performance pay. The intuition for the relatively low level of performance pay in the data is simple from an asset pricing perspective. Since performance pay, to reward effort, is high whenever output is high and so news about ability and future compensation are positive, workers employed under performance-pay contracts effectively hold a portfolio of state-contingent claims to output, whose value increase with a worker's ability. In particular, this portfolio pays out more in good times—when signals about ability are high—and less in bad times—when signals about ability are low. However, risk-averse investors prefer, and are willing to pay a premium for, assets that diversify their risk and correspondingly demand contracts that reduce it. As a result, in equilibrium performance pay tends to be low to hedge workers against the correlated risk in lifetime wages induced by uncertainty about ability. Note that this argument confirms and extends the early intuition of Harris and Hölmstrom [1982] on the role of the dynamic insurance provided by wage contracts for the evolution of wages with experience.[5]

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. Using our decomposition of estimated piece rates, we show that insurance against uncertainty about ability and human capital acquisition are quantitatively the most important determinants of the sensitivity of pay to performance. On the contrary, career-concerns incentives and the life-cycle variation in 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.

We also estimate that performance incentives are critical to life-cycle wage growth because they encourage workers to exert effort, which in turn crucially contributes to the accumulation of human capital. Our findings imply that the effort expended to produce output is complementary to that spent investing in human capital—that is, our estimates support the notion that human capital is accumulated through learning-by-doing. As the variance of performance pay amounts to a large fraction of the variability of wages over time, especially during the first half of

[5] Our estimated degree of risk aversion falls within the range of existing estimates. See Section 7 for details.

the life cycle, performance incentives are key to wage dispersion too. Specifically, when we take into account the impact of effort on human capital acquisition, we estimate that performance incentives provided through variable pay account for more than 30% of wage growth and for no less than 44% of the variability of wages over the first 30 years of labor market experience in our data. Thus, by supporting effort and so indirectly the accumulation of human capital, 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.

An important lesson from our work is that common statistical decompositions of wage dispersion among workers (as in Abowd et al. [1999]) can be misleading. In particular, one might be interested in decomposing the variance of wages into that of "worker" and "firm" effects as well as of a residual. These terms are often interpreted to capture, respectively, differences in ability among workers, in firm productivity including output risk, and in other unmeasured factors. According to our framework, such an exercise would attribute a large portion of the observed variation in wages to dispersion in workers' ability. But our model also predicts that performance pay declines with the uncertainty about individual ability. Indeed, if it were possible to eliminate differences in ability among workers, then our model would imply that the variance of wages would substantially *increase*. Intuitively, without uncertainty about ability, equilibrium contracts would feature much higher piece rates, since workers would no longer demand insurance against 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 accounting for the endogeneity of the wage structure to the degree of risk and uncertainty in the labor market when assessing the role of different sources of wage dispersion.[6]

This result also provides a cautionary note for the debate on inequality, as it 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. In particular, lower dispersion in initial ability, for instance, through better schooling, may induce firms to offer wages more sensitive to performance. Then, more homogeneous groups of workers in terms of 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 papers on: *i*) the importance of human capital acquisition with experience for wage growth (Heckman et al. [1998], Gladden and Taber [2009], and Sanders and Taber [2012]); *ii*) distinguishing the impact of uncertainty and heterogeneity among individuals on wage dispersion (Cunha et al. [2005] and Cunha and Heckman [2016]); *iii*) measuring the role of uncertainty and learning about ability for careers (Miller [1984], Kahn and Lange [2014], and Pastorino [2019]); 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

[6] 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.

Golan et al. [2015]).[7] In related work, Golan et al. [2015] analyze how moral hazard and human capital acquisition determine wages. While we focus on the life-cycle dynamics of wages and their components for (supervisory) workers, those authors study the relationship between firm size and executive pay. In their work, executives acquire general and firm-specific human capital, and choose among jobs and firms that differ in the pecuniary and non-pecuniary benefits they offer. We refrain from studying the assignment of workers to jobs but we incorporate in our framework unobserved worker ability, allow for persistent shocks to ability and so productivity, and consider a richer agency problem with multiple possible effort levels for workers to capture workers' varying labor supply and investment choices over the life cycle. Golan et al. [2015] rely on bond prices to recover executives' preferences. In contrast, our model is identified just from data on wages and their components.[8]

Much work emphasizes the importance of unobserved heterogeneity, which is at the heart of our learning and dynamic incentive mechanisms, for the wage process. For instance, Geweke and Keane [2000] provide evidence from the PSID on the role of transitory shocks and individual heterogeneity for the dynamics of individual wages. Also based on the PSID, Meghir and Pistaferri [2004] document the importance of idiosyncratic transitory and permanent components for the wage process, in particular of unobserved heterogeneity for the variance of wages. For related evidence 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].[9] Dustmann and Meghir [2005] show the importance of match-specific effects and heterogeneous returns to human capital, in the form of a correlated random-coefficients model, for the impact of experience on wages. Adda and Dustmann [2020] estimate the contribution of human capital and unobserved ability to wage growth using a rich dynamic model of workers' occupational choice.

The paper proceeds as follows. We introduce our data in Section 2, where we also document that performance pay relative to total pay has a hump-shaped pattern over the life cycle. Section 3 presents 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 discusses the estimation of the model and its results, and Section 8 explores the implications for wage growth and dispersion. Section 9 concludes. The appendices contain all omitted details.

[7] 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 [2019] 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.

[8] In their framework, only two effort levels are possible, "effort" or "shirking," which simplifies issues of incentive compatibility of compensation contracts. Our model is identified up to the second derivative of the effort cost function and a level normalization. The model in Golan et al. [2015] is identified up to the non-pecuniary utility and human capital acquired upon shirking.

[9] 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 earnings residuals.

## 2 Evidence on the Sensitivity of Pay to Performance

In this section, we provide evidence on the experience profile of the ratio of variable (or performance) pay to total pay using public data from the PSID as well as proprietary data from the personnel records of two firms. These records were first described 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 have the advantage that they contain information on fixed pay $f_{it}$ and variable pay $v_{it}$, which together account for the total compensation or *wage* of worker $i$, $w_{it} = f_{it} + v_{it}$. In models with variable pay proportional to output, $v_{it} = b_i y_{it}$, and with free entry of firms such as ours, the ratio $\mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$ of average variable pay to average total pay then measures the piece rate $b_t$, that is, the sensitivity of pay to performance.[10] Based on these data spanning across 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 career-concerns models.

**PSID.** We focus on the main PSID sample, excluding the poverty, latino, and immigrant sub-samples, and consider male heads of households aged 21 to 65 observed between 1993 and 2013 with valid education information, that is, with more than 0 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 and years of education (minus six). 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 receiving a performance pay of zero—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.

In Figure 1, we show how the sensitivity of pay to performance varies with experience by broad industry groups, that is, in manufacturing, transport, services, and in the financial, insurance, and real estate (FIRE) industry for three cohorts of individuals, respectively, 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 qualitatively and quantitatively 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.[11] The PSID data thus

[10] With linear incentive contracts, variable pay is given by $b_t y_{it}$, where $b_t$ is the contract piece rate that measures the sensitivity of pay to a worker's output, $y_{it}$. By the assumption of free entry of firms in the labor market, average wages equal average output at each $t$. Hence, we can recover the piece rate in each $t$ as the ratio $\mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$.

[11] Not all individuals in the sample are employed in the four industry groups shown, but the sample size for the remaining industries is so

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 Personnel Records.** We use firm-level 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 identify them by the names of the authors who first analyzed these data and so refer to these firms as the Baker-Gibbs-Holmström (BGH) firm and the Gibbs-Hendricks (GH) firm. For both firms, we only have information about white-collar workers—managers 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, was not separately reported prior to 1981. The BGH data contain 36,695 person-year observations and 9,800 unique individuals. Since we only have information about managers at this firm, the average salary is fairly high, namely, \$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 instead cover the years from 1989 to 1993—we cannot reveal the industry the 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 contain information about all white-collar employees of the firm, the average salary is lower than in the BGH data and close to \$40,000 (also 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. We estimate the model using the BGH data because its fairly long panel covering 8 years with information on fixed and variable pay enables us to better trace life-cycle patterns. The BGH data also represents a touchstone in the personnel literature and is therefore useful to connect our findings to extant papers, including some of our own such as Kahn and Lange [2014] and Pastorino [2019].

The left panel of Figure 2 reports the experience profile of the sensitivity of pay to performance in the BGH data for managers with 21 to 65 years of age. This profile is hump-shaped: it increases over the first 20 years of labor market experience and then unambiguously decreases over the remaining 20 years. These data thus reject the basic implication of the standard career-concerns model based on reputational concerns and explicit contracts developed by Gibbons and Murphy [1992]. As shown in Figure 3, performance pay in the GH data is likewise hump-shaped over the life cycle and therefore at odds with the standard career-concerns model. Analogous patterns arise if we restrict attention to skilled (college) or unskilled (no college) workers; see the center and right panels, respectively, of Figures 2 and 3. As we will discuss in Subsection 5.2, the location of the peak of these profiles will prove informative about the relative importance of learning about ability and human capital acquisition early in the life cycle as well as the speed of the two processes.

small that discernible patterns cannot be easily detected.

# 3 Model

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

## 3.1 Environment

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 $\theta_{it}$, which is unobserved and subject to persistent shocks. When employed, workers exert effort $e_{it}$ and acquire human capital $k_{it}$ that depends on their effort $e_{it}$. Ability $\theta_{it}$ is not directly observed by any market participant, including workers. Workers, unlike firms, directly observe their effort and human capital. Finally, all firms observe output $y_{it}$ as well as the terms of wage contracts. Note that since a worker's ability is unknown to all whereas output is observable to all, ours is a model of symmetric learning about ability.

**Production.** The production technology is common to all potential producers and entry in this market is free.[12] Worker $i$ in period $t$ produces output $y_{it}$ according to:[13]

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

The shock to output $\varepsilon_{it}$ can be interpreted as either true risk in the worker's output or noise in its measurement. Worker $i$'s ability evolves over time according to the process $\theta_{it+1} = \theta_{it} + \zeta_{it}$, where $\zeta_{it}$ is an unobserved shock to the worker's ability between periods $t$ and $t+1$. A worker's initial ability is normally distributed with mean $m_\theta$ and variance $\sigma^2_\theta$. Similarly, output noise and shocks to ability are normally distributed with mean zero and variances $\sigma^2_\varepsilon$ and $\sigma^2_\zeta$, respectively. When $\sigma^2_\zeta = 0$, ability is fixed over time. Allowing for ability shocks implies that uncertainty about ability need not decline over time and therefore implicit incentives from career concerns do not necessarily decline with experience.

**Human Capital.** The human capital of any worker $i$ evolves with the worker's effort according to the process

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

[12] This market could be one of many markets segmented by location, occupation, or industry, in each of which the matching of workers and firms is subject to informational frictions. In particular, a labor market is defined by the distribution of a single index of unknown worker productivity as well as common learning and human capital processes across firms. What is important is that these markets are sufficiently separate that employment opportunities in other markets are irrelevant for workers' decisions in a given market. In our empirical application, we focus on the market for managers in a service industry.

[13] Like Gibbons and Murphy [1992], p. 476, we allow effort to be negative and so workers to destroy output because positive effort might not be optimal for a worker. We can then conveniently 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, which is the empirically relevant range, and derive conditions for equilibrium piece rates to belong to this interval in the Appendix.

where $1-\lambda \in [0,1]$ is the depreciation rate, $\gamma_t \in \mathbb{R}$ is the rate at which effort in period $t$ changes the stock of human capital in period $t+1$, and $\beta_t \geq 0$ is a deterministic term common to all workers.[14] Workers have a common stock of human capital at entry in the labor market, $k_0$. By absorbing $k_0$ into $m_\theta$, we let $k_0 = 0$ without loss.[15] This formulation of the human capital process encompasses both the situation in which human capital acquisition is complementary to current production ($\gamma_t > 0$), as in standard learning-by-doing models, and the situation in which human capital acquisition is rival to it ($\gamma_t < 0$), as in models à la Ben-Porath [1967]. Also note that this specification of the amount of efficient labor that a worker supplies in a period as $\theta_{it} + g_{it} + e_{it} + \varepsilon_{it}$, where $g_{it} = k_{it}$, extends that of Bagger et al. [2014], who specify it as $\theta_i + g_{it} + \varepsilon_{it}$. Differently from these authors, we allow: *i*) the individual heterogeneity parameter $\theta_{it}$ to be unknown to workers and firms and vary over time; *ii*) the stock of human capital acquired on the job $g_{it}$ to evolve endogenously as a function of a worker's past effort; and *iii*) 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 the law of motion of human capital depends nonparametrically on effort and workers differ in terms of their ability distribution and human capital process.

**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

$$-\exp\left\{-r\left[\sum_{\tau=0}^{T-t}\delta^\tau\left(w_{t+\tau}-e_{t+\tau}^2/2\right)\right]\right\},$$

where $r > 0$ is the coefficient 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.[16]

**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 focus on linear contracts so that worker $i$'s wage in period $t$ is $w_{it} = a_{it} + b_{it}y_{it}$, where $a_{it}$ is worker $i$'s fixed pay, $b_{it}y_{it}$ is worker $i$'s variable pay, and $b_{it}$ is worker $i$'s piece rate. Note that the assumption of one-period contracts is equivalent to the assumption of renegotiation-proof long-term contracts.[17] We focus on linear contracts for three reasons. First, the 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

[14] Note that $\beta_t$ can capture additions to human capital that depend on observable investment activities such as formal training. Our model focuses on investments through effort on the job that are unobservable and convolute the production of human capital and output.

[15] Our production function of skills $h_{it+1} \equiv \theta_{it+1} + k_{it+1}$ with $k_{it+1} = \lambda k_{it} + \gamma_t e_{it}$ can be interpreted as a log form of the production function $\tilde{h}_{it+1} = \tilde{a}_{t+1}(\theta_{it}, \theta_{it+1})\tilde{h}_{it}^{\lambda}\tilde{e}_{it}^{\gamma_t}$ with $h_{it} = \ln(\tilde{h}_{it})$, $a_{t+1}(\theta_{it}, \theta_{it+1}) = \ln(\tilde{a}_{t+1}(\theta_{it}, \theta_{it+1})) = \theta_{it+1} - \lambda\theta_{it}$, $e_{it} = \ln(\tilde{e}_{it})$, and $h_{i0} = \theta_{i0}$. We can allow for heterogeneous initial stocks of human capital by assuming that $k_{i0}$ is known but random. We can also let $k_{it}$ fluctuate stochastically over time, which is equivalent to $\sigma_\varepsilon^2$ increasing over time.

[16] 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 dynamic moral hazard models.

[17] See Gibbons and Murphy [1992] for a proof of this result. Their proof immediately extends to our environment.

strength of contractual incentives, which is a key feature of interest in our analysis, through a single one-dimensional measure, namely, the piece rate $b_{it}$.

**Wages.** Competition among firms implies that expected wages in every period equal expected output. Hence, worker $i$'s wage in period $t$ satisfies

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

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 $t$, which 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 depends on the worker's conditional expected ability in $t$.

**Strategies and Equilibria.** A worker's history in period $t$ consists of the worker's effort and output histories in the period. These are, respectively, 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 their 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 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 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 leaves the current employer with probability equal to the corresponding empirical probability of separation of workers from the firm in our data.[18] Hence, the model is consistent with worker turnover in equilibrium.

## 3.2 Remarks

Our model nests a number of well-known models in the literature on 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 either ??? or ???. When effort is not a choice variable but fixed, the model further specializes to one of "passive" human capital acquisition with experience. If, instead, effort does not contribute to human capital accumulation and ability is not subject to shocks ($\gamma_t = 0$ at each $t$ and $\sigma^2_\zeta = 0$), then our model simplifies to the career-concerns model with explicit incentives developed by 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

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

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 (Farber and Gibbons [1996]).

Our functional-form assumptions are common in the literature and allow us to completely characterize equilibrium. In particular, because output is linear in its components, contracts are linear in output, shocks to ability are additive, and ability, ability shocks, and output are normally distributed, our model with CARA preferences admits a mean-variance representation as in Gibbons and Murphy [1992]. This feature further 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 only dependent on time. In Section 6 and the appendices, we consider more general versions of the baseline 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 for the baseline model presented below.

# 4 Informal Equilibrium Derivation

We now informally discuss the 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 the equilibrium piece rates. A more formal characterization of the equilibrium follows in Section 5.

## 4.1 Learning About Ability

Firms and workers learn about a worker's ability 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 $z_{it} = y_{it} - e_t^* - k_t^*$ the portion of the worker's output in period $t$ that cannot be explained by the worker's effort or human capital. Then,

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

is the signal about a worker's ability in period $t$ that firms and workers extract from the worker's output. Since initial ability as well as 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 so fully described by their posterior mean and variance, respectively, $m_{it} = \mathbb{E}[\theta_{it}|I_{it}]$ and $\sigma_{it}^2 = \mathrm{Var}[\theta_{it}|I_{it}]$ where $m_{i0} = m_\theta$ and $\sigma_{i0}^2 = \sigma_\theta^2$. We refer to $m_{it}$ as worker $i$'s *reputation* in period $t$. 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 $\sigma_{it}^2$ in (5), respectively, describe how a worker's reputation and the variance of pos-

terior beliefs about a worker's ability change over time.[19] Observe that the variance $\sigma_{it}^2$ evolves independently of the realization of $z_{it}$ and so is common to all individuals in $t$. Thus, we can suppress the subscript $i$ and simply denote this variance by $\sigma_t^2$. Since signals do not perfectly reveal ability and ability is subject to shocks, uncertainty about ability persists throughout a worker's career and, as captured by $\sigma_t^2$, eventually reaches the nonnegative fixed point $\sigma_\infty^2$ by (5). The variance $\sigma_t^2$ monotonically decreases to $\sigma_\infty^2$ if $\sigma_\theta^2 > \sigma_\infty^2$ and monotonically increases to $\sigma_\infty^2$ 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.[20] By iterating on the law of motion for $m_{it}$ in (5), we can trace out the evolution of a worker's reputation as 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:

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

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

## 4.2 Dynamic Returns to Effort

We now discuss how workers' desire to increase their reputation—the market expectation of their ability—and to acquire human capital affect the returns to effort. As we show in Section 5, the expressions derived here apply to the unique equilibrium of our model, which has the property that piece rates only depend on the time index $t$, not on individual output histories. Consider then an individual who faces a sequence of piece rates $\{b_t\}_{t=0}^{T}$. Since equilibrium piece rates do not depend on the worker's identity, we suppress the index $i$ to keep the notation simple. In what follows, we first present the worker's problem and derive the first-order conditions determining the worker's choice of effort in each period. We then show that we can decompose this first-order condition into distinct terms that capture career-concerns 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. 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 signals about ability are normally distributed, it follows from (3) and Lemma 1 that 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 $\mu$ and variance $\sigma^2$, then

[19] 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.

[20] See Hölmstrom [1999] for a proof of these facts. Straightforward algebra shows that $\sigma_\infty^2 = [\sigma_\zeta^2 + (\sigma_\zeta^4 + 4\sigma_\zeta^2\sigma_\varepsilon^2)^{1/2}]/2$. Kahn and Lange [2014] refer to this process as "learning about a moving target" and find evidence for it from the correlation patterns between performance ratings and total pay in the BGH data.

$\mathbb{E}\big[\exp\{rX\}\big] = \exp\big\{r\mu - r^2\sigma^2/2\big\}$. Thus, $e_t$ maximizes $U_{it}(e_t)$ if, and only if, it maximizes

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

**First-Order Conditions for Effort.** The wage contract in (3) implies that $\partial\mathbb{E}[w_{it}|h_i^t]/\partial e_t = b_t$. 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 through its effect on the worker's future human capital stock $k_{it+\tau}$, which impacts both the *fixed* and *variable* components of future pay.[21] The first-order condition for worker $i$'s effort in period $t$ is then

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

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 consists of the impact 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 the Appendix, we show that we can express the first-order condition in (7) as

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

where[22]

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

It follows from (8) and (9) that effort is positive if piece rates are in the unit interval, which is the empirically relevant case.[23] The terms $R_{CC,t}$ and $R_{LBD,t}$ describe a worker's dynamic marginal benefit of effort that arises from the impact of effort on the worker's reputation and human capital, respectively. They capture the effect of effort in period $t$ on a worker's future wages through its effect on their *fixed* and *variable* components.

