NBER WORKING PAPER SERIES

ON THE ROLE OF LEARNING, HUMAN CAPITAL, AND PERFORMANCE INCENTIVES FOR WAGES

Braz Camargo
Fabian Lange
Elena Pastorino

Working Paper 30191
http://www.nber.org/papers/w30191

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2022, Revised August 2025

We thank Peter Arcidiacono, Nick Bloom, Fabrice Collard, Dean Corbae, Jan Eeckhout, Eugenia Gonzalez-Aguado, Cristina Gualdani, Jonathan Heathcote, Christian Hellwig, Erik Hurst, Mike Keane, Philipp Kircher, Pete Klenow, Thomas Lemieux, Luigi Pistaferri, Sergio Salgado, Kjetil Storesletten, Chris Taber, Michela Tincani, Petra Todd, Gabriel Ulyssea as well as participants at various seminars and conferences for helpful comments. Braz Camargo gratefully acknowledges financial support from CNPq, Fabian Lange from the Canada Research Chairs Program, and Elena Pastorino from SIEPR (Stanford) and the NSF. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

On the Role of Learning, Human Capital, and Performance Incentives for Wages
Braz Camargo, Fabian Lange, and Elena Pastorino
NBER Working Paper No. 30191
June 2022, Revised August 2025
JEL No. D8, D86, J24, J3, J31, J33, J41, J44

ABSTRACT

Although performance pay is a key source of wage dispersion, little is known about why it typically accounts for only a small portion of pay or how it shapes wages over the life cycle. As we document, standard models of performance incentives fail to explain its declining importance later in the life cycle. We fill this gap and resolve this puzzle by proposing a new framework that integrates well-known models of dynamic moral hazard, uncertainty about worker productivity, and human capital acquisition, and proving it is identified based on its equilibrium characterization. Once we parameterize the model using well-known firm-level data, we find that performance pay is central to life-cycle wages. Strikingly, its level and life-cycle profile are not driven by the canonical risk-incentive tradeoff of moral-hazard models. Instead, they reflect workers' desire to insure against the productivity risk arising from uncertainty about their productivity and to acquire human capital.

Braz Camargo
Sao Paulo School of Economics
Department of Economics
braz.camargo@fgv.br

Fabian Lange
McGill University
Department of Economics
and NBER
fabian.lange@mcgill.ca

Elena Pastorino
Stanford University
and Hoover Institution
and also NBER
epastori@stanford.edu

1 Introduction

To align workers' incentives with firms' objectives, firms often link wages to performance on the job through bonuses, commissions, and piece rates. This practice has become increasingly prevalent in the United States and, since Lemieux [2008] and Lemieux, MacLeod, and Parent [2009], has been found to account for a large share of the rise in wage inequality among workers. For most workers, however, performance pay amounts to only a small fraction of pay. One reason is that workers already face strong implicit incentives for performance. For example, they may be motivated to work hard to convince current and future employers, uncertain about their talent, that their productivity is high. These implicit incentives for effort can then substitute for the explicit incentives from performance pay. This well-known *career-concerns* argument provides a common explanation for why performance pay often makes up only a small portion of pay (Holmström [1999]). It also has implications for how performance pay varies over the life cycle. Namely, as workers' experience accumulates and their productivity becomes better known, implicit incentives for performance weaken. To compensate, explicit incentives from performance pay should become more important (Gibbons and Murphy [1992]). Although intuitive, this popular explanation raises a puzzle though: as we document, the *opposite* pattern is common in the data—relative to total pay, performance pay eventually declines with experience. Thus, whether performance incentives matter for pay and, if so, why performance pay is typically so small remain open questions.[1]

In this paper, we reevaluate the importance of performance incentives for wages starting from the notion that workers face other powerful implicit incentives that can support, or discourage, effort on the job. For instance, the prospect of acquiring new human capital may affect the desire to exert effort. Indeed, standard models of human capital postulate that effort on the job complements the effort to acquire human capital, as in models of learning-by-doing (Becker [1962]), or substitutes for it, as in models of on-the-job training or learning-or-doing (Ben-Porath [1967]). But then compensation meant to incentivize work effort also influences how much human capital workers accumulate. We argue that the uncertainty about workers' productivity—central to the career-concerns logic—and the possibility of acquiring human capital during employment are key

[1] In the PSID, variable pay accounts for less than 5% of pay and does not represent a major component of it at any point over the life cycle (Frederiksen et al. [2017]); see the appendix for similar evidence from the NLSY. For an application of the career-concerns argument to firms' unobserved investment decisions, see Atkeson et al. [2015].

to understanding why performance pay is relatively low and how it evolves over time. In particular, we find that uncertainty about workers' productivity generates substantial wage risk over the life cycle, which is primarily responsible for the low level of performance pay. Yet, performance pay crucially shapes life-cycle wages through its indirect impact on workers' process of human capital acquisition, which in turn explains its peculiar life-cycle profile. Taken together, these findings call into question the canonical risk-incentive trade-off emphasized by the literature as the main determinant of performance incentives, and suggest that, although often ignored, performance pay is in fact a major source of the growth and dispersion of wages over the life cycle.

We formalize these intuitions by proposing a tractable model for the multiple incentives for effort on the job that arise from performance pay, career concerns, and the opportunity to accumulate human capital through employment. Our model thus offers a unified framework to investigate how these forces together determine the level and dynamics of wages and of their fixed and variable (performance-pay) components. In doing so, we accomplish three goals. First, as this framework allows us to analytically decompose the returns to effort and the sensitivity of pay to performance into the mechanisms we nest, we shed light on the forces governing how performance pay evolves with experience and so on the environments in which alternative life-cycle patterns of performance pay are likely to emerge. Conversely, we show that these patterns provide rich information that can be used to identify the different determinants of fixed and variable pay that we integrate.

Second, we resolve the puzzle that performance pay follows a hump-shaped pattern over the life cycle relative to total pay—it first increases and then decreases with experience—contrary to the prediction of leading models of performance incentives. For this hump-shaped pattern to arise, human capital is essential. In particular, when human capital is acquired through learning-by-doing, as we estimate, so that effort to produce output increases human capital, explicit incentives from performance pay *complement* the implicit ones from human capital acquisition: they support greater effort and thus human capital accumulation early on, when human capital is most valuable. As acquiring new human capital becomes less valuable over time, explicit incentives for performance progressively weaken. Thus, performance pay eventually declines relative to total pay.

Third, using our parameterized model, we demonstrate that, unlike what is often conjectured, performance pay plays a crucial role for the dynamics of wages over the life cycle through both its direct effect on total pay and its indirect effect on workers' accumulation of human capital.

We proceed to revisit the role of performance pay for wages by nesting and extending prominent models of performance incentives, learning about ability, and human capital acquisition—including those cited above. We integrate the canonical moral-hazard and job-assignment frameworks by building on the notion that workers can exert effort on *simple* tasks that are contractable—the typical dimension of labor supply—and on *complex* tasks that are not contractable—the typical source of moral hazard. We interpret different jobs as distinct bundles of simple and complex tasks. Workers differ in their productive ability, which is unobserved to all, and can accumulate human capital when employed. Effort in simple and complex tasks contributes to a worker's output and human capital—we do not restrict human capital to be accumulated through learning-by-doing or learning-or-doing. Whereas a worker's ability is unobserved to all and effort on complex tasks and human capital are observed only by the worker, effort on simple tasks and output (or performance) are publicly observed. Output then serves as a noisy signal of ability, effort on complex tasks, and human capital. Firms compete for workers through contracts that allow for variable pay contingent on a worker's output. Thus, workers face both explicit incentives for effort, as their pay is linked to performance, and implicit ones, as performance influences the market's perception of their ability and human capital and, in turn, their future wages. By making explicit how contracting considerations affect pay, its components, and the type of activities that workers perform in firms over time, we can then account for the dynamics of wages, their structure, and the known shift in workers' responsibilities from simple to complex tasks as their careers within firms progress.

Through the lens of this framework, we can decompose performance pay into distinct terms that capture the mechanisms we focus on: *i*) the trade-off between risk and incentives typical of moral-hazard settings; *ii*) the career-concerns incentives for performance due to the uncertainty about workers' ability; *iii*) the insurance firms provide against the risk due to this uncertainty; and *iv*) the incentives for performance due to human capital acquisition. Through this characterization, we identify the conditions that give rise to alternative life-cycle patterns of performance pay relative to total pay and corresponding profiles of job assignment. Crucially, we find that a human capital motive for effort rationalizes the hump-shaped profile of performance pay, which, as we show, is characteristic of both well-known firm-level data in personnel economics (Baker et al. [1994a,b], henceforth BGH, and Gibbs and Hendricks [2004]) and public data (PSID and NLSY).

Empirically, evaluating the incentive power of wage contracts and, in general, multiple mechanisms for the variability of wages across individuals and over time entails a difficult measurement exercise, since the underlying sources of the variation in wages are unobserved and mediated by firm and worker behavior. We proceed by first establishing that the primitive forces we examine can be easily recovered from the life-cycle profile of mean wages, their covariance structure, and the ratio of variable to total pay under intuitive conditions. We then estimate our model by matching how mean wages, their variance, and the ratio of variable to total pay evolve with experience in the BGH data, which our model replicates well—along with the untargeted pattern of the task complexity of workers' jobs. We do so under alternative restrictions on the parameters of our model, which reduce it to important special cases studied in the literature. We can thus measure the strength of the incentives we focus on in prototypical data and examine how they shape performance pay and total pay over time. By comparing our estimated full model to estimated special cases of it, which amount to classic models of performance incentives, learning about ability, and human capital, we can also assess the extent to which integrating known frameworks offers novel insights about the impact of performance incentives, and the other forces we explore, on wages.

We find that existing models, when they have predictions for performance pay, yield a life-cycle profile for it at odds with the data, with important implications for the nature of wage risk they imply. For instance, compared with our model—which instead closely matches the profile of performance pay—these models imply that ability risk is much greater yet far less persistent. This would suggest that performance incentives matter little for pay over most of the life cycle, since otherwise performance pay would be far larger than observed. Moreover, given the substantial risk arising from the variability of life-cycle wages, existing models counterfactually predict performance pay to be *negative* early on. By contrast, the productivity gains from accumulating human capital in our model lead to positive performance pay even early in the life cycle, as in the data.

Although workers receive little of their compensation as performance pay, we estimate that performance pay is central to life-cycle wage growth because it encourages workers to exert effort, which increases *both* output and human capital, as we find that human capital accumulates through learning-by-doing. Indeed, performance pay, through the effort it induces, accounts for more than 30% of life-cycle wage growth, once the cumulative impact of effort on human capital is taken into account. Performance pay is also critical for wage dispersion: it accounts for a large portion of the variability of wages over the first 10 to 20 years of experience. To the best of our knowledge, all these estimates of the role of performance pay for life-cycle wages—validated by our model's implications for workers' progression across jobs over their careers—are new to the literature.

Interestingly, we find that the key force depressing performance pay throughout the life cycle is workers' desire to insure against the wage risk stemming from the uncertainty about their ability rather than the pure productivity (output) risk stressed in the literature (see also Low et al. [2010] on the importance of persistent individual productivity risk for life-cycle income risk). From an asset pricing perspective, the intuition is simple. To reward effort, performance pay is large whenever output is high and so news about ability, and future pay, are good. But then workers paid according to performance-pay contracts effectively hold a portfolio of state-contingent claims to their output whose value comoves with their perceived ability. This portfolio pays out *persistently* more when output and thus signals about ability are high and less when output and thus signals about ability are low, thereby compounding the idiosyncratic risk that workers already face because their output fluctuates over time. Like any risk-averse investor, workers are willing to pay a premium for assets that diversify their risk. Accordingly, they favor contracts that reduce the wage risk generated by the variability of the beliefs about their ability, as firms learn about it. Indeed, in the absence of human capital considerations, firms and workers would agree to negative performance pay as a hedge against the risk in base pay. Performance pay then tends to be small to partially shield workers against the risk in lifetime wages induced by the persistent uncertainty about their ability.[2]

Since this uncertainty increases the variability of wages, a natural conjecture—and a common reading of learning models—is that the variance of wages would be lower if ability was known. Indeed, this would be the case if wage contracts did *not* respond to changes in the uncertainty about ability. But performance pay optimally increases when uncertainty declines, as workers demand less insurance against the lower implied wage risk, which increases the variance of wages. Hence, a trade-off exists between ex-ante wage risk due to the uncertainty about workers' productivity and ex-post wage risk due to the variability of wages induced by performance pay. In the setting we consider, lower dispersion in workers' productivity early in the life cycle—say, through improved schooling or better occupational sorting—turns out to lead firms to offer wages that are much more sensitive to performance. Then, *more* homogeneous workers in terms of their initial skills end up experiencing *more* wage inequality. These findings confirm once more that understanding the

[2] These results extend the intuition of Harris and Holmström [1982] on the insurance value of wage contracts to a framework with moral hazard, explicit performance incentives, and human capital acquisition.

structure of pay is central to understanding wage inequality.

Our paper relates to multiple strands of literature, including work on *i*) exploring the role of incentive pay for wage inequality and productivity (Lemieux, MacLeod, and Parent [2009] and Bloom and Van Reenen [2010]); *ii*) measuring the impact of learning about ability for education choices and for job and wage mobility (Gibbons et al. [2005], Arcidiacono et al. [2010], Kahn and Lange [2014], Kircher et al. [2015], Keane et al. [2017], Eeckhout and Weng [2022], Aryal et al. [2022], Pastorino [2024], and Tincani [2025]); and *iii*) assessing the role of human capital acquisition for wage growth (Gladden and Taber [2009] and Taber and Vejlin [2020]). Many studies emphasize the role for wages of persistent unobserved worker heterogeneity (Keane and Wolpin [1997], Geweke and Keane [2000], Meghir and Pistaferri [2004] and Low et al. [2010]), which is at the heart of the mechanisms we study. As for the rest of the paper, in Section 2, we present evidence on the life-cycle patterns of pay, performance pay, and their relationship with performance from the BGH data, which much of the personnel economics literature has drawn from. We use these data to estimate our model, introduced in Section 3 and analyzed in Sections 4 and 5. In Section 6, we estimate the model and demonstrate how performance incentives, learning about ability, and human capital acquisition shape the life-cycle profile of pay and its components. Section 7 concludes. See the the appendix and the supplementary appendix for omitted details.

2 Performance Pay over the Life Cycle

Here, we begin by describing our data and documenting that performance pay first increases and then decreases relative to total pay as labor market experience accumulates. We then present the rest of the moments we use to discipline our model in later sections. We conclude by arguing that the data is suggestive that both learning about ability and performance incentives matter for pay.

We use the data of Baker et al. [1994a,b], two seminal papers that laid the foundation of the literature on careers in firms by illustrating patterns of job and wage mobility that have been replicated by multiple studies.[3] The original dataset consists of the personnel records of all manage-

[3]Several papers—including DeVaro and Waldman [2012], Kahn and Lange [2014], Frederiksen et al. [2017], Ekinci et al. [2018], and Pastorino [2024] just to name a few—have exploited the rich information on compensation, job assignment, and worker performance in these data, which were graciously shared with us by Michael Gibbs. Frederiksen et al. [2017] report many regularities in terms of the distribution of wages and performance management systems across the BGH and five other firm-level data sets; see also the review by Waldman [2012]. In the appendix, we provide evidence on the hump-shaped experience profile of performance pay using data from another firm analyzed in Gibbs and Hendricks [2004] as well as public data from the PSID, the NLSY79, and the NLSY97.

Table 1: Summary Statistics 1981-1988

Panel A: Main Variables					
Variable	Mean	Std. Dev.	Min.	Max.	Observations
Age (years)	40.10	9.27	22	65	22,609
Experience (years)	18.58	9.81	1	47	22,609
Base salary (1988 $1,000s)	57.10	29.12	11.27	650.00	22,609
Bonus (1988 $1,000s)	2.41	8.82	0	296.13	22,609
Total compensation (1988 $1,000s)	58.91	29.17	22.71	857.84	22,609
Residual compensation	0.00	6.63	-148.18	148.18	22,609

Panel B: Education and Race			
Education	Frequency	Percent	Observations
High School	3,938	17.42%	
Some College	3,984	17.62%	
College	8,228	36.39%	
Advanced Degree	6,459	28.57%	
Total	22,609	100.00%	22,609
Race	Frequency	Percent	Observations
Minority (0)	2,433	10.78%	
White (1)	20,139	89.22%	
Total	22,572	100.00%	22,572

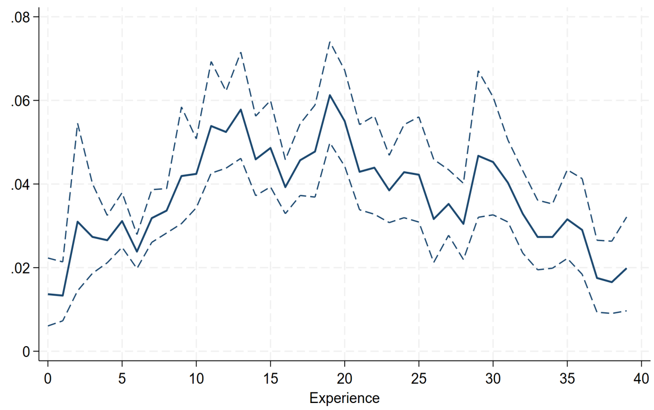
Notes: Panel A reports statistics on continuous variables such as age, experience, base salary, bonus, total compensation, and residual compensation obtained by projecting compensation on year effects, experience, and individual fixed effects. Total compensation is windsorized within each level of experience. All compensation measures are in thousands of 1988 dollars. Panel B reports frequencies for categorical variables including education level and race.

ment (supervisory) employees of a medium-sized U.S. firm in the service industry between 1969 and 1988; from 1981, information on both base and performance (bonus) pay is available. Since over this period supervisory positions were mostly filled by males, we exclude females from our analysis and retain 22,609 observations from 5,364 males in supervisory positions at the firm between 1981 and 1988. Being a sample of supervisory workers, as Table 1 shows, our sample mostly consists of white and highly educated individuals—total compensation for this group significantly exceeds that of the average worker.[4] Two other key features emerge from the table. First, performance pay makes up only about 4% of total pay, which mirrors the well-known fact in the personnel literature that for most workers, performance pay amounts to only a relatively small fraction of pay. Our analysis below elucidates why this is the case and why performance pay nevertheless is an important force governing the life-cycle profile of wages. Second, a large portion of the variation in pay can be accounted for by permanent differences among workers in terms of their observable characteristics. Specifically, the standard deviation of total pay in our data is $29.2K, but once we control for year effects, experience, and, crucially, individual fixed effects,

[4]Compensation is reported in 1988 dollars. Adjusted to 2025, average total compensation amounts to $161K.

the residual standard deviation is just $6.6K. We use this residualized measure of the variance of wages to discipline our model. In Section 6, we show how this relatively modest degree of wage dispersion nonetheless masks a high degree of income risk that workers face due to the uncertainty about their productivity.

Figure 1: Life-Cycle Ratio of Performance Pay to Total Pay in BGH Data



Our analysis is motivated by how the level and composition of pay varies over the life cycle, as firms learn about workers' unobserved skills and compensate them for their effort. Our data contain information about fixed f_{it} and variable (bonus) pay v_{it}, which together sum to worker i's pay or *wage* $w_{it} = f_{it} + v_{it}$ in period t. When variable pay v_{it} is proportional to output y_{it} and firms make on average zero profits, average variable pay is directly informative about the piece rate that links compensation w_{it} to performance y_{it}. Since these two conditions hold in our model, using that $\mathbb{E}[v_{it}] = \mathbb{E}[b_t y_{it}]$ and $\mathbb{E}[y_{it}] = \mathbb{E}[w_{it}]$, we can back out the piece rate b_t as $\mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$, which measures the sensitivity of pay to performance.[5]

Figure 1 shows that this ratio in the BGH data markedly varies as experience increases—it first increases and then decreases. In particular, the importance of performance pay relative to total pay *declines* over the second half of workers' careers: it eventually becomes 1/3 as large compared to its peak. This feature, which we document in multiple data sets as discussed in the appendix, runs counter to the core prediction of models of career concerns and performance incentives. According to them, as implicit reputational incentives for performance weaken over time, explicit incentives from performance pay should strengthen to induce the appropriate effort, leading performance pay to *increase* relative to total pay. By augmenting the standard model of career concerns and performance incentives with human capital acquisition, we propose in the next section a model

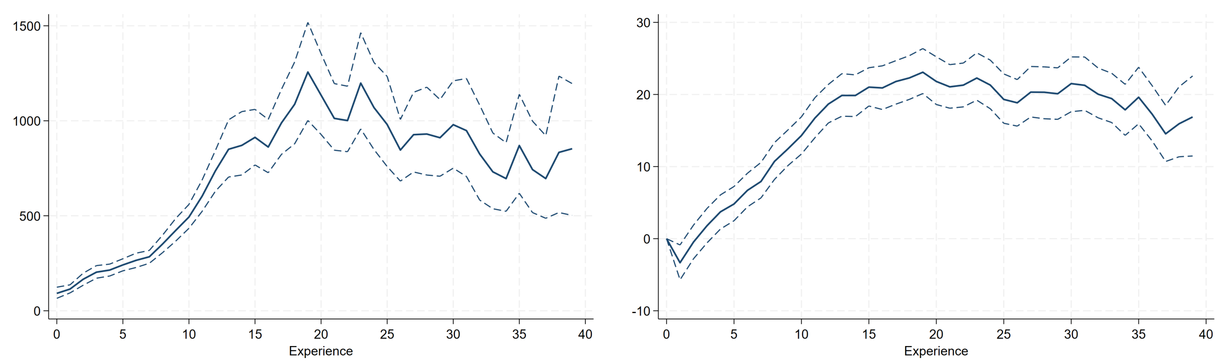
[5] A similar argument applies when wages are marked down relative to expected output; see the appendix.

that accounts for this hump-shaped pattern of performance pay together with the overall evolution of wages over the life cycle.[6] We discipline this model by targeting the experience profiles of the variance and growth of wages over the life cycle reported in Figure 2, which are *also* hump-shaped. We return to this point in Section 6, where we discuss the connection between the patterns of performance pay and of the variance of wages over the life cycle.

Figure 2: Life-Cycle Variance and Mean of Wages in BGH Data

(a) Life-Cycle Variance of Wages

(b) Life-Cycle Growth of Mean Wages



We now turn to present evidence supporting the notion that performance pay rewards effort on the job by documenting that performance pay strongly correlates with performance, as measured by the subjective ratings of performance assigned by managers to their team members.[7] To start, Kahn and Lange [2014] report that variation in performance induces lasting variation in pay in our data, which they conclude is consistent with the presence of learning about workers' ability. Intuitively, when learning is present, performance ratings, being a measure of realized output, persistently affect beliefs about ability and so present and future wages. Those authors further estimate a higher correlation of *pay* with recent ratings—1 to 3 years out—than with past ones, in line with ratings providing new signals about productivity each period that are used to update beliefs and thus wages. We additionally find that *variable pay* correlates more strongly with current ratings than with past or future ones, in line with the idea that bonus pay rewards contemporaneous performance. Taken together, these findings support the premise of our model that firms learn about workers' productivity and, consistent with the presence of moral hazard, link pay to performance.

Table 2 provides further evidence that moral-hazard considerations matter for pay based on the relationship between performance, performance pay, and the overall level of pay, once we account

[6]As argued in Section 6, a model of human capital acquisition alone would be silent about performance pay.

[7]See Kahn and Lange [2014] for an extensive analysis of the comovements between this performance measure and pay and Pastorino [2024] for an examination of how performance impacts wages and job assignment at the firm.

Table 2: Effect of Top Performance Rating on Pay Components

	Total Pay		Performance Pay (Bonus)		Base Pay	
	OLS	FE	OLS	FE	OLS	FE
Top Rating	11.36***	3.83***	2.73***	2.15***	9.03***	2.38***
	(0.46)	(0.18)	(0.15)	(0.18)	(0.46)	(0.17)
R^2	0.23	0.95	0.07	0.45	0.22	0.95
Observations	22,609	22,609	22,609	22,609	22,609	22,609

Notes: Standard errors in parentheses. All models include year effects and control for experience; the OLS models also include education effects and race whereas the FE models include individual worker fixed effects. Compensation is measured in thousands of 1988 dollars. Asterisks denote conventional significance levels.

for workers' observed and unobserved characteristics. Specifically, the table reports results from regressions of total pay, performance pay, and base (or fixed) pay on an indicator variable for whether a worker receives the top performance rating on a 5-point scale in any given year. All specifications control for year effects, experience, and race; the results labeled *OLS* also include education effects, whereas the results labeled *FE* additionally include individual fixed effects.

The OLS estimates show that performance measures are highly predictive of base and performance pay, as the estimates of the corresponding top rating coefficients are large and significant. This strong relationship between a top rating and both base and performance pay suggests that a close link exists not just between performance and performance pay, as a model of moral hazard would imply, but also between performance and base pay, as a model of learning would imply, in which firms use performance to update beliefs about a worker's ability, which influence base pay.

Once we control for permanent individual heterogeneity, the effect of a top rating on total pay and base pay becomes substantially *smaller*—compare the columns labeled FE with those labeled OLS—whereas its effect on performance pay remains largely *unchanged*. The positive FE estimate of the impact of a top rating on total pay and base pay implies that variation in performance ratings in a period is correlated with variation in pay and its main component in the same period even after *fixed* unobserved individual characteristics are accounted for, reinforcing the case for the presence of learning. That is, this evidence suggests that *time-varying* unobserved characteristics correlated with performance, as beliefs are, still have an impact on pay. That the FE estimate of the impact of a top rating on performance pay is close to the OLS estimate—unlike the corresponding FE estimate for base pay, which is much smaller than the OLS one—further indicates that any idiosyncratic variation in performance translates directly and sizably into variation in total pay through performance pay itself, just as predicted by a moral-hazard model.

Overall, these findings support the basic assumptions of a model of learning about ability and moral hazard, such as the one we develop next, for two main reasons. First, total pay, not just performance pay, systematically varies with performance—as consistent with a learning model, since performance affects beliefs about ability, which determine base pay. Second, even once permanent unobserved differences across workers are controlled for, performance in a period strongly correlates with performance pay in the period, suggesting that the BGH firm conditions variable pay on performance to incentivize workers—as consistent with a moral-hazard model.

3 A Model of Effort Incentives, Learning, and Human Capital

We now present the environment, define equilibrium, and discuss our setup.

3.1 Environment

The labor market consists of heterogeneous risk-averse workers and homogeneous risk-neutral firms that can freely enter the market.[8] Time is discrete and ranges from 0 to T. We index workers by i and time by t. Workers differ in their ability, which is subject to persistent shocks and is not observed by any market participant, including workers themselves. There are two types of tasks or activities that workers can perform in a firm, a *simple* task that requires observable and contractable effort and a *complex* task that requires unobservable and non-contractable effort—in the remarks below, we discuss how we interpret a worker's job as a bundle of these two tasks. Effort in both tasks augments output and influences human capital acquisition. All firms observe workers' output and employment contracts and infer a worker's unobserved ability based on this information. Since all firms share the same information, (employer) learning about workers' ability is common.

Production. The common output technology is such that worker i's output in period t is

$$y_{it} = \theta_{it} + \xi_k k_{it} + \xi_1 e_{i1t} + \xi_2 e_{i2t} + \varepsilon_{it}, \tag{1}$$

where θ_{it} is the worker's unobserved ability, k_{it} is the worker's human capital, e_{i1t} is the worker's effort in the simple task, e_{i2t} is the worker's effort in the complex task, and ε_{it} captures idiosyncratic variation in the worker's output. The parameter ξ_k describes the contribution of human capital to

[8]This labor market can be viewed as one of many sufficiently segmented by location, occupation, or industry that employment opportunities in other markets are irrelevant for workers' and firms' decisions in any given market.

output, which we can set to one without loss (see below), whereas the parameters ξ_1 and ξ_2 capture the contribution of each type of effort to output.[9] Worker i's initial ability θ_{i0} is drawn from a normal distribution with mean m_θ and variance σ_θ^2 and evolves over time according to the process $\theta_{it+1} = \theta_{it} + \zeta_{it}$, where ζ_{it} is an unobserved idiosyncratic shock realized at the end of t. The shocks ε_{it} and ζ_{it} are normally distributed with mean zero and variances σ_ε^2 and σ_ζ^2, respectively.

Human Capital. Human capital evolves according to the law of motion

$$k_{it+1} = \lambda k_{it} + \gamma_1 e_{i1t} + \gamma_2 e_{i2t} + \beta_t, \tag{2}$$

where $(1 - \lambda) \in [0, 1]$ is the depreciation rate, $k_{i0} \equiv k_0$ is the initial stock of human capital, γ_1 and γ_2 are, respectively, the rates at which effort in the simple and complex tasks affect future human capital, and β_t is a time-varying constant.[10] Note that we can set $\xi_k = 1$ in (1), since we can absorb it into γ_1, γ_2, and β_t, and redefine human capital accordingly—just consider the change of variables $\gamma_1 \mapsto \gamma_1/\xi_k$, $\gamma_2 \mapsto \gamma_2/\xi_k$, $\beta_t \mapsto \beta_t/\xi_k$, and $k_{it} \mapsto k_{it}/\xi_k$. This formulation of the human capital process encompasses the case in which the effort to acquire human capital *complements* the effort to produce output in both tasks ($\gamma_1, \gamma_2 > 0$), as in models of learning-by-doing, and the case in which the effort to acquire human capital *substitutes* for the effort to produce output in both tasks ($\gamma_1, \gamma_2 < 0$), as in models of learning-or-doing. In the former case, the investments in human capital in t are e_{i1t} and e_{i2t} with corresponding rates of human capital accumulation γ_1 and γ_2. In the latter case, the investments in t are $\overline{e}_{1t} - e_{i1t}$ and $\overline{e}_{2t} - e_{i2t}$ with corresponding rates of human capital accumulation $|\gamma_1|$ and $|\gamma_2|$, where $\overline{e}_{1t}$ and $\overline{e}_{2t}$ are a worker's endowment of time or efficiency units in the two tasks in t and we absorb $\gamma_1\overline{e}_{1t} + \gamma_2\overline{e}_{2t}$ into β_t. Cases in which the effort to acquire human capital complements the effort to produce output in one task and substitutes for it in the other task are also possible. We will refer to γ_1 and γ_2 as the rates of human capital accumulation. In the appendix, we consider more general formulations of the human capital process in which human capital differs unobservably across workers or depends nonparametrically on effort.

Worker Preferences. The lifetime utility from t on of a worker receiving wages $\{w_{t+\tau}\}_{\tau=0}^{T-t}$ and exerting the efforts $\{e_{1t+\tau}\}_{\tau=0}^{T-t}$ and $\{e_{2t+\tau}\}_{\tau=0}^{T-t}$ in the simple and complex tasks, respectively, is $-\exp\{-r \sum_{\tau=0}^{T-t} \delta^\tau [w_{t+\tau} - c(e_{1t+\tau}, e_{2t+\tau})]\}$, where $r > 0$ and δ are a worker's coefficient of (ab-

[9] That the coefficient ξ_θ multiplying θ_{it} in (1) is set to one is also without loss since it can be incorporated into θ_{it}.

[10] We can allow for heterogeneous initial stocks of human capital provided that they are observable.

solute) risk aversion and discount factor, respectively, and $c(e_1, e_2) = (\rho_1 e_1^2 + 2\eta e_1 e_2 + \rho_2 e_2^2)/2$ with $\rho_1, \rho_2 > 0$ is the monetary cost of the effort pair (e_1, e_2).[11] In what follows, we assume that $\rho_1 = \rho_2 = 1$ and $\eta = 0$, and consider the general case in the appendix.[12] In the supplementary appendix, we consider a version of our model in which both tasks feature non-contractable effort.

Contracts. Each period firms offer workers one-period employment contracts. A contract for worker i in period t is a pair (e_{i1t}, w_{it}) consisting of the worker's effort in the simple task, e_{i1t}, and wage schedule in the complex task, $w_{it} = c_{it} + b_{it} y_{it}$, where c_{it} is the fixed component of worker i's wage and b_{it} is worker i's *piece rate* in t.[13] We consider wage schedules that are linear in output for three reasons. First, this assumption is standard and so allows us to compare our framework to existing ones. Second, contracts are often linear in output or approximately so in practice. Third, linear contracts allow us to summarize the strength of contractual incentives for effort in the complex task through a simple one-dimensional continuous measure, the piece rate b_{it}.

3.2 Equilibrium

A worker's history in t consists of the sequence of the worker's effort choices in the complex task, employment contracts, and output realizations up to $t-1$. A strategy for a firm specifies contract offers to workers conditional on the public portion of their histories. A strategy for a worker specifies a choice of contract and effort in the complex task for each history for the worker and contract offers by firms. We consider pure-strategy perfect Bayesian equilibria. Free entry of firms implies that in equilibrium firms make zero expected profits each period. Thus, if (e_{i1t}, w_{it}) is worker i's equilibrium contract in period t when the public portion of the worker's history is I_{it}, then $c_{it} = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}]$, where $\mathbb{E}[y_{it}|I_{it}]$ is expected output in t given I_{it}. Hence,

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

and worker i's equilibrium contract in t can be described by the pair (e_{i1t}, b_{it}). By (1), $\mathbb{E}[y_{it}|I_{it}]$ depends on worker i's prescribed equilibrium behavior up to t, which pins down the worker's human capital and effort choices in t, and on worker i's conditional expected ability, $\mathbb{E}[\theta_{it}|I_{it}]$.

[11] Our arguments extend to a more general cost function. For reasons of identification, we treat it as known.

[12] There, we show that we can renormalize the model parameters so that $\rho_1 = \rho_2 = 1$ and key objects of interest such as equilibrium piece rates do not depend on η—since effort in the simple task is contractable, the provision of explicit incentives for effort in the complex task is not influenced by effort in the simple task.

[13] Gibbons and Murphy [1992] show that restricting attention to one-period contracts is equivalent to considering renegotiation-proof long-term contracts. Their proof extends to our environment.

3.3 Discussion

We conclude by discussing the main features of our model and dimensions along which it can be extended; see also the appendix and the supplementary appendix. To start, our model can be interpreted as the log version of one in levels with output and human capital technologies of the usual Cobb-Douglas form, namely, in which *i*) the period-t output of a worker with ability Θ_t and human capital K_t who exerts efforts E_{1t} in the simple task and E_{2t} in the complex task is $Y_t = \Theta_t K_t^{\xi_k} E_{1t}^{\xi_1} E_{2t}^{\xi_2} \Omega_t$, with ξ_k, ξ_1, and ξ_2 as in (1) and Ω_t a mean-one shock; and *ii*) the human capital in $t+1$ of a worker with human capital K_t in t who exerts efforts E_{1t} in the simple task and E_{2t} in the complex task in t is $B_t K_t^{\lambda} E_{1t}^{\gamma_1} E_{2t}^{\gamma_2}$, with B_t a positive time-varying constant and λ, γ_1, and γ_2 as in (2).

As in Gibbons and Murphy [1992], workers have constant absolute risk aversion preferences over present-discounted streams of wage payments, net of monetary effort costs. This preference specification, which is common in models of dynamic moral hazard for its tractability, allows us to abstract from wealth effects. Since, as is also common in the literature, output is linear in inputs, wages are linear in output, shocks to ability are additive, and initial ability, ability shocks, and output shocks are normally distributed, worker preferences admit a mean-variance representation. This feature, in turn, implies that a worker's trade-off between consumption or wages and leisure does not depend on a worker's history, which enables us to completely characterize equilibrium.

Our model extends existing dynamic moral-hazard models by allowing for multiple worker activities or tasks so as to micro-found the notion of a worker's job and the resulting assignment process based on the responsibilities that a job entails, which may be contractable to different degrees and can vary with a worker's experience. For instance, as discussed in Section 6.5, workers in the BGH data progress over time to more complex jobs for which general management duties—such as general administration or planning—that require workers to perform activities difficult to contract become increasingly more important. Simpler activities that are easier to contract—such as creating or selling products—correspondingly decrease in importance.

Our model nests well-known models of learning about ability, human capital accumulation, and performance incentives. When ability is known and there exists a single task requiring contractable effort ($\xi_2 = \gamma_2 = 0$), our model reduces to one of dynamic labor supply and human capital accumulation through investments that can complement or substitute for the effort expended to produce output. When effort is not a choice variable, the model specializes to one of human capital acquisition with experience through learning-by-doing (if $\gamma_1 > 0$) or learning-or-doing (if $\gamma_1 < 0$). When there exists a single task requiring non-contractable effort ($\xi_1 = \gamma_1 = 0$), effort does not contribute to human capital ($\gamma_2 = 0$), and ability is not subject to shocks ($\sigma_\zeta^2 = 0$), the model simplifies to the career-concerns model with explicit incentives of Gibbons and Murphy [1992]. Without performance pay, the model further reduces to the career-concerns model of Holmström [1999]. When, in addition, effort is not a choice variable, our model is a symmetric learning model with ability general across firms as in Farber and Gibbons [1996].[14]

As for the type of labor market competition that we incorporate, note that our analysis applies essentially unaltered if instead of capturing the entire surplus from their matches with firms, workers capture only a fraction of it, so that wages are marked down relative to workers' output; see the appendix. Also, by reinterpreting the term β_t in the law of motion for human capital in (2) as a firm productivity parameter, our model extends to settings in which firms differ in their productivity. To see how, suppose that firms are characterized by a productivity level p so that the output of worker i in t when employed by a firm of productivity $p = p_{it}$ is $y_{it} = p_{it} + \theta_{it} + \xi_k k_{it} + \xi_1 e_{i1t} + \xi_2 e_{i2t} + \varepsilon_{it}$—in our baseline model with homogeneous firms, p is absorbed in β_t. Assume that in each period, a worker is matched with a set of heterogeneous firms that Bertrand-compete for workers; see Pastorino [2024]. If the firm at which a worker is most productive changes over time, say, because of productivity shocks, then the wage equation in our model is analogous to that in Bagger et al. [2014]. We do not explicitly consider such heterogeneity in our analysis just for simplicity, as we use data from one firm in our empirical exercises.

4 Learning about Ability and Effort in the Complex Task

In this section, we present key results that allow us to characterize the equilibrium of the model. We first describe the process of learning about ability. We then determine a worker's choice of effort in the complex task—the task with non-contractable effort—for given employment contracts and specify how it depends on career-concerns and human capital incentives.

[14]When firms can commit to long-term contracts without performance pay, our framework extends that of Harris and Holmström [1982] to a setting with moral hazard and human capital acquisition.

4.1 Learning about Ability

Firms and workers learn about a worker's ability over time by observing a worker's output. Consider worker i in period t, whose equilibrium effort choices and human capital in t are e_{1t}^*, e_{2t}^*, and k_t^*, respectively—for simplicity, we omit the dependence of effort choices and human capital on i. Let $z_{it} = y_{it} - k_t^* - \xi_1 e_{1t}^* - \xi_2 e_{2t}^*$ be the portion of the worker's output in t that is not explained by the worker's human capital and efforts. By (1), $z_{it} = \theta_{it} + \varepsilon_{it}$ is the signal about the worker's ability in t extracted from the worker's output. Since initial ability and shocks to ability and output are normally distributed, it follows that posterior beliefs about a worker's ability in any period are normally distributed and so fully described by their mean $m_{it} = \mathbb{E}[\theta_{it}|I_{it}]$—worker i's *reputation* in t—and variance $\sigma_{it}^2 = \mathrm{Var}[\theta_{it}|I_{it}]$, with $m_{i0} = m_\theta$ and $\sigma_{i0}^2 = \sigma_\theta^2$. By standard results,

$$m_{it+1} = \sigma_\varepsilon^2 m_{it}/(\sigma_{it}^2 + \sigma_\varepsilon^2) + \sigma_{it}^2 z_{it}/(\sigma_{it}^2 + \sigma_\varepsilon^2) \text{ and } \sigma_{it+1}^2 = \sigma_{it}^2 \sigma_\varepsilon^2/(\sigma_{it}^2 + \sigma_\varepsilon^2) + \sigma_\zeta^2. \quad (4)$$

The recursions for m_{it} and σ_{it}^2 describe how a worker's reputation and its variability, as captured by the variance of posterior beliefs about a worker's ability, evolve over time.[15] Since σ_{it}^2 evolves independently of z_{it} and so is common to all workers, we can suppress the subscript i and simply denote this variance by σ_t^2. By iterating on (4), we can trace out 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 sequence $\{a_k\}$, worker i's reputation in $t+\tau$ with $1 \le \tau \le T-t$ given reputation m_{it} in t is

$$m_{it+\tau} = \left(\prod_{k=0}^{\tau-1} \mu_{t+k}\right) m_{it} + \sum_{s=0}^{\tau-1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right) (1 - \mu_{t+s}) z_{it+s}. \quad (5)$$

4.2 Effort in the Complex Task

As it turns out, the equilibrium is unique, symmetric, and such that effort choices and piece rates depend only on time. To understand workers' problem and the determinants of their effort choices in the complex task, suppose that workers face a sequence of employment contracts $\{(e_{1t}, b_t)\}_{t=0}^T$ such that efforts in the simple task and piece rates depend only on time. Consider worker i's period-t choice of effort in the complex task, e_{2t}, when the worker's future effort choices in this task depend only on time. Let $w_{it+\tau}$ be worker i's wage in period $t+\tau$ with $0 \le \tau \le T-t$.

[15] These expressions are valid even when effort in the complex task deviates from the equilibrium path as any output realization is possible for any given effort choice, so firms cannot infer effort in the complex task from realized output.