To understand the term $R_{CC,t}$, observe that higher effort in period $t$ increases the period-$t$ signal about worker $i$'s ability, which raises the worker's reputation in all future periods. By equation by (3), the fixed component of worker $i$'s wages in all these periods also increases. The term $R_{CC,t}$ thus captures the standard career-concerns incentive

[21] Note that effort in $t$ does not affect the variance of future wages. This standard feature (see Gibbons and Murphy [1992]) is implied by our normal prior-signal information structure, the linearity of output in ability and effort, and the exponential-linear setup for preferences and contracts. To see why effort does not affect the variance of future wages, recall that $b_t$ is 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 the worker's future reputation. Similarly, a worker's stock of human capital has no impact on the variance of output or wages.

[22] The marginal benefit of effort in $t$ does not vary with $e_t$ since $R_{CC,t}$ and $R_{LBD,t}$ 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 are the same. Hence, individuals facing the same current and future piece rates, and choosing the same effort in the future, also behave identically at $t$. This feature is key to establishing that the unique equilibrium is symmetric and that effort choices and piece rates depend only on time.

[23] 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 $b_{t+\tau} \geq 0$.

from 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. Formally,

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

for all $1 \leq \tau \leq T-t$ by Lemma 1. 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 $t+\tau$ by

$$(1-b_{t+\tau})\left(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t).$$

The term $R_{CC,t}$ is simply the present-discounted value of all these marginal increases.

To understand next the term $R_{LDB,t}$, 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. In addition, by changing the worker's stock of human capital, effort in period $t$ affects future output signals about the worker's ability, and so the worker's future reputation and *fixed* pay.

To elaborate, a marginal increase in effort in period $t$ leads to an increase in worker $i$'s undepreciated stock of human capital and output in period $t+\tau$ by $\gamma_t\lambda^{\tau-1}$. This increase in output, in turn, leads to an increase in the variable component of the worker' wage in period $t+\tau$ by the amount $b_{t+\tau}\gamma_t\lambda^{\tau-1}$. But an increase in worker $i$'s human capital in period $t+\tau$, by increasing the worker's output in $t+\tau$, also increases the magnitude of the signal $z_{it+\tau}$ about the worker's ability observed at the end of $t+\tau$. Then, by the same argument for the derivation of the term $R_{CC,t}$, such an increase in signals correspondingly increases worker $i$'s expected future reputation—as larger signals induce the market to infer that a worker is of higher ability. As a result, the increase in worker $i$'s output in period $t+\tau$ by $\gamma_t\lambda^{\tau-1}$ leads to an increase in the fixed component of the worker's wages from $t+\tau$ on equal to $\gamma_t\lambda^{\tau-1}R_{CC,t+\tau}$. This term captures the impact of human capital on future career-concerns incentives and so on future fixed pay. Thus, the total increase in worker $i$'s expected value of wages from period $t+\tau$ on resulting from a marginal increase in effort in period $t$ is $\gamma_t\lambda^{\tau-1}\big(b_{t+\tau}+R_{CC,t+\tau}\big)$. The term $R_{LBD,t}$ is the present-discounted value of all these marginal increases.

### 4.3 Equilibrium Piece Rates

The first-order condition in (8) determines a worker's choice of effort in any period $t$ as a function of the piece rate in $t$, when the future path of piece rates is taken as given. We now determine the last-period piece rate and then proceed backwards to determine the remaining ones. With this characterization of equilibrium piece rates at hand, we can rely on the results from the previous subsection to derive equilibrium effort choices provided that equilibrium piece

rates depend only on time, which will be the case.

**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 noise in output. Thus, when $t = T$, our model is equivalent to the canonical static linear-normal model of incentives with quadratic effort cost, in which output noise is 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.** In order to determine equilibrium piece rates in period $t < T$, suppose that equilibrium efforts and piece rates from $t + 1$ on depend only on time, not on a worker's history—we just showed that this property holds when $t = T - 1$. For each $0 \leq \tau \leq T - t$, let $b_{t+\tau}^*$ be the equilibrium piece rate in period $t + \tau$ and define $R_{CC,t}^*$ and $R_{LBD,t}^*$ as in (9) with $b_{t+\tau} = b_{t+\tau}^*$ for each $\tau$. Then, a worker's choice of effort in period $t$ when the worker's contract piece rate in $t$ is $b$ is

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

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 $0 \leq \tau \leq T - t$ and the present-discounted value of these 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 the 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 in period $t$ maximizes

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

where $I_t$ is the public information about the worker in period $t$.

We now show that the problem of maximizing (11) admits a unique solution, that this solution is independent of $I_t$ and so the same for all workers, and characterize it. First, observe that

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

The first term on the right side of (12) captures the marginal impact of $b$ on the expected wage in period $t$ and equals one since $\partial e_t/\partial b = 1$ by (10). As for the second term, note that by increasing effort in period $t$ by one unit, a worker not only increases expected output in period $t$ by one unit but also expected output in period $t+\tau$ with $1 \leq \tau \leq T-t$ by $\gamma_t \lambda^{\tau-1}$ units, which amount to the increase in the worker's stock of human capital in $t+\tau$. Then, the second term on the right side of (12) is simply the present-discounted value of these expected output increases, as the competition

among firms yields that the worker fully captures this value.

Consider now the derivative of the second term in (11) with respect to $b$. Since, as discussed, a worker's effort in period $t$ does not affect the variance of wages, $\mathrm{Var}[W_t^*|I_t]$ depends on $b$ only through the direct effect of $b$ on the variance of wages. Given that $b$ affects the variance of $w_{t+\tau}^*$ only when $\tau = 0$, we can write

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

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

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

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

Using once again the fact that $\partial e_t/\partial b = 1$, it follows that 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

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

by (10).[24] Expression (14) is the equilibrium piece rate in period $t$, which is independent of a worker's history in $t$ and thus is the same across workers.[25] We provide a detailed analysis of $b_t^*$ next.

# 5 Equilibrium Characterization and Properties

Here we characterize the equilibrium and examine the implied 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

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

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

[25] Since the equilibrium piece rates in period $t$ are the same for every worker, 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 have the same properties from period $t$ on. So, by induction, the equilibrium efforts and piece rates are symmetric and depend only on time.

with initial condition $\sigma_0^2 = \sigma_\theta^2$. Recall that for each $0 \leq t \leq T$

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

and our convention that $\prod_{k=1}^{0} a_k = 1$ for any numeric sequence $\{a_k\}$.[26]

**Proposition 1.** *The equilibrium is unique, symmetric, and such that effort choices and piece rates depend only on time. Let $e_t^*$ and $b_t^*$, respectively, be the equilibrium effort and piece rate in period $0 \leq t \leq T$. For each $t$, define $b_t^0$, $R_{CC,t}^*$, $R_{LBD,t}^*$, and $H_t^*$ as*

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

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

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

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

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

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

*and*

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

In the Appendix we state and prove the equilibrium characterization in 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 $F_t$ strictly increasing and concave. There we also provide simple conditions for equilibrium piece rates to lie in the unit interval. Since we have discussed expression (22) for equilibrium effort choices in Section 4, in what follows we discuss expression (21) for equilibrium piece rates, which consists of 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 a worker's expected present-discounted value of lifetime output resulting from a marginal increase in effort in $t$. As discussed, the third and fourth terms $R_{CC,t}^* + R_{LBD,t}^*$ capture the dynamic marginal benefit of effort in period $t$, which amounts to the increase in a worker's expected present-discounted value of lifetime wages associated with a marginal increase in effort in $t$. The fifth term $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$ is the increase in the variance of the present-discounted value of lifetime wages following a marginal increase in the piece rate in $t$.

[26] As discussed in the previous section, (15) admits a unique nonnegative fixed point $\sigma_\infty^2$, which is positive if, and only if, $\sigma_\zeta^2 > 0$. Also, $\sigma_t^2$ strictly decreases to $\sigma_\infty^2$ when $\sigma_\theta^2 > \sigma_\infty^2$ and strictly increases to $\sigma_\infty^2$ when $\sigma_\theta^2 < \sigma_\infty^2$.

One way to understand expression (21) is by comparing it to the piece rate that would induce workers to exert the first-best level of effort. In the canonical static linear-normal model of incentives, a piece rate equal to one leads a worker to choose the first-best level of effort, which, by definition, equates the marginal cost of effort, here $e_t^*$, to its social marginal return given by the corresponding marginal increase in output. In our setting, a piece rate equal to $\chi_t = 1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R^*_{CC,t} - R^*_{LBD,t}$ would lead a worker to exert the first-best level of effort by (22), since $b_t^* = \chi_t$ implies that $e_t^* = 1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$. Note that $\chi_t = 1$ only when $t = T$ and no dynamic considerations influence a worker's choice of effort.

The equilibrium piece rate in (21) differs from $\chi_t$ in two ways. First, it subtracts from $\chi_t$ 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, as firms learn about a worker's ability based on realized output. The equilibrium contract partially insures 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 $\sigma_t^2$ declines or does not increase too fast with $t$, the term $rH_t^*$ also declines with $t$: as the present-discounted value of 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 $r$—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 $b_t^0 < 1$.The term $b_t^0$ adjusts the piece rate to account for the trade-off between risk and incentives familiar from static models of moral hazard. Namely, the equilibrium contract weighs the output gain from larger piece rates, which induce higher effort, against the cost of increasing the 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, $\sigma_t^2 + \sigma_\varepsilon^2$, due to the uncertainty about ability and output noise.[27]

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

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

This decomposition, which separates the equilibrium piece rates into four components, differs from that in (21) as it isolates the component of the returns to human capital accumulation that does not accrue to the worker, $\gamma_t \sum_{\tau=1}^{T-1} \delta^\tau \lambda^{\tau-1} - R^*_{LBD,t}$, and is useful for three reasons. First, it helps illustrate how the economic forces at play in our model shape the provision of incentives for effort over time. Second, based on this decomposition, we can analytically determine conditions under which these forces give rise to alternative life-cycle profiles of piece rates depending on

[27] As in Gibbons and Murphy [1992], the inefficiency in the provision of incentives relative to the first-best is due risk aversion. Despite the uncertainty about ability, if workers were risk neutral, then piece rates would equal one and so workers' choices of effort would equate the marginal cost to the social marginal return of effort in each period. Indeed, if $r = 0$, then $b_T^* = 1$. This, in turn, implies that $R^*_{CC,T-1} = 0$ and $R^*_{LBD,T-1} = \gamma_{T-1}\delta$, so that $b_{T-1}^* = 1$. It follows by induction that $b_t^* = 1$, $R^*_{CC,t} = 0$, and $R^*_{LBD,t} = \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ for all $t$. In particular, with risk-neutral workers, implicit incentives for effort would only arise from human capital considerations.

their relative strength, as we discuss in the next subsection. Finally, this decomposition will prove important for our identification arguments, as we show in Section 6.

The first term in (23) is the equilibrium piece rate $b_t^0$ in the static linear-normal model of incentives with exponential utility and quadratic cost of effort when the variance of output is $\sigma_t^2 + \sigma_\varepsilon^2$. Absent dynamic considerations, firms would offer the piece rate $b_t^0$ in each period. The difference between the static piece rate and the equilibrium piece rate implied by our model reflects two forces: the implicit incentives for effort arising from workers' desire to affect the process of learning about their ability and to invest in human capital, and workers' demand for insurance against the wage risk due to the uncertainty about ability.

The second and third terms in (23) capture the contribution of uncertainty and learning about ability to the equilibrium 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 $t < T$. As in Gibbons and Murphy [1992], these implicit incentives induce workers to exert effort even in the absence of any explicit link between performance and pay: by partially substituting for explicit incentives, they lead to lower piece rates.[28] The third term in (23) discounts the piece rate so as to provide workers with insurance against the risk in life-cycle wages generated by the uncertainty about ability.

The last term in (23), which is novel, captures the contribution of human capital acquisition to the explicit incentives 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 increase in lifetime output that results from the increase in a worker's human capital following a marginal increase in the effort investent in it in period $t$. The second part, which is proportional to $R^*_{LBD,t}$, reflects the implicit incentives for effort arising from the prospect of human capital acquisition, which substitute for explicit incentives and so lower piece rates. This last term adjusts piece rates so as 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 $\{\gamma_t\}$. When $\gamma_t$ is greater than zero, and human capital and output production are complements, this term, unlike the previous two, contributes positively to equilibrium piece rates.[29]

## 5.2 Piece Rates Over the Life Cycle

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

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

[29] We implicitly maintain that $R^*_{LDB,t}$ is positive when $t < T$, which holds when piece rates are in the unit interval.

**Pure Learning-About-Ability Case.** If we mute human capital acquisition with $\gamma_t = 0$ each $t$, then (21) reduces to

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

The next result describes the evolution of piece rates in this case.

**Lemma 2.** *Let $\gamma_t = 0$ for each $t$. There exists $T_0 > 0$ 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 $t$ for all $t$ if $\sigma_\theta^2 > \sigma_\infty^2$.*

When ability is not subject to shocks, 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 $\sigma_t^2$ converges to $\sigma_\infty^2$, the degree of uncertainty about ability eventually becomes constant. At this stage, the only force governing how piece rates evolve over time is the reduction of an individual's working horizon as experience accumulates. Naturally, a shorter working horizon weakens the implicit incentives for effort provided by career concerns, since workers can benefit from a higher reputation only for a shorter period of time. Firms then optimally compensate for weaker career-concerns 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 pieces to strictly increase with experience.[30]

**Pure Human-Capital-Acquisition Case.** By setting $\sigma_\theta^2 = \sigma_\zeta^2 = 0$, we eliminate uncertainty, and thus learning about ability, from the model. When this is the case, the equilibrium piece rate in (21) becomes

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

where $b^0 = 1/(1 + r\sigma_\varepsilon^2)$. In this case, piece rates are solely governed by human capital acquisition through the difference between $\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, the dynamic social marginal benefit of effort, and $\gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} b_{t+\tau}^*$, the dynamic private marginal benefit of effort. Note that the contribution of human capital acquisition to piece rates is positive when $t < T$ whenever $\gamma_t > 0$ and piece rates are smaller than one, which is intuitive. Indeed, if effort to produce output and to acquire human capital are complements, then workers do not fully capture the returns to their investments in human capital when future piece rates are smaller than one, which reduces their willingness to exert effort. Current piece rates help offset this undersupply of effort but they do so imperfectly because of the risk-incentives trade-off. More generally, it is possible to show that piece rates are nonnegative as long as $\gamma_t > \lambda - 1/\delta$

[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 $\sigma_\theta^2$ is positive but small.

for all $t$. That is, even when effort to produce output and to acquire human capital are rival, it is still optimal to induce workers to exert effort provided that the trade-off between output and human capital production is not too severe.

The variation of the rate $\gamma_t$ of human capital accumulation over the life cycle clearly affects the experience profile of piece rates. Here we state and discuss three results—their proofs are in the Appendix—that illustrate how for alternative life-cycle profiles of the rates $\{\gamma_t\}$, 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 if the rates of human capital accumulation eventually (weakly) decrease over time and are not too large, piece rates eventually decline.[31]

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

An immediate corollary of Lemma 3 is that piece rates are bounded above by one and globally strictly decreasing if the rates $\gamma_t$ are constant, positive, and bounded above by $(1-\delta\lambda)(1+r\sigma_\varepsilon^2)/\delta[1-(\delta\lambda)^T]$. Piece rates can also be initially increasing. This will be the case, for instance, when the rates of human capital accumulation are initially small but increase rapidly over time, as we prove next.

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

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

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

**General Case.** Suppose now that both learning about ability and human capital acquisition are present. In this case, naturally, the stronger force shapes the experience profile of piece rates. For instance, when $\sigma_\zeta^2$ is small so that ability is effectively known in the long run, human capital acquisition eventually prevails if workers are sufficiently long-lived. Intuitively, at some point, the residual uncertainty about ability becomes small enough that learning about ability 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 governs the profile of piece rates in the long run, thus leading piece rates to be eventually strictly increasing with experience. The next result confirms these intuitions.

[31] To understand why the rates of human capital accumulation cannot be too large 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 $\gamma_{T-1}$, $b_{T-2}^*$ is *smaller* than $b_{T-1}^*$ when $\gamma_{T-1}$ is sufficiently large. More generally, if $b_{t+1}^*$ to $b_T^*$ are smaller than one, then $b_t^*$ is strictly increasing with $\gamma_t^*$ and unbounded above by (24), in which case $b_{t-1}^*$ can be smaller than $b_t^*$ by the same equation.

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

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 $\sigma_\theta^2$ and $\sigma_\zeta^2$ 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 $\sigma_\theta^2$ is large and $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ 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 the two conditions hold: (i) $\sigma_\theta^2$ and $\sigma_\zeta^2$ 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, $\sigma_\theta^2$ is large, and $\sigma_\varepsilon^2$ and $\sigma_\zeta^2$ 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 results 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, instead, learning about ability matters early on. However, a consequence of Lemma 3 and the logic leading to Proposition 2 is that if the rates of human capital accumulation are constant (and not large), then piece rates are hump-shaped only if learning about ability is important early on. Intuitively, then, once the process of learning about ability is known, the observed pattern of piece rates allows to pin down the process of human capital acquisition. We show that this is indeed the case in the next section. Since, as we also show in the next section, our model has differing implications for the experience profile of the second moments of the distribution of wages depending on the characteristics of the learning process about ability, we can recover the relative importance of learning about ability and human capital

acquisition at different stages of a worker's career by combining information on the life-cycle profile of piece rates and the covariance structure of wages. We make this point formal in the next section.

# 6 Identification

We start by establishing simple conditions under which the model is identified from panel data on wages and their fixed or variable components in Section 6.1.[32] In sum, the identification of our model relies on combining information on the level of wages, their covariance structure, and the ratio of variable to total pay over the life cycle. Intuitively, from this latter ratio, as discussed, we can recover the experience profile of piece rates. Conditional on the profile of piece rates, the second moments of the distribution of wages allow us to identify the process of learning about ability and the distribution of ability shocks. Once these primitives are identified, we can recover the remaining ones by exploiting our characterization of equilibrium piece rates, which, given knowledge of the learning process, maps piece rates into worker risk preferences and the parameters governing the human capital process. In particular, the time variation of piece rates is informative about workers' evolving productivity as measured by $\{\gamma_t\}$ so it allows us to recover a worker's rates of human capital accumulation and depreciation without the need for any instruments.

In Section 6.2, we show that our identification results extend to the case in which unobserved heterogeneity exists among workers, even when wages are measured with error. We further show that, provided information about worker performance is available, our identification results can be adapted to the case in which human capital evolves nonparametrically with effort according to the law of motion $k_{it+1} = \lambda k_{it} + F_t(e_{it})$.

In these identification arguments, we treat the worker time discount factor $\delta$ as known and impose that the drift terms $\beta_0$ to $\beta_{T-1}$ are zero to simplify the exposition. In the Appendix, we discuss the identification of the more general case in which $\beta_0$ to $\beta_{T-1}$ are unknown.[33]

## 6.1 Identification of the Baseline Model

We show that piece rates $b_0^*$ to $b_T^*$, the variance of the initial distribution of ability $\sigma_\theta^2$, of the output noise $\sigma_\varepsilon^2$, and of the ability shocks $\sigma_\zeta^2$, the risk aversion coefficient $r$, and the rates of human capital depreciation $1-\lambda$ and accumulation $\gamma_0$ to $\gamma_{T-1}$ are identified from a panel of wages and their variable components, up to the mean initial ability $m_\theta$.

**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. The risk aversion parameter $r$ is identified from the piece rate $b_T^*$ and the variances*

[32] We have normalized the second derivative of the effort cost function to one. In the more general case in which this derivative 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 the same. Alternatively, one can show that $r$ and $c$ are separately identified if $\gamma_{T-1}$ 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.