Worker i chooses e_{2t} to maximize $U_{it}(e_{2t}) = \mathbb{E}[-\exp\{-r[W_{it} - c(e_{1t}, e_{2t})]\}|h_i^t]$, where $W_{it} = \sum_{\tau=0}^{T-t} \delta^\tau w_{it+\tau}$ and the expectation in $U_{it}(e_{2t})$ is conditional on worker i's period-t history h_i^t. Yet, as we will see, the optimal choice of e_{2t} is independent of h_i^t; it is also independent of e_{1t}. Since signals about ability are normally distributed, it also follows from (3) and (5) that wages, $\{w_{it+\tau}\}_{\tau=0}^{T-t}$, are normally distributed, and so is their present-discounted value W_{it}. Thus, e_{2t} maximizes $U_{it}(e_{2t})$ if, and only if, it maximizes $\mathbb{E}[W_{it}|h_i^t] - r\text{Var}[W_{it}|h_i^t]/2 - e_{2t}^2/2$.[16] As for the first-order condition for effort in the complex task, note that $\partial\mathbb{E}[w_{it}|h_i^t]/\partial e_{2t} = \xi_2 b_t$ by (1) and (3). But worker i's choice of e_{2t} also affects future wages through its impact on the worker's future reputation—which affects the *fixed* component of future pay—and future human capital—which affects both the *fixed* and *variable* components of future pay. As $\mathbb{E}[W_{it}|h_i^t] = \sum_{\tau=0}^{T-t} \delta^\tau \mathbb{E}[w_{it+\tau}|h_i^t]$ and, as shown in the appendix (see the discussion after Lemma A.1), effort in the complex task does not affect the variance of future pay, the first-order condition for worker i's choice of e_{2t} reduces to[17]

$$e_{2t} = \xi_2 b_t + \sum_{\tau=1}^{T-t} \delta^\tau \partial\mathbb{E}[w_{it+\tau}|h_i^t]/\partial e_{2t}. \tag{6}$$

The right side of (6), which describes the marginal benefit of effort in the complex task in t, is the sum of two terms. The first term captures the *static* marginal benefit of effort. The second term captures its *dynamic* marginal benefit, which is nonzero as long as $t < T$ and can be further manipulated as follows. Let $\widehat{\mu}_{t,\tau} = (\prod_{k=1}^{\tau-1} \mu_{t+\tau-k})(1 - \mu_t)$ and define $R_{CC,t}$ and $R_{HK,t}$ as

$$R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau})\widehat{\mu}_{t,\tau} \text{ and } R_{HK,t} = \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(b_{t+\tau} + R_{CC,t+\tau}). \tag{7}$$

In the appendix, we show that we can express the first-order condition in (6) as

$$e_{2t} = \xi_2 b_t + \xi_2 R_{CC,t} + R_{HK,t}. \tag{8}$$

The terms $\xi_2 R_{CC,t}$ and $R_{HK,t}$ describe the dynamic marginal benefit of effort in the complex task arising from its effect on a worker's future reputation and human capital.[18] To understand $\xi_2 R_{CC,t}$, note that at the margin, a higher e_{2t} increases the expected period-t signal z_{it} about a worker's

[16] Recall that $\mathbb{E}[\exp\{rX\}] = \exp\{r\mu - r^2\sigma^2/2\}$ if X is normally distributed with mean μ and variance σ^2.

[17] As in Gibbons and Murphy [1992], we allow for negative effort so as to be able to use first-order conditions to characterize workers' effort choices. We later provide conditions under which effort is positive.

[18] By (8), effort in the complex task is identical across workers. This fact is key for the symmetry of equilibrium and for the result that equilibrium piece rates, and thus effort choices, depend only on t.

ability by ξ_2. By (5), this increase raises a worker's expected reputation in $t+\tau$, with $1 \leq \tau \leq T-t$, by $\xi_2 \widehat{\mu}_{t,\tau}$. In turn, at the margin, a higher reputation in $t+\tau$ increases the *fixed* component of the wage in $t+\tau$ by $1-b_{t+\tau}$. The term $R_{CC,t}$ is just the present-discounted value of all these marginal increases. Therefore, even without 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 and increase their future fixed pay.

To understand the term $R_{HK,t}$, observe that worker i's choice of effort in period t directly affects the variable component of the worker's wage in all subsequent periods by affecting the worker's stock of human capital and thus output in each such period. By changing the worker's stock of human capital, effort in period t additionally affects future output signals about the worker's ability, and so the worker's future reputation and fixed pay. To elaborate, note that at the margin, a higher e_{2t} changes worker i's output in $t+\tau$ by $\gamma_2 \lambda^{\tau-1}$, which amounts to the change in the stock of human capital in $t+\tau$. This change in output affects the *variable* component of the wage in $t+\tau$ by $b_{t+\tau}\gamma_2\lambda^{\tau-1}$. It also affects the magnitude of the signal about ability in $t+\tau$ by $\gamma_2\lambda^{\tau-1}$, which, by the same argument as for $R_{CC,t}$, increases the present-discounted value of the fixed component of the wages from $t+\tau$ on by $\gamma_2\lambda^{\tau-1}R_{CC,t+\tau}$. Hence, greater incentives for effort from career concerns also imply greater incentives for effort from human capital acquisition. The term $R_{HK,t}$ is simply the present-discounted value of all these marginal changes.

5 Equilibrium and Identification

We now characterize the equilibrium, describe the implied life-cycle patterns of piece rates and effort, and establish that the model is identified from panel data on wages and their structure.

5.1 Equilibrium Characterization

Recall that $\mu_t = \sigma_\varepsilon^2/(\sigma_t^2 + \sigma_\varepsilon^2)$. Let $\{\sigma_t^2\}_{t\geq 0}$, which describes the uncertainty about a worker's ability in any period t, be such that $\sigma_0^2 = \sigma_\theta^2$ and $\sigma_{t+1}^2 = \mu_t\sigma_t^2 + \sigma_\zeta^2$. Since output signals do not perfectly reveal ability, this uncertainty persists over time and converges to a non-negative value σ_∞^2, which is positive if $\sigma_\zeta^2 > 0$.[19] The next result summarizes our equilibrium characterization.

[19] The variance σ_t^2 monotonically decreases to σ_∞^2 if $\sigma_\theta^2 > \sigma_\infty^2$ and monotonically increases to σ_∞^2 if $\sigma_\theta^2 < \sigma_\infty^2$. See Holmström [1999] for a proof that $\sigma_\infty^2 = [\sigma_\zeta^2 + (\sigma_\zeta^4 + 4\sigma_\zeta^2\sigma_\varepsilon^2)^{1/2}]/2$ and for the properties of σ_t^2.

Proposition 1. *In the unique equilibrium, piece rates and effort choices are the same for all workers and depend only on time. Let e^*_{1t} and e^*_{2t} be, respectively, the equilibrium efforts in the simple and complex task in period t and let b^*_t be the equilibrium piece rate in the same period. For each t, let $b^0_t = 1/[1 + (r/\xi_2^2)(\sigma_t^2 + \sigma_\varepsilon^2)]$, $R^*_{CC,t}$ and $R^*_{HK,t}$ be given by (7) with $b_t = b^*_t$ for all t, and $H^*_t = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$. Then, $e^*_{1t} = \xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, $e^*_{2t} = \xi_2 b^*_t + \xi_2 R^*_{CC,t} + R^*_{HK,t}$, and*

$$b^*_t = b^0_t \left[1 + (\gamma_2/\xi_2) \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (1/\xi_2) R^*_{HK,t} - R^*_{CC,t} - (r/\xi_2^2) H^*_t\right]. \tag{9}$$

As discussed in Section 4, effort in the complex task in any period t, e^*_{2t}, equates its marginal cost to its marginal private benefit, which features a static and a dynamic component. The latter arises from the impact of effort in the complex task on a worker's future reputation and human capital. By contrast, in any period t, effort in the simple task equates its marginal cost to its marginal social (output) benefit, $\xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$.[20] In the appendix, we extend Proposition 1 to the more general case in which the law of motion of human capital depends nonparametrically on effort. For simplicity, in this case we assume that workers perform only the complex task.

To understand the expression in (9), first note that b^0_t is the piece rate of canonical static linear-normal models of incentives with quadratic effort costs when the variance of output is $\sigma_t^2 + \sigma_\varepsilon^2$ and the coefficient of risk aversion is r/ξ_2^2—the variance of output is σ_ε^2 in standard moral-hazard models, as ability is assumed to be known. By (8), the term $1 + (\gamma_2/\xi_2) \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (1/\xi_2) R^*_{HK,t} - R^*_{CC,t}$ is the piece rate that equates the marginal cost of effort in the complex task to its marginal social benefit in period t, $\gamma_2 + \xi_2 \sum_{\tau=1}^{T-1} \delta^\tau \lambda^{\tau-1}$. As is well known, when effort is unobserved, equilibrium piece rates deviate from first-best piece rates since risk-averse workers are unwilling to bear all output risk.[21] In a static setting, this distortion results in piece rates being adjusted by the factor $b^0_t < 1$. In our dynamic setting, an additional distortion arises because of the risky process through which learning about ability occurs: any variation in output in $t < T$ leads to variation not only in wages in t but also in future wages, as the latter depend on a worker's reputation, which changes with a worker's realized output. The insurance (or hedge) term $(r/\xi_2^2)H^*_t$ mitigates this risk by reducing the correlation between a worker's performance and pay whenever $t < T$.[22]

[20] Note that e^*_{1t} is positive if $\gamma_1 \geq 0$. When $\gamma_1 < 0$, e^*_{1t} is positive if $\gamma_1 > \xi_1(\lambda - 1/\delta)$ and e^*_{10} is positive. By the expressions for $R^*_{CC,t}$ and $R^*_{HK,t}$, e^*_{2t} is positive if $\gamma_2 > 0$ and piece rates are between zero and one. Also, if piece rates are strictly positive and bounded above by one, then e^*_{2t} is positive even if $\gamma_2 < 0$ as long as $|\gamma_2|/\xi_2$ is small.

[21] Indeed, $b^*_T = 1$ if $r = 0$. So, $R^*_{CC,T-1} = 0$ and $R^*_{HK,T-1} = (\gamma_2/\xi_2)\delta$, so $b^*_{T-1} = 1$. By induction, $b^*_t = 1$, $R^*_{CC,t} = 0$, and $R^*_{HK,t} = (\gamma_2/\xi_2) \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ for all t, so implicit incentives for effort arise solely from human capital.

[22] That the term $(r/\xi_2^2)H^*_t$ depends only on σ_t^2 and not on total output risk $\sigma_t^2 + \sigma_\varepsilon^2$ follows from the fact that the

By re-arranging (9), equilibrium piece rates can be expressed as

$$b_t^* = b_t^0 - b_t^0 R^*_{CC,t} - b_t^0 (r/\xi_2^2) H_t^* + (b_t^0/\xi_2) \left(\gamma_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R^*_{HK,t} \right). \tag{10}$$

This decomposition helps illustrate how the economic forces nested by our model shape the provision of explicit incentives for effort in the complex task over time. The first term in (10) is the piece rate that firms would offer in a static setting, whereas the second and third terms in (10) capture the contribution of uncertainty and learning about ability to piece rates and are familiar from Gibbons and Murphy [1992]. Namely, the second term lowers the explicit incentives for effort provided by piece rates in light of the implicit reputational incentives resulting from the uncertainty about ability. As discussed, the third term in (10) lowers piece rates to provide workers with insurance against the life-cycle wage risk due to the variability of beliefs about ability as learning takes place.

The last term in (10), which captures the effect of human capital on piece rates, is novel and consists of two further terms. The first is proportional to $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, which is the present-discounted change in lifetime output following the change in human capital after a marginal increase in effort in the complex task in t. The second term, which is negatively proportional to $R^*_{HK,t}$, reflects the implicit incentives for effort from the prospect of human capital acquisition, which substitute for explicit incentives, and so decrease piece rates, when γ_2 is greater than zero, and complement explicit incentives, and so increase piece rates, when γ_2 is smaller than zero. Intuitively, in this latter case, e_{2t} negatively contributes to human capital acquisition, which discourages workers from exerting effort to produce output—higher piece rates help support this effort. This last term in (10) simply corrects piece rates to better align the private marginal returns to effort in the complex task from human capital with the corresponding social returns.

5.2 Piece Rates and Effort over the Life Cycle

We now discuss how learning about ability and human capital acquisition affect the life-cycle profile of piece rates and effort choices. We first consider the cases in which either human capital acquisition or learning about ability are not present, which leads to counterfactual implications for piece rates, and then turn to the general case, which is consistent with the data.

life-cycle wage risk due to learning about ability is primarily due to the correlation between current and future wages through a worker's ability, whereas output shocks—the other source of output risk—are purely idiosyncratic.

Learning and Moral-Hazard Case. Suppose that $\gamma_1 = \gamma_2 = 0$ so workers do not accumulate human capital. Furthermore, assume that $\xi_2 = 1$, which is without loss since the case with $\xi_2 \neq 1$ is equivalent to the case with $\xi_2 = 1$ and the coefficient of risk aversion $r' = r/\xi_2^2$. Then, (9) becomes $b_t^* = b_t^0(1 - R_{CC,t}^* - rH_t^*)$. This setup generalizes the model in Gibbons and Murphy [1992]—under the assumption of quadratic effort cost—in two ways. First, we endogenize job assignment by allowing workers to perform two tasks, one requiring contractable effort, the simple task, and one requiring non-contractable effort, the complex task, whereas in Gibbons and Murphy [1992] workers perform only one task that requires non-contractable effort. Second, unlike Gibbons and Murphy [1992], we allow ability to stochastically change over time. As in Gibbons and Murphy [1992], the insurance against life-cycle wage risk provided by the term rH_t^* can be strong enough that piece rates are negative. This is the case early in a career if T is large and δ is close to one.[23]

Since equilibrium piece rates do not depend on either ξ_1 or γ_1, they reduce to the ones in Gibbons and Murphy [1992] when $\sigma_\zeta^2 = 0$ and ability is constant over time. To understand how shocks to ability affect piece rates, note that a worker's career-concerns incentive to exert effort in the complex task increases not only with the uncertainty about the worker's ability but also with the worker's time horizon—the shorter this horizon, the smaller the gain from a higher reputation, and so the smaller the return from effort in the complex task. When ability is constant, and so uncertainty about ability decreases monotonically to zero over time, the two forces shaping implicit incentives for effort in the complex task—the degree of uncertainty about ability and the length of the remaining working horizon—work in the same direction and weaken over time. Gibbons and Murphy [1992] show that in this case, firms compensate for the decline in the implicit incentives for effort by increasing the strength of explicit incentives over time. In the appendix, we show that the same logic applies when $\sigma_\theta^2 \geq \sigma_\infty^2$ and uncertainty about ability decreases over time. When, instead, $\sigma_\theta^2 < \sigma_\infty^2$ and uncertainty about ability increases over time, the two forces shaping implicit incentives for effort move in opposite directions. However, if the working horizon is long enough, then at some point the only force governing the evolution of piece rates is the decrease in the working horizon, as uncertainty about ability eventually becomes constant (σ_t^2 converges to σ_∞^2). We thus have the following result.

[23] Also as in Gibbons and Murphy [1992], piece rates are smaller than one. Indeed, $b_t^* < b_t^0$ if $R_{CC,t}^* > 0$. Moreover, $b_T^* < 1$ implies that $R_{CC,T-1}^* > 0$. An induction argument then shows that $R_{CC,t}^* > 0$ for all t.

Lemma 1. *Piece rates eventually strictly increase over time if T is large enough. Moreover, piece rates strictly increase over time if $\sigma_\theta^2 \geq \sigma_\infty^2$.*

Consider now how workers' effort choices and, correspondingly, their task allocation vary over the life cycle. When $\gamma_1 = 0$, effort in the simple task, e_{1t}^*, is constant over time. Since $e_{2t}^* = b_t^* + R_{CC,t}^*$, the life-cycle profile of effort in the complex task is ambiguous, though. When $\sigma_\theta^2 \geq \sigma_\infty^2$, piece rates strictly increase over time as $R_{CC,t}^*$ strictly decreases. A similar tension arises when $\sigma_\theta^2 < \sigma_\infty^2$. Thus, a priori, workers' task allocation can change in different ways over the life cycle. When $R_{CC,t}^*$ is small for all t—the empirically relevant case, as we will discuss—the life-cycle pattern of effort in the complex task is determined by the pattern of piece rates. In this case, workers progress to more complex tasks over time in the sense that $e_{2t}^* - e_{1t}^*$ strictly increases with t whenever $\sigma_\theta^2 \geq \sigma_\infty^2$.[24]

Human Capital and Moral-Hazard Case. Suppose now that $\sigma_\theta^2 = \sigma_\zeta^2 = 0$ so there exists no uncertainty about workers' ability. In this case, $b_t^0 \equiv b^0 = 1/[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$ and (9) reduces to $b_t^* = b^0[1 + (\gamma_2/\xi_2)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(1 - b_{t+\tau}^*)]$. Piece rates thus vary over time only because of firms' desire to influence workers' accumulation of human capital. This motive contributes positively to piece rates when human capital is acquired through learning-by-doing, that is, $\gamma_2 > 0$, and piece rates are smaller than one, which holds if γ_2 is not too large.[25] Indeed, when the effort to produce output in the complex task complements the effort to acquire human capital and piece rates are smaller than one, workers do not fully capture the returns to their investments in human capital, $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, and so their willingness to exert effort in the complex task is reduced. Piece rates partially offset this undersupply of effort. More generally, piece rates are positive if $\gamma_2 \geq \xi_2(\lambda - 1/\delta)$. Hence, even when the effort to produce output in the complex task and the effort to acquire human capital are rival, that is, $\gamma_2 < 0$, it is optimal to induce workers to exert more effort for human capital reasons if the trade-off between output and human capital production is not too severe.

The sign of γ_2 also determines the evolution of piece rates over time. When $\gamma_2 < 0$, piece rates strictly *increase* over time. By contrast, when $\gamma_2 > 0$, piece rates strictly *decrease* over time if γ_2

[24]In Section 6.5, we define the task complexity of a worker's job in period t as $(1 + e_{2t}^*)/(1 + e_{1t}^*)$ or, equivalently, as $\ln((1 + e_{2t}^*)/(1 + e_{1t}^*))$. When e_{1t}^* and e_{2t}^* are not too large, as we estimate, the pattern of task complexity over time is governed by the pattern of $e_{2t}^* - e_{1t}^*$—just note that $\ln(1 + e) \approx e$.

[25]That piece rates can be greater than one when γ_2 is positive and large follows from b_t^* linearly increasing with γ_2 whenever $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(1 - b_{t+\tau}^*)$ is positive.

is not too large. Intuitively, firms wish to encourage human capital acquisition early in a worker's career, when the return from doing so is largest. When the effort to produce output in the complex task substitutes for the effort to acquire human capital, firms can do so by discouraging effort in the complex task early on. On the contrary, when the effort to produce output in the complex task complements the effort to acquire human capital, firms support human capital acquisition by encouraging effort in the complex task early on. The reason why γ_2 cannot be too large for this latter result to hold is that equilibrium piece rates in one period decrease with equilibrium piece rates in the following period when γ_2 is positive.[26] Since $b^*_{T-1} = b^0[1 + (\gamma_2/\xi_2)\delta(1 - b^0)]$ linearly increases with γ_2, then $b^*_{T-2} < b^*_{T-1}$ if γ_2 is above a certain threshold. In this case, piece rates oscillate over time in that $b^*_{T-1} > b^*_T$, $b^*_{T-2} < b^*_{T-1}$, $b^*_{T-3} > b^*_{T-2}$, and so on.

Lemma 2. *There exists $\overline{\gamma}_2 > 0$ such that $b^*_t \in (0, 1)$ for all t if $\xi_2(\lambda - 1/\delta) \leq \gamma_2 \leq \overline{\gamma}_2$. Moreover, piece rates strictly increase over time when $\gamma_2 < 0$, strictly decrease over time when $0 < \gamma_2 < \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$, and (weakly) oscillate over time otherwise.*

As for how efforts in the two tasks evolve over time, since $e^*_{1t} = \xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, effort in the simple task strictly decreases over time if $\gamma_1 > 0$ and strictly increases over time if $\gamma_1 < 0$. Given that $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} b^*_{t+\tau} = \xi_2 + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - \xi_2(b^*_t/b^0)$ by (9), effort in the complex task e^*_{2t} equals $\xi_2 + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_2)\sigma_\varepsilon^2 b^*_t$, which is the socially optimal level of effort in this task net of a term proportional to piece rates.[27] When piece rates are small, as in the data, the life-cycle profile of effort in the complex task is largely shaped by the life-cycle profile of the term $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$. In this case, whether $e^*_{2t} - e^*_{1t}$ increases or decreases over time depends on whether γ_2 is smaller or greater than γ_1: workers progress towards more complex tasks in the first case and towards simpler tasks in the second case. In our data, we find the opposite pattern—namely, task complexity increases with experience and $\gamma_2 > \gamma_1$—which lends support to the general case we consider next.

General Case. When uncertainty and learning about ability and human capital acquisition are both present, naturally the stronger of these two forces shapes the experience profile of piece rates.

[26] Intuitively, when $\gamma_2 > 0$, an increase in piece rates in subsequent periods increases the return to investments in human capital in the current period, thus reducing the need to incentivize effort in the current period.

[27] It might appear counterintuitive that effort in the complex task decreases with piece rates. Note, however, that piece rates help align the private and social marginal returns to effort in the complex task. Then, it is precisely when workers' incentives to exert effort are low that piece rates are high.

For instance, when shocks to ability are small enough that ability is effectively known in the long run, human capital incentives eventually govern piece rates provided that the working horizon is long enough. Intuitively, at some point the residual uncertainty about ability becomes so small that learning about it no longer matters for the evolution of piece rates. Thus, towards the end of workers' career, piece rates strictly decrease over time when $0 < \gamma_2 < \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$ and oscillate over time when $\gamma_2 > \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$.[28] By contrast, when not much human capital is acquired in the complex task, learning about ability shapes the life-cycle profile of piece rates. In particular, when the working horizon is long enough, piece rates eventually strictly increase over time. The next proposition summarizes this discussion.[29]

Proposition 2. *For a fixed $\gamma_2 > 0$, piece rates either eventually strictly decrease over time, when $\gamma_2 < \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$, or eventually oscillate over time, when $\gamma_2 > \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$, provided that σ_ζ^2 is small enough and T is sufficiently large. By contrast, piece rates eventually strictly increase over time if $|\gamma_2|$ is small enough and T is sufficiently large.*

An implication of Proposition 2 is that as long as the working horizon is long enough, the fact that piece rates decrease with experience as workers approach the end of their careers suggests that human capital acquisition matters for the complex task. Furthermore, when shocks to ability are small—as we estimate—Proposition 2 and the fact that piece rates eventually decrease with experience suggest that the effort to produce output complements the effort to acquire human capital in the complex task.

Another consequence of Proposition 2 is that when shocks to ability are sufficiently small and the working horizon is long enough, piece rates are not maximized at the end of a worker's career when the rate at which effort in the complex task increases human capital is positive but not too large. If, in addition, the initial uncertainty about ability is not too small and piece rates are between zero and one, then piece rates are not maximized at the start of a worker's career either. Indeed, if piece rates are between zero and one, then both the career-concerns and human capital motive contribute negatively to piece rates. This, in turn, implies that when the initial uncertainty

[28]For ease of exposition, we ignore the knife-edge case in which $\gamma_2 = \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2)]$. In this case, piece rates are eventually approximately constant if σ_ζ^2 is sufficiently small and T is large enough.

[29]In light of Lemmas 1 and 2, a natural conjecture is that piece rates eventually strictly increase over time when $\gamma_2 < 0$, since in this case learning and human capital acquisition influence piece rates in the same way in the long run. We show in the appendix that this is true when human capital depreciation is small.

about ability is not too small, the insurance against life-cycle wage risk at the start of a worker's career is strong enough to make piece rates lower than the static ones, and thus lower than the last-period piece rate. Therefore, there exist conditions under which piece rates are maximized at an intermediate level of experience—namely, they display a hump-shaped pattern with experience, as we observe in the data.

Proposition 3. *Let $0 < \gamma_2 < \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$ and suppose that piece rates are between zero and one. Then, piece rates are maximized at an intermediate level of experience if σ_θ^2 is sufficiently large, σ_ζ^2 is sufficiently small, and T is sufficiently large.*

As discussed, that piece rates eventually strictly decrease with experience suggests that human capital acquisition matters for effort in the complex task. Since, by continuity, σ_θ^2 and σ_ζ^2 small imply that the experience profile of piece rates is shaped by human capital considerations, which, by Lemma 2, cannot generate a hump-shaped profile, such a profile in the data also suggests that uncertainty and learning about ability matters for piece rates.

We conclude this section by discussing how workers' effort choices vary over the life cycle in the general case. Since the life-cycle profile of effort in the simple task depends only on the sign of γ_1, the discussion in the human capital and moral-hazard case applies here without change. As for effort in the complex task, first note from (9) that $\xi_2(b_t^*/b_t^0) = \xi_2 + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^* - \xi_2 R_{CC,t}^* - (r/\xi_2)H_t^*$. Thus, since $\xi_2 R_{CC,t}^* + R_{HK,t}^* = e_{2t}^* - \xi_2 b_t^*$, it follows that $e_{2t}^* = \xi_2 + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_2)[(\sigma_t^2 + \sigma_\varepsilon^2)b_t^* + H_t^*]$, and the expression for effort in the complex task is similar to that in the human capital and moral-hazard case, except that b_t^* is now multiplied by $\sigma_t^2 + \sigma_\varepsilon^2$ and an additional negative term proportional to H_t^* appears. Intuitively, the life-cycle wage risk due to uncertainty and learning about ability further depresses effort in the complex task relative to the first-best level. Since H_t^* strictly decreases with t, the additional negative term slows down the decrease of effort in the complex task over time compared to the human capital and moral-hazard case. Hence, unlike in that case, when piece rates are small, workers progress over time to more complex tasks even when γ_2 is greater than γ_1, precisely as we estimate.

5.3 Identification

To measure the importance of the incentives we focus on using our framework, we must first establish its empirical content. The challenge in doing so is that the competing mechanisms it features are all unobserved. Key to our approach is exploiting the information provided by the life-cycle profile of wages and their fixed and variable components, which we next argue is sufficient to recover the primitives of our model up to a level and scale normalization.

Intuitively, as in any learning model, the experience profile of the variance of wages is informative about the evolution of the uncertainty about workers' ability and so the learning process about it. In our setting, the experience profile of both mean wages and, as illustrated in Section 5.2, piece rates is revealing of the importance of workers' human capital process. The pattern of piece rates in particular is central to pinning down the distinct mechanisms of our model. As discussed, piece rates initially increase with experience *only* if learning about ability is important early in a worker's career. They are also positive early on, and declining later on, *only* if human capital acquisition occurs, whenever uncertainty about ability is substantial over the life cycle—as is the case in our data. Since workers' risk preferences have a direct effect on the level of piece rates by (9), piece rates further provide crucial information to recover them.

Formally, we prove that our model is identified based on panel data on wages and their fixed or variable components. The key steps of the argument are as follows. The ratio of variable pay to total pay identifies piece rates at each year of experience. With piece rates known, intuitively, the second moments of the distributions of wages at each experience pin down the distributions of initial ability and of ability and output shocks, which completely determine the learning process. Then, differences in mean wages and piece rates over time identify the degree of human capital depreciation. Once these primitives are recovered, by exploiting our characterization of piece rates, we show that workers' degree of risk aversion can be recovered from piece rates in the last period (T) and the rates of human capital accumulation in the complex task from piece rates in the previous periods.[30] Finally, the rate of human capital accumulation in the simple task and the drift terms $\{\beta_t\}$ are residually determined from mean wages. In establishing these results, we treat the discount factor and the sensitivity of output to effort in the simple and complex tasks as known.[31] We discuss in Section A.9 in the appendix how the latter restriction can be relaxed. Since we can absorb k_0 into m_θ, which we normalize, we set $k_0 = 0$.[32]

[30]The model provides a natural exclusionary restriction in that dynamic and static piece rates coincide in period T.

[31]See Margiotta and Miller [2000], Gayle and Miller [2009, 2015], and Gayle, Golan, and Miller [2015] on the identification of moral-hazard models for executive pay. Unlike these authors, we consider a model that also features learning about ability and persistent shocks to it, and rely only on information on wages and their structure.

[32]Indeed, by rewriting (2) with $\widehat{\beta}_t = \beta_t - (1 - \lambda^t)k_0$ in place of β_t, we can absorb k_0 into m_θ.

Proposition 4. *The piece rates* $\{b_t^*\}_{t=0}^T$ *are identified from mean wages and performance pay. The parameters* $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ *are identified from the second moments of the distributions of wages. Once piece rates and* $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ *are identified, the risk aversion parameter* r*, the rate of human capital accumulation in the complex task* γ_2*, and the depreciation rate* $1-\lambda$ *are identified from piece rates. Once piece rates and* $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, r, \gamma_2, \lambda)$ *are identified, the rate of human capital accumulation in the simple task* γ_1 *and the drift terms* $\{\beta_t\}_{t=0}^{T-1}$ *are identified from mean wages up to* m_θ*.*

We conclude with four remarks. First, our argument does not require any exogenous variation external to the model even to recover preference parameters—intuitively, time variation in piece rates plays a role analogous to that of an instrument in standard IV settings. Second, the parameters of the learning process are identified independently of those of the human capital process so the recovery of the former is robust to the specification of the latter. Third, as we show in the appendix, a richer version of the model, in which parameters unobservably differ across workers, is also identified even when wages are measured with error. Fourth, as we also show in the appendix, the version of our model in which human capital evolves nonparametrically with effort is also identified provided that performance measures are available, which is often the case; see Frederiksen et al. [2017].

6 The Role of Performance Pay, Learning, and Human Capital

We now empirically investigate how performance pay, learning about ability, and human capital together shape the life-cycle profiles of wages and their fixed and variable components discussed in Section 2. To probe how our model captures the trade-offs behind the mechanisms we nest, we consider in Section 6.1 three alternative parameterizations that differ in the restrictions they impose on the model. The first one assumes that piece rates are exogenous so they are fixed in the model at their values in the data reported in Figure 1. The remaining parameters are estimated using only the mean and variance of wages over the life cycle without imposing any of the constraints from optimal contracting on performance pay that inform piece rates. The second parameterization allows piece rates to be endogenously determined as in Proposition 1 and corresponds to our unrestricted (full) model. Finally, the third parameterization imposes a lower variance of output shocks and so a higher speed of learning about ability than implied by the first two, more in line with existing estimates. We show that this parameterization has implications for piece rates that are clearly rejected by our data. Together, these parameterizations illustrate not only the importance of performance pay in governing life-cycle wages, but also the discipline that accounting for it introduces when distinguishing among alternative models of wage growth and dispersion.

In Section 6.2, we contrast our model to four prominent models in the literature that are nested by it, each of which addresses only some of the aspects of the data we consider. This comparison shows how our model combines existing frameworks to offer a novel and more comprehensive account of the wage process and the structure of wages. We then explore the implications of our three parameterizations for the relative importance of the primitive determinants of piece rates and effort incentives that we analyze (Section 6.3). We next examine the contribution of performance pay to the growth and dispersion of wages over the life cycle (Section 6.4). This analysis reveals that uncertainty about worker productivity is a powerful force depressing piece rates and yet the life-cycle pattern of piece rates is a critical source of the dynamics of wages.

In Section 6.5, we explore an extended version of our model with both contractible and non-contractible effort to examine how the allocation of effort between easily contractible activities (simple tasks) and difficult-to-contract ones (complex tasks) changes with experience in a firm. We find that the effort paths implied by our model are in line with the evolution of the complexity of workers' jobs over their careers at the BGH firm. We also find that even after incorporating a standard dimension of labor supply—contractable effort—performance pay and the effort it sustains still play a key role for the growth and dispersion of wages throughout the life cycle.

6.1 Three Model Parameterizations

To focus on the key mechanisms we integrate, we first consider three parameterizations of our model in which effort is only non-contractable ($\xi_1 = \gamma_1 = 0$); see Section 6.5 for the general case. By contrasting these three estimated versions of our model, we can illustrate its workings in a transparent way and shed light on how the multiple sources of incentives we nest shape compensation and its components over the life cycle. We start by describing our estimation approach.

Estimation. We estimate each model using an equally-weighted minimum-distance estimator and the moments from the BGH data in Figures 1 and 2 presented in Section 2. The first and third parameterizations with exogenous piece rates are estimated using the 80 moments consisting of the means and variances of wages over the first 40 years of labor market experience. The second

Table 3: Parameter Estimates Based on BGH Data

Parameters	Exogenous Piece Rates	Endogenous Piece Rates	Faster Learning ($K_0 = 0.2$ and $K_\infty = 0.05$)
σ_θ: std. dev. of initial ability (1,000 of 1988 $)	49.1	49.2	28.9
σ_ζ: std. dev. of ability shocks (1,000 of 1988 $)	0.0	0.0	2.97
σ_ε: std. dev. of output shocks (1,000 of 1988 $)	439.4	522.9	57.8
γ_2: human capital accumulation rate (complex task)	0.939	0.804	0.461
λ: human capital depreciation rate	0.967	0.991	0.974
r/ξ_2^2: effective risk aversion	N/A	0.00024	N/A

Notes: All models feature $\xi_1 = \gamma_1 = 0$, $\xi_2 = 1$, and $T = 40$. Parameters are estimated by equally-weighted minimum distance at very high levels of precision not reported here; details are available upon request.

parameterization is estimated by also matching the ratio of (average) performance pay to (average) total pay at each of these experience years for a total of 120 moments.[33] Since we discipline a tightly specified theoretical framework of 5 to 6 parameters by targeting 80 to 120 moments, the model is vastly overidentified. Also, as our data comprise more than 22,000 observations, the moments informing our exercises are estimated with high precision (Figures 1 and 2). Our model fits these moments well, as shown below, and its parameter estimates are likewise very precise.[34]

Results. Our first parameterization matches the means and variances of wages from the BGH data for the first 40 years of experience without imposing the restrictions that piece rates $\{b_t^*\}$ are obtained through optimal contracting—rather, they are treated as exogenously given by their empirical counterparts in the BGH data. The estimated parameters of this version of the model are the three learning parameters $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ describing the variance of ability and of output and ability shocks, and the two parameters (γ_2, λ) for the accumulation rate of human capital in the complex task and its depreciation rate. Their estimates are reported in column 1 of Table 3.[35]

Figure 3 shows how well the model (red lines) reproduces the life-cycle profile of the variance of wages (panel a) and of wage growth (panel b) in the data (blue lines). This version of the model fits the data best when $\sigma_\zeta = 0$—which we then maintain—so ability is constant over the life cycle. We estimate a standard deviation of shocks to output σ_ε (per worker) close to half a million dollars and a standard deviation of ability across workers σ_θ of about 50 thousand dollars. The remarkably good fit of the model to the profile of the variance of wages in the data derives from the fact that this variance in our model, $\mathrm{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$, reflects piece rates—namely,

[33] We also estimate a version of the third parameterization with endogenous piece rates; see below.

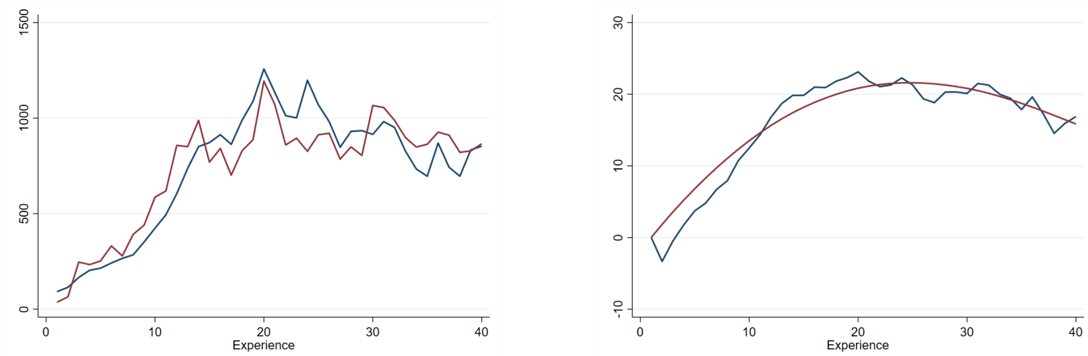
[34] Given that the standard errors are very small, we refrain from presenting them for brevity.

[35] Since the risk aversion parameter affects the variance of wages only through piece rates and does not directly affect the mean of wages, without performance pay, we lack crucial information to identify the (scaled) risk aversion parameter r/ξ_2^2. Thus, we cannot estimate it under this parameterization.

Figure 3: Fit of Model with Exogenous Piece Rates

(a) Variance of Wages

(b) Wage Growth



the square of the period piece rate b_t^* multiplied by the variance of output $\sigma_t^2 + \sigma_\varepsilon^2$, as shown in Lemma A.2 in the appendix. Then, the large variance of output shocks σ_ε^2, which scales piece rates in the expression for $\text{Var}[w_{it}]$ and dominates $\text{Var}[w_{it}]$, implies that the hump-shaped pattern of piece rates in the data leads to a hump-shaped pattern for the variance of wages in the model.[36] Thus, performance pay crucially shapes how the variance of wages evolves over time.

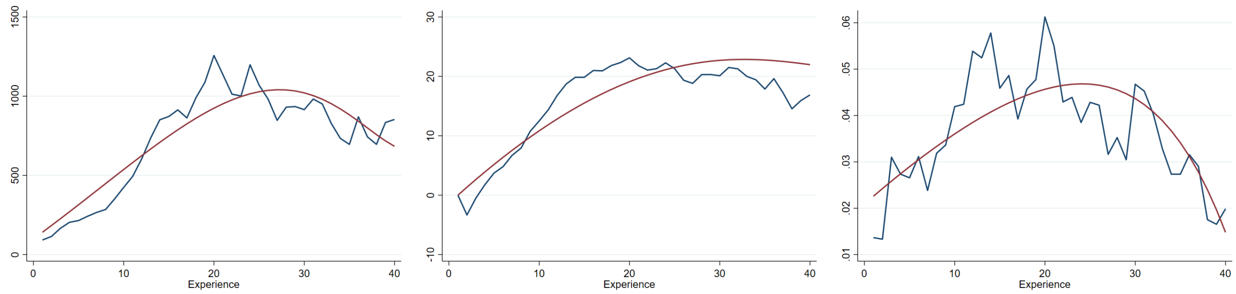
Our second parameterization endogenizes piece rates according to our optimal contracting framework. We then impose the 40 additional constraints given by (9), which determine how piece rates depend on the model parameters and vary over time. We now use information on performance pay to pin down the parameters, including the parameter r/ξ_2^2 for the curvature of worker utility with respect to consumption and effort that governs the trade-off between risk and incentives. Although it features only the additional parameter r/ξ_2^2, this parameterization fits the data very well. As panel c of Figure 4 shows, our full model (red line) well matches the life-cycle profile of the ratio of performance pay to total pay in the data (blue line). It also largely reproduces the variance and growth of wages over the life cycle with parameter estimates that are quite similar to those obtained with exogenous piece rates.

It is instructive to consider the aspects of the data behind the large estimate of the variance σ_ε^2 of output shocks. A large value of σ_ε^2 implies low piece rates at the end of working life (T), but so does a high degree of risk aversion if σ_ε^2 were small. A small σ_ε^2, however, would imply that output signals are very informative about ability so firms rapidly learn about it. The life-cycle risk due to the uncertainty about ability would rapidly decrease over the first half of workers' careers and so would the insurance term (see H_t^*) in (9). As a result, piece rates would rapidly increase over the

[36]By contrast, a high value for σ_θ^2 and so for σ_t^2 would counterfactually imply a high variance of wages early on.

Figure 4: Fit of Model with Endogenous Piece Rates

(a) Variance of Wages (b) Wage Growth (c) Piece Rates



life cycle. The data do not conform to this prediction though—piece rates are small throughout the life cycle. Thus, a large value for σ_ε^2 rationalizes the life-cycle level and range of piece rates in the data—and so helps account for the hump-shape of the variance of wages.

As for the human capital process, columns 1 and 2 of Table 3 report positive estimates of γ_2, which are consistent with human capital being acquired through learning-by-doing. As established in Section 5.2, since piece rates decline late in working life, the data favors a positive value for γ_2. For the parameterization with exogenous piece rates, the values of 0.939 for γ_2 and 0.967 for the depreciation factor λ in column 1 imply that a marginal increase in effort that induces a one-dollar increase in output this year leads to 0.94 dollars of additional output next year. Effort increasing output by one dollar at the beginning of a worker's career raises the present-discounted value of output over the life cycle by 11.2 dollars. For the parameterization with endogenous piece rates, the estimates of γ_2 and λ in column 2 have similar implications, albeit less extreme.[37]

Although these productivity gains from learning-by-doing imply output or social returns to effort that may seem large, they do not lead to implausibly large private returns in terms of wage gains, which are considerably smaller. Recall from (7) that the marginal return to effort in terms of future wages due to human capital acquisition is $R_{HK,t} = \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(b_{t+\tau} + R_{CC,t+\tau})$. The term $b_{t+\tau} + R_{CC,t+\tau}$ would equal one if workers were risk neutral, in which case social and private returns to human capital would coincide. But with risk-averse workers, the largest value of $b_{t+\tau}$ is about 0.05 and, as shown in panels a and b of Figure 9, the value of $R_{CC,t+\tau}$ (red lines) is small so the term $b_{t+\tau} + R_{CC,t+\tau}$ is also small. That is, the output gains to acquiring human capital are only

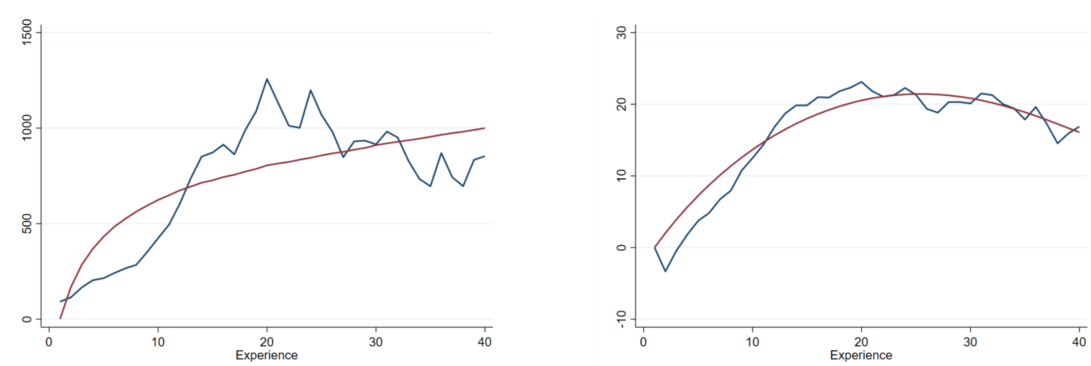
[37] Allowing for a shorter horizon ($T = 30$) or worker exit from the labor market—so for a higher effective discount rate $\widehat{\delta} = \delta s$ of future wages, where s is the exogenous exit rate—would imply lower values for σ_ε^2 and γ_2, due to the lower insurance workers would demand, and a higher speed of learning. Details are available upon request.

slowly and partially capitalized into wages, as firms insure workers against output (σ_ε^2) and ability risk (σ_t^2), which leads to low piece rates and relatively low cumulative returns to human capital.