[33] The proof of Proposition 4 below only relies on information on piece rates, the second moments of wages, and the change in average wages between $T-1$ and $T$. We can then use the remaining $T-1$ moments from the average wage profile and piece rates to identify $\{\beta_t\}_{t=0}^{T-1}$.

$(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$. *The human capital depreciation rate* $1-\lambda$ *and accumulation rates* $\{\gamma_t\}_{t=0}^{T-1}$ *are identified from the piece rates* $\{b_t^*\}_{t=0}^{T}$*, the variances* $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$*, and average wages in* $T-1$ *and* $T$ *up to mean ability* $m_\theta$*.*

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 identified—in the Appendix, we argue how to dispense with the normalization of $m_\theta$. The structure 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 absolute 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 last 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 show in more detail the proof of Proposition 4.

**Part I: Piece Rates and Variances of Unobservables.** 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 the fixed and variable components, respectively. Since contracts are linear in output, variable pay is $v_{it} = b_t^* y_{it}$ and $\mathbb{E}[w_{it}] = (1-b_t^*)\mathbb{E}[\mathbb{E}[y_{it}|I_t]] + b_t^*\mathbb{E}[y_{it}] = \mathbb{E}[y_{it}]$ by (3). Thus, the piece rate in $t$ 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 from the second moments of the distributions of wages in the first and second years of experience as follows. We show in the Appendix 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 $\mathrm{Var}[w_{i1}] = \sigma_\theta^2 + \sigma_\zeta^2 - \sigma_1^2 + (b_1^*)^2(\sigma_1^2 + \sigma_\varepsilon^2)$.[34] Hence, $\sigma_\theta^2$ and $\sigma_\varepsilon^2$ are identified from the variance of wages in the first year and the covariance of wages in the first two years. Then, $\sigma_\zeta^2$ is identified from the variance of wages in the second year since $\sigma_1^2 = \sigma_\zeta^2 + \sigma_\theta^2\sigma_\varepsilon^2/(\sigma_\theta^2 + \sigma_\varepsilon^2)$.[35]

**Part II: Risk Aversion and Human Capital Parameters.** We now establish that the parameters $r$, $\lambda$, and $\gamma_0$ to $\gamma_{T-1}$ are identified from average wages and the identified vector $(b_0^*, \ldots, b_T^*, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ up to $m_\theta$. First, note that once $(b_0^*, \ldots, b_T^*, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ is identified, so are the terms $\sigma_t^2$, $R_{CC,t}^*$, and $H_t^*$ for all $t$ by (15) to (18) and (20) in Proposition 1. Therefore, $r$ is identified from $b_T^*$, $\sigma_T^2$, and $\sigma_\varepsilon^2$ given that $b_T^* = 1/[1 + r(\sigma_T^2 + \sigma_\varepsilon^2)]$. Likewise, $\gamma_{T-1}$ 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 the depreciation rate, we know from the characterization of equilibrium effort in 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 $b_{T-1}^*$, $R_{CC,T-1}^*$, $\gamma_{T-1}$, and $b_T^*$. Since $\mathbb{E}[w_{it}] = m_\theta + e_t^* + k_t^*$ and human capital evolves as $k_t^* = \lambda k_{t-1}^* + \gamma_{t-1}e_{t-1}^*$ 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 $\beta_{T-1}$ is zero. Hence, $\lambda$ is identified from average wages in the last two years, $e_T^*$, $\gamma_{T-1}$, and $e_{T-1}^*$ up to $m_\theta$.

[34] There we show that $\mathrm{Var}[w_{it}] = \mathrm{Var}[\mathbb{E}[\theta_{it}|I_t]] + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ and $\mathrm{Cov}[w_{it}, w_{it+s}] = \mathrm{Var}[\mathbb{E}[\theta_{it}|I_t]] + b_t^*\sigma_t^2$ for all $0 \leq t \leq T$ and $1 \leq s \leq T-t$, where $\mathrm{Var}[\mathbb{E}[\theta_{it}|I_t]] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2$ and $\sigma_0^2 = \sigma_\theta^2$. Thus, $\mathrm{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$.

[35] Note that the parameters of the learning process are identified independently of the parameters of the human capital process. Thus, the parameters of the learning process can be identified and estimators of them can be constructed on the basis of the above identification arguments regardless of the specification of the human capital process.

We conclude by showing that $\gamma_0$ to $\gamma_{T-2}$ are identified from the vector $(b_0^*, \ldots, b_T^*, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r, \lambda)$. Note that

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

for all $0 \leq t \leq T-2$ by (19) and (21). Since all the terms in (25) but $\gamma_t$ are known from $(b_0^*, \ldots, b_T^*, \sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r, \lambda)$ in each $0 \leq t \leq T-2$, the parameters $\gamma_0$ to $\gamma_{T-2}$ are identified from this vector by (25). Intuitively, the right side of (25) is the portion of piece rates that cannot be explained by learning about ability alone and so is informative about the degree of human capital accumulation over a worker's career.[36]

**Remark.** In these arguments, we have assumed that the coefficient of absolute risk aversion is constant whereas the rates of human capital accumulation change 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 $\gamma$ is identified from the piece rate in $T-1$ and $\lambda$ 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).[37]

## 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, even when wages are measured with error, and to the case in which the law of motion of human capital depends nonparametrically on effort.

**Unobserved Heterogeneity and Measurement Error.** Suppose there exist $J$ groups or types of workers who differ in their initial distributions of ability, output noise, and shocks to ability, degree of risk aversion, and human capital process. Mirroring the identification argument of the baseline model, assume that the drift terms $\beta_0$ to $\beta_{T-1}$ in the human capital process are zero for all $J$ groups—this assumption can be relaxed. Each group is observable to model agents but not to the econometrician. Denote a generic group by $j$ and the probability that a worker is of type $j$ by $\pi_j$. Let $\sigma_{\theta j}^2$, $\sigma_{\varepsilon j}^2$, $\sigma_{\zeta j}^2$, $r_j$, $1-\lambda_j$, and $\gamma_{jt}$ be, respectively, the variance of the initial distribution of ability, the variance of the output noise, the variance of shocks to ability, the risk aversion parameter, the depreciation rate of human capital, and the period-$t$ rate of accumulation of human capital for type-$j$ workers.[38] The equilibrium characterization of Proposition 1 holds for each type. Let then $e_{jt}^*$, $k_{jt}^*$, and $b_{jt}^*$ be, respectively, the equilibrium effort, stock of human

[36] Alternative normalizations are possible. For instance, the parameters $\lambda$, $\gamma_0$ to $\gamma_{T-1}$, and $m_\theta$ are all identified from the piece rates $b_0^*$ to $b_T^*$ and the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r)$ if $\gamma_{T-2} = \gamma_{T-1}$. As for $\beta_t$, only the assumption that $\beta_{T-1} = 0$ is used in the identification argument.

[37] 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 $\gamma_T$ does not appear in any expression.

[38] We abstract from heterogeneity in $\delta$ or $m_\theta$ since they were fixed in the argument in the previous subsection. As discussed, alternative normalizations in lieu of that of $m_\theta$ are possible.

capital, and piece rate in period $0 \leq t \leq T$ for type-$j$ workers. By (3), the wage of worker $i$ of type $j$ in $t$ is

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

which is normally distributed. Thus, the distribution of wages in each period is a finite mixture of normal distributions. Since these mixtures are identifiable (Teicher [1963]), both the mixture weights $(\pi_j)$ and the component distributions are identified in each period, and so are their component means $(\mathbb{E}_j[w_{ijt}])$. Likewise, the variable component of pay of worker $i$ of type $j$ in $t$, $v_{ijt} = b^*_{jt}(\theta_{it} + k^*_{jt} + e^*_{jt} + \varepsilon_{ijt})$, is normally distributed. Then, its distribution in each period is also a finite mixture of normal distributions with the same component weights as the corresponding mixture distribution of wages. Thus, for each worker type $j$ and period $t$, mean variable pay $\mathbb{E}_j[v_{ijt}]$ is identified so that, as in the baseline model, the piece rate of type-$j$ workers in $t$ is identified as $b^*_{jt} = \mathbb{E}_j[v_{ijt}]/\mathbb{E}_j[w_{ijt}]$.[39]

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 $\{b^*_{jt}\}_{t=0}^T$ and the second moments of the distributions of wages in the first two years of experience. Then, for each type $j$, the preference and human capital parameters $(r_j, \lambda_j, \gamma_{j0}, \ldots, \gamma_{jT-1})$ are identified from the vector $(b^*_{j0}, \ldots, b^*_{jT}, \sigma^2_{\theta j}, \sigma^2_{\varepsilon j}, \sigma^2_{\zeta j})$ up to mean initial ability $m_{j\theta}$. We have thus established:

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

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. Through this latent-type formulation in which workers differ in their ability distribution and human capital process in an unrestricted way, the model is capable of accommodating alternative scenarios 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, human capital, uncertainty, risk, and workers' risk attitudes.[40]

**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 all firm-level data sets we examine and in many others commonly used (Frederiksen et al. [2017]). For ease of exposition, we start by

[39] The correct pairing of the components of the mixtures of total and variable pay in each $t$ is possible by their mixing weights, since these weights 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.

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

assuming that the performance measure is a noisy measure of a worker's effort; we later discuss the case in which the performance measure provides a noisy signal of a worker's effort and human capital.[41] Let $p_{it} = e_{it} + \eta_{it}$ be the performance measure of worker $i$ in period $t$ observed by the econometrician, where $\eta_{it}$ is a continuously distributed noise term independent across workers and over time with cumulative distribution function $G$ with known mean.

Suppose the equilibrium is unique, symmetric, and such that effort choices and piece rates depend only on time, and let $e_t^*$ and $k_t^*$ be, respectively, the worker's equilibrium effort and stock of human capital in period $t$—we present conditions under which this is the case in the Appendix. It follows again from (3) that $\mathbb{E}[w_{it}] = k_t^* + e_t^* + m_\theta$. Since $\mathbb{E}[p_{it}] = e_t^* + \mathbb{E}[\eta_{it}]$ with $\mathbb{E}[\eta_{it}]$ known, both $e_t^*$ and $k_t^*$ in each $t$ are identified from average wages and average performance in $t$ up to $m_\theta$. 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, $\lambda$ is identified from the vector $(k_{T-1}^*, e_{T-1}^*, k_T^*, e_T^*)$ up to $m_\theta$. To conclude the identification of the human capital process, note that since $k_{t+1}^* = \lambda k_t^* + F_t(e_t^*)$ for each $t$, we can identify $(F_0(e_0^*), \ldots, F_{T-1}(e_{T-1}^*))$ from $\lambda$ and the sequence of equilibrium efforts and human capital from 0 to $T$. Thus, if the investment functions $F_t$ do not depend on experience or, alternatively, if they do and any of the parameters $\sigma_\theta^2$, $\sigma_\varepsilon^2$, $\sigma_\zeta^2$, $r$, and $\gamma_t$ vary across observable groups of individuals, so as to induce different choices of effort among workers of different groups at each $t$, then these functions are identified from $(F_0(e_0^*), \ldots, F_{T-1}(e_{T-1}^*))$ and $(e_0^*, \ldots, e_{T-1}^*)$.[42] The identification of piece rates and the remaining parameters follows by the same argument as in the proof of Proposition 4. See the Appendix for the case in which the econometrician observes only a discrete version of $p_{it}$.

The argument so far has relied on a specific functional form for the performance measure $p_{it}$. In the Appendix, we show that we can extend this argument to the more general case in which $p_{it} = f(e_{it}, k_{it}) + \eta_{it}$, where $f : \mathbb{R}^2 \mapsto \mathbb{R}$ is a known differentiable function nondecreasing in each of its arguments such that $f(\cdot, k_{it})$ is surjective for each $k_{it} \in \mathbb{R}$ and $\partial f(e_{it}, k_{it})/\partial e_{it} \neq \partial f(e_{it}, k_{it})/\partial k_{it}$ 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.[43]

[41] That this additional outcome measure is informative about effort or human capital is a key step to separately recover the evolution of effort and human capital over time in this more general case. Like in a standard factor model, the path of effort and human capital can be identified provided that 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 Cunha et al. [2010]). This is the case here since observed wages and performance depend on effort and human capital up to independent measurement errors.

[42] Our identification argument works regardless of the length of the time interval between two consecutive periods in our model. So when the functions $F_t$ are the same in every period, 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 $\sigma_\theta^2$, $\sigma_\varepsilon^2$, $\sigma_\zeta^2$, $r$, or $\gamma_t$ among workers, say, with different age at entry or year of entry in the firm would induce variation in effort in each $t$ that would allow us to identify $F_t$ at every possible equilibrium choice of effort in $t$.

[43] We can extend the analysis to the case in which the performance measure provides information about 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 $\mathbb{E}_\theta[f(e, k, \theta)]$ is the expectation of $f(e, k, \theta)$ with respect to $\theta$, 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 the information on wages and their variable component up to $m_\theta$, we can treat $\widehat{f}_t(e_{it}, k_{it}) = \mathbb{E}_\theta[f(e_{it}, k_{it}, \theta_{it})]$ as a known function. It is easy to provide conditions on the functions $f_t$ under which the functions $\widehat{f}_t$ satisfy the conditions for identification.

The availability of an additional performance measure prompts the question of why firms would not offer contracts in which they condition wages not only on output but also on the realization of this measure—as argued by Hölmstrom [1979], firms should do so as long as output is not a sufficient statistic for the 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, for instance, in Baker [1992]. A common explanation for this feature of contracts is that although observable, these performance measures are not verifiable or are manipulable by workers. However, in the presence of learning about ability, although firms cannot or may not want to explicitly link wages to 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 these measures. In the Supplementary Appendix, 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 discuss the estimation of the model, the fit of the estimated model to the data, and compare our parameter estimates to analogous ones in the literature.

**Estimation Sample.** We estimate the model using the well-known BGH data discussed in Section 2. This firm-level data set has been extensively studied in the literature and, as such, provides a natural starting point for investigating how wages and, in particular, performance pay vary over the life cycle.[44] Being of administrative nature and high quality (Baker et al. [1994a]), the BGH data are less likely to be contaminated by measurement error than commonly used survey data such as the PSID. Crucially, the rich panel dimension of the data, coupled with their coverage of individuals across all experience years, permits a meaningful life-cycle analysis, which is the focus of our exercise. Indeed, in the BGH data, experience reaches 47 years. However, since sample size declines rapidly after 40 years of experience, 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 deviation of 9 years. 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. Their average tenure at the firm is of 5 years with a standard deviation of 4 years and a maximum of 18 years.

Wage profiles in our data are comparable with those documented in the literature. For instance, we find that the log wages of male college-educated workers, who are the majority of workers in our sample, increase by 0.67 log

[44] Frederiksen et al. [2017] report several regularities in terms of the distribution of wages and performance management systems across BGH and five other firm-level data sets, which suggests that the BGH data provide a sample of standard compensation and performance management practices. Indeed, many features of the BGH data, which have been replicated by other studies, are now considered stylized facts about careers in firms. See, for instance, Waldman [2012] on this point.

points during the first 30 years of labor market experience. This estimate is consistent with that of an average wage growth of about 1 log point reported on the basis of cross-sectional census data between 1960 and 2000 by Elsby and Shapiro [2012]. Rubinstein and Weiss [2006] find estimates of similar magnitude using both the PSID and the National Longitudinal Survery of Youth.

**Parameterization.** In estimation, we fix workers' discount factor $\delta$ at 0.95. Recall that we have specified the effort cost function as $g(e) = e^2/2$. 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 assume that the drift terms $\beta_0$ to $\beta_{T-1}$ are zero.[45] Then, we estimate eight parameters: the parameters $\sigma_\theta^2$, $\sigma_\varepsilon^2$, and $\sigma_\zeta^2$ of the learning process about ability, those governing the human capital acquisition process, namely, the coefficients $\psi_0$, $\psi_1$, and $\psi_2$ 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 one hundred and twenty 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, computed as 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.[46]

Table 1: Estimates of Model Parameters

| Parameters | Estimates | Standard Errors |
|---|---|---|
| $\sigma_\theta^2$, variance of initial ability | 2,024.099 | 0.0009736 |
| $\sigma_\varepsilon^2$, variance of noise in output | 267,019.845 | 0.0638429 |
| $\sigma_\zeta^2$, variance of shock to ability | 29.458 | 0.0000632 |
| $\psi_0$, coefficient of degree 0 of $\gamma_t$ | 0.892 | 1.68E-07 |
| $\psi_1$, coefficient of degree 1 of $\gamma_t$ | 0.035 | 6.10E-08 |
| $\psi_2$, coefficient of degree 2 of $\gamma_t$ | -0.001 | 1.25E-09 |
| $\lambda$, fraction of undepreciated human capital | 0.955 | 3.48E-08 |
| $r$, coefficient of relative risk aversion | 0.0002 | 9.75E-13 |

**Parameter Estimates.** Table 1 reports the estimates of the model parameters together with their asymptotic standard errors. All parameters are precisely estimated and statistically different from zero at conventional significance levels—for a sense of magnitudes, note 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 related to the uncertainty about workers' ability and the process of learning about it. We estimate that the standard deviation of the distribution of initial ability among workers $\sigma_\theta$ is 44.99 thousand dollars and the standard deviation of the distribution of shocks to ability $\sigma_\zeta$ is 5.43 thousand dollars. Together, these estimates imply that after

[45] See below the discussion of the estimates for the case in which $g(e)2/2$ 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 $\beta_t = \beta \iota_{it}$ and $\iota_{it}$ 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 different versions of the model and model fit only slightly improved, given the high quality of the fit of the baseline model.

[46] We compute these statistics after winsorizing 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].

40 years of labor market experience, the standard deviation of the distribution of ability is 56.33 thousand dollars, that is, about 25% larger than when workers enter the labor market. The estimate of the standard deviation of the noise in output $\sigma_\varepsilon$ of 516.74 thousand dollars is an order of magnitude larger than the estimate of $\sigma_\theta$. Thus, learning would occur very slowly even in the absence of shocks to ability.[47] In particular, without shocks to ability ($\sigma_\zeta = 0$), uncertainty about it as measured by the variance of posterior beliefs $\sigma_t^2$ would monotonically decline over time but decrease only by 23% over 40 years of experience. Our estimates, however, imply that ability evolves stochastically over time. The presence of shocks to ability combined with the slow speed of learning implies that uncertainty about ability actually *increases* with labor market experience. Indeed, after 40 years of experience, the variance of posterior beliefs is more than 20% higher than when workers enter the labor market. Since we estimate $\sigma_\theta^2$ to be smaller than the limiting value of $\sigma_t^2$ given by $\sigma_\infty^2 = [\sigma_\zeta^2 + (\sigma_\zeta^4 + 4\sigma_\zeta^2\sigma_\varepsilon^2)^{1/2}]/2 = 2{,}819.38$ reached as experience becomes (arbitrarily) large, uncertainty about ability eventually becomes 1.4 times larger than at entry in the labor market.

Consider next the parameters that govern the process of human capital acquisition, that is, the marginal human capital product of effort $\gamma_t$ and the depreciation rate of human capital $1 - \lambda$. The estimated law of motion of human capital implies a path of human capital accumulation that is concave in experience, as apparent from the solid blue line in panel (f) of Figure 5. Our estimate of $\gamma_t$ suggests that the return to effort in terms of additional human capital is fairly large. 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.[48] We estimate that the depreciation rate of human capital $1 - \lambda$ equals 4.5%, which 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 $r$ as equal to $2 \times 10^{-4}$ is consistent with the estimates in Handel [2013] of the coefficient of absolute risk aversion in the interval $[1.9, 3.25] \times 10^{-4}$ from data on health insurance and medical utilization choices, 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

[47] For $T$ large enough, $\text{Var}[w_{iT}] \approx (\sigma_\theta^2 + T\sigma_\zeta^2) - [1-(b_T^*)^2]\sigma_\infty^2 + (b_T^*)^2\sigma_\varepsilon^2$. 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 $\Delta_{it} \equiv \text{Var}[w_{it}] - \text{Var}[w_{it-1}]$ and $\sigma_\infty^2$ is independent of $\sigma_\theta^2$, changes in the variance of wages late in the life cycle are informative about $\sigma_\varepsilon^2$. As we argued, the variance of wages and its increase early in the life cycle, by contrast, are informative about $\sigma_\theta^2$ and $\sigma_\zeta^2$. The literature offers a range of estimates of the speed of learning. Ours suggest that learning occurs more slowly than found in Lange [2007] or Kahn and Lange [2014] but are comparable to those in Pastorino [2019].