Third Parameterization: Comparison with Existing Estimates. With just six parameters, our model captures well how the mean and variance of wages as well as piece rates evolve with experience. Rather than claim success and conclude here, we consider one additional parameterization motivated by the fact that our estimates of the learning parameters differ from those in the literature that suggest that firms rapidly learn about worker productivity, in particular Altonji and Pierret [1997], Lange [2007], Arcidiacono et al. [2010], and Aryal et al. [2022]. Our estimates, which are more in line with those in Pastorino [2024], imply instead that firms learn slowly about worker productivity. This final parameterization forces a faster learning speed about the ability of both young and old workers but allows worker productivity to change over the life cycle, thus accounting for the evidence in Kahn and Lange [2014] that firms continue to learn about this "moving target"—for this parameterization, the variance of shocks to ability turns out to be non-negligible.

Figure 5: Fit of Model with Faster Learning

(a) Variance of Wages (b) Wage Growth



Specifically, we impose two additional restrictions on the learning process, namely, that the speed of learning, defined as the weight $K_t = \sigma_t^2/(\sigma_t^2 + \sigma_\varepsilon^2)$ placed on output signals in the updating of beliefs about ability, is 0.2 at the beginning of a worker's career ($t = 0$) and 0.05 at the end of it ($t \to \infty$). A speed of learning of 0.2 early in life is consistent with Lange [2007] and Aryal et al. [2022], whereas a speed of learning of 0.05 late in life is consistent with learning continuing throughout the life cycle as in Kahn and Lange [2014]. We then first re-estimate our version of the model with exogenous piece rates taken from the data—so as not to constrain this version of the model to fit them—and with the restrictions that $K_0 = \sigma_\theta^2/(\sigma_\theta^2 + \sigma_\varepsilon^2) = 0.2$ and $K_\infty =$

$\sigma^2_\infty/(\sigma^2_\infty+\sigma^2_\varepsilon)=0.05$. Column 3 of Table 3 reports the resulting estimates, which imply a smaller rate of accumulation of human capital, γ_2. Figure 5 displays the fit of this version of the model to the data. Naturally, the fit worsens when we impose the restrictions described on the speed of learning. In particular, the model fits quite well the growth of wages over the life cycle. As panel a of Figure 5 shows, however, this parameterization does not capture the decline in the variance of wages over the second half of the life cycle, thus failing to reproduce its hump shape. In fact, rapid learning about ability implies that the variance of wages tends to increase over time.

When we allow piece rates to be endogenously determined as in (9), fast learning has also implications for piece rates that are highly counterfactual: it leads to piece rates that are very large in magnitude, negative when workers are young, and rapidly increasing with experience—all features (unreported) at odds with the data. Intuitively, when learning is fast, much new information is revealed early on. Workers then demand insurance against the risk that negative output realizations reveal them to be of low ability, permanently lowering their future wages. Firms partially insure workers against this risk by offering them negative piece rates when young. But as the remaining working life shortens and posterior beliefs about ability become more precise, lifetime risk decreases, the insurance component of piece rates (see H^*_t in (9)) decreases as a result, and so piece rates rapidly increase. Hence, a model with fast learning is hard to reconcile with the experience profile of the variance of wages or of piece rates in the data.

Summary. We have proposed three alternative parameterizations emphasizing different aspects of the data and incorporating priors about parameter values that reflect a number of estimates in the literature. As such, they offer a useful contrast that underscores both the challenges of matching key features of the dynamics of wages and their components, and the limitations of existing frameworks in accounting for them. We use these parameterizations below to address key questions about the determinants of piece rates, the magnitude of the sources of wage risk we consider, and the returns to effort as well as the importance of performance pay for life-cycle wages. Before doing so, we examine four prominent models in the literature that are nested by our model, which help isolate the mechanisms we integrate. By analyzing them one by one, we can explore these mechanisms in more detail and illustrate why none of these models in isolation can account for the data.

6.2 Comparison with Leading Models of Wage Growth and Dispersion

Our model nests several models central to labor and personnel economics. We focus on four representative such models to illustrate the features of the data that *cannot* be matched by these nested models, thus validating our integrated approach. This exercise also builds further intuition for the features of the data behind the estimates of the parameters of our model. Table 4 lists these models, their main features, the parameter restrictions that reduce our model to them, and the moments that each model does not explicitly account for. We reestimate the free parameters of each model using the moments it is designed to match and examine how each one fits the data.

Table 4: Nested Models

Model	Economic Content	Restrictions	Moments of Interest
Human Capital (HK)	Full Information, HK Investment	$\sigma_\varepsilon = \sigma_\theta = 0$	Wage variance, PP
Learning	No Hidden Effort, no HK	$\gamma_2 = e_{2t} = 0$	Wage growth, PP
Career Concerns (CC)	Learning, no PP, no HK	$\gamma_2 = b_t = 0$	Wage growth, PP
CC and Performance Pay (PP)	Learning, PP, no HK	$\gamma_2 = 0$	Wage growth

Human Capital Model. This model (Ben-Porath [1967] and Becker [1962]) allows for two free parameters, the accumulation rate γ_2 and the depreciation rate $1 - \lambda$ of human capital.[38] Contrasting panel a of Figure 6 with panel b of Figure 4 reveals that this model fits life-cycle wage growth better than our baseline model. A depreciation rate $1-\lambda$ of roughly 4% reproduces the decline in wage growth late in a worker's career, whereas an accumulation rate γ_2 of 0.471generates the rapid growth in wages early on. This result is not surprising as the parameters (γ_2, λ) are chosen to match only the growth of wages over the life cycle—in our model, they are also constrained to reproduce the life-cycle profile of the variance of wages and of piece rates. In our model, these estimates would imply much lower piece rates, given the magnitudes of the learning parameters that reproduce the variance of wages.[39] In this basic form, instead, the human capital model is essentially silent about the variance of wages—conditional on acquired capital—and their structure.

Learning Model. Farber and Gibbons [1996] propose a tractable log-linear formulation of the standard model of learning about (general) ability without any effort choice, human capital ac-

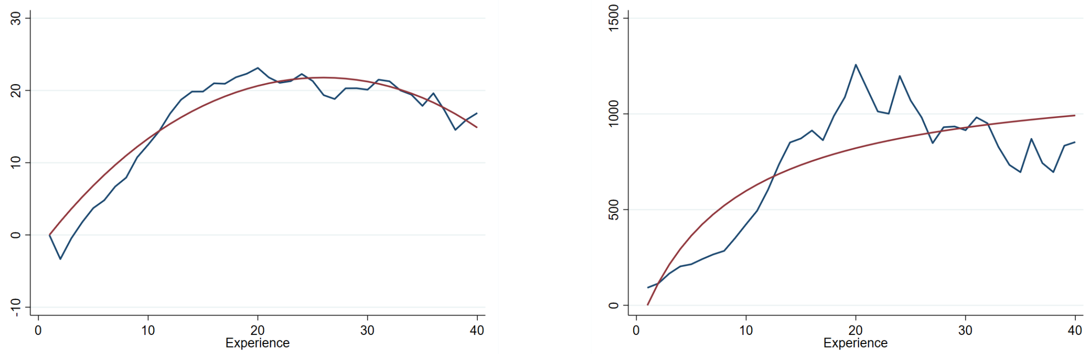
[38] Note that in this version of our model featuring just the task requiring non-contractable effort, known ability, and no shock to ability or output, piece rates are equal to one in each period, so workers are paid their output as they would be in the version of our model with known ability and just the task requiring contractable effort ($\xi_2 = \gamma_2 = 0$).

[39] Intuitively, a high σ_ε matches the low level of piece rates in the data—a high σ_θ would do so too but would also lead to too high a variance of wages early on. A high σ_ε implies a large σ_t, which in turn leads to a high insurance term H_t^* in piece rates; see (9). Such a term requires a value for γ_2 higher than 0.471 for piece rates to be positive.

Figure 6: Fit of Nested Models

(a) Human Capital Model: Wage Growth

(b) Learning Model: Variance of Wages



quisition, or contracting on performance, to account for the variance of wages over the life cycle. Their model allows for heterogeneity in ability but predicts no wage growth on average and has no implications for performance pay (see Table 4). Panel b of Figure 6 shows the fit of this model (red line) to the variance of wages over the life cycle (blue line), once we estimate it using the two parameters $(\sigma_\theta^2, \sigma_\varepsilon^2)$ governing the variance of ability and output shocks. The model does quite well at capturing the increase in the variance of wages over the first half of workers' careers. Yet, it implies a monotone concave profile for it that is at odds with its hump-shaped profile in the BGH data, which reveals an increasing and convex pattern for the variance of wages at low levels of experience (less than 20 years) with a peak at around 20 years, followed by a period of moderate decline. Our model substantially improves on this fit because piece rates exhibit a hump-shaped pattern with experience, which the variance of wages inherits as discussed.

Career-Concerns Model. Learning about ability also governs the dynamics of the model of implicit incentives for performance developed by Holmström [1999] in the early 1980s in response to the *Fama conjecture*. This conjecture holds that a reputation for high productivity in the labor market can substitute for explicit incentives for performance, thus eliminating the need for contracting. Holmström [1999] disproves it by showing that in the absence of explicit incentives for performance, effort cannot be sustained over time in general. The core mechanism of this model is that uncertainty about ability induces workers to exert effort so as to improve the market's expectation about their ability. As in our model, in equilibrium all workers choose the same effort in any period so the variance of wages is entirely determined just by the learning process. Thus, the fit of this model to the variance of wages (unreported) is almost precisely the fit in panel b of Figure 6.

Career-Concerns and Performance-Pay Model. The last model in Table 4 is that of Gibbons and Murphy [1992] of uncertainty and learning about ability as in Farber and Gibbons [1996], in which workers unobservably exert effort when employed as in Holmström [1999]. The model also features explicit contracting on performance, thus allowing us to explore the interplay between implicit and explicit incentives for effort. We obtain it as a special case of our model by restricting γ_2 to zero and λ to one so that free parameters are the variance of (initial) ability and output shocks $(\sigma_\theta^2, \sigma_\varepsilon^2)$ and workers' effective risk aversion r/ξ_2^2. We estimate it by matching the profile of the variance of wages and of piece rates over the life cycle.

Figure 7: Fit of Nested Career-Concerns Model of Gibbons and Murphy (1992)

(a) Variance of Wages (b) Piece Rates

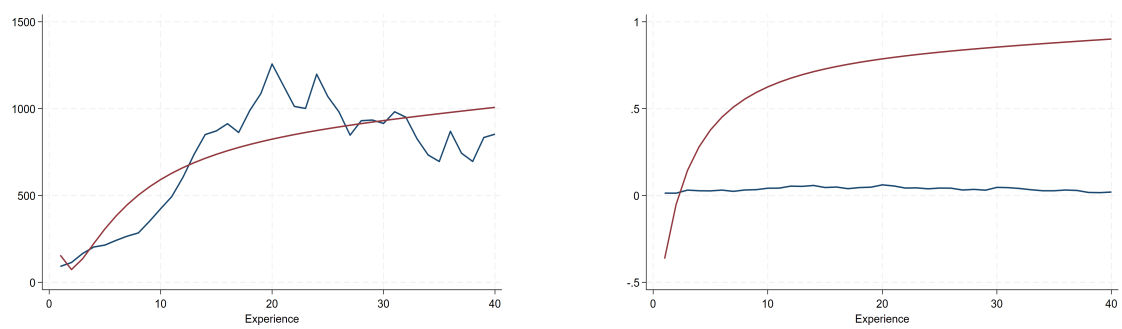


Figure 7 shows the fit of this model (red lines) to the targeted moments (blue lines). This standard career-concerns model with explicit contracting is unable to simultaneously reproduce the evolution of the variance of wages and of piece rates over the life cycle. In particular, its implications for the level and experience profile of piece rates are starkly at odds with the data. Without a human capital motive, piece rates are predicted to start at a negative level—because of the large career-concerns and insurance terms in (10)—and rapidly increase over time, as ability is revealed and the wage risk due to the uncertainty about it declines.[40] Specifically, the estimated standard deviation of output shocks σ_ε is 29.4 thousand dollars—much smaller than for the full model in columns 1 or 2 of Table 3—so ability is effectively known after just a few years. As learning quickly takes place, both workers' career concerns and desire to insure against the risk of low ability realizations become less and less important, leading firms and workers to agree to higher and higher piece rates since the performance incentive problem becomes progressively easier to solve. As is common to learning models, workers are characterized by large differences

[40]Gibbons and Murphy [1992] report evidence that performance pay increases over time but just for CEOs.

in ability relative to σ_ε—the estimated standard deviation of ability σ_θ is 17 thousand dollars.

Taken together, this ability heterogeneity, the rapid market learning about it, and the variability of beliefs about ability as information about it is acquired generate substantial wage risk. Such risk, which leads the variance of wages to increase over the life cycle, and firms' desire to insure workers against it are crucial determinants of the magnitude and variation of piece rates, as in our model. But this standard model has highly counterfactual implications for piece rates. We next show how these forces operate in our model as well and yet in a manner consistent with the data.

6.3 Determinants of Life-Cycle Piece Rates and Effort in the Complex Task

We now discuss the determinants of piece rates and the returns to effort. Our main findings are that the key forces shaping them are the prospect of human capital acquisition and the uncertainty about ability. Indeed, wage risk is considerable over the life cycle due to this uncertainty: both explicit and implicit incentives for effort are depressed by it—substantially more than by the pure output risk emphasized in moral-hazard problems.[41] This discussion also helps set the stage for the analysis in Section 6.4, where we argue that the effort induced by performance pay is nonetheless crucial for the dynamics of wages through its effect on the human capital process. We also show there how the impact of ability risk on wage risk is critically mediated by piece rates.

Piece Rates. Using the decomposition in (10) describing how the economic forces nested by our model shape the provision of explicit incentives for effort in the complex task demonstrates how lifetime wage risk and acquired human capital crucially govern piece rates. Figure 8 displays the contribution of each term to piece rates across 40 years of experience for our baseline parameterization in column 2 of Table 3 with endogenous piece rates—we place the four terms in two panels given their different scale. We focus here on this parameterization, since, as argued, it is the only one able to reproduce the profile of piece rates in the data.

Panel a displays the static piece-rate, b_t^0, and career-concerns, $b_t^0 R^*_{CC,t}$, terms, both of which account for a small portion of piece rates and their evolution over time—recall that $\xi_2 = 1$. That the static piece rate $b_t^0 = 1/[1 + (r/\xi_2^2)(\sigma_t^2 + \sigma_\varepsilon^2)]$—the blue line in panel a—is small throughout the life cycle reflects the large estimated variance of the noise in output σ_ε^2 and the sizable estimated degree of uncertainty about ability, as captured by the variance of posterior beliefs σ_t^2 and further

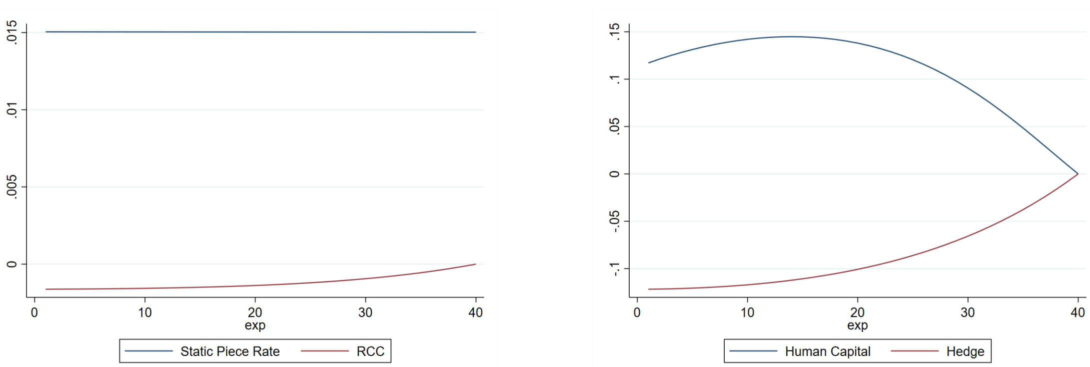
[41] See Appendix A.10 on dynamic risk due to the uncertainty about ability and static risk due to performance pay.

discussed below. The career-concerns term with $R^*_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b^*_{t+\tau})(\prod_{s=1}^{\tau-1} \mu_{t+\tau-s})(1 - \mu_t)$ by (7)—the red line in panel a—is likewise small. Intuitively, the large estimate of σ^2_ε implies not only that the static piece rate is small, but also that the signal-to-noise ratio governing the speed of learning about ability is low. As the weights $1 - \mu_t = \sigma^2_t/(\sigma^2_t + \sigma^2_\varepsilon)$ on output signals in the belief updating rule in (4) are correspondingly small, the learning process is not very sensitive to new observations of a worker's output and, as a result, effort has little effect on beliefs about ability. Thus, career-concerns incentives are small and so have a limited impact on piece rates.

The human capital, $(b^0_t/\xi_2)(\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R^*_{HK,t})$, and insurance, $b^0_t(r/\xi^2_2)H^*_t$, terms in panel b account for the bulk of piece rates at each level of experience. As these two sources of incentives roughly offset each other, piece rates are on average quite small. As apparent from its form, the insurance or hedge term $H^*_t = -\sigma^2_t \sum_{\tau=1}^{T-t} \delta^\tau$ is proportional to the uncertainty about ability as measured by the dispersion in posterior beliefs σ^2_t, which turns out to be substantial. Workers thus face a high degree of life-cycle risk induced by the variability of the beliefs about their ability and, accordingly, of their expected output and wages. Because of the magnitude of this risk, workers have a strong desire to insure themselves against it. Without markets providing insurance against low output realizations and so the future revelation of low ability, firms offer partial insurance through low piece rates—the more so, the more risk-averse workers are. As the horizon shortens, this insurance motive naturally weakens, which explains why the red line in panel b of Figure 8 representing this term eventually declines in absolute value with experience.

Figure 8: Decomposition of Piece Rates

(a) Static Piece Rate and Career Concerns   (b) Human Capital and Insurance

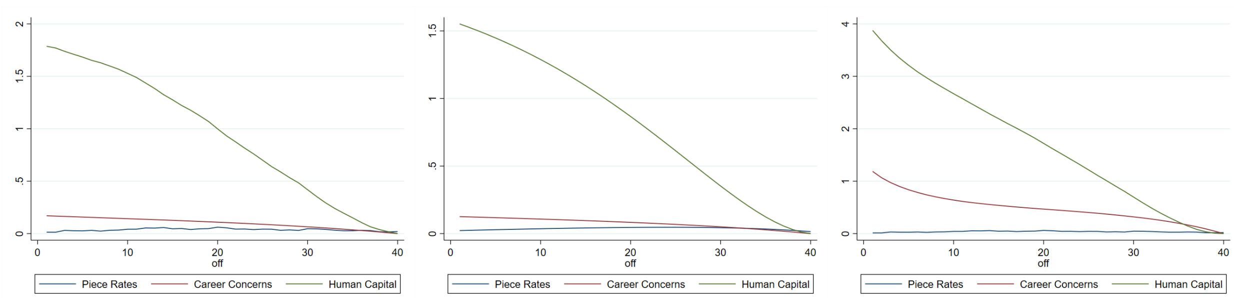


Finally, the human capital term $(b^0_t/\xi_2)(\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R^*_{HK,t})$ is sizable and is the key force that outweighs the insurance term. This result follows from the high estimated productivity of effort in terms of human capital production, as noted, and the low estimated rate of human capital depreciation. Hence, human capital has an important effect not only on the implicit incentives for effort via higher future wages resulting from a worker's higher productivity—as shown next—but also on the explicit incentives for effort via higher piece rates. As workers accumulate experience, however, this term eventually declines, imparting to piece rates their characteristic hump shape—a pattern at odds with the predictions of standard models of career concerns and performance pay. Thus, human capital resolves the puzzle of the hump-shaped profile of performance pay.

Effort in the Complex Task. We can also decompose the marginal returns to effort in the complex task e^*_{2t} into its determinants using (6), (7), and that, by (8), $e^*_{2t} = \xi_2 b^*_t + \xi_2 R^*_{CC,t} + R^*_{HK,t}$ in equilibrium. Figure 9, which reports this decomposition for the three parameterizations in Table 3, shows that the implicit returns from acquiring new human capital ($R^*_{HK,t}$) exceed those from both performance pay ($\xi_2 b^*_t$) and career concerns ($\xi_2 R^*_{CC,t}$) by a large margin over most of a worker's career. Contrasting panel c with panels a and b reveals, perhaps surprisingly, that this finding is robust to very different values of the speed of learning about ability, although the parameterization imposing fast learning implies much higher returns from both career concerns, $R_{CC,t}{=}\sum_{\tau=1}^{T-t}\delta^\tau(1-b_{t+\tau})\widehat{\mu}_{t,\tau}$, and human capital, $R_{HK,t}{=}\gamma_2\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(b_{t+\tau}+R_{CC,t+\tau})$; see the range of the panels.

Figure 9: Returns to Effort in the Complex Task

(a) With Exogenous Piece Rates (b) With Endogenous Piece Rates (c) With Faster Learning



To see why, note that a higher learning speed amplifies the impact of effort on a worker's future reputation by increasing the weight $\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)$ on new information, z_{it}, in the updating of beliefs about ability in (4), which raises the marginal benefit of effort due to career concerns—compare the red line in panel c with those in panels a and b. But higher incentives for effort from career concerns also imply higher incentives for effort from human capital, as captured by the second component of the human capital return to effort, $\gamma_2\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}R_{CC,t+\tau}$. Additionally, when learning is faster, increments to human capital are more rapidly capitalized into wages, since a rapidly declining variance of posterior beliefs about ability leads to higher piece rates by (10), as b_t^0 increases and H_t^* decreases. Higher future piece rates in turn increase the first component of the human capital return to effort, $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} b_{t+\tau}$. This explains why returns to effort from acquiring human capital exceed those from career concerns. As we argue next, effort choices and so performance pay are central to life-cycle wages because of their impact on workers' human capital.

6.4 The Importance of Performance Pay

We have illustrated how the risk induced by the process of learning about ability is an important force leading to small piece rates. The question then arises as to whether performance pay, and the effort it induces, can simply be abstracted from when studying the wage process, as is often argued. We now demonstrate how such an approach would miss a qualitatively and quantitatively important source of the dynamics of wages with experience.

Wage Growth. According to our model, average wages evolve over time as effort in the complex task and human capital change, since $\mathbb{E}[w_{it}] - \mathbb{E}[w_{i1}] = (e_{i2t}^* - e_{i21}^*) + (k_{i2t}^* - k_{i21}^*)$. In the two panels of Figure 10 and panel a of Figure A.4, we show how changes in effort $(e_{i2t}^* - e_{i21}^*)$ and human capital $(k_{i2t}^* - k_{i21}^*)$ contribute to life-cycle wage growth under our three parameterizations. Clearly, the accumulation of human capital governs wage growth, as effort tends to moderately decline. Under our third parameterization with a rapid speed of learning in panel a of Figure A.4, effort declines even more sharply than under the other two parameterizations, since ability is learned quickly and career concerns dissipate fast.

Figure 10: Dynamics of Effort, Human capital, and Wages

(a) With Exogenous Piece Rates (b) With Endogenous Piece Rates

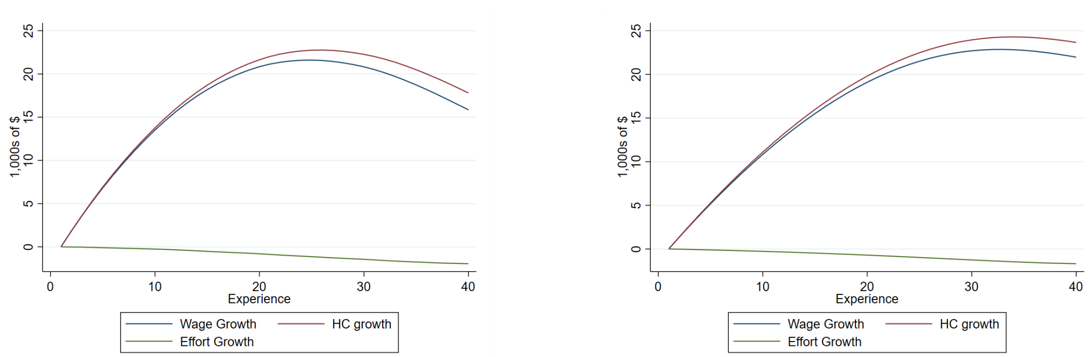
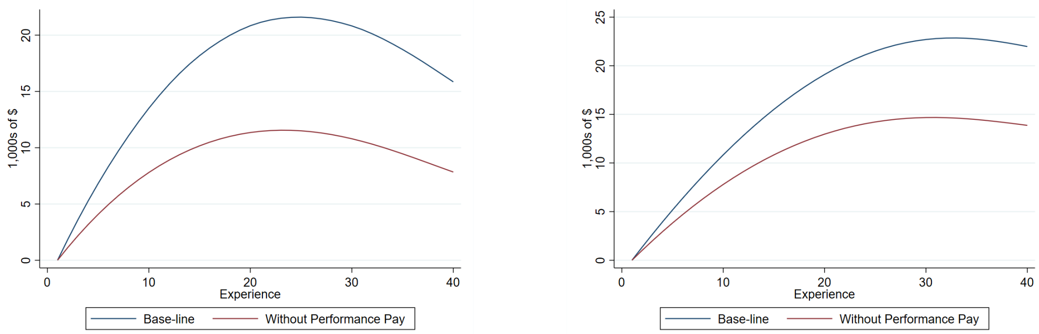


Figure 11: Wage Growth With and Without Performance Pay

(a) With Exogenous Piece Rates (b) With Endogenous Piece Rates



Given that effort is small and changes little, it might be tempting to conclude that performance pay cannot significantly affect wage growth. However, the decomposition in Figure 10 and panel a of Figure A.4 only uncovers the *direct* effect of effort on wage growth. According to our model, effort also contributes to human capital. Indeed, this is the key channel through which effort affects wage growth. One way to measure this *indirect* effect of effort on wages is to constrain firms to offer contracts without variable pay ($b_t^* \equiv 0$) as in Holmström [1999]. Figure 11 and panel b of Figure A.4 show the wage profiles that emerge with (red lines) and without (blue lines) this restriction—since our third parameterization, as discussed, has implications for piece rates at odds with the data, we focus here on our first two. Without performance pay, firms lack an important instrument to reward performance and thereby encourage workers to exert effort. Relative to the baseline, much less effort is exerted and so much less human capital is acquired. Lower effort and human capital in turn imply lower wage growth over the life cycle (red lines) relative to our baseline (blue lines), as panel a and b of Figure 11 show. By the 20th year of experience, wage growth is at least 30% lower than in the baseline. Thus, although performance pay is small, it has a substantial impact on wage growth because it indirectly affects workers' human capital.[42]

Wage Inequality. To measure how much performance pay contributes to the variability of wages, we begin by decomposing the variance of wages into the variance of their fixed and variable components—the latter is $(b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ in t—at the estimated parameter values. Figure 12 shows this decomposition for our first two parameterizations, on which we focus here; see panel c of Figure A.4 for our third one. According to our baseline parameterization in column 2 of Table 3, since

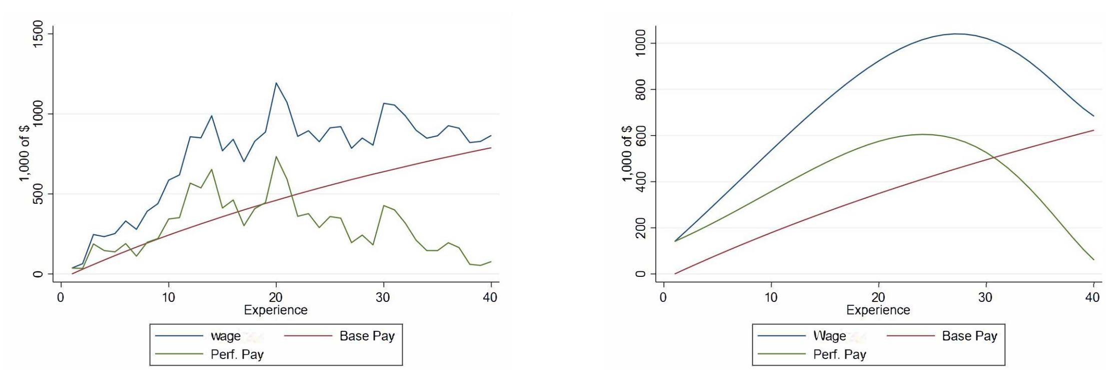
[42] As discussed, with a rapid speed of learning, learning about ability has a large impact on effort. Correspondingly, performance pay has a relatively smaller effect on wage growth; see panel b of Figure A.4.

shocks to output (σ_ε^2) are large, the variance of performance pay is large relative to the variance of fixed (base) pay for most of the life cycle—it accounts for more than 30% of the variance of wages over the first 30 years in the labor market—as shown in panel b of Figure 12. Fixed pay, which is revised each period based on realized output and thus provides implicit incentives for effort, increases slowly over time as information about ability is revealed. Similar conclusions can be drawn when piece rates are exogenous from panel a of Figure 12.[43,44] Although performance pay accounts for only a small fraction of pay at any given time, it is then responsible for a large share of the variability of wages over the life cycle.[45]

It turns out that through performance pay $b_t^* y_t$, uncertainty about ability is a major source of wage dispersion—due to its direct impact on the variability of beliefs (σ_t^2) and so output ($\sigma_t^2 + \sigma_\varepsilon^2$) and, as Figure A.5 illustrates, to its indirect impact on the level of piece rates (b_t^*). To elaborate, in panel a of the figure, we compare the variance of wages under our baseline parameterization (blue line) with the counterfactual one that would result at the *estimated* piece rates without any heterogeneity in ability (lavender line)—that is, when $\sigma_\theta^2 = \sigma_\zeta^2 = 0$—which is much lower. Intuitively, as is often maintained, lower dispersion in ability leads to lower wage dispersion.

Figure 12: Decomposition of the Variance of Wages

(a) With Exogenous Piece Rates (b) With Endogenous Piece Rates



Such a comparison, however, ignores that firms may offer different wage contracts in the absence of uncertainty about ability. In fact, as panel b of Figure A.5 shows, when we take into

[43]See in the figure also a decomposition of the variance of wages in the data, as fixed pay correlates little with performance pay so its variance can be backed out as the vertical difference between the blue and green lines.

[44]As panel c of Figure A.4 shows, imposing a fast speed of learning leads to a small variance of performance pay. In this case, the life-cycle increase in the variance of wages is almost all due to the increase in the variance of fixed pay, which is at odds with the data over the first half of the life cycle; see panel a of Figure 12.

[45]Similarly, Lemieux, MacLeod, and Parent [2009] find that the increased prevalence of performance pay from the 1970s to the 1990s accounts for about 21% of the increase in the variance of (log) wages over this period.

account firms' incentives to adjust piece rates in response to the lower uncertainty about ability, the variance of wages becomes up to six times larger (lavender line) than that in the baseline model (blue line). As panel c further shows, this increase is due to piece rates becoming much higher—up to three times as high (lavender line) as those in the baseline (blue line). Higher piece rates amplify any residual productivity risk, leading on balance to much higher wage dispersion over most of the life cycle. Hence, reducing differences in ability *ex ante* induces firms to offer contracts with a greater sensitivity of pay to performance *ex post*, which more than offsets the lower dispersion in ability and thus results in much more variable wages overall. Thus, lower dispersion in ability can actually lead to much *higher* wage dispersion.

6.5 Task Complexity over the Life Cycle

So far we have focused on the incentives for effort in the complex task. We now consider the general case of our model in which workers can perform simple and complex tasks, devoting possibly different amounts of effort to each, respectively e_{i1t} and e_{i2t}, in any period. This case illustrates how our results so far extend to a setting in which, unlike in usual job-assignment models, workers engage in multiple activities, and these activities expose them to different degrees of risk—recall that effort in the simple task is optimally remunerated without variable pay. A natural horse-race then emerges in terms of the importance of performance incentives for life-cycle wages: if life-cycle wage growth can be accounted for by human capital accumulation in simple contractible tasks, does performance pay still matter for life-cycle wages? This general case helps validate our findings by showing that our model can jointly account for the patterns of workers' wages and tasks at the BGH firm and that performance pay still plays a key role for the dynamics of wages.

By interpreting the efforts e_{i1t} and e_{i2t} as proxies of the task content of a job, the ratio $(1+e_{i2t})/(1+e_{i1t})$ can be viewed as capturing the *task complexity* of worker i's job in t. We can then measure task complexity in our data as follows. In the BGH data, an occupation or *job* is defined at the granular level of the occupation's *title*—there are 276 in our data. The nature of a job, and so the complexity of the tasks it involves, can be inferred from the description of its cost center, which is the organizational unit a job belongs to. BGH construct the firm's job hierarchy from workers' transition across job titles, which are aggregated into eight job levels, and find it to be divided into two parts: the bottom rungs, corresponding to job levels 1 to 4, at which nearly all workers (managers) start their careers at the firm, and the top rungs, corresponding to job levels 5 to 8 (chairperson-CEO). As BGH remark, higher-level jobs of this hierarchy, to which workers progress over time, require "*managing large groups, coordinating across business units, and strategic planning, while lower level jobs depend more on specialized functional knowledge and performing less complex tasks*" (Baker et al. [1994a], p. 893).

At job levels 1 to 4, about 60% of the jobs relate to specific *line* (revenue-generating) business units—positions that involve direct contact with customers or creating and selling products. Approximately 35% are overhead positions in areas such as accounting, finance, or human resources. At job levels 5 and 6, the two percentages of line business-unit and overhead activities decrease to 45% and 25%, whereas general management descriptions such as general administration or planning increase to about 30%—job levels 5 and 6 are the highest ones managers reach in our sample. At job levels 7 or 8, all activities are general management or planning. By defining a complex task as related to managing large groups, coordinating across business units, and strategic planning as BGH suggest, we can measure the degree to which a job requires non-contractable effort by the proportion of its general management or overhead activities. Similarly, we can measure the degree to which a job requires contractable effort by the proportion of its specific and easier-to-monitor activities involving direct contact with customers or creating and selling products.

We estimate the version of our model with effort in simple and complex tasks so as to match, as before, the variance of wages, average wages, and piece rates over the life cycle, now with the additional parameter γ_1 for human capital accumulation in the simple task—effort in the simple task affects output, human capital, and wages by (1)-(3).[46] Naturally, this version of our model better fits life-cycle wage growth; see panel b of Figure A.6. The estimates of the parameters that are common to the baseline parameterization in column 2 of Table 3 (unreported) are very similar, though. For instance, the estimate of γ_2, 0.75, is now only slightly lower; the estimate of γ_1 is 0.14.

This version of our model is in line with the range of task complexity of workers' jobs (0.67 to 1.22 in the data and 0.71 to 1.01 in the model) as well as its mean (0.84 in the data and 0.79 in the model) in the BGH data, even though none of these moments has been targeted. We also estimate that, as experience accumulates, workers eventually engage in more complex and harder-

[46] Although, as in our baseline, we maintain that workers are homogenous in their efficiency in performing tasks—$\xi_1 = \xi_2 = 1$—and in the rate at which their ability increases output—ξ_θ—these assumptions can be relaxed. See Section A.9 in the appendix for the case in which workers are heterogeneous in these parameters.

to-contract activities, as BGH document—our estimated measure of task complexity tends to increase with workers' tenure at the same rate as in the data. Furthermore, the model's implications about the importance of performance pay for life-cycle wages remain virtually unchanged relative to our baseline parameterization without the simple task. In particular, effort in the complex task and performance pay are still central to the growth of wages over the life cycle because of their impact on workers' accumulation of human capital. Overall, this exercise illustrates how the life-cycle profile of workers' tasks in the BGH data validates our notion of jobs and the resulting job assignment process. Importantly, it also reinforces the conclusion that incentives from performance pay, learning about ability, and human capital are key drivers of the wage process at the BGH firm.

7 Conclusion

We propose a tractable model of the labor market to analyze how performance pay, uncertainty and learning about ability, and human capital acquisition together determine life-cycle wages and their fixed and variable components. This framework reproduces key features of the dynamics of wages and of workers' career progression with experience, and is highly flexible in that it both nests many leading models of wage growth and dispersion and can be extended in several dimensions.

We find that two motives—namely, workers' demand for insurance against the substantial risk due to the uncertainty about their productivity and their desire to invest in human capital—are key determinants of performance pay that have sizable effects of *opposite* sign on its experience profile relative to total pay: the former negative, the latter positive. This tension rationalizes the low level of performance pay observed throughout the life cycle for most workers and, contrary to the prediction of influential models of performance incentives, its hump-shaped pattern relative to total pay. Although performance pay accounts for a small fraction of total pay, our analysis illustrates its centrality to the dynamics of wages through its indirect effect on workers' accumulation of human capital and its direct effect on the variability of wages—amplified by the endogenous response of piece rates to the degree of skill dispersion in a labor market. We hope that our results offer a first step toward richer models of incentives that can shed light on the multiple sources of the variation of wages across individuals and over time.

References

Altonji, J.G., and C. Pierret (1997): "Employer Learning and the Signaling Value of Education," *Industrial Relations, Incentives, and Employment*, Isao Ohashi and Toshiaki Tachibanaki, Eds., London, UK, Palgrave Macmillan, 159–195.

Arcidiacono, P., P. Bayer, and A. Hizmo (2010): "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2 (4), 76–104.

Aryal, G., M. Bhuller, and F. Lange (2022): "Signaling and Employer Learning with Instruments," *American Economic Review*, 112 (5), 1669–1702.

Atkeson, A., C. Hellwig, and G. Ordonez (2015): "Optimal Regulation in the Presence of Reputation Concerns," *Quarterly Journal of Economics*, 130 (1), 415–464.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104 (6), 1551–1596.

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

Baker, G.P., M. Gibbs, and B. Hölmstrom (1994): "The Wage Policy of a Firm," *Quarterly Journal of Economics*, 109 (4), 921–955.

Becker, G.S. (1962): "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, 70 (5, Part 2), 9–49.

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75 (4, Part 1), 352–365.

Bloom, N., and J. Van Reenen (2010): "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives*, 24 (1), 203–224.

Camargo, B., Lange, F., and E. Pastorino (2024): "On the Role of Learning, Human Capital, and Performance Incentives for Wages," NBER Working Paper 30191.

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

Eeckhout, E., and X. Weng (2022): "Assortative Learning," *Economica*, 89 (355), 647–688.

Ekinci, E., A. Kauhanen, and M. Waldman (2018): "Bonuses and Promotion Tournaments: Theory and Evidence," *Economic Journal*, forthcoming.

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

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

Gayle, G.-L., and R.A. Miller (2009): "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review*, 99 (5), 1740–1769.

Gayle, G.-L., and R.A. Miller (2015): "Identifying and Testing Models of Managerial Compensation," *Review of Economic Studies*, 82 (3), 1074–1118.

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

Geweke, J., and M. Keane (2000): "An Empirical Analysis of Earnings Dynamics among Men in the PSID: 1968-1989," *Journal of Econometrics*, 96 (2), 293–356.

Gibbons, R., and K.J. Murphy (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 100 (3), 468–505.

Gibbons, R., L.F. Katz, T. Lemieux, and D. Parent (2005): "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 23 (4), 681–723.

Gibbs, M., and W. Hendricks (2004): "Do Formal Salary Systems Really Matter?" *ILR Review*, 58 (1), 71–93.

Gladden, T., and C. Taber (2009): "The Relationship between Wage Growth and Wage Levels," *Journal of Applied Econometrics*, 24 (6), 914–932.

Groes, F., P. Kircher, and I. Manovskii (2015): "The U-Shapes of Occupational Mobility," *Review of Economic Studies*, 82 (2), 659–692.

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

Holmström, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66 (1), 169–182.

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

Keane, M.P., A. Ching, and T. Erdem (2017): "Empirical Models of Learning Dynamics: A Survey of Recent Developments," *Handbook of Marketing Decision Models*, Berend Wierenga and Ralf van der Lans, Eds., New York City, NY, Springer, 223–257.

Keane, M.P., and K.I. Wolpin (2007): "The Career Decisions of Young Men," *Journal of Political Economy*, 105 (3), 473–522.

Lange, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, 25 (1), 1–35.

Lemieux, T. (2008): "The Changing Nature of Wage Inequality," *Journal of Population Economics*, 21 (1), 21–48.

Lemieux, T., W.B. MacLeod, and D. Parent (2009): "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 124 (1), 1–49.

Low, H., C. Meghir, and L. Pistaferri (2010): "Wage Risk and Employment Risk over the Life Cycle," *American Economic Review*, 100 (4), 1432–67.

Margiotta, M., and R.A. Miller (2000): "Managerial Compensation and the Cost of Moral Hazard," *International Economic Review*, 41 (3), 669–719.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72 (1), 1–32.

Pastorino, E. (2024): "Careers in Firms: The Role of Learning about Ability and Human Capital Acquisition," *Journal of Political Economy*, 132 (6), 1994–2073.

Taber, C., and R. Vejlin (2020): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Econometrica*, 88 (3), 1031–1069.

Tincani, M., F. Kosse, and E. Miglino (2025): "Beliefs and the Incentive Effects of Preferential College Admissions: Evidence from an Experiment and a Structural Model," mimeo.

Waldman, M. (2012): "Theory and Evidence in Internal Labor Markets," in *Handbook of Organizational Economics*, Robert Gibbons and John Roberts, Eds., Princeton, NJ, Princeton University Press, 520–572.

A Appendix

We begin by describing the data we use (Section A.1). Next, we derive the equilibrium (Section A.2), prove the results of Section 5.2 concerning the life-cycle profile of piece rates (Section A.3), and establish our identification results (Section A.4). Then, we discuss some extensions of our framework: *i*) the model with wage markdowns (Section A.5); *ii*) the general cost-function case (Section A.6); *iii*) the model with Cobb-Douglas technology (Section A.7); *iv*) the model in which human capital evolves nonparametrically with effort (Section A.8); and *v*) the model with heterogeneous workers, either in their ability at the complex task or in how their ability affects output (Section A.9). We conclude by reporting omitted empirical results (Section A.10).