[48] These calculations rely on marginal increases in effort. Based on equilibrium effort levels, we find that human capital increases by nearly the same amount as output, for instance, roughly by more than 20 thousand dollars by experience 20.

lottery with a 50 percent chance of gaining 100 dollars and a 50 percent 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 workers' absolute risk aversion is to convert it to an estimate of *relative* risk aversion (RRA) using the result that 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 $R(w) = wA(w)$. Then, our estimate of $r$ corresponds to a coefficient of relative risk aversion 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 of workers' relative risk aversion is consistent with the range of estimates in the literature; see, for instance, Chetty [2006], who documents an upper bound of 2.[49]

**Decomposing Piece Rates.** Expression (23) in Section 5 decomposes piece rates into five terms, each of which captures a distinct economic force that determines how performance pay evolves over the life cycle relative to total pay. This decomposition, shown in panel (a) of Figure 5 at the estimated values of the model parameters, 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. That the static piece rate is small and marginally increases over the life cycle reflects the large estimated variance of the noise in output $\sigma_\varepsilon^2$ and the sizable estimated degree of uncertainty about ability, given by the variance of posterior beliefs $\sigma_t^2$, which increases over time.

The second term of the piece-rate decomposition that is relatively small is $-b_t^0 R_{CC,t}^*$ given by the dashed green line in panel (a) of Figure 5. Recall that $R_{CC,t}^* = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau}^*)(\prod_{s=1}^{\tau-1} \mu_{t+\tau-s})(1 - \mu_t)$ is the dynamic marginal benefit of effort due to the impact of effort on a worker's reputation. This term adjusts the equilibrium piece rates for the career-concerns incentives for effort due to uncertainty and learning about ability. Intuitively, the large estimate of $\sigma_\varepsilon^2$ implies not only that the static piece rate $b_t^0$ is small, as noted above, but also that the signal-to-noise ratio governing the speed of learning about ability is low. In particular, as the weights $1 - \mu_t = \sigma_t^2/(\sigma_t^2 + \sigma_\varepsilon^2)$ on the signals about ability in the belief updating equation in (5) are small, the learning process is not very sensitive to a worker's output and, as a result, effort has a small effect on beliefs about ability. Thus, the career-concerns incentives arising from uncertainty and learning about ability are small and so have a limited impact on piece rates.

The third term of the piece-rate decomposition that is relatively small is $-b_t^0 R_{LBD,t}^*$ given by the dash-dotted teal line in panel (a) of Figure 5; recall that $R_{LBD,t}^* = \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} (b_{t+\tau}^* + R_{CC,t+\tau}^*)$ is the dynamic marginal benefit of effort due to the impact of effort on a worker's human capital. This term, which is the adjustment to the

[49] As a robustness exercise, we also estimated a version of the model with $r$ fixed at the value of 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 as large as estimated for our baseline model, the estimated model implies a lower degree of uncertainty about ability, consistent with intuition. However, the estimates of the parameters of the law of motion of human capital are very similar to those for the baseline model.

equilibrium piece rates that arises from the career-concerns incentives for effort due to human capital acquisition, is a relatively minor contributor to piece rates because the estimated piece rates $b^*_{t+\tau}$ and, as just discussed, the terms $R^*_{CC,t+\tau}$ are small. Intuitively, the private returns to accumulating human capital are relatively small because the slow speed of the learning process about ability combined with the small sensitivity of pay to performance implies that acquired human capital has a small impact on future pay. Thus, the career-concerns incentives arising from human capital acquisition are also small and so have a limited impact on piece rates as well.

The remaining two terms of the piece-rate decomposition are quantitatively important. The term $-b^0_t r H^*_t$ given by the dashed orange line in panel (a) of Figure 5 describes the insurance that the equilibrium contract provides against uncertainty about ability; recall that $H^*_t = \sigma^2_t \sum_{\tau=1}^{T-t} \delta^\tau$. This term is negative and explains the relatively low level of piece rates throughout the life cycle. The large estimated dispersion in posterior beliefs $\sigma^2_t$ implies that workers face a large degree of risk resulting from the variation in performance pay induced by the variation in their ability, as this ability risk translates into performance risk in each period. Piece rates are thus lowered to provide workers with insurance against this risk. This insurance motive tends to weaken as the working horizon shortens, despite the increase in uncertainty about ability over time. This explains 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 reference 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 whose payouts are instead low when investors value consumption relatively more. Namely, assets whose payouts positively correlate with the market portfolio are less desirable. In our framework, performance pay in any period is positively correlated with output in the current period and, through the process of learning about ability, future periods as high performance pay is positive news about future compensation and low performance pay is negative news. As a result, contracts that specify a large pay in periods with high performance realizations and a small pay in periods with low performance realizations are less attractive to workers since performance pay amplifies the variability of wages. Thus, equilibrium contracts will tend to have relatively low performance-pay components. 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 provided by wage contracts for the evolution of wages and provides an explanation for the low level of performance pay relative to total pay 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 remaining term $b^0_t \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ given by the dashed lavender line in panel (a) of Figure 5 describes the adjustment to the equilibrium piece rates that arises from the alignment of the private marginal return to effort with the social marginal return to effort provided by the equilibrium contracts. This term is positive and explains the

hump-shaped profile of piece rates over the life cycle. The large estimated rates of human capital accumulation $\gamma_t$ and the small estimated rate of human capital depreciation $1-\lambda$ imply that this term outweighs the negative insurance term discussed in the previous paragraph, while the initially increasing and then decreasing estimated rates of human capital accumulation lead to the initially increasing and then decreasing piece rates.

The decomposition of the piece rates implied by our estimated model is consistent with our characterization of the life-cycle profile of piece rates 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.

**Estimated Piece Rates, Wage Dispersion, and Wage Growth.** Figure 4 shows how our estimated model fits the targeted moments, successfully reproducing 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 implications of our estimates for the pattern of piece rates, we now turn to discuss their implications for the life-cycle profiles of the variance of wages and wage growth.

Our model implies that the variance of wages, $\mathrm{Var}[w_{it}] = (\sigma_\theta^2 + t\sigma_\zeta^2) - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$, is the sum of three terms. The first term $\sigma_\theta^2 + t\sigma_\zeta^2$ is the variance of worker ability, which increases over the life cycle because of the accumulating shocks to ability. The second term $-\sigma_t^2$ accounts for the fact that learning about ability takes place over time. In the absence of shocks to ability ($\sigma_\zeta^2 = 0$), the first two terms sum to $\sigma_\theta^2 - \sigma_t^2$, which increases to $\sigma_\theta^2$ in the long run as the dispersion of wages increasingly reflects the dispersion of worker abilities in the population. In the presence of shocks to ability ($\sigma_\zeta^2 > 0$), which is the case for our estimated parameters, the law of motion for $\sigma_t^2$ in (5) implies that $\sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2$ increases without bound because the shocks to ability increase the dispersion of worker abilities in the population. The final term $(b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ describes the impact on the variance of wages of the variability of performance pay due to the dispersion of worker abilities in the population and the variability of the noise in (or shocks to) output. Since the uncertainty about ability $\sigma_t^2$ eventually stops changing over time, the large estimated value of $\sigma_\varepsilon^2$ and the declining portion of the profile of piece rates towards the end of the life cycle together imply that the third term is eventually decreasing, allowing the model to reproduce the first increasing and then decreasing pattern of the variance of wages as shown in the middle panel of Figure 4.

Consider now wage growth. Since in our model firms make zero expected profits, the average wage of workers is $\mathbb{E}[w_{it}] = m_\theta + e_t^* + k_t^*$, their expected output. Therefore, the growth in average wages stems from the growth in either effort or human capital. We know from above that the rates of human capital accumulation $\gamma_t$ peak in the middle of the life cycle and subsequently decline. Effort is relatively more constant, though. So early in a worker's career, the accumulation of human capital is rapid. However, the depreciation of the stock of human capital, together with the decline in $\gamma_t$ over the second half of the life cycle, reduces and eventually reverses the growth in human

capital. The result is a concave experience profile for mean wages, reproduced in the right panel of Figure 4, which eventually plateaus. This is consistent with much evidence in the literature.

# 8 Performance Incentives and Wages

Our estimated model provides a laboratory that we can use to perform counterfactual exercises to explore how performance pay affects the distribution of wages over the life cycle. Based on these exercises, we illustrate the importance of performance pay for human capital acquisition and for the resulting dynamics of wages and their components.

## 8.1 Performance Incentives and Wage Growth

In our model, average wages are $\mathbb{E}[w_{it}] = m_\theta + e_t^* + k_t^*$ in any period $t$ so their cumulative growth, $\mathbb{E}[w_t] - \mathbb{E}[w_1]$, can be decomposed into the growths in effort and in the stock of human capital. In panel (b) of Figure 5, we show how the evolution of effort as measured by $e_t^* - e_1^*$ (the dot-dashed green line) and of human capital as measured by $k_t^* - k_1^*$ (the dashed red line) contribute to wage growth over the life cycle (the solid blue line). A comparison of the dashed red line to the solid blue line illustrates how closely the profile of human capital tracks the profile of wages. By contrast, changes in effort do not appear to substantially contribute to the growth in wages.

Such a decomposition, however, only measures the *direct* effect of effort on wages. In particular, it does not account for the *indirect* effect of effort on the dynamics of wages through its impact on the process of human capital acquisition. Panel (c) of Figure 5 demonstrates the importance of this indirect channel by comparing the estimated wage growth (the solid blue line) with the counterfactual wage growth that would result if we held effort constant over the life cycle (the dashed red line) at the average level implied by our model estimates. As can be seen from panel (c), wages would grow much less during the first 30 years of labor market experience in this counterfactual case than in the baseline model. Intuitively, in the baseline model, workers exert greater effort early in the life cycle because the career concerns arising from the workers' desire to influence their reputation and acquire human capital are largest over those experience years. Only later in the life cycle, when career-concerns incentives and piece rates decline, effort in the counterfactual scenario exceeds that in the baseline. Panel (c) thus illustrates that the variation in incentives and effort over the life cycle is central to observed wage growth.

This counterfactual exercise has further implications for understanding the determinants of the human capital process. Specifically, this exercise implies that standard models of human capital acquisition that do not allow for an intensive margin of investment—effort here—may be misspecified. The human capital process in these "passive" human-capital-acquisition models is usually just a function of the number of periods of employment and so conflates variation in investment, $e_t$, and variation in its marginal product, $\gamma_t$. But panel (c) shows that it is important to account for variation in $e_t$ to correctly infer the role of human capital acquisition for wage growth.[50]

[50]This logic applies to other counterfactual settings that have implications for workers' effort over the life cycle, for instance, when changes

Another way to assess the importance of effort for wages is to explore the impact of performance pay on their dynamics. Specifically, suppose that firms were restricted, say, for administrative or regulatory reasons, to offer 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 to reward performance and thereby encourage workers to exert effort and acquire human capital. Panels (e) and (f) of Figure 5 indeed show that the resulting equilibrium profiles of effort and human capital accumulation (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 lead to a much lower wage growth over the life cycle as shown in panel (d) of Figure 5, namely, 30% lower by the twentieth year of labor market experience (see the dashed red line relative to the solid blue line in the panel). Hence, although performance pay is quantitatively small relative to total pay, it has a substantial impact on wage growth through its impact on workers' effort and human capital acquisition.

### 8.2 Performance Incentives and Wage Inequality

To explore the impact of performance incentives on the dispersion of wages across workers over the life cycle, we start by decomposing the variance of wages into the variance of fixed and variable pay. By performing this decomposition at the estimated parameter values, we find that the variance of variable pay accounts for 44% to 100% of the variance of wages over the first 30 years of experience; see panel (a) of Figure 6.[51] So, even though performance pay represents only a small fraction of total pay at each experience, it is highly variable and thus responsible for a large share of the variance of wages over the life cycle.

Panel (b) of Figure 6 further shows that uncertainty about ability is a major source of the variance of wages. In this panel, we compare the variance of wages implied the model (the solid blue line) with the counterfactual variance of wages that would result *at the estimated piece rates* if ability was identical across individuals and over time (the dashed lavender line), that is, when $\sigma_\theta^2 = \sigma_\zeta^2 = 0$. Note that the difference between the solid and the dashed line significantly increases over the life cycle because shocks to ability accumulate with experience thereby contributing to the increase in the variance of wages over time, according to our model.

This decomposition, however, ignores the possibility that wage contracts may be very different 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 equilibrium wage contracts would change in response to the new distribution of workers' 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 a firm's production or organizational structure induce a different exposure of a worker's output to firm risk, which in our framework would lead to a different degree of output variability as measured by $\sigma_\varepsilon^2$, different piece rates, and so different profiles of effort and human capital.

[51] 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 that period.

in the form of higher piece rates, which could lead to an overall increase, rather than a decrease, in the variance of wages. This suggests a potential tension between ex-ante wage risk, which arises because of the initial dispersion in ability among workers, and ex-post wage risk, which arises because of the variability of performance pay.

This is precisely what we find when we mute uncertainty about ability altogether by setting $\sigma_\theta^2 = \sigma_\zeta^2 = 0$ and determine the resulting equilibrium piece rates. The resulting variance of wages is shown by the dashed lavender line in panel (c) of Figure 6, which exceeds the variance of wages in the baseline model, represented by the solid blue line, by up to six times. Panel (d) of Figure 6 reports the profile of equilibrium piece rates in the absence of uncertainty about ability (the dashed lavender line), which are up to three times as large as those in the baseline model (the solid blue line). These much higher piece rates, in turn, amplify any residual productivity risk faced by workers leading overall to a much larger wage dispersion. Hence, compressing the dispersion in ability among workers induces firms to offer contracts with a higher sensitivity of pay to performance, which more than compensates for the lower dispersion 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 sources of wage dispersion among workers, in particular heterogeneity in ability. Specifically, this result implies that popular reduced-form linear 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, with the level of uncertainty and risk in the labor market (see Abowd et al. [1999], Card et al. [2013], and the subsequent literature, and specifically Guiso et al. [2005] on the importance of the pass-through of firm-level shocks to wages). Here we have shown that the premise of these exercises may not always be warranted, since the wage structure is a key endogenous dimension through which firms' 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 separate primitive components linearly affecting wages may be inaccurate when performance incentives matter. In fact, our analysis implies that these components are highly interdependent and their impact on wages is subtly mediated by their dispersion. Specifically, once firms' incentives to offer contracts with different sensitivities of wages to performance, in response to different levels of uncertainty or output risk, are taken into account, *lower* dispersion in ability (or output risk) can be associated with *greater* dispersion in wages—although small decreases in the dispersion of ability or output risk for given contract form indeed lead to lower wage dispersion. Since the variance of wages is given by $\mathrm{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ according to our model, it is easy to show that this result holds at any level of experience for given dispersion in initial ability $\sigma_\theta^2$, variance of ability shocks $\sigma_\zeta^2$, and degree of workers' risk

aversion $r$, provided that the variance of the noise in output $\sigma_\varepsilon^2$ is large enough.[52]

# 9 Conclusion

Workers tend to acquire more skills that are valuable in the labor market while working. That is, by exerting effort on the job, workers may not only produce more output in a period but also eventually become more productive. This simple insight is the starting point for our exploration of how performance incentives influence the dynamics of wages with labor market experience both directly, through the impact of current effort on current output, and indirectly, through the impact of effort on the human capital process. Our goal is to examine theoretically and empirically how incentives for effort are affected by human capital considerations and in turn shape the structure and evolution of wages over the life cycle.

To this purpose, we develop and estimate a tractable model of the labor market that integrates three key sources of the dynamics of wages, namely, human capital acquisition, uncertainty and learning about individual ability, and performance incentives, in order 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 standard models of investment in human capital, models of dynamic moral hazard, and so-called career-concerns models of learning about ability and performance incentives. We characterize the optimal wage contract in this framework and analytically decompose the implied sensitivity of pay to performance into the relative contribution of the basic forces we incorporate: the trade-off between output risk and incentives for effort characteristic of moral hazard, the insurance that firms provide to workers against output risk and uncertainty ability through wage contracts, and the career-concerns incentives arising the workers' desire to affect the market assessment of their ability and accumulate human capital. Based on this characterization, we prove that the model is identified just from panel data on wages and their components, and obtain simple estimators of the model primitives.

Although performance pay accounts for a small fraction of total pay, the estimates of the model illustrate the centrality of performance pay to the dynamics of wages and their components. In particular, we find that through the cumulative impact of effort on human capital acquisition, incentives for performance are a critical source of wage growth and dispersion over the life cycle. We also show the importance of the wage structure as an endogenous mechanism for the transmission of output risk to wages. That is, although lower dispersion in ability decreases wage dispersion for a *given* combination of fixed and variable pay, lower dispersion in ability may lead firms to offer contracts with higher-powered incentives, in the form of a higher fraction of performance pay to total pay, and so may give rise to a greater variability of wages across individuals and over time.

[52] When $\sigma_\theta^2$ and $\sigma_\zeta^2$ are lowered 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," with $b_{t,n}^* > b_t^*$ and $\sigma_{t,n}^2 < \sigma_t^2$. Then, a sufficient condition for the variance of wages to be higher in the absence of uncertainty about ability is that $[(b_{t,n}^*)^2 - (b_t^*)^2]\sigma_\varepsilon^2$ exceeds the initial value of $\sigma_\theta^2 + t\sigma_\zeta^2$, which can be guaranteed when $\sigma_\varepsilon^2$ is large enough.

Our augmented model further allows us to rationalize a novel empirical finding on the life-cycle profile of variable pay. Namely, we document that the ratio of variable pay to total pay for most U.S. workers tends to decline over the second half of their careers, precisely when standard models of career concerns predict that variable pay should become more and more important. Our estimates suggest that two motives, namely, a desire for insurance against uncertainty about ability and considerations of human capital acquisition, are especially important in determining performance pay over the life cycle and have effects of opposite sign on the experience profile of performance pay relative to total pay, which explains its peculiar hump-shaped pattern. The role of insurance emerges from the substantial wage risk arising from the uncertainty about ability that workers face, due to the combination of dispersion in ability at entry in the labor market and persistent shocks to ability that accumulate with experience. Furthermore, in order to reward effort, performance pay is by design high precisely when output is high and workers receive positive news about their ability and so future compensation, and is correspondingly low when output is low and workers receive negative news. Then, wage contracts that specify large performance-pay components are unattractive to workers because their payouts are positively correlated with lifetime income, that is, performance pay amplifies the risk due to the uncertainty about ability. Accordingly, we find that the insurance component of wage contracts is quantitatively large and *depresses* performance pay relative to total pay throughout the life cycle, especially early in workers' careers. We believe this rationale to be a leading explanation for why performance pay tends to be empirically small for most workers. As for the role of human capital, note that, at the same time, performance incentives provided through variable pay promote workers' investments in human capital by encouraging effort on the job. This force tends to *increase* performance pay relative to total pay but naturally becomes less relevant with experience. This balance between the insurance and human capital determinants of performance pay rationalizes the observed hump-shaped pattern of the ratio of performance pay to total pay over the life cycle, according to our model. Compared to insurance and human capital motives, career-concerns incentives and the life-cycle variation in the strength of the static trade-off between risk and incentives—a key component of variable pay in static moral-hazard models—are empirically much less important.

Our analysis has sidestepped questions related to how individuals sort, for instance, into distinct markets so as to transparently integrate the alternative mechanisms of wage growth and dispersion we study within a framework that can be analytically characterized and has empirical content.[53] 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.