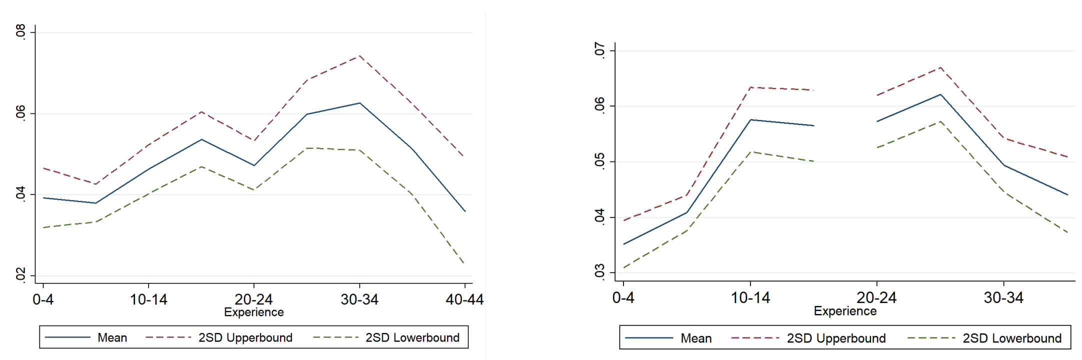
A.1 Data Samples

We provide here additional evidence on the hump-shaped profile of performance pay relative to total pay from both public worker survey data and proprietary firm-level administrative data.

Public Data: PSID and NLSY. We focus on the main sample of the PSID (Panel Study of Income Dynamics), excluding the poverty, Latino, and immigrant subsamples, and consider male heads of households aged 21 to 65 observed between 1993 and 2013 with valid education information—that is, with more than zero and up to 17 years of education, the largest value. We further restrict attention to those who work more than 45 weeks each year in any industry except for the government and the military, have non-missing positive total labor income, and are not self-employed. The resulting sample consists of more than 24,000 person-year observations. We calculate labor market *experience* as potential experience, defined as the difference between an individual's age (minus six) and years of education. We refer to an individual's labor income as the individual's *wage*. Although three measures of variable pay—namely, tips, bonuses, and commissions—are available in the PSID from 1993 onward, we focus here only on bonus pay for consistency across the data sets we examine. Bonus pay, though, is by far the most important component of variable pay, making up 80% of variable pay in our sample. We regularize the sample by excluding observations on bonus pay larger than total labor income and by winsorizing labor income at the 1st and 99th percentiles and bonus pay at the 99th percentile of their respective distributions. Finally, we restrict attention to workers who ever receive variable pay (bonus pay in our case) in their current job, that is, workers in *performance-pay jobs*; this definition of a performance-pay job is the same as in Lemieux, MacLeod, and Parent [2009]. In the resulting sample of workers in performance-pay jobs, the average salary is \$80,000 (in 2009 dollars), with a standard deviation of \$67,000, and the average bonus pay is \$4,000, with a standard deviation of \$8,000. Panel a of Figure A.1 shows how the ratio of (average) variable pay to (average) total pay, which also measures the sensitivity of pay to performance, follows a hump-shaped pattern with experience. Analogous profiles emerge if we divide the sample into workers with and without a college degree—the hump-shape of the experience profile of the sensitivity of pay to performance is most pronounced in the college sample. The PSID data thus suggest that the sensitivity of pay to performance increases early in the life cycle, peaks around its middle, and then subsequently declines.

Figure A.1: Life-Cycle Ratio of Performance Pay to Total Pay in Public Data

(a) PSID (b) NLSY97 and NLSY79



We perform an analogous exercise in the NLSY79 (National Longitudinal Survey of Youth–1979 Cohort) and NLSY97 (National Longitudinal Survey of Youth–1997 Cohort) by applying the same sample selection criteria applied to the PSID, finding analogous results.[47] The surveys are administered biyearly over the period of interest; since the question about bonus pay is retrospectively asked about the previous year, we can measure the amount of bonus pay that a worker receives for the years 2001, 2003, and so on, up to 2015. By virtue of the design of the two data sets, in the NLSY79, we can observe only individuals with 20 to 39 years of experience; in the NLSY97, we can observe only individuals with 0 to 19 years of experience. In the sample of workers with performance-pay jobs, the average salary is \$72,000 in the NLSY79 and \$48,000 in the NLSY97, with a standard deviation of \$56,000 and \$32,000, respectively. Average variable

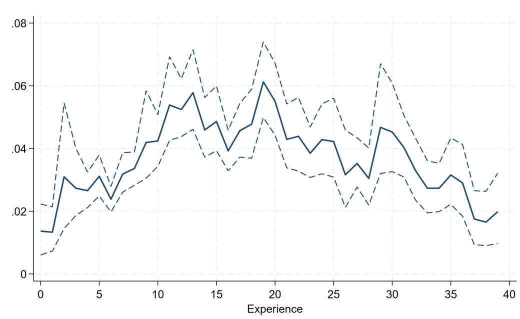
[47] We note that only bonus pay is recorded for both the 1979 and the 1997 NLSY cohorts, as tips and commissions are reported only for the latter cohort. Moreover, the amount of bonus pay received by a worker is effectively available in the NLSY79 only for the waves between 2002 and 2016.

pay is \$3,800 and \$2,000, with a standard deviation of \$12,000 and \$4,000. Thus, the statistics for the NLSY79 and the PSID are very similar. That the mean and standard deviation of total pay and bonus pay in the NLSY97 are lower than in the NLSY79 is intuitive, since the former samples younger workers. Panel b of Figure A.1 shows that the sensitivity of pay to performance exhibits a hump-shaped pattern in both the NLSY79 and the NLSY97, which is quite similar to that from the PSID—the break is due to the different cohorts tracked by the NLSY79 and the NLSY97.

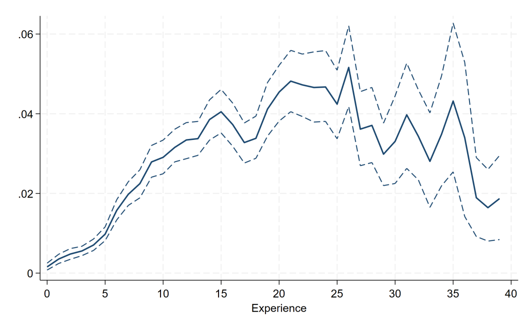
Proprietary Data: BGH and GH Data. We now contrast our findings on the profile of performance pay relative to total pay from the BGH data to those from the proprietary data from another U.S. firm, studied by Gibbs and Hendricks [2004] (henceforth GH). Both the BGH and the GH data are described in detail in Frederiksen et al. [2017]. For both firms, we have information only on white-collar workers—managers in the case of the BGH data.

Figure A.2: Life-Cycle Ratio of Performance Pay to Total Pay in BGH and GH Data

(a) BGH



(b) GH



The GH data cover the years from 1989 to 1993—we cannot reveal the industry that the firm belongs to. For the GH firm, we have information about 15,618 individuals for a total of 47,603 person-year observations. As these data pertain to all white-collar employees of the firm, the average salary is naturally lower than in the BGH data at \$39,215 (in 1988 dollars), with a standard deviation of \$27,968. Bonus pay on average amounts to \$1,537, with a standard deviation of about \$9,346. The two panels of Figure A.2 report the experience profile of the ratio of (average) variable pay to (average) total pay from both the BGH and the GH data for managers with up to 40 years of experience. In both firms, performance pay is hump-shaped relative to total pay, peaking at about 20 years of experience. Analogous patterns emerge if we focus on college-educated or non-college-educated workers, much like in the BGH data. Hence, a hump-shaped profile for performance pay relative to total pay is clearly a characteristic feature of both data sets.

A.2 Equilibrium Derivation

We first derive effort choices in the complex task for workers facing a sequence of employment contracts such that effort choices in the simple task and piece rates depend only on time when workers' future effort choices in the complex task also depend only on time. We then determine the equilibrium employment contracts and show that they are the same for all workers and are as described above. Finally, we derive the equilibrium.

A.2.1 First-Order Conditions for Effort in the Complex Task

We start with the following auxiliary result. Recall that if u and v are vectors in an Euclidean space, then $\langle v, u\rangle$ denotes their scalar product.

Lemma A.1. *Fix* $\{a_t\}_{t=1}^T$. *For each* $0 \leq t \leq T-1$,

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

Proof. The result is trivially true for $t = T-1$, since $R_{CC,T} = 0$. Fix $0 \leq t \leq T-2$, and let $u, v \in \mathbb{R}^{T-t-1}$ be such that $u = (a_t, \ldots, a_{T-t-1})$ and $v = (\delta^2(1-b_{t+2}), \ldots, \delta^{T-t}(1-b_T))$. Moreover, let A be the square matrix of order $T-t-1$ such that $A_{ij} = 0$ if $i < j$ and $A_{ij} = (\prod_{k=1}^{i-j} \mu_{t+i+1-k})(1-\mu_{t+j})$ if $i \geq j$. Then

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

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

$$\langle v, Au\rangle = \langle (AD^{-1})'v, Du\rangle = \langle (D^{-1})'A'v, Du\rangle = \langle D^{-1}A'v, Du\rangle. \tag{11}$$

On the other hand, given that $(D^{-1}A'v)_i = \delta^{-i}(A'v)_i = \delta^{-i}\sum_{j=1}^{T-t-1} A_{ji}v_j$, it follows that

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

for each $1 \leq i \leq T-t-1$; note the change of variables $j \mapsto j+i-1$. Thus, by (11),

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

where we again used the fact that $R_{CC,T} = 0$. This establishes the desired result. □

Suppose workers face a sequence $\{(e_{1t}, b_t)\}_{t=0}^{T}$ of employment contracts such that e_{1t} and b_t depend only on time and consider worker i's choice of period-t effort in the complex task, e_{2t}, when the worker's future choices of effort in this task also depend only on time. We claim that e_{2t} does not affect the variance of future wages. Indeed, since the variance of signals about ability does not depend on current effort choices, (5) implies e_{2t} does not affect the variance of future reputations. Moreover, future effort choices and piece rates do not depend on e_{2t}, being dependent only on time. Finally, a worker's stock of human capital has no impact on the variance of output or wages. The argument in the main text then shows that the first-order condition for worker i's optimal choice of effort in the complex task in period t is given by (6); recall that $w_{it+\tau}$ and h_i^t are, respectively, the worker's wage in period $t+\tau$ with $0 \leq \tau \leq T-t$ and history in period t. We claim that (6) reduces to (8).

First, by (3), $w_{it+\tau} = (1-b_{t+\tau})\mathbb{E}[y_{it+\tau}|I_{it+\tau}] + b_{t+\tau} y_{it+\tau}$ for all $1 \leq \tau \leq T-t$, where $y_{it+\tau}$ is worker i's output in period $t+\tau$ and $I_{it+\tau}$ is the public information about the worker available in $t+\tau$. Let $m_{it+\tau}$ be the worker's reputation in $t+\tau$; note that $m_{it+\tau}$ depends on $I_{it+\tau}$. Since for each $1 \leq \tau \leq T-t$, the effort e_{2t} affects $\mathbb{E}[y_{it+\tau}|I_{it+\tau}]$ only through its impact on $m_{it+\tau}$, as the other terms in the conditional expectation depend on the worker's conjectured effort and stock human capital in $t+\tau$ and the worker's future effort choices depend only on time,

$$\frac{\partial \mathbb{E}[w_{it+\tau}|h_i^t]}{\partial e_{2t}} = (1-b_{t+\tau})\frac{\partial \mathbb{E}[m_{it+\tau}|h_i^t]}{\partial e_{2t}} + b_{t+\tau}\frac{\partial \mathbb{E}[y_{it+\tau}|h_i^t]}{\partial e_{2t}}$$

for all $1 \leq \tau \leq T-t$. Now note that $\partial \mathbb{E}[y_{it+\tau}|h_i^t]/\partial e_{2t} = \gamma_2 \lambda^{\tau-1}$ for all $1 \leq \tau \leq T-t$, again since worker i's behavior from $t+1$ on depends only on time. Moreover, from (5),

$$\begin{aligned}
\frac{\partial \mathbb{E}[m_{it+\tau}|h_i^t]}{\partial e_{2t}} &= \sum\nolimits_{s=0}^{\tau-1} \left(\prod\nolimits_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1-\mu_{t+s})\frac{\partial \mathbb{E}[z_{it+s}|h_i^t]}{\partial e_{2t}} \\
&= \left(\prod\nolimits_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t)\frac{\partial \mathbb{E}[z_{it}|h_i^t]}{\partial e_{2t}} \\
&\quad + \sum\nolimits_{s=1}^{\tau-1} \left(\prod\nolimits_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1-\mu_{t+s})\frac{\partial \mathbb{E}[z_{it+s}|h_i^t]}{\partial e_{2t}},
\end{aligned}$$

where z_{it+s} is the signal about worker i's ability in $t+s$. Given that $\partial\mathbb{E}[z_{it}|h_i^t]/\partial e_{2t} = \xi_2$ and $\partial\mathbb{E}[z_{it+s}|h_i^t]/\partial e_{2t} = \gamma_2\lambda^{s-1}$ for all $1 \leq s \leq T-t$, we can rewrite (6) as

$$\begin{aligned} e_{2t} \;=\; & \xi_2 b_t + \xi_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau (1-b_{t+\tau}) \left(\prod\nolimits_{k=1}^{\tau-1} \mu_{t+\tau-k}\right)(1-\mu_t) \\ & + \gamma_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \left\{(1-b_{t+\tau}) \sum\nolimits_{s=1}^{\tau-1} \left(\prod\nolimits_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1-\mu_{t+s})\lambda^{s-1} + b_{t+\tau}\lambda^{\tau-1}\right\}. \end{aligned}$$

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

Condition (8) is necessary for optimality. It is also sufficient since the marginal benefit of effort in the complex task—the right side of (8)—is independent of the effort exerted, while the marginal cost—the left side of (8)—is increasing with the effort exerted.

A.2.2 Equilibrium Employment Contracts

We now solve for the last-period equilibrium employment contracts and then proceed backwards to determine the equilibrium employment contracts in previous periods. With this characterization of employment contracts at hand, we use (8) to derive the equilibrium choices of effort in the complex task, provided that equilibrium efforts and piece rates depend only on time, which is the case.

Last-Period Employment Contracts. The absence of dynamic considerations in the last period implies that a workers' choice of effort in the complex task is $e_2 = \xi_2 b$ if the piece rate is b. Then, by the mean-variance representation of worker preferences and free entry of firms, a worker's equilibrium employment contract in T is the pair (e_1, b) that maximizes $V_T = \mathbb{E}[w_T|I_T] - r\text{Var}[w_T|I_T]/2 - (e_1^2 + e_2^2)/2$, where w_T and I_T are a worker's wage and public information in T, respectively. Competition between firms further implies that $\mathbb{E}[w_T|I_T] = \mathbb{E}[y_T|I_T]$—this follows from (3) and the law of iterated expectations. Since $\mathbb{E}[y_T|I_T] \propto \xi_1 e_1 + \xi_2 e_2 = \xi_1 e_1 + \xi_2^2 b$ and $\text{Var}[w_T|I_T] = b^2(\sigma_T^2 + \sigma_\varepsilon^2)$, the pair maximizing V_T is $(e_1, b) = (e_{1T}^*, b_T^*)$ with $e_{1T}^* = \xi_1$ and $b_T^* = 1/[1 + (r/\xi_2^2)(\sigma_T^2 + \sigma_\varepsilon^2)]$. Note that the employment contract (e_{1T}^*, b_T^*) is independent of I_T and so the same for all workers. In turn, this implies that in equilibrium workers' choices of effort in the complex task are independent of their private histories and so also the same for all of them.

Employment Contracts in Previous Periods. Let $0 \leq t < T$ and suppose that equilibrium efforts and piece rates from period $t+1$ on depend only on time; this is true for $t = T-1$. For each $1 \leq \tau \leq T-t$, let $b_{t+\tau}^*$ be the equilibrium piece rate in period $t+\tau$, and define $R_{CC,t}^*$ and $R_{HK,t}^*$

as in (7) with $b_{t+\tau} = b^*_{t+\tau}$ for each τ. Then, a worker's period-t effort in the complex task as a function of the piece rate b in t is

$$e_{2t} = e_{2t}(b) = \xi_2 b + \xi_2 R^*_{CC,t} + R^*_{HK,t}. \tag{12}$$

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 the wages from t on as functions of b. An equilibrium employment contract for a worker in period t is a pair (e_1, b) that maximizes $V_t = \mathbb{E}[W_t|I_t] - r\mathrm{Var}[W_t|I_t]/2 - (e^2_{1t} + e^2_{2t})/2$, where I_t is the public information about the worker in t. We determine the pair (e_1, b) that maximizes V_t in what follows. As it turns out, this pair is independent of I_t and so the same for all workers in t.

First, note that

$$\frac{\partial \mathbb{E}[W_t|I_t]}{\partial b} = \xi_2^2 + \xi_2 \gamma_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}. \tag{13}$$

Indeed, if $y_{t+\tau}$ is the worker's output in period $t+\tau$ with $0 \leq \tau \leq T-t$, then competition between firms implies that $\mathbb{E}[w_{t+\tau}|I_t] = \mathbb{E}[y_{t+\tau}|I_t]$ for all $0 \leq \tau \leq T-t$. By (1) and (12), $\partial \mathbb{E}[y_t|I_t]/\partial b = \xi_2 \partial e_{2t}/\partial b = \xi_2^2$, which corresponds to the first term on the right side of (13). Regarding the second term on the right side of (13), note that by increasing effort in the complex task in t by ξ_2 units, a marginal increase in b also changes expected output in $t+\tau$ with $1 \leq \tau \leq T-t$ by $\xi_2\gamma_2\lambda^{\tau-1}$ units, the change in the worker's stock of human capital in $t+\tau$. The second term is the present-discounted value of these expected output changes. Now observe that since $\partial \mathbb{E}[y_t|I_t]/\partial e_1 = \xi_1$ and $\partial \mathbb{E}[y_{t+\tau}|I_t]/\partial e_1 = \gamma_1 \lambda^{\tau-1}$ for all $1 \leq \tau \leq T-1$, it follows that $\partial \mathbb{E}[W_t|I_t]/\partial e_1 = \xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$. We show below that

$$\frac{\partial \mathrm{Var}[W_t|I_t]}{\partial b} = 2b(\sigma_t^2 + \sigma_\varepsilon^2) + 2H_t^*, \tag{14}$$

where $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$. Since $\mathrm{Var}[W_t|I_t]$ is independent of e_{2t}—as effort in the complex task in t does not affect the variance of current and future wages—and $\partial e_{2t}/\partial b = \xi_2$, it then follows that the first-order conditions for the problem of maximizing V_t are

$$\xi_1 + \gamma_1 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - e_1 = 0 \text{ and}$$
$$\xi_2^2 + \xi_2\gamma_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - rb(\sigma_t^2 + \sigma_\varepsilon^2) - rH_t^* - \xi_2 e_{2t} = 0. \tag{15}$$

We now prove (14). Given that effort in the complex task does not affect the variance of future wages, $\mathrm{Var}[W_t|I_t]$ depends on b only through its effect on the variance of w_t. Hence,

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

where $\mathrm{Var}[w_t|I_t] = b^2(\sigma_t^2 + \sigma_\varepsilon^2)$ and Var_0 is a term that does not depend on b. We claim that $\mathrm{Cov}[w_t, w_{t+\tau}|I_t] = b\sigma_t^2$ for all $1 \leq \tau \leq T-t$, from which (14) follows. Since the worker's reputation in t is nonrandom conditional on I_t, $\mathrm{Cov}[w_t, w_{t+\tau}|I_t] = b\mathrm{Cov}[y_t, w_{t+\tau}|I_t]$ for all $1 \leq \tau \leq T-t$ by (3). Now note, once again from (3), that

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

for all $1 \leq \tau \leq T-t$, where $m_{t+\tau} = m_{t+\tau}(b)$ is a worker's reputation in $t+\tau$ as a function of the period-t piece rate. Like $y_{t+\tau}$, the reputation $m_{t+\tau}$ depends on b only through the impact of b on the workers' effort in the complex task in t. Thus, if $z_{t+s} = z_{t+s}(b)$ with $0 \leq s \leq T-t$ is the signal about ability in period $t+s$ as a function of b, then

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

for all $1 \leq \tau \leq T-t$ by (5). Now note that since $\mathrm{Cov}[y_t, y_{t+\tau}|I_t] = \sigma_t^2$ for all $1 \leq \tau \leq T-t$, $\mathrm{Cov}[y_t, z_{t+s}|I_t] = \sigma_t^2 + \sigma_\varepsilon^2$ if $s = 0$, and $\mathrm{Cov}[y_t, z_{t+s}|I_t] = \sigma_t^2$ if $1 \leq s \leq T-t$,

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

To conclude, note 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 $\mathrm{Cov}[y_t, w_{t+\tau}|I_t]$ is equal to

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

The desired result follows since the weights in the law of motion for a worker's reputation in (5) must sum up to one, so $\sum_{s=0}^{\tau-1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1 - \mu_{t+s}) = 1 - \prod_{k=1}^{\tau} \mu_{t+\tau-k}$ and the term in square brackets equals one.

The unique solution to (15) is $(e_1, b) = (e^*_{1t}, b^*_t)$ with $e^*_{1t} = \xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ and

$$b^*_t = b^0_t \left[1 + (\gamma_2/\xi_2) \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (1/\xi_2) R^*_{HK,t} - R^*_{CC,t} - (r/\xi_2^2) H^*_t \right],$$

by (12). Clearly, e^*_{1t} is the choice of e_1 that maximizes V_t no matter the choice of b. That b^*_t is that choice of b that maximizes V_t follows from the fact that V_t is strictly concave as a function of b. Note that (e^*_{1t}, b^*_t) is independent of I_t and so the same for all workers. The pair (e^*_{1t}, b^*_t) is the equilibrium employment contract in t under the induction hypothesis that equilibrium efforts and piece rates from period $t + 1$ on depend only on time.