[53]Note that if we interpreted worker $i$'s output as the log of $\tilde{y_{it}} = \tilde{\alpha}_{ijt}\tilde{\theta}_{it}\tilde{k}_{it}\tilde{e}_{it}\tilde{\varepsilon}_{it}$, where $\tilde{\alpha}_{ijt}$ depends on the worker's job $j$ in period $t$ at the firm we study, the empirical analysis so far would apply since we remove time, demographic, and fixed effects from wages and the ratio of average variable pay to average total pay—the contract piece rate—does not differ much across the firm's jobs.

## References


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

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

Jérôme Adda and Christian Dustmann. Sources of wage growth. Bocconi University, Mimeo, 2020.

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

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

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

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

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

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

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

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

David Card, Jörg Heining, and Patrick Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015, 2013.

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

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

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

Flavio Cunha and James J. Heckman. Formulating, identifying, and estimating the technology for the formation of skills. *Journal of Human Resources*, 43(4):738–782, 2008.

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

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

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

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

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

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

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

Anders Frederiksen, Fabian Lange, and Ben Kriechel. Modeling earnings dynamics. *Journal of Economic Behavior & Organization*, 134(C):408–429, 2017.

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

George-Levi Gayle and Robert A. Miller. Identifying and testing models of managerial compensation. *The Review of Economic Studies*, 82(3):1074–1118, 2015.

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

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

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

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

Limor Golan, George-Levi Gayle, and Robert A. Miller. Promotion, turnover and compensation in the executive labor market. *Econometrica*, 83(6):2293–2369, 2015.

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

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

Benjamin R. Handel. Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–2682, 2013.

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

James J. Heckman, Lance Lochner, and Christopher 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.

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

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

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

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

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

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

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

Thomas Lemieux, W. Bentley MacLeod, and Daniel Parent. Performance pay and wage inequality. *The Quarterly Journal of Economics*, 124(1):1–49, 2009.

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

Lance Lochner, Youngmin Park, and Youngki Shin. Wage dynamics and returns to unobserved skill. NBER Working Paper 24220, 2018.

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

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

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

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

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

Elena Pastorino. Careers in firms: The role of learning and human capital. Stanford University, Mimeo, 2019.

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

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

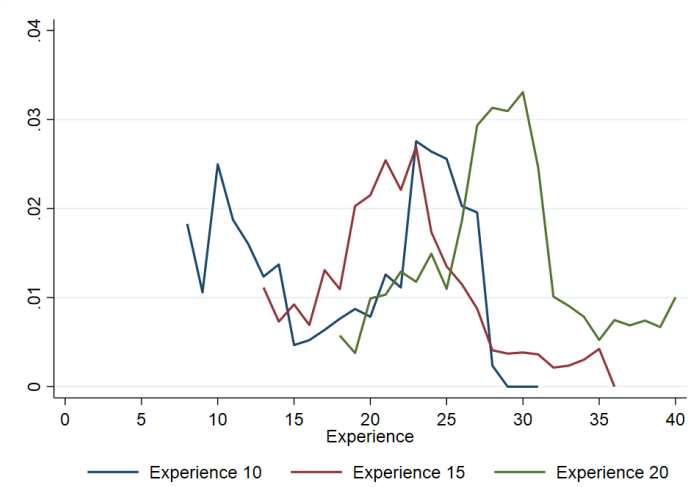
Carl Sanders and Christopher Taber. Life-cycle wage growth and heterogeneous human capital. *Annual Review of Economics*, 4(1):399–425, 2012.

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

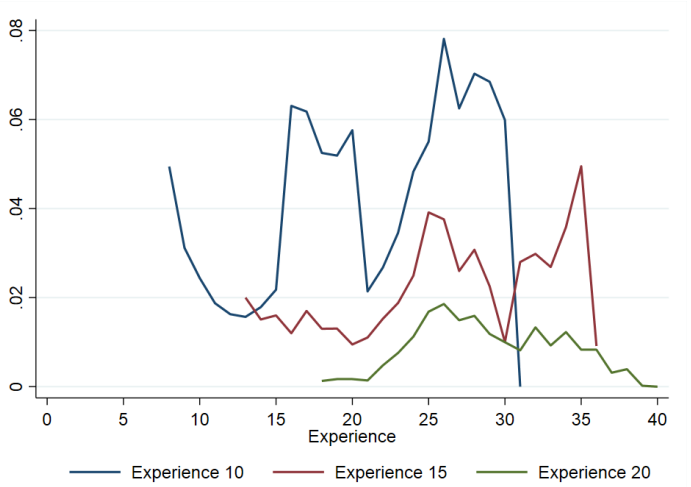
Michael Waldman. Theory and evidence in internal labor markets. *Handbook of Organizational Economics*, Robert Gibbons and John Roberts, Eds., Princeton, NJ:Princeton University Press, 520–572, 2012.

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

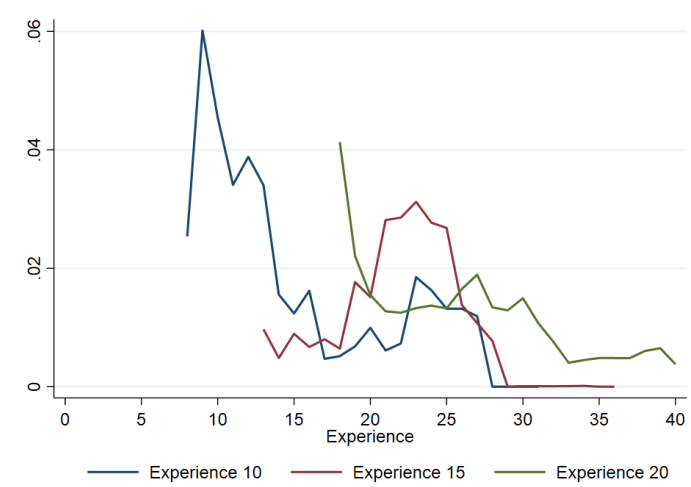
(a) Manufacturing




(b) Transport




(c) Services




(d) FIRE

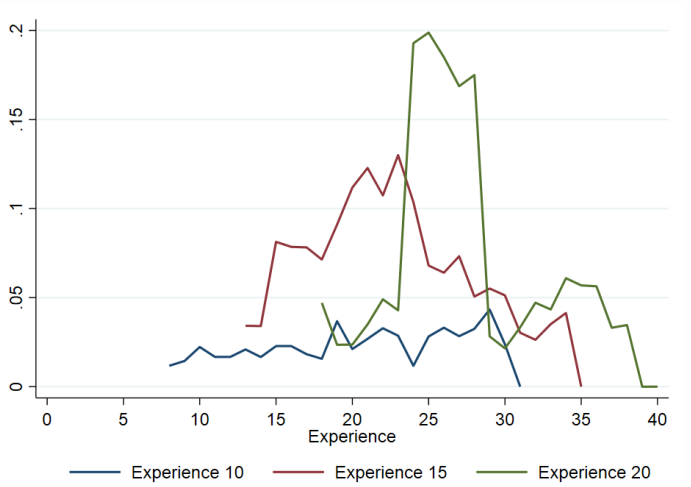


Figure 2: Ratio of Performance Pay to Total Pay in BGH Data

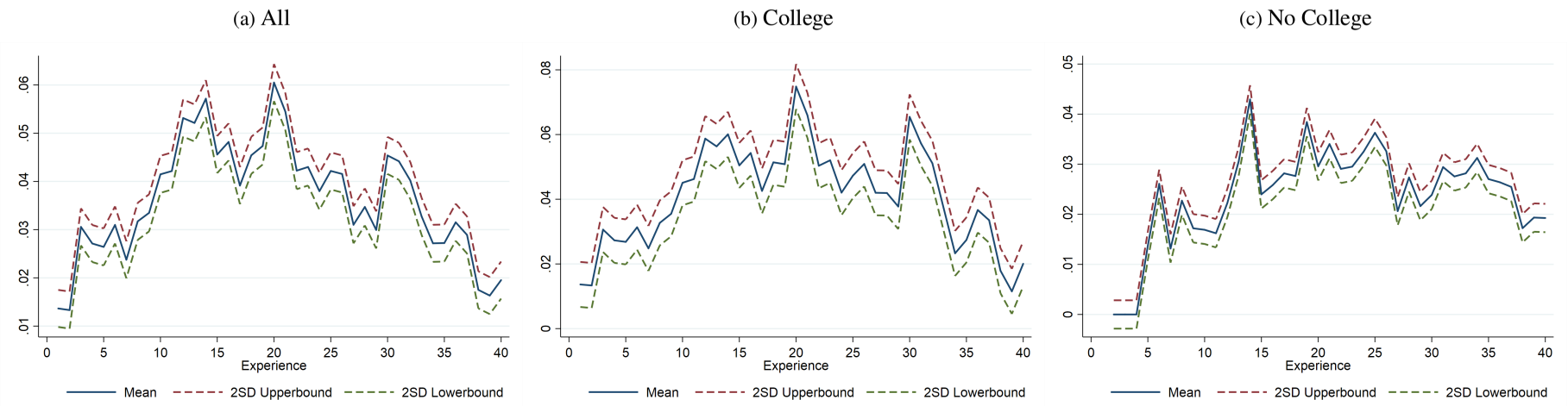



Figure 3: Ratio of Performance Pay to Total Pay in GH Data

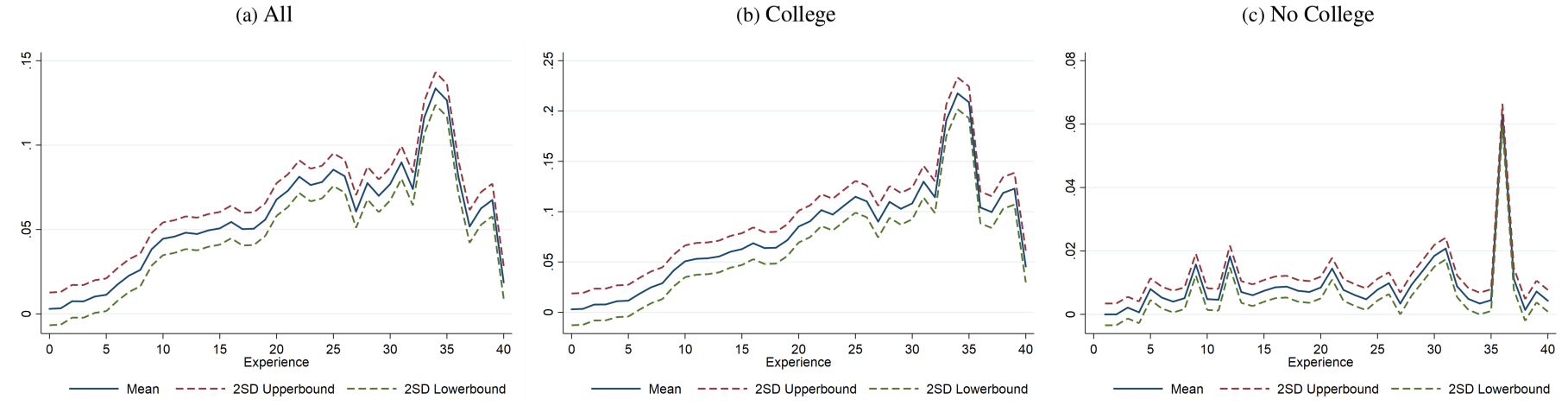
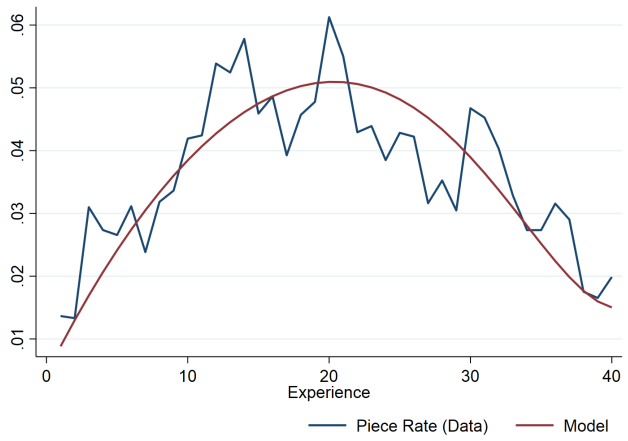
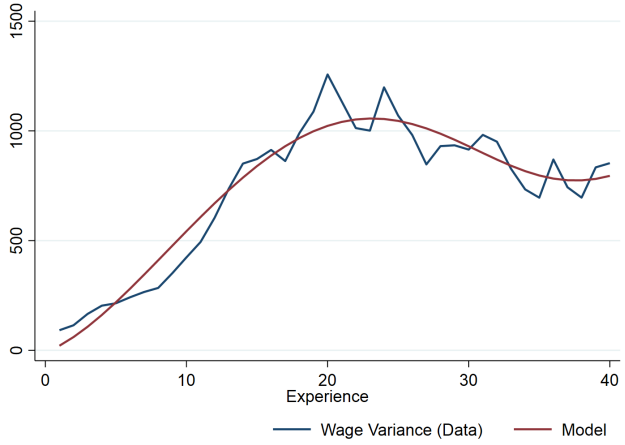


Figure 4: Model Fit to Targeted Moments from BGH Data

(a) Ratio of Performance Pay to Total Pay




(b) Wage Variance

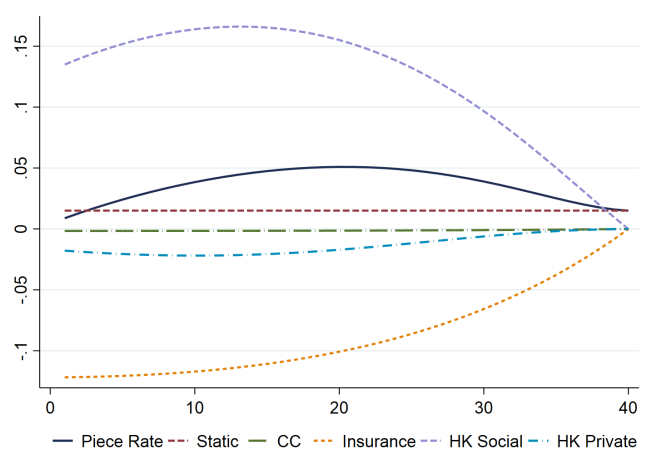



(c) Wage Growth Relative to Initial Wage

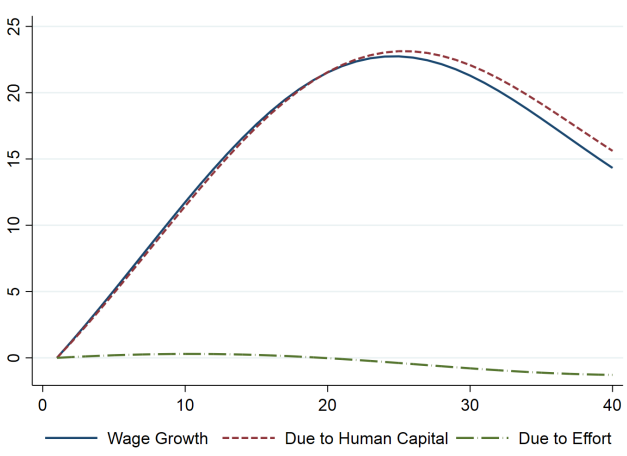


Figure 5: Piece Rates, Wage Growth, Effort, and Human Capital

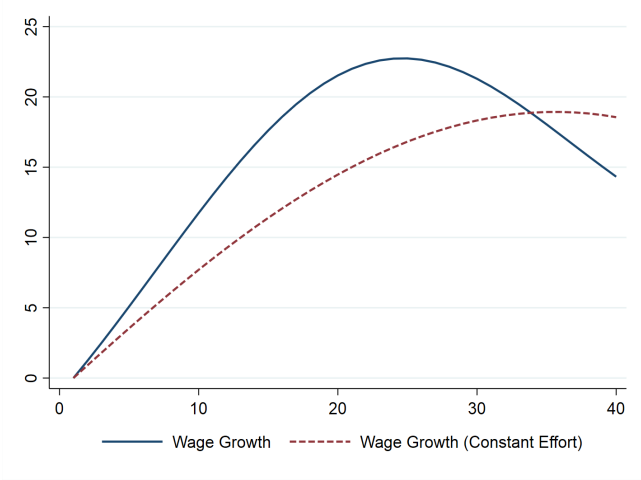
(a) Piece Rate Decomposition



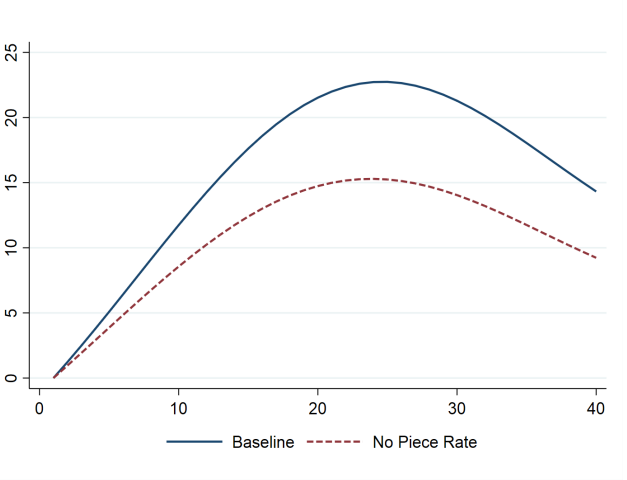

(b) Wage Growth Decomposition



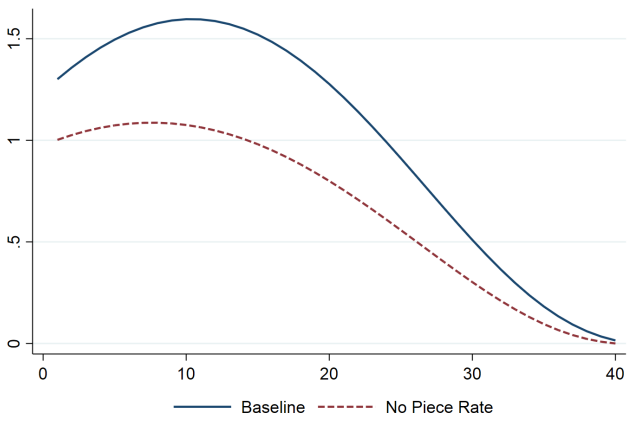

(c) Wage Growth with Constant Effort




(d) Wage Growth Without Piece Rates




(e) Effort Without Piece Rates




(f) Human Capital Without Piece Rates

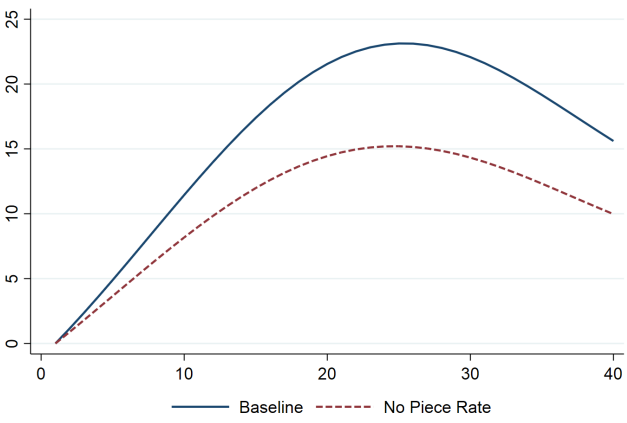
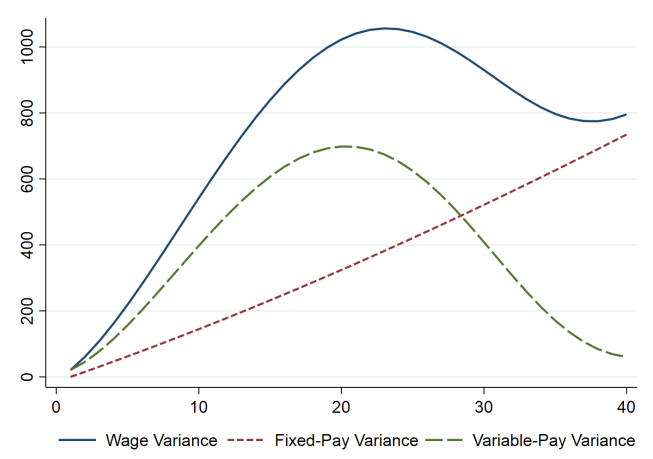


Figure 6: Variance of Wages, Piece Rates, and Role of Ability Uncertainty

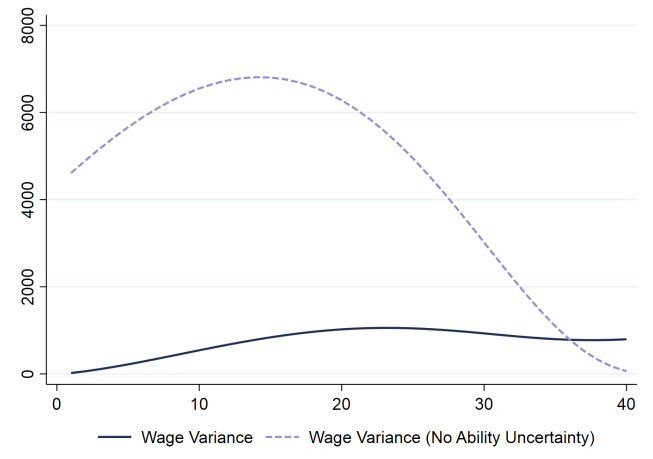
(a) Wage Variance Decomposition



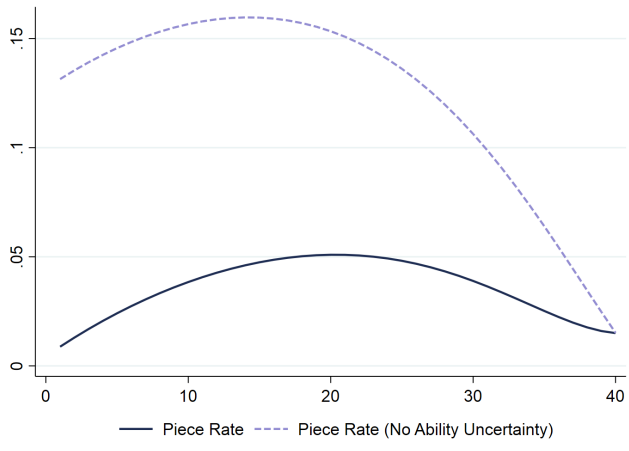

(b) Wage Variance Contribution of Ability Uncertainty




(c) Wage Variance Without Ability Uncertainty




(d) Piece Rates Without Ability Uncertainty



# A Appendix: Equilibrium Derivation

Here we construct the equilibrium. 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 the functions $F_t : \mathbb{R} \to \mathbb{R}$ are thrice differentiable, strictly increasing, and weakly concave with $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$, $F_t'''$ nonpositive and nondecreasing, and $F_t''(\infty) = \inf_{e\in\mathbb{R}} F_t''(e) > -\infty$. This case reduces to the case considered in the main text when $F_t(e) = \gamma_t e$, with $\gamma_t \geq 0$, for all $t$.

We first derive the first-order conditions for the optimal choices of effort when piece rates are noncontingent and future effort choices are noncontingent as well. We then determine the equilibrium piece rates and show that they are symmetric, noncontingent, and uniquely defined. 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 the piece rates for a worker are noncontingent and given by $\{b_t\}_{t=0}^T$, then the first-order condition for the worker's optimal choice of effort in period $0 \leq t \leq T$ when the worker's future behavior is noncontingent is

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

where

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

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 $0 \leq t \leq T-1$, we have that*

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

*Proof.* The result is trivially true when $t = T-1$ as $R_{CC,T} = 0$. Fix $0 \leq t \leq T-2$ and let $u, v \in \mathbb{R}^{T-t-1}$ be such that $u = (\xi_t, \ldots, \xi_{T-t-1})$ and $v = \big(\delta^2(1 - b_{t+2}), \ldots, \delta^{T-t}(1 - b_T)\big)$. 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 $i \geq j$. If we let $\langle v, Au \rangle$ denote the scalar product of the vectors $v$ and $Au$, then

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

the second equality follows from the change of variables $i \mapsto i-1$ and the fact that the term corresponding to $i = 1$ 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 $1 \leq i \leq T-t-1$. Therefore, (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. □

Suppose piece rates for a worker are noncontingent and given by $\{b_t\}_{t=0}^T$. The argument in the main text—whether the functions $\{F_t\}_{t=0}^T$ are linear or not does not matter for the argument—shows that the first-order condition for the worker's choice of effort in period $0 \leq t \leq T$ when the worker's future behavior is noncontingent is

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

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

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

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

for all $\tau \geq 1$. Now note from the law of motion (A1) and the non-contingency of behavior from period $t+1$ on that

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

for all $\tau \geq 1$. Finally, note from Lemma 1 that

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

where $z_{t+s}$ is the signal about the worker's ability in period $t+s$. Given that $\partial \mathbb{E}[z_t|h^t]/\partial e_t = 1$, and the noncontingency of behavior from period $t+1$ on implies that

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

for all $s \geq 1$, we can rewrite (A5) as

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

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

The first-order condition (A2) is necessary of optimality. In the benchmark case in which the functions $\{F_t\}_{t=0}^T$ are linear, this first-order condition is also sufficient for optimality. Indeed, the marginal benefit of effort to the worker, which is the right-hand side of (A2), is independent of the worker's effort, whereas the marginal cost of effort to the worker, the left-hand side of (A2), is increasing in effort. When the functions $\{F_t\}_{t=0}^T$ are nonlinear, (A2) need not be sufficient for optimality, though. 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, in this case $R_{LBD,t}(e)$ is nonincreasing in $e$, so that the marginal benefit of effort is nonincreasing in effort. Condition (A6) holds if piece rates are in the unit interval. To see why, note that $R_{CC,t+\tau} \geq 0$ if $b_{t+\tau+s} \leq 1$ for all $1 \leq s \leq T-t-\tau$, in which case $b_{t+\tau} + R_{CC,t+\tau} \geq 0$ if $b_{t+\tau} \geq 0$.

## A.2 Equilibrium Piece Rates

We now derive the equilibrium piece rates and show that they are symmetric, noncontingent, and uniquely defined. We consider the linear and nonlinear cases separately. We will see that whereas the characterization of the equilibrium piece rates in the first case is valid in general, the characterization of the equilibrium piece rates in the second case requires some restrictions on the model's primitives.

### A.2.1 Linear Case

Suppose that $F_t(e) = \gamma_t e$, with $\gamma_t > 0$, for all $t$. We first derive the last-period equilibrium piece rates and then derive the equilibrium piece rates in previous periods under the assumption that future equilibrium piece rates and effort choices are symmetric and noncontingent. We conclude the linear case by showing that equilibrium piece rates are symmetric and noncontingent in every period, and using this fact to derive a recursive characterization of the equilibrium piece rates. The uniqueness of the equilibrium piece rates follows immediately from this recursive characterization.

**Last-Period Piece Rates.** A standard argument shows that the period-$T$ equilibrium piece rates are symmetric, a

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

and so are noncontingent. Since it follows from the previous part that the period-$T$ effort choice of a worker with piece rate $b$ is $e_T = b$ no matter the worker's history, the period-$T$ equilibrium effort choices are symmetric and noncontingent as well.

**Piece Rates in Previous Periods.** Let $0 \leq t \leq T-1$ and suppose the equilibrium piece rates and effort choices from period $t+1$ on are symmetric and noncontingent; this is true for $t = T-1$. We show that the equilibrium piece rates and effort choices in period $t$ are also symmetric and noncontingent, and derive an expression for the equilibrium piece rates in this period. For the argument that 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 $R^*_{LBD,t}(e)$ to be given by (A3) with $b_{t+\tau} = b^*_{t+\tau}$ for all $\tau$. Note that $R^*_{LBD,t}(e)$ is independent of $e$, that is, $R^*_{LBD,t}(e) \equiv R^*_{LBD,t}$.

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

$$e_t = b + R^*_{CC,t} + R^*_{LBD,t}. \tag{A7}$$

In particular, given that (A7) does not depend on the worker's history, the worker's equilibrium choice of effort in period $t$ is symmetric and noncontingent if the same is true for worker's equilibrium piece rate in the same period. It follows immediately from (A7) that $e_t = e_t(b)$ is strictly increasing in $b$ and such that $\partial e_t(b)/\partial b = 1$.

Now let $w_{t+\tau}(b)$ be the 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 $W_t(b)$ to be such that $W_t(b) = \sum_{\tau=0}^{T-t} \delta^\tau w_{t+\tau}(b)$. The argument in the main text shows that the worker's equilibrium piece rate in period $t$ is the choice of $b$ that maximizes

$$\mathbb{E}[W_t(b)|I_t] - (r/2)\mathrm{Var}[W_t(b)|I_t] - e_t(b)^2/2, \tag{A8}$$

where $I_t$ is the information firms have about the worker in period $t$. In what follows, we establish that there exists a unique value of $b$ that maximizes (A8) and that this value of $b$ is independent of $I_t$. Thus, the equilibrium piece rates in period $t$ are the same for all workers and noncontingent.

We start by deriving the first-order condition for the problem of maximizing (A8). Let $y_{t+\tau}(e)$ be the worker's output in period $t+\tau$ as a function of the worker's effort in period $t$. It follows from (3) that $\mathbb{E}[w_{t+\tau}(b)|I_t] = \mathbb{E}[y_{t+\tau}(e_t(b))|I_t]$. Given that $\partial \mathbb{E}[y_t(e_t)|I_t]/\partial e_t = 1$ and $\partial \mathbb{E}[y_{t+\tau}(e_t)|I_t]/\partial e_t = F'_t(e_t)\lambda^{\tau-1}$ for all $\tau \geq 1$, we then have that

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

We show in the next paragraph that

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

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

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

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

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

where $\mathrm{Var}_0$ is independent of $b$. We claim that $\mathrm{Cov}[w_t(b), w_{t+\tau}(b)|I_t] = b\sigma_t^2$ for all $\tau \geq 1$, from which (A10) follows immediately. Given that the worker's reputation in period $t$ is nonrandom conditional on $I_t$, it follows that

$$\mathrm{Cov}[w_t(b), w_{t+\tau}(b)|I_t] = b\,\mathrm{Cov}[y_t(e_t(b)), w_{t+\tau}(b)|I_t]$$

for all $\tau \geq 1$ from (3). Now observe, once again from (3), that

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

for all $\tau \geq 1$, where $m_{t+\tau}(e_t)$ is the worker's reputation in period $t+\tau$ as a function of the worker's effort in period $t$. Hence, if $z_{t+s}(e_t)$ with $s \geq 0$ is the signal about the worker's ability in period $t+s$ as a function of the worker's effort in period $t$, then Lemma 1 implies that

$$\begin{aligned}\mathrm{Cov}[y_t(e_t(b)), w_{t+\tau}(b)|I_t] \;&=\; b_{t+\tau}\mathrm{Cov}[y_t(e_t(b)), y_{t+\tau}(e_t(b))|I_t] \\ &\quad + (1-b_{t+\tau})\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})\mathrm{Cov}[y_t(e_t(b)), z_{t+s}(e_t(b))|I_t]\end{aligned}$$

for all $\tau \geq 1$. Since $\mathrm{Cov}[y_t(e_t(b)), y_{t+\tau}(e_t(b))|I_t] = \sigma_t^2$ for all $\tau \geq 1$ and

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

we then have that

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

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 $\mathrm{Cov}[y_t(e_t(b)), w_{t+\tau}(b)|I_t]$ as

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

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

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

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

which is noncontingent, and thus the same for all workers; note that $b_T^* = 1/[1+r(\sigma_T^2+\sigma_\varepsilon^2)]$. That $b_t^*$ maximizes (A8), and so is the equilibrium piece rate in period $t$, follows from the fact that $\partial\mathbb{E}[W_t(b)|I_t]/\partial b$, the marginal benefit to a worker of an increase in $b$, is constant in $b$, while

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

the marginal cost to a worker of an increase in $b$, is strictly increasing in $b$. 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 $0 \leq t \leq T-1$ such that from period $t+1$ on the equilibrium piece rates and effort choices are symmetric and noncontingent, then the equilibrium piece rates and effort choices in period $t$ are symmetric and noncontingent as well. Since the last-period equilibrium piece rates and effort choices are symmetric and noncontingent, it follows by induction that

the equilibrium piece rates are symmetric and noncontingent in every period. From this it further follows, once again by the above argument, that for each $0 \leq t \leq T$ the equilibrium piece rate in 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 $t \leq T-1$ and such that

$$\frac{\sigma_t^2}{\sigma_\varepsilon^2} < F_t'(e) < \frac{\sigma_t^2}{\sigma_\varepsilon^2}\big[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)\big] \text{ for all } e \in \mathbb{R} \text{ and } 0 \leq t \leq T-1.$$

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 symmetric and noncontingent; and (*ii*) future equilibrium piece rates are in the interval $(0,1)$. Following that we show that the equilibrium piece rates and effort choices are symmetric and noncontingent and the equilibrium piece rates are in the interval $(0,1)$ if $r$ is small enough, and use this fact to derive a recursive characterization of the equilibrium piece rates. As in the linear case, the uniqueness of the equilibrium piece rates follows from this recursive characterization. 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 $\lambda$ is smaller than but close to one 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. In particular, they are symmetric and noncontingent.

**Piece Rates in Previous Periods.** Let $0 \leq t \leq T-1$ and suppose the equilibrium piece rates and effort choices from period $t+1$ on are symmetric and noncontingent, and piece rates belong to the interval $(0,1)$; this is true when $t = T-1$. We show that the equilibrium piece rates in period $t$ are also symmetric and noncontingent, and derive an expression for the equilibrium piece rates in this period. Once again, 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 $R_{LBD,t}^*(e)$ to be given by (A3) with $b_{t+\tau} = b_{t+\tau}^*$ for all $\tau$.

Consider first a worker's optimal choice of effort in period $t$. Since $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b_{t+\tau}^* + R_{CC,t+\tau}^*\big) \geq 0$ when $b_{t+\tau}^* \in (0,1)$ for all $1 \leq \tau \leq T-t$, we have that if the 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

$$e_t = b + R_{CC,t}^* + R_{LBD,t}^*(e_t); \tag{A13}$$

recall that $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\big(b_{t+\tau}^* + R_{CC,t+\tau}^*\big) \geq 0$ implies that $R_{LBD,t}^*(e)$ is monotone decreasing in $e$. As in the linear case, given that (A13) does not depend on the worker's history, the worker's equilibrium choice of effort in period $t$ is noncontingent, and so independent of the worker's identity, if the same holds for worker's equilibrium piece rate in period $t$.

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

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

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

Once again, let $W_t(b)$ be the present-discounted value of the wage payments to the worker 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 $t$ 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 and noncontingent solution.

We know from (A9) that

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

Since the functions $\{F_t\}_{t=0}^T$ do not matter for the derivation of $\mathrm{Var}[W_t(b)|I_t]$, it follows from (A10) that

$$\frac{\partial \mathrm{Var}[W_t(b)|I_t]}{\partial b} = 2\big[b(\sigma_t^2 + \sigma_\varepsilon^2) + H_t^*\big];$$

recall that $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^{\tau-1}$. So, the first-order condition for the problem of maximizing (A8) is

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

Using the first-order condition for effort (A13) and the definition of $R_{LBD,t}^*(e)$, we can rewrite (A15) as

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

Note that the solutions to (A16), if they exist, are noncontingent, and so are the same for every worker.

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

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

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

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

be the marginal cost to the worker of an increase in $b$. Given that $e_t(b)$ is nondecreasing in $b$ and $\partial e_t(b)/\partial b$ is nonincreasing in $b$, it follows that $MB_t$ is nonincreasing in $b$. Now note that since $F_t'''$ nonpositive and nondecreasing implies that $F_t'''(e)e \geq F_t''(e)$ for all $e \in \mathbb{R}$, we then have from (A14) that[54]

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

[54]The desired inequality is immediately true if $e \leq 0$. Suppose then that $e > 0$. Then $F_t''(e) = F_t''(0) + \int_0^e F_t'''(s)ds$ implies that $F_t''(e) \leq \int_0^e F_t'''(s)ds \leq \int_0^e F'''(e)ds = eF_t'''(e)$, where 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 $e > 0$.

Thus, $MC_t$ is strictly increasing in $b$, 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 is noncontingent. First note that $MB_t$ is bounded given that $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$ and $\partial e_t(b)/\partial b$ belongs to the unit interval. On the other hand, since $e_t(b)\partial e_t(b)/\partial b$ is strictly increasing in $b$, 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 $\partial e_T(b)/\partial b = 1$.

**Recursive Characterization of Equilibrium Piece Rates.** The above argument shows that if there exists $0 \leq t \leq T-1$ such that from period $t+1$ on the equilibrium piece rates and effort choices are symmetric and noncontingent, and the equilibrium piece rates are in the unit interval, then the equilibrium piece rates and effort choices in period $t$ are symmetric and noncontingent. 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 sufficient small. From this we are able to show that when $\lambda = 1$ in every period the equilibrium piece rates are symmetric, noncontingent, and belong to the interval $(0,1)$ as long as $r$ 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

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

we then have that

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

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

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

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

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

The above inequality holds since $F_t'(e) < (\sigma_t^2/\sigma_\varepsilon^2)\left[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)\right]$ for all $e \in \mathbb{R}$ and, by (A3),

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

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

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

the second equality follows from (A17) and the inequality follows since $b^*_{t+\tau}<1$ for all $1\leq\tau\leq T-t$. Therefore,

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

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

$$F'_t(e_t(b))\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b^*_{t+\tau}-R^*_{CC,t+\tau})-R^*_{CC,t}=R^*_{CC,t}\left(\frac{\sigma_\varepsilon^2}{\sigma_t^2}F'_t(e_t(b))-1\right)>0;$$

since $b^*_{t+\tau}<1$ for all $1\leq\tau\leq T-t$, we have that $R^*_{CC,t}>0$. It then follows from (A20) that there exists $\overline{r}>0$ such that

$$1+F'_t(e_t(b))\sum_{\tau=1}^{T-t}\delta^{\tau}(1-b^*_{t+\tau}-R^*_{CC,t+\tau})-R^*_{CC,t}-\frac{rH^*_t}{\partial e_t(b)/\partial b}>0 \tag{A21}$$

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 the argument so far, there exists $\overline{r}>0$ such that $b^*_t\in(0,1)$ provided $r\in(0,\overline{r})$. Note that since $\sigma_t^2$ is monotonically decreasing if $\sigma_\theta^2>\sigma_\zeta^2$ and monotonically increasing if $\sigma_\theta^2<\sigma_\zeta^2$, it follows that $\sigma_t^2\leq\max\{\sigma_\theta^2,\sigma_\zeta^2\}$. 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 $0\leq t\leq T-1$ such that from period $t+1$ on the equilibrium piece rates and effort choices are symmetric and noncontingent 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 in period $t$ have the same properties provided $r\in(0,\overline{r})$ and $\lambda=1$. Since the last-period equilibrium piece rates and effort choices are symmetric and noncontingent and the last-period equilibrium piece rates are in the interval $(0,1)$, a straightforward induction argument shows that if $\lambda=1$ and $r\in(0,\overline{r})$, then in every period the equilibrium piece rates and effort choices are symmetric and noncontingent and the equilibrium piece rates are in the interval $(0,1)$. Moreover, as

$$\frac{\partial e_t(b)}{\partial b}=\left(1-\frac{F''_t(e_t(b))}{F'_t(e_t(b))}R^*_{LBD,t}(e_t(b))\right)^{-1}$$

by (A3) and (A14), equations (A13) and (A16) imply that the equilibrium piece rate in $t$ is defined recursively as

$$\begin{aligned}b^*_t = \frac{1}{1+r(\sigma_t^2+\sigma_\varepsilon^2)\left[1-\frac{F''_t(e^*_t)}{F'_t(e^*_t)}R^*_{LBD,t}(e^*_t)\right]}\Bigg\{&1+F'_t(e^*_t)\sum_{\tau=1}^{T-t}\delta^{\tau}\lambda^{\tau-1}-R^*_{LBD,t}(e^*_t)-R^*_{CC,t}\\ &-r\left[1-\frac{F''_t(e^*_t)}{F'_t(e^*_t)}R^*_{LBD,t}(e^*_t)\right]H^*_t\Bigg\},\end{aligned}$$

where $e^*_t$ is the unique solution to $e^*_t=b^*_t+R^*_{CC,t}+R^*_{LBD,t}(e^*_t)$ and $\lambda=1$. This concludes the equilibrium derivation in the nonlinear case.

**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 $\lambda$.[55] So, for each $0 \leq t \leq T$, 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 $\lambda$ 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 restrictions on the marginal rates of human capital accumulation are natural. The marginal rates of human capital accumulation cannot be too large, otherwise piece rates can be greater than one. Likewise, the marginal rates of human capital accumulation cannot be too small, otherwise the learning-about-ability motive dominates, and we know from Gibbons and Murphy [1992] that it can lead to negative piece rates. Workers cannot be too risk averse as well, otherwise the demand for insurance against the lifetime risk in compensation due to 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 equilibrium characterization. It includes Proposition 1 as a special case.

**Proposition 6.** *For each $0 \leq t \leq T$, let $H^*_t = \sigma^2_t \sum_{\tau=1}^{T-t} \delta^{\tau}$ and define $R^*_{CC,t}$, $R^*_{LBD,t}(e)$, $b^*_t$, and $e^*_t$ recursively as:*

$$R^*_{CC,t} = \sum_{\tau=1}^{T-t} \delta^{\tau}(1 - b^*_{t+\tau}) \left( \prod_{k=1}^{\tau-1} \mu_{t+\tau-k} \right) (1 - \mu_t);$$

$$R^*_{LBD,t}(e) = F'_t(e) \sum_{\tau=1}^{T-t} \delta^{\tau} \lambda^{\tau-1} \big( b^*_{t+\tau} + R^*_{CC,t+\tau} \big);$$

$$b^*_t = \frac{1}{1 + r(\sigma^2_t + \sigma^2_{\varepsilon}) \left[ 1 - \dfrac{F''_t(e^*_t)}{F'_t(e^*_t)} R^*_{LBD,t}(e^*_t) \right]} \Bigg\{ 1 + F'_t(e^*_t) \sum_{\tau=1}^{T-t} \delta^{\tau} \lambda^{\tau-1} - R^*_{LBD,t}(e^*_t) - R^*_{CC,t} - r \left[ 1 - \frac{F''_t(e^*_t)}{F'_t(e^*_t)} R^*_{LBD,t}(e^*_t) \right] H^*_t \Bigg\};$$

$$e^*_t = b^*_t + R^*_{CC,t} + R^*_{LBD,t}(e^*_t).$$

*1. Suppose that $F_t(e) = \gamma_t e$, with $\gamma_t > 0$, for all $0 \leq t \leq T$. The equilibrium is unique, symmetric, and noncontingent, with the equilibrium piece rate and effort choice in period $t$ given by $b^*_t$ and $e^*_t$, respectively.*
*2. Suppose the functions $\{F_t\}_{t=0}^{T}$ are nonlinear for at least one $t \leq T - 1$ and satisfy*

$$\frac{\sigma^2_t}{\sigma^2_{\varepsilon}} < F'_t(e) < \frac{\sigma^2_t}{\sigma^2_{\varepsilon}} \big[ 1 + r(\sigma^2_t + \sigma^2_{\varepsilon}) \big] \text{ for all } e \in \mathbb{R} \text{ and } 0 \leq t \leq T - 1.$$

*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 equilibrium is unique, symmetric, and noncontingent, with the equilibrium piece rate and effort choice in period $t$ given by $b^*_t$ and $e^*_t$, respectively. Moreover, the equilibrium piece rates are in the interval $(0, 1)$.*

Clearly, the conditions under which equilibrium piece rates in the nonlinear case are in the unit interval also apply in the linear case with $\gamma_t$ in place of $F'_t(e)$. We conclude this part by providing an alternative set of conditions under which equilibrium piece rates in the linear case are in the unit interval. Unlike the conditions of Proposition 6, they apply even in the absence of learning about ability.[56] It is easy to extend Corollary 2 to the nonlinear case, thus obtaining alternative sets of conditions under which the equilibrium characterization of Proposition 6 holds.

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

[56] The upper bound for $\gamma_t$ in Corollary 2 is tighter than the upper bound for $\gamma_t$ provided by Proposition 6 when $\sigma^2_t$ is large.

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

$$\frac{\sigma_t^2}{\sigma_\varepsilon^2} < \gamma_t < \frac{1-\delta\lambda}{\delta} r(\sigma_t^2 + \sigma_\varepsilon^2) \text{ for all } 0 \leq t \leq T-1.$$

*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* $0 \leq t \leq T$.

*Proof.* We only need to prove that piece rates are bounded above by one if $\gamma_t < \delta^{-1}(1-\delta\lambda) r(\sigma_t^2 + \sigma_\varepsilon^2)$ for all $0 \leq t \leq t \leq T-1$. Recall that

$$\begin{aligned} b_t^* &= \frac{1}{1 + r(\sigma_t^2 + \sigma_\varepsilon^2)} \left[ 1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{CC,t}^* - R_{LBD,t}^* - r H_t^* \right] \\ &= \frac{1}{1 + r(\sigma_t^2 + \sigma_\varepsilon^2)} \left[ 1 + \gamma_t \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} (1 - b_{t+\tau}^* - R_{CC,t+\tau}^*) - R_{CC,t}^* - r H_t^* \right] \end{aligned}$$

for all $0 \leq t \leq T$. First note that $b_T^* = 1/[1 + r(\sigma_T^2 + \sigma_\varepsilon^2)] < 1$. Suppose then, by induction, that there exists $1 \leq t \leq T$ such that $b_{t+\tau}^* < 1$ for all $0 \leq \tau \leq T-t$. Let $s = t-1$. We are done if we show that $b_s^* < 1$. Since $R_{CC,s+\tau}^* > 0$ for all $0 \leq \tau \leq T-s$ by the induction hypothesis and $H_s^* > 0$, it follows from the equation for the equilibrium piece rates that $b_s^* < 1$ if

$$\gamma_s \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1} (1 - b_{s+\tau}^*) < r(\sigma_s^2 + \sigma_\varepsilon^2).$$

Using the induction hypothesis one more time, we have that the left-hand side of the above inequality is bounded above by $\gamma_s \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1} < \gamma_s \delta (1-\delta\lambda)^{-1}$. This implies the desired result. □

# 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 $\sigma_t^2$ is nonincreasing in $t$. 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,

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

where the weak inequality follows since $\sigma_t^2$ is nonincreasing in $t$, and so $b_t^0$ is nondecreasing in $t$. 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

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

the first inequality follows from the fact that $b_T^* \in (0,1)$ and $\mu_t \in (0,1)$ for $0 \leq t \leq T$ whereas the second inequality follows since $b_{s+1+\tau}^* > b_{s+\tau}^*$ for all $1 \leq \tau \leq T-s-1$ by the induction hypothesis. Hölmstrom [1999] shows that $(1-\mu_s) \prod_{k=1}^{\tau-1} \mu_{s+\tau-k} = (1-\mu_s) \prod_{k=1}^{\tau-1} \mu_{s+\tau}$ is strictly increasing in $\mu_s$ for all $\tau \geq 1$. So, given that $\sigma_t^2$

nonincreasing in $t$ implies that $\mu_t$ is nondecreasing in $t$, 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

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

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 $T_0 > 0$ and $K > 0$ and let $T = T_0 + K$; 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 that would prevail in period $t$ if uncertainty about ability from period $T_0$ on were constant and equal to $\sigma^2_\infty$. 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, note from the first part of the proof that $\widehat{b}_t$ is strictly increasing in $t$ for all $T_0 \leq t \leq T$. So, 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 in $t$ for all $T_0 \leq t \leq T$ as well. Since $\lim_{T_0 \to \infty} \sigma^2_{T_0} = \sigma^2_\infty$, it follows from the argument in the previous paragraph that there exists $T_0 \geq 0$ such $|b^*_t - \widehat{b}_t| \leq \eta$ for all $T_0 \leq t \leq T$. This concludes the proof.

## B.2 Proof of Lemma 3

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

$b^*_{s+\tau} \geq b^0$ for all $1 \leq \tau \leq T-s$, which also holds by the induction hypothesis, implies that

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

Thus, a sufficient condition for $b^*_s \leq 1$ is that

$$1+\gamma_s r\sigma_\varepsilon^2 b^0\sum_{\tau=1}^{T-s}\delta^\tau\lambda^{\tau-1} \leq (b^0)^{-1} = 1+r\sigma_\varepsilon^2,$$

which holds if $\gamma_s \leq (1-\delta\lambda)(1+r\sigma_\varepsilon^2)/\delta(1-(\delta\lambda)^{T-T_0})$.

We now show that $\gamma_t$ nonincreasing in $t$ for all $T_0 \leq t \leq T-1$ implies that $b^*_t$ is strictly decreasing in $t$ 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

$$\begin{aligned} b^*_s \;&>\; b^0\left(1+\gamma_s\sum_{\tau=1}^{T-s-1}\delta^\tau\lambda^{\tau-1}(1-b^*_{s+\tau})\right) > b^0\left(1+\gamma_s\sum_{\tau=1}^{T-s-1}\delta^\tau\lambda^{\tau-1}(1-b^*_{s+1+\tau})\right) \\ &\geq\; b^0\left(1+\gamma_{s+1}\sum_{\tau=1}^{T-s-1}\delta^\tau\lambda^{\tau-1}(1-b^*_{s+1+\tau})\right) = b^*_{s+1}; \end{aligned}$$

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 there exists $0 < T_0 < T$ such that $b^*_t < 1$ for all $T_0 \leq t \leq T$ and set $\gamma_{T_0} > 0$; the assumption that $\gamma_{T_0}$ is positive is consistent with the assumption on the equilibrium piece rates from period $T_0$ on. 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 $\gamma_{T_0-1}$ if necessary, we can ensure that $\gamma_{T_0} > \gamma_{T_0-1}$. From (24) it follows that $b^*_{T_0-1} \in (b^0, b^*_{T_0})$. Since $b^*_{T_0-1} < 1$, 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 backwards we obtain the desired result.

### B.4 Proof of Proposition 2

We first show that when $\sigma_\zeta^2$ is small, there exists $T_0 \geq 0$ such that if $T > T_0$, $\gamma_t$ is nonincreasing in $t$ for all $T_0 \leq t \leq T-1$, and $0 < \gamma_{T-1} \leq \gamma_{T_0} \leq (1-\delta\lambda)(1+r\sigma_\varepsilon^2)/\delta(1-(\delta\lambda)^{T-T_0})$, then $b^*_t$ is strictly decreasing in $t$ 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 $\sigma_\zeta^2$ 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 $\sigma_\zeta^2$ is positive but small. Fix $T_0 > 0$ and $K > 0$ and let $T = T_0 + K$; we pin down $T_0$ below. 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 that would prevail in period $T_0 \leq t \leq T$ if only learning by doing were present. The same argument as in the proof of Lemma 2 shows that $\lim_{\sigma^2_{T_0}\to 0} b^*_t = \widehat{b}_t$ for all $T_0 \leq t \leq T$.

Since, by Lemma 3, $\widehat{b}_t$ is strictly increasing in $t$ for all $T_0 \leq t \leq T$ and $\lim_{T_0 \to \infty} \sigma^2_{T_0} = 0$, it then follows that there exists $T_0 \geq 0$ such that $b^*_t$ is strictly decreasing in $t$ for all $T_0 \leq t \leq T$.

We now show that there exists $T_0 \geq 0$ and $\underline{\gamma} > 0$ such that if $T > T_0$ and $\gamma_t < \underline{\gamma}$ for all $T_0 \leq t \leq T-1$, then $b^*_t$ is strictly increasing in $t$ 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 $T_0 \leq t \leq T-1$; 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 $\gamma_{T_0}$ to $\gamma_{T-1}$ are positive but small. 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 in $t$ for all $T_0 \leq t \leq T$ if $T_0$ is sufficiently large.

## B.5 Proof of Proposition 3

The first part of the proposition follows immediately by continuity and Corollary 1. 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. Suppose that $\lambda = 1$. It follows from the proof of Proposition 6, see (A18), that

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

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 $\sigma^2_\theta > \sigma^2_\infty$; we know from the proofs of Lemma 2 and Proposition 2 that both assumptions can be relaxed. Then $b^*_t$ is strictly increasing in $t$ 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 $R^*_{CC,T_0-1} > 0$. From (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 $\gamma_{T_0-2}$ for which $b^*_{T_0-1} < b^*_{T_0-2} < 1$. Continuing backwards we obtain the desired result.

## B.6 Initially Increasing Piece Rates

Here we show that equilibrium piece rates are initially increasing when $\sigma^2_\theta$ is large and $\sigma^2_\varepsilon$ and $\sigma^2_\zeta$ are small. For simplicity, assume that $\sigma^2_\varepsilon = 0$. Since the equations for the equilibrium piece rates are continuous in $\sigma^2_\varepsilon$, the results extend to the case in which $\sigma^2_\varepsilon$ is positive but small. Note that $\sigma^2_\varepsilon = 0$ implies that $\sigma^2_t = \sigma^2_\zeta$ for all $1 \leq 1 \leq T$ and, since $\mu_t \equiv 0$, that $R^*_{CC,t} = \delta(1 - b^*_{t+1})$ for all $0 \leq t \leq T-1$. So,

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

for all $1 \leq t \leq T-1$ with $b^*_T = 1/(1 + r\sigma^2_\zeta)$; the second equation 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 $\sigma^2_\zeta$ is sufficiently small. Indeed, in the limit as $\sigma^2_\zeta$ converges to zero the above equations for $b^*_t$ reduce to $b^*_t = 1 - \delta + \gamma_t \delta + (1 - \gamma_t)\delta b^*_{t+1}$ for all $1 \leq t \leq T-1$ with $b^*_T = 1$. In this limiting case it follows immediately that $b^*_t = 1$ for all $t \geq 1$. The desired result follows since the equations for $b^*_t$ depend continuously on $\sigma^2_\zeta$. Now note that

$$b^*_0 = \frac{1}{1 + r\sigma^2_\theta} \left( 1 - \delta + \gamma_0 \delta + (1 - \gamma_0)\delta b^*_1 - r\sigma^2_\theta \sum_{\tau=1}^{T} \delta^\tau \right),$$

and so $b_0^*$ is smaller than $\eta$ if $\sigma_\theta^2$ is sufficiently large. Since $b_1^*$ does not depend on $\sigma_\theta^2$, it then follows that $b_0^* < b_1^*$ if $\sigma_\theta^2$ is sufficiently large and $\sigma_\zeta^2$ is sufficiently small.

# C Appendix: Identification

## C.1 Second Moments of the Wage Distribution

Here we compute the second moments of the wage distribution. From (3), we can write worker $i$'s wage in period $t$ as $w_{it} = \overline{w}_{it} + r_{it}$, where $r_{it} = (1 - b_t^*)\mathbb{E}[\theta_{it}|I_t] + b_t^*(\theta_{it} + \varepsilon_{it})$. By construction, $r_{it}$ is the random part of $w_{it}$. Assume without loss that $m_\theta = 0$, in which case $\mathbb{E}[\theta_{it}] = 0$, and so $\mathbb{E}[r_{it}] = 0$. In what follows we use repeatedly the fact that $\mathbb{E}[\theta_{it}|I_t] \perp \theta_{it} - \mathbb{E}[\theta_{it}|I_t]$.[57]

**Variances of Wage Residuals.** We claim that

$$\mathrm{Var}[r_{it}] = \mathrm{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2).$$

Since

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

we have that

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

Now observe that $\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_t]] = \mathrm{Var}[\theta_{it}] - \mathrm{Var}[\mathbb{E}[\theta_{it}|I_t]]$.[58] Moreover, given that $\theta_{it}|I_t = \iota_t$ is normally distributed with mean $\mathbb{E}[\theta_{it}|I_t = \iota_t]$ and variance $\sigma_t^2$, the random variable $(\theta_{it} - \mathbb{E}[\theta_{it}|I_t])|I_t = \iota_t$ is normally distributed with mean zero and variance $\sigma_t^2$. Therefore, $\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_t]] = \sigma_t^2$, and so, since $\mathrm{Var}[\theta_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2$, it follows that $\mathrm{Var}[\mathbb{E}[\theta_{it}|I_t]] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2$.[59] The desired result follows from (C24).

**Covariances of Wage Residuals.** We claim that

$$\mathrm{Cov}[r_{it}, r_{it+s}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + b_t^*\sigma_t^2$$

for all $1 \leq s \leq T - t$. Let $\eta_{it}^s = \mathbb{E}[\theta_{it+s}|I_{t+s}] - \mathbb{E}[\theta_{it}|I_t]$. We claim that $\mathbb{E}[\eta_{it}^s|I_t] = 0$.[60] Indeed, given that $\mathbb{E}[\theta_{it+s}|I_t] = \mathbb{E}[\theta_{it}|I_t]$, the law of iterated expectations for conditional expectations implies that

$$\mathbb{E}[\eta_{it}^s|I_t] = \mathbb{E}[\mathbb{E}[\theta_{it+s}|I_{t+s}]|I_t] - \mathbb{E}[\theta_{it}|I_t] = \mathbb{E}[\theta_{it+s}|I_t] - \mathbb{E}[\theta_{it}|I_t] = 0.$$

Since

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

we then have that

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

[57] This is a consequence of the fact that the conditional expectation is an orthogonal projection.

[58] Indeed, $\mathrm{Var}[A - B] = \mathrm{Var}[A] + \mathrm{Var}[B] - 2\mathrm{Cov}[A, B]$ and $\mathrm{Cov}[\theta_{it}, \mathbb{E}[\theta_{it}|I_t]] = \mathrm{Var}[\mathbb{E}[\theta_{it}|I_t]]$.

[59] Notice that $\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_t]] = \mathbb{E}[(\theta_{it} - \mathbb{E}[\theta_{it}|I_t])^2] = \mathbb{E}[\mathbb{E}[(\theta_{it} - \mathbb{E}[\theta_{it}|I_t])^2|I_t]] = \mathbb{E}[\mathbb{E}[\sigma_t^2|I_t]] = \sigma_t^2$.

[60] This result is intuitive. By definition, $\eta_{it}^s$ is the change in the conditional expectation about worker $i$'s ability from period $t$ to period $t + s$. Since shocks to ability have zero mean and the information about the worker's ability learned in period $t$ and after is "orthogonal," that is new, to the information contained in $I_t$, Bayes' rule implies that $\eta_{it}^s$ has mean zero conditional on $I_t$.

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

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

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

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

where $\Lambda_t^s$ is a zero-mean random variable. Thus, given that $\mathbb{E}[(\theta_{it}+\varepsilon_{it})\big(\theta_{it}+\varepsilon_{it}-\mathbb{E}[\theta_{it}|I_t]\big] = \sigma_t^2 + \sigma_\varepsilon^2$ and $\mathbb{E}[\theta_{it}(\theta_{it}-\mathbb{E}[\theta_{it}|I_t])] = \sigma_t^2$, it follows 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 second equality follows from (A17) and the fact that $(\sigma_t^2+\sigma_\varepsilon^2)(1-\mu_t) = \sigma_t^2$.

**Summing Up.** The following lemma summarizes the results we obtained. It provides a complete characterization of the second moments of the wage distribution.

**Lemma 6.** *For all $0 \le t \le T$ and $1 \le s \le T-t$, we have that:*
*(i)* $\mathrm{Var}[r_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2+\sigma_\varepsilon^2)$;
*(ii)* $\mathrm{Cov}[r_{it}, r_{it+s}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + b_t^*\sigma_t^2$.

## C.2 More General Human-Capital Process

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

**The Case of Discrete Performance.** Consider now the case in which the econometrician observes only a discrete version of $p_{it}$. Namely, assume that for each $0 \le t \le T$, there exist thresholds $\overline{p}_{1t} < \ldots < \overline{p}_{Kt}$ and that the econometrician observes the performance measure $p_{it}^o$ given by

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

Note that this is a plausible representation of performance scales in firms; see, for instance, the discussion in Baker et al. [1994a]. Since $\mathbb{P}\{p_{it}^o = K\} = 1 - \mathbb{P}\{p_{it} \le \overline{p}_{Kt}\}$ and $\mathbb{P}\{p_{it}^o = k\} = \mathbb{P}\{p_{it} \le \overline{p}_{k+1t}\} - \mathbb{P}\{p_{it} \le \overline{p}_{kt}\}$ for all $k \in \{1,\ldots,K-1\}$ by definition of $p_{it}^o$, it is immediate that the probabilities $\mathbb{P}\{p_{it} \le \overline{p}_{1t}\}$ to $\mathbb{P}\{p_{it} \le \overline{p}_{Kt}\}$ are identified from the probabilities $\mathbb{P}\{p_{it}^o = 1\}$ to $\mathbb{P}\{p_{it}^o = K\}$, that is, from the distribution of the discrete performance measure in period $t$.

Once again, it is easy to show that the equilibrium is unique, symmetric, and noncontingent; see the Appendix for details. Let $k_t^*$ and $e_t^*$ be, respectively, a worker's equilibrium stock of human capital and effort in $t$. Then, using $\mathbb{E}[w_{it}] = m_\theta + k_t^* + e_t^*$ and that $\mathbb{P}\{p_{it} \le \overline{p}_{kt}\} = \mathbb{P}\{\eta_{it} \le \overline{p}_{kt} - e_t^*\} = G(\overline{p}_{kt} - e_t^*)$ for each $k$ with $G$ strictly increasing

and so invertible, we obtain a linear system of $K+1$ equations

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

in the $K+3$ unknowns $(e_t^*, k_t^*, \overline{p}_{1t}, \ldots, \overline{p}_{Kt})$ for each $t$. This system has a unique solution up to $m_\theta$ and $\overline{p}_{1t}$. 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 $m_\theta$ 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 the mean wage and the distribution of workers' performance up to $m_\theta$ and $\overline{p}_{1t}$ in each period $t$. The rest of the argument proceeds as in the case of continuous performance in the main text.

**General Performance Function.** Here we consider 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 $0 \leq t \leq T$ 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 is clear that we can extend the analysis to the case in which the econometrician observes the truncated version $p_{it}^0$ by following the same approach used in the main text. Suppose that the equilibrium is unique, symmetric, and noncontingent and let $e_t^*$ and $k_t^*$ be, respectively, the workers' equilibrium effort and stock of human capital in period $t$. For each $0 \leq t \leq T$ we have the following system of equations:

$$\begin{array}{rcl} e_t^* + k_t^* & = & \mathbb{E}[w_{it}] - m_\theta \\ f_t(e_t^*, k_t^*) & = & \mathbb{E}[p_{it}] - \mathbb{E}[\eta_{it}] \end{array}, \tag{C25}$$

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

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

First notice that since $e \mapsto f_t(e, \alpha - e)$ is surjective for all $\alpha \in \mathbb{R}$, equation (C26) has a solution regardless of $m_\theta$, $\mathbb{E}[w_{it}]$, $\mathbb{E}[p_{it}]$, 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.[61]

We have thus established that if the functions $f_1$ to $f_T$ have the properties described in the previous paragraph, then the workers' effort and stock of human capital in each period $0 \leq t \leq T$ are identified from mean wages and mean performance measures in this period up to $m_\theta$. The remainder of the identification argument is identical to the argument in the main text.

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 $0 \leq t \leq T$ and $e \mapsto f_t(e, \alpha - e)$ surjective for all $\alpha \in \mathbb{R}$ and $0 \leq t \leq T$ are necessary for identification. Fix $0 \leq t \leq T$ and let $F_t : \mathbb{R}^2 \to \mathbb{R}^2$ be such that $F_t(e,k) = (e+k, f_t(e,k))$. A necessary condition for identification is that $F_t(e,k) = v$ has a solution for any $v \in \mathbb{R}^2$. Given that $F_t$ is continuously differentiable, it follows from Haddamard's global inverse function theorem, see, e.g., Gordon [1972], that $F_t$ is a ($C^1$) diffeomorphism if, and only if, $DF_t(e,k)$, the Jacobian matrix of $F_t$ evaluated at $(e,k)$, has non-zero determinant for all $(e,k) \in \mathbb{R}^2$ and $\lim_{||(e,k)|| \to \infty} ||F_t(e,k)|| = \infty$, where $||\ ||$ is the Euclidian norm.[62] So, a necessary condition for identification is that $\det DF_t(e,k) \neq 0$ for all $(e,k) \in \mathbb{R}^2$ and

[61] If there exist $e_1 < e_2$ with $h(e_1) = h(e_2)$, then the intermediate value theorem implies that there exists $e^* \in [e_1, e_2]$ such that $h_t'(e^*) = 0$.

[62] 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

$\lim_{||(e,k)||\to\infty} ||F_t(e,k)|| = \infty$.

Since $\det DF_t(e,k) = \partial f_t(e,k)/\partial k - \partial f_t(e,k)/\partial e$, it then follows that $\partial f_t(e,k)/\partial k \neq \partial f_t(e,k)/\partial e$ for all $(e,k) \in \mathbb{R}^2$ is necessary for identification. Now note 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 argument is the same when the former condition holds. Hence, $f(e, \alpha - e)$ is strictly increasing in $e$ for all $\alpha \in \mathbb{R}$. Given that $||F_t(e, \alpha - e)|| = \sqrt{\alpha^2 + f_t(e, \alpha - e)^2}$ 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 $\lim_{||(e,k)||\to\infty} ||F_t(e,k)|| = \infty$ 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.

# D Supplementary Appendix

We provide here omitted model and estimation details.

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

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

$$p_{it} = \gamma_t^e e_{it} + \gamma_t^k k_{it} + \theta_{it} + \eta_{it},$$

where $\gamma_t^e$ and $\gamma_t^k$ are known constants and $\eta_{it}$ is an unobserved idiosyncratic shock to worker $i$'s performance measure in period $t$ that is normally distributed with mean zero and variance $\sigma_\eta^2$ and is orthogonal to all other shocks in the environment. For ease of exposition, assume that $\gamma_t^e \equiv 1$ and $\gamma_t^k \equiv 0$; our analysis extends to the more general case if, and only if, $\gamma_t^e \neq \gamma_t^k$ for all $0 \leq t \leq T$.

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 again given by $w_{it} = (1 - b_{it})\mathbb{E}[y_{it}|I_t] + b_{it}y_{it}$, where $b_{it}$ is the worker's piece rate in period $t$ and $I_t$ is the public information about the worker that is available in period $t$. However, unlike the case without the performance measure, $I_t$ not only contains the worker's output realizations before period $t$ but also contains the realizations of the worker's performance measure before period $t$. The definition of an equilibrium is the same as before and so is the definition of a noncontingent equilibrium. We still consider pure-strategy equilibria.

**Learning About Ability.** We first discuss how the presence of the performance measure affects learning about the 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.[63] 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 the period. Since in equilibrium agents correctly anticipate a worker's effort and stock of human capital at any point in time, the same argument as in the main text shows that the posterior belief about worker $i$'s ability in period $t$ is normally distributed with some mean $m_{it}$ and variance $\sigma_{it}^2$. 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 $\sigma_{it}^2$ evolve over time according to

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

continuously differentiable.

[63] As in the main text, $e_{it}^*$ and $k_{it}^*$ can depend on the worker's history in period $t$.

the equations for the evolution of $m_{it}$ and $\sigma_{it}^2$ 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.[64]

Now let $\sigma_{\varepsilon\eta}^2 = \sigma_\varepsilon^2\sigma_\eta^2/(\sigma_\varepsilon + \sigma_\eta^2)$ and

$$z_{it} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} z_{it}^y + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} z_{it}^p. \tag{D27}$$

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

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

So, the evolution of the posterior means and variances of the workers' abilities follow the *same* laws of motion as in the case without the performance measure, except that $\sigma_{\varepsilon\eta}^2$ plays the role of the variance of the noise in output and $z_{it}$ given by (D27) plays the role of the signal about worker $i$'s ability in period $t$.[65] 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 workers' reputation is still given by the expression in Lemma 1 with $z_{it}$ now given by (D27).

**Dynamic Returns to Effort.** We now consider the first-order conditions for worker effort when piece rates and future behavior are noncontingent. Since $\partial\mathbb{E}[z_{it}|h^t]/\partial e_t = 1$ for any period $t$ and any period-$t$ private history $h^t$ for a worker, it follows that the expressions for $R_{CC,t}$ and $R_{LBD,t}$ are the same as 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 are noncontingent.[66]

**Equilibrium Piece Rates.** Since the first-order conditions for worker effort when piece rates and future behavior are noncontingent are the same as 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 the computation of the derivative $\partial\mathrm{Var}[W_t(b)I_t]/\partial b$, as the presence of the performance measure potentially affects the covariance of wage payments across periods. We claim that $\partial\mathrm{Var}[W_t(b)I_t]/\partial b$ 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 is in the evolution of the posterior variance of the workers' ability.

It still follows that

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

where $\mathrm{Var}_0$ does not depend on $b$; recall that $w_{t+\tau}(b)$ with $0 \leq \tau \leq T-t$ is a worker's wage in period $t+\tau$ when the piece rate in period $t$ is $b$. We claim that $\mathrm{Cov}[w_t(b), w_{t+\tau}(b)|I_t] = b\sigma_t^2$ for all $\tau \geq 1$, which implies the desired

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

[65] When $\sigma_\eta^2 = \infty$ and the performance measure is uninformative, the laws of motion for $m_{it}$ and $\sigma_t^2$ reduce to the laws of motion in the absence of the performance measure.

[66] More generally, $\partial\mathbb{E}[z_{it}|h^t]/\partial e_t = (\sigma_\eta^2 + \sigma_\varepsilon^2)^{-1}(\sigma_\eta^2 + \gamma_t^y\sigma_\varepsilon^2)$, 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 $R_{LBD,t}$ remains the same. Since, as we show below, we can identify the vector of 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 and 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 $m_\theta$ and satisfies the conditions for identification in 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.

result. As in Appendix A, $\text{Cov}[w_t(b), w_{t+\tau}(b)|I_t] = b\,\text{Cov}[y_t(e_t(b)), w_{t+\tau}(b)|I_t]$ and

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

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

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

The presence of the performance measure does not change the fact that $\text{Cov}[y_t(e_t(b)), y_{t+\tau}(e_t(b))|I_t] = \sigma_t^2$ for all $\tau \geq 1$. Now observe that

$$\text{Cov}[y_t(e_t(b)), z_{t+s}(e_t(b))|I_t] = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\text{Cov}[y_t(e_t(b)), z^y_{t+s}(e_t(b))|I_t] + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2+\sigma_\varepsilon^2}\text{Cov}[y_t(e_t(b)), z^p_{t+s}(e_t(b))|I_t].$$

Since $\text{Cov}[y_t(e_t(b)), z^p_{t+s}(e_t(b))|I_t] \equiv \sigma_t^2$ and

$$\text{Cov}[y_t(e_t(b)), z^y_{t+s}(e_t(b))|I_t] = \begin{cases} \sigma_t^2+\sigma_\varepsilon^2 & \text{if} \quad s=0 \\ \sigma_t^2 & \text{if} \quad s \geq 1 \end{cases},$$

we then have that

$$\begin{aligned}\text{Cov}[y_t(e_t(b)), w_{t+\tau}(b)|I_t] \;=\; & \sigma_t^2\left((1-b_{t+\tau})\sum_{s=0}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s}) + b_{t+\tau}\right) \\ & + \sigma^2_{\varepsilon\eta}(1-b_{t+\tau})\left(\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}\right)(1-\mu_t).\end{aligned}$$

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

**Identification.** As before, we can identify the equilibrium piece rates from a panel of wages by experience with information on their fixed and variable components. Now notice that 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 of variances $(\sigma_\theta^2, \sigma_\eta^2, \sigma_\varepsilon^2)$ 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 $\eta_{it}$ is known in order to obtain identification. The variance $\sigma_\zeta^2$ 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 $\sigma_1^2$ is known given the vector $(\sigma_\theta^2, \sigma_\eta^2, \sigma_\varepsilon^2)$. 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 $m_\theta$ and satisfies the conditions for identification in the case with the more general human capital process, the rest of the identification argument proceeds as in the main text.

## D.2 Additional Estimation Results

Here we present and discuss the estimates of two augmented versions of our model in which we allow, respectively, for a more flexible human capital process and for 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 $\iota_{it}$ represents a pure learning-by-doing investment in human capital that accrues for any period a worker spends in the labor market in that $\iota_{it}$ equals 1 in $t$ if worker $i$ is employed and equals 0 otherwise—the estimates reported in Table 1 are for

the baseline model in which the law of motion of human capital features $\beta \equiv 0$. As 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 mdodel is virtually unchanged. For instance, the estimated standard deviation of the initial distribution of ability $\sigma_\theta$, the noise in output $\sigma_\varepsilon$, and ability shocks $\sigma_\zeta$ are, respectively, 44.99, 516.74, and 5.43 for the baseline and are 48.15, 529.50, and 5.77 for the augmented model, whereas the estimates of $\gamma_1$, $\gamma_2$, and $r$ are virtually identical across the two models. The estimate of $\gamma_0$ is somewhat lower for the augmented model, 0.739, than for the baseline model, 0.892, much like the estimate of $\lambda$, which is also slightly lower for the augmented model, 0.932, than for the baseline model, 0.955. 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 wage growth.

Table 2: Estimates of Augmented Model Parameters

| Parameters | Estimates | Standard Errors |
|---|---|---|
| $\sigma^2_\theta$, variance of initial ability | 2,318.081 | 0.0013288 |
| $\sigma^2_\varepsilon$, variance of noise in output | 280,372.479 | 0.1114008 |
| $\sigma^2_\zeta$, variance of shock to ability | 33.286 | 0.0000792 |
| $\psi_0$, coefficient of degree 0 of $\gamma_t$ | 0.739 | 0.0000006 |
| $\psi_1$, coefficient of degree 1 of $\gamma_t$ | 0.035 | 0.0000001 |
| $\psi_2$, coefficient of degree 2 of $\gamma_t$ | -0.001 | 1.45E-09 |
| $\lambda$, fraction of undepreciated human capital | 0.932 | 0.0000001 |
| $r$, coefficient of relative risk aversion | 0.0002 | 1.52E-10 |
| $\beta$, coefficient on experience investment | 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. Like in the baseline model, the contribution of effort to human capital acquisition is sizable in all years, increasing with experience for younger workers and declining with experience for older workers after peaking at a marginal return of 1.01 dollars at experience 17.

In terms of identification, the variance and risk aversion parameters are identified as before. As for the remaining parameters, consider first the case in which the rate of human capital depreciation $1-\lambda$ is known. By the argument in the main text, $e^*_{T-1}$ and $\gamma_{T-1}$ are identified. As a result, $k^*_{T-1}$ and $k^*_T$ are identified from $k^*_{T-1} = \mathbb{E}[w_{iT-1}] - e^*_{T-1}$ and $k^*_T = \mathbb{E}[w_{iT}] - e^*_T$, respectively. Hence, the law of motion of human capital in $T$, $k^*_T = \lambda k^*_{T-1} + \beta + \gamma_{T-1} e^*_{T-1}$, identifies $\beta$ as $\beta = k^*_T - \lambda k^*_{T-1} - \gamma_{T-1} e^*_{T-1}$. Consider now the case in which the rate of human capital depreciation $1-\lambda$ is unknown. To see how the parameters $\lambda$ and $\beta$ can be separately identified, note first 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

$$\underbrace{\mathbb{E}[w_{i0}] - b^*_0 - R^*_{CC,0}}_{A_0} = R^*_{LBD,0} = \frac{\gamma_0}{\delta} \sum\nolimits_{\tau=1}^{T} (\delta\lambda)^{\tau-1} \big(b^*_\tau + R^*_{CC,\tau}\big). \tag{D28}$$

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

$$\underbrace{\frac{b^*_0}{b^0_0} - \big(1 - R^*_{CC,0} - rH^*_0\big)}_{B_0} = \frac{\gamma_0}{\delta} \sum\nolimits_{\tau=1}^{T} (\delta\lambda)^{\tau-1} \big(1 - b^*_\tau - R^*_{CC,\tau}\big). \tag{D29}$$

Denote the left side of (D28) by $A_0 > 0$ and that of (D29) by $B_0$. Both $A_0$ and $B_0$ are known once the variance and

risk aversion parameters are identified. Taking the ratio of (D28) and (D29) side by side yields

$$\frac{\sum_{\tau=1}^{T}(\delta\lambda)^{\tau-1}\left(1-b_{\tau}^{*}-R_{CC,\tau}^{*}\right)}{\sum_{\tau=1}^{T}(\delta\lambda)^{\tau-1}\left(b_{\tau}^{*}+R_{CC,\tau}^{*}\right)}=\frac{B_0}{A_0},$$

which can be further manipulated to obtain

$$\sum_{\tau=1}^{T}(\delta\lambda)^{\tau-1}\left[A_0(1-b_{\tau}^{*}-R_{CC,\tau}^{*})-B_0(b_{\tau}^{*}+R_{CC,\tau}^{*})\right]=\sum_{\tau=1}^{T}(\delta\lambda)^{\tau-1}\left[A_0-(A_0+B_0)(b_{\tau}^{*}+R_{CC,\tau}^{*})\right]=0. \quad \text{(D30)}$$

By Descartes' rule of signs, it is possible to determine the number of positive roots of this polynomial equation for $\delta\lambda$ by counting the number of sign changes in the coefficients of the polynomial proceeding from lower to higher powers, $n$. Then, $n$ is the maximum number of positive roots, which in our case is one as apparent from the inspection of (D30), since $A_0+B_0$ and $b_{\tau}^{*}+R_{CC,\tau}^{*}$ at each $\tau$ are positive and $A_0$ and $(A_0+B_0)(b_{\tau}^{*}+R_{CC,\tau}^{*})$ cross only once if $T$ is not too large. Hence, $\delta\lambda$ is identified and so is $\lambda$ since $\delta$ is assumed to be known. Once $\lambda$ is identified, the same argument as the one for the case in which $\lambda$ is known establishes that $\beta$ is identified. The rest of the identification argument proceeds as in the main text.

**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 $\tilde{r}_{it}=r_{it}+u_{it}$, where

$$u_{it+1}=\rho u_{it}+\nu_{it+1}, \;\; \nu_{it} \text{ i.i.d. with variance } \sigma_{\nu}^{2}, \;\; \text{and} \;\; \text{Var}(u_{it})=\frac{\sigma_{\nu}^{2}}{1-\rho^{2}}.$$

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 $(\rho,\sigma_{\nu}^{2})$ 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 $(\rho,\sigma_{\nu}^{2})$ are identified. Our task is thus to show how we can recover $(\rho,\sigma_{\nu}^{2})$ from the covariance matrix of observed wages. Once this is established, we can recover the covariance matrix of "true" wages and proceed with the identification arguments presented in Section 6. To this purpose, consider

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

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

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

which implies that

$$\frac{\text{Cov}(\tilde{r}_{it},\tilde{r}_{it+3}-\tilde{r}_{it+1})}{\text{Cov}(\tilde{r}_{it},\tilde{r}_{it+2}-\tilde{r}_{it+1})}=\frac{\rho^{2}-1}{\rho-1}=1+\rho. \quad \text{(D31)}$$

This ratio identifies $\rho$. To recover $\sigma_{\nu}^{2}$, we can use

$$\text{Cov}(\tilde{r}_{it},\tilde{r}_{it+2}-\tilde{r}_{it+1})=-\rho\sigma_{\nu}^{2}/(1+\rho). \quad \text{(D32)}$$

The estimates of the parameters of this version of the model are available upon request.