Equilibrium Characterization. The above reasoning shows that if there exists $t < T$ such that from period $t + 1$ on equilibrium piece rates and effort choices are the same for all workers and depend only on time, then equilibrium employment contracts in period t are such that piece rates and effort choices in the simple task are the same for all workers. In turn, by (12), equilibrium choices of effort in the complex task are the same for all workers, and thus depend only on t. Since last-period equilibrium piece rates and effort choices are the same for all workers and (trivially) depend only on T, it then follows by induction that the equilibrium piece rates and effort choices are the same for all workers and depend only on time. From this, it further follows that the equilibrium is characterized by Proposition 1.

A.3 Piece Rates over the Life Cycle

We now prove the results in Section 5.2 concerning the life-cycle profile of piece rates.

A.3.1 Proof of Lemma 1

Consider first the case in which $\sigma^2_\theta \geq \sigma^2_\infty$ and σ^2_t is nonincreasing with t. Since $H^*_{T-1} > 0$ and $R^*_{CC,T-1} = \delta(1 - b^*_T)(1 - \mu_{T-1}) > 0$, $b^*_{T-1} = b^0_{T-1}(1 - R^*_{CC,T-1} - rH^*_{T-1}) < b^0_{T-1} \leq b^0_T = b^*_T$. Suppose, by induction, that there exists $1 \leq t \leq T - 1$ with $R^*_{CC,t+\tau} > R^*_{CC,t+\tau+1}$ and $b^*_{t+\tau} < b^*_{t+\tau+1}$ for all $0 \leq \tau \leq T - t - 1$; the induction hypothesis holds for $t = T - 1$. Thus,

$$\begin{aligned} R^*_{CC,s} &> \sum\nolimits_{\tau=1}^{T-s-1} \delta^\tau (1 - b^*_{s+\tau}) \left(\prod\nolimits_{k=1}^{\tau-1} \mu_{s+\tau-k}\right) (1 - \mu_s) \\ &> \sum\nolimits_{\tau=1}^{T-s-1} \delta^\tau (1 - b^*_{s+1+\tau}) \left(\prod\nolimits_{k=1}^{\tau-1} \mu_{s+\tau-k}\right) (1 - \mu_s), \end{aligned}$$

where the first inequality follows since $b^*_T \in (0,1)$ and $\mu_t \in (0,1)$ for $0 \leq t \leq T$ and 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. Holmström [1999] shows that $(1-\mu_s)\prod_{k=1}^{\tau-1}\mu_{s+\tau-k}$ is a decreasing function of μ_s (see argument in p. 174). Given that $\mu_{s+1} \geq \mu_s$, we then have that

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

Now note that $1-R^*_{CC,t}-rH^*_t-b\big[1+r(\sigma^2_t+\sigma^2_\varepsilon)\big] \leq 0$ if $b \geq b^*_t$. Since $R^*_{CC,s} > R^*_{CC,t}$, $H^*_s \geq H^*_t$, and $\sigma^2_s \geq \sigma^2_t$, it then 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 our equilibrium derivation that the first-order conditions in (15) are necessary and sufficient for the equilibrium employment contracts. Hence, $b^*_s = b^*_{t-1} < b^*_t$ and equilibrium piece rates strictly increase over time by induction.

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

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

for $T_0 \leq t \leq T$. By construction, b^∞_t is the equilibrium piece rate in period $T_0 \leq t \leq T$ if uncertainty about ability from period T_0 on were constant and equal to σ^2_∞. We claim that $\lim_{\sigma^2_{T_0}\to\sigma^2_\infty} b^*_t = b^\infty_t$ for all such t. First observe that $\lim_{\sigma^2_{T_0}\to\sigma^2_\infty} b^*_T = b^\infty_T$ as $\sigma^2_{T_0} < \sigma^2_T < \sigma^2_\infty$. Now suppose, by induction, that there exists $T_0 < t \leq T$ such that $\lim_{\sigma^2_{T_0}\to\sigma^2_\infty} b^*_{t+\tau} = b^\infty_{t+\tau}$ for all $0 \leq \tau \leq T-t$; the induction hypothesis is true for $t=T$. The desired result holds if $\lim_{\sigma^2_{T_0}\to\sigma^2_\infty} b^*_s = b^\infty_s$ for $s=t-1$. Given that $\sigma^2_{T_0} \leq \sigma^2_{s+\tau} < \sigma^2_\infty$ for all $0 \leq \tau \leq T-s$, it then follows that $\lim_{\sigma^2_{T_0}\to\sigma^2_\infty}\sigma^2_{s+\tau} = \sigma^2_\infty$, and thus $\lim_{\sigma^2_{T_0}\to\sigma^2_\infty}\mu_{s+\tau} = \mu_\infty$, for all such τ. This, in turn, implies that

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

by the induction hypothesis and the fact that the piece rate b^*_s is jointly continuous in $(b^*_{s+1},\ldots,b^*_T,$- $\sigma^2_s,\mu_s,\ldots,\mu_T)$. To conclude, note that since b^∞_t is strictly increasing with t for all $T_0 \leq t \leq T$ by the first case in the proof, there exists $\eta > 0$ such that if $|b_t^* - b_t^\infty| \leq \eta$ for all $T_0 \leq t \leq T$, then b_t^* is also strictly increasing with t for all such t. The desired result follows since $\lim_{T_0 \to \infty} \sigma_{T_0}^2 = \sigma_\infty^2$ and $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_t^* = b_t^\infty$ for all $T_0 \leq t \leq T$, and so, by taking T_0 large enough, we can ensure that $|b_t^* - b_t^\infty| \leq \eta$ for all $T_0 \leq t \leq T$.

A.3.2 Proof of Lemma 2

Let $\overline{\gamma}_2 = \xi_2(1 - \delta\lambda) r \sigma_\varepsilon^2 / \delta$. We claim that $b_t^* \in (0, 1)$ for all t if $\xi_2(\lambda - 1/\delta) \leq \gamma_2 \leq \overline{\gamma}_2$. Suppose, by induction, that there exists $1 \leq t \leq T$ such that $b_{t+s}^* \in (0, 1)$ for all $0 \leq s \leq T - t$; the induction hypothesis is true if $t = T$. We are done if $b_s^* \in (0, 1)$ for $s = t - 1$. First note that if $\gamma_2 \leq \overline{\gamma}_2$, then

$$b_s^* < b^0 \left[1 + \frac{\overline{\gamma}_2}{\xi_2} \sum\nolimits_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1} \right] < b^0 \left[1 + \frac{\overline{\gamma}_2}{\xi_2} \frac{\delta}{1 - \delta\lambda} \right] = 1,$$

where the first inequality follows since $b_t^* > 0$ for all $s < t \leq T$ by the induction hypothesis and the equality follows from the definition of $\overline{\gamma}_2$. Moreover, if $\gamma_2 \geq \xi_2(\lambda - 1/\delta)$, then

$$b_s^* > b^0 \left[1 + \left(\lambda - \frac{1}{\delta} \right) \sum\nolimits_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1} \right] > 0,$$

where the first inequality follows since $b_t^* < 1$ for all $s < t \leq T$ by the induction hypothesis and the second inequality follows because $(\lambda - 1/\delta) \sum_{\tau=1}^{T-s} \delta^\tau \lambda^{\tau-1} \geq -1$.

We now establish the properties of the experience profile of piece rates given γ_2. Since

$$b_t^* = b^0 \left[1 + \frac{\gamma_2}{\xi_2} \delta (1 - b_{t+1}^*) + \frac{\gamma_2}{\xi_2} \sum\nolimits_{\tau=2}^{T-t} \delta^\tau \lambda^{\tau-1} (1 - b_{t+\tau}^*) \right].$$

and $(\gamma_2/\xi_2) \sum_{\tau=2}^{T-t} \delta^\tau \lambda^{\tau-1} (1 - b_{t+\tau}^*) = \delta\lambda (b_{t+1}^*/b^0 - 1)$, we have that

$$b_t^* = b^0 \left(1 - \delta\lambda + \frac{\gamma_2}{\xi_2} \delta \right) + \delta b_{t+1}^* \left(\lambda - b^0 \frac{\gamma_2}{\xi_2} \right) \tag{16}$$

for all $0 \leq t \leq T - 1$. Given that $b_{T-1}^* = b_0[1 + (\gamma_2/\xi_2)\delta(1 - b_T^*)] < b_0 = b_T^*$ when $\gamma_2 < 0$, it follows from (16) that b_t^* strictly increases with t in this case—just note from (16) that $b_{T-2}^* < b^0[1 - \delta\lambda + (\gamma_2/\xi_2)\delta] + \delta b_T^*(\lambda - b^0\gamma_2/\xi_2) = b_{T-1}^*$ and apply a straightforward induction argument. Consider now the case in which $\gamma_2 > 0$. As $b_{T-1}^* > b_T^*$ when $\gamma_2 > 0$ and the coefficient of b_{t+1}^* in (16) is positive (respectively, negative) when $\gamma_2 < \widetilde{\gamma}_2 = \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$ (respectively, $\gamma_2 > \widetilde{\gamma}_2$), it then follows by induction that b_t^* is strictly decreasing (respectively, oscillating) with t when $\gamma_2 < \widetilde{\gamma}_2$ (respectively, $\gamma_2 > \widetilde{\gamma}_2$).

A.3.3 Proof of Proposition 2 and Extension

Let $\gamma_2 > 0$ and, for simplicity, assume that $\sigma_\zeta^2 = 0$. Since the equations for the equilibrium piece rates depend continuously on σ_ζ^2 and σ_t^2 eventually becomes small when σ_ζ^2 is small, we can extend the argument to the case in which σ_ζ^2 is positive but small. Fix $T_0 > 0$ and let $T > T_0$; we pin down T_0 below. Now consider the difference equation

$$b_t^{hc} = \frac{1}{1 + r\sigma_\varepsilon^2}\left[1 + \gamma_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(1 - b_{t+1}^{hc})\right]$$

for $T_0 \leq t \leq T$. By definition, b_t^{hc} is the piece rate in period $T_0 \leq t \leq T$ if only human capital acquisition were present. The same argument as that in the proof of Lemma 1 shows that $\lim_{\sigma_{T_0}^2 \to 0} b_t^* = b_t^{hc}$ for all $T_0 \leq t \leq T$. Since, by Lemma 2, b_t^{hc} either strictly decreases with t or oscillates with t for all $T_0 \leq t \leq T$ and $\lim_{T_0 \to \infty} \sigma_{T_0}^2 = 0$, it then follows that we can choose $T_0 \geq 0$ so that b_t^* behaves in the same way as a function of t for all $T_0 \leq t \leq T$.

We now show that there exist $T_0 \geq 0$ such that if $T > T_0$, then b_t^* is strictly increasing with t for all $T_0 \leq t \leq T$ provided that $|\gamma_2|$ is sufficiently small. Fix $T_0 \geq 0$ and let $T > T_0$. By Lemma 1, if $\gamma_2 = 0$, then piece rates are strictly decreasing with t for all $T_0 \leq t \leq T$ provided that T_0 is large enough. Since the equations for equilibrium piece rates depend continuously on γ_2, we can adapt the argument in the proof of Lemma 1 to show that if $|\gamma_2|$ is sufficiently small, then piece rates are also strictly increasing with t for all $T_0 \leq t \leq T$. This concludes the proof.

We now extend the second part of Proposition 2 to show that when $\gamma_2 < 0$, piece rates eventually strictly increase when the depreciation rate of human capital is sufficiently small provided that T is large enough. Suppose $\lambda = 1$; since the equations for the equilibrium piece rates depend continuously on λ, the argument extends to the case in which λ is sufficiently close to one. Given that $\sum_{s=0}^{\tau 1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1 - \mu_{t+s}) = 1 - \prod_{k=1}^{\tau} \mu_{t+\tau-k}$, as the coefficients in the law of motion for a worker's reputation in (5) sum up to one, and so $\sum_{s=1}^{\tau-1} \left(\prod_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right)(1\mu_{t+s}) = 1 - \prod_{k=1}^{\tau-1} \mu_{t+\tau-k}$ by straightforward algebra, it follows from Lemma A.1 that

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

Therefore,

$$b_t^* = b_t^0 \left(1 + \frac{\gamma_2}{\xi_2}(1 - \mu_t)^{-1} R_{CC,t}^* - R_{CC,t}^* - \frac{r}{\xi_2^2} H_t^*\right).$$

Now let $T_0 \geq 0$, suppose $T > T_0$, and consider the difference equation

$$b_t^\infty = \frac{1}{1 + r(\sigma_\infty^2 + \sigma_\varepsilon^2)} \left[1 - \widehat{\xi} \sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau}^\infty) \mu_\infty^{\tau-1} (1 - \mu_\infty) - \widehat{r} \sigma_\infty^2 \sum_{\tau=1}^{T-t} \delta^\tau \right]$$

for all $T_0 \leq t \leq T$, where $\widehat{\xi} = 1 + |\gamma_2|/\xi_2(1 - \mu_\infty) > 0$, $\widehat{r} = r/\xi_2^2$, and σ_∞^2 and μ_∞ are as in the proof of Lemma 1. By construction, b_t^∞ is the equilibrium piece rate in period $T_0 \leq t \leq T$ if uncertainty about ability from period T_0 on were constant and equal to σ_∞^2. It follows from Lemma 1 that b_t^∞ is strictly increasing with t for $T_0 \leq t \leq T$. Indeed, by redefining δ appropriately, we can absorb $\widehat{\xi}$ into δ. Then, by adjusting $\widehat{r}$ appropriately, the equation for b_t^∞ becomes that of the equilibrium piece rates in the pure learning case when $\sigma_\theta^2 = \sigma_\infty^2$, which strictly decrease over time by Lemma 1. The same argument as that in the proof of Lemma 1 shows that $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_t^* = b_t^\infty$. So, by taking T_0 large enough that $\sigma_{T_0}^2 \approx \sigma_\infty^2$, we have that b_t^* strictly increases with t for $T_0 \leq t \leq T$.

A.3.4 Proof of Proposition 3

Let $0 < \gamma_2 < \lambda \xi_2 (1 + r\sigma_\varepsilon^2)$. We know from Proposition 2 that piece rates eventually strictly decrease over time, and thus are not maximized at the end of a worker's career, if σ_ζ^2 is sufficiently small and T is large enough. Now assume that piece rates are between zero and one, so that $R_{HK,t}^*$ and $R_{CC,t}^*$ are non negative for all t. Since $\lambda \leq 1$, it follows from (9) that $b_0^* < b_0^0[1 + (\gamma_2/\xi_2 - r\sigma_\theta^2/\xi^2) \sum_{\tau=1}^{T} \delta^\tau]$. Thus, $b_0^* < b_0^0$ provided that σ_θ^2 is large enough. By increasing σ_θ^2 further if necessary, we can ensure that σ_t^2 strictly decreases with t, and so $b_0^* < b_T^0 = b_T^*$ and piece rates are also not maximized at the start of a worker's career.

A.4 Identification

Here, we prove Proposition 4 and extend our identification results to the case in which there is unobserved heterogeneity and measurement error. We start with the following result.

Lemma A.2. *For all* $0 \leq t \leq T$ *and* $1 \leq s \leq T - t$, $\text{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ *and* $\text{Cov}[w_{it}, w_{it+s}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + b_t^* \sigma_t^2$.

Proof. Note from (3) that $w_{it} = \overline{w}_{it} + r_{it}$, where $r_{it} = (1 - b_t^*)\mathbb{E}[\theta_{it}|I_{it}] + b_t^*(\theta_{it} + \varepsilon_{it})$ is the random part of w_{it}. Since we can incorporate m_θ into $\overline{w}_{it}$, it is without loss to assume that $\mathbb{E}[\theta_{it}] \equiv 0$, so that $\mathbb{E}[r_{it}] \equiv 0$. Therefore, $\text{Var}[w_{it}] = \mathbb{E}[r_{it}^2]$ and $\text{Cov}[w_{it}, w_{it+s}] = \mathbb{E}[r_{it} r_{it+s}]$. Also note that $\mathbb{E}[\theta_{it}|I_{it}] \perp \theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]$, as the conditional expectation is an orthogonal projection.

Variances of Wages. Since $r_{it} = \mathbb{E}[\theta_{it}|I_{it}] + b_t^*\big(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}] + \varepsilon_{it}\big)$, we have that

$$\mathrm{Var}[w_{it}] = \mathrm{Var}[r_{it}] = \mathrm{Var}[\mathbb{E}[\theta_{it}|I_{it}]] + (b_t^*)^2\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] + (b_t^*)^2\sigma_\varepsilon^2. \quad (17)$$

Now note that $\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] = \mathrm{Var}[\theta_{it}] - \mathrm{Var}[\mathbb{E}[\theta_{it}|I_{it}]]$. Indeed, $\mathrm{Var}[A - B] = \mathrm{Var}[A] + \mathrm{Var}[B] - 2\mathrm{Cov}[A, B]$ and $\mathrm{Cov}[\theta_{it}, \mathbb{E}[\theta_{it}|I_{it}]] = \mathrm{Var}[\mathbb{E}[\theta_{it}|I_{it}]]$. Moreover, given that $\theta_{it}|I_{it}$ is normally distributed with mean $\mathbb{E}[\theta_{it}|I_{it} = \iota_t]$ and variance σ_t^2 when $I_{it} = \iota_t$, the random variable $(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}])|I_{it}$ is normally distributed with mean zero and variance σ_t^2. Consequently, $\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] = \mathbb{E}[\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]]|I_{it}] = \sigma_t^2$, and so, since $\mathrm{Var}[\theta_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2$, it follows that $\mathrm{Var}[\mathbb{E}[\theta_{it}|I_{it}]] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2$. The desired result follows from (17).

Covariances of Wages. Let $\eta_{it}^s = \mathbb{E}[\theta_{it+s}|I_{t+s}] - \mathbb{E}[\theta_{it}|I_t]$. Since

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

we then have that

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

We are done if we show that $\mathbb{E}[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s] = 0$ and $\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta_{it}^s] = \sigma_t^2$. First, note that

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

by (5). Since $\theta_{it+k} = \theta_{it} + \zeta_{it} + \cdots + \zeta_{it+k-1}$, we have that $\mathbb{E}[\mathbb{E}[\theta_{it}|I_{it}]\eta_{it}^s] = 0$. Moreover,

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

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

$$\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta_{it}^s] = \sigma_t^2\left\{\prod\nolimits_{j=1}^{s-1}\mu_{t+s-j} + \sum\nolimits_{k=1}^{s-1}\left(\prod\nolimits_{j=1}^{s-1-k}\mu_{t+s-j}\right)(1 - \mu_{t+k})\right\}.$$

The desired result follows from the fact that the term in braces is one. $\square$

We now turn to the proof of Proposition 4.

Piece Rates and Variances. The wage of worker i in period t can be expressed as $w_{it} = f_{it} + v_{it}$, where f_{it} and v_{it} are its fixed and variable components, respectively. Since competition among firms implies that $\mathbb{E}[w_{it}] = \mathbb{E}[y_{it}]$ and $v_{it} = b_t^* y_{it}$, as contracts are linear in output, it follows that $b_t^* = \mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$. With piece rates recovered, the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified as follows. First, σ_θ^2 and σ_ε^2 are identified from b_0^*, $\text{Var}[w_{i0}]$, and $\text{Cov}[w_{i0}, w_{i1}]$. In turn, σ_ζ^2 is identified from $\text{Var}[w_{i1}]$, b_1^*, σ_θ^2, and σ_ε^2, since $\sigma_1^2 = \sigma_\zeta^2 + \sigma_\theta^2\sigma_\varepsilon^2/(\sigma_\theta^2 + \sigma_\varepsilon^2)$.

Risk Aversion, Human Capital in Complex Task, and Depreciation. First note that if $\{b_t^*\}_{t=0}^T$, σ_θ^2, σ_ε^2, and σ_ζ^2 are identified, so are σ_t^2, $R_{CC,t}^*$, and H_t^* for all t. Thus, r is identified from b_T^*, σ_T^2 and σ_ε^2, as $b_T^* = 1/[1 + (r/\xi_2^2)(\sigma_T^2 + \sigma_\varepsilon^2)]$, and so b_t^0 is identified for all t from r, σ_t^2, and σ_ε^2, as $b_t^0 = 1/[1 + (r/\xi_2^2)(\sigma_t^2 + \sigma_\varepsilon^2)]$. In turn, γ_2 is identified from b_{T-1}^*, b_{T-1}^0, b_T^*, $R_{CC,T-1}^*$, and H_{T-1}^*, as $b_{T-1}^* = b_{T-1}^0[1 + (\gamma_2/\xi_2)\delta(1 - b_T^*) - R_{CC,T-1}^* - (r/\xi_2^2)H_{T-1}^*]$. Finally, λ is identified from b_{T-2}^*, b_{T-2}^0, γ_2, b_{T-1}^*, $R_{CC,T-1}^*$, b_T^*, $R_{CC,T-2}^*$, and H_{T-2}^*, as $b_{T-2}^* = b_{T-2}^0\{1 + (\gamma_2/\xi_2)[\delta(1 - b_{T-1}^* - R_{CC,T-1}^*) + \delta^2\lambda(1 - b_T^*)] - R_{CC,T-2}^* - (r/\xi_2^2)H_{T-2}^*\}$.

Human Capital in Simple Task and Drift Terms. Note that once $\{b_t^*\}_{t=0}^T$, σ_θ^2, σ_ε^2, σ_ζ^2, r, γ_2, and λ are identified, so is $R_{HK,t}^*$ for all t, and thus is e_{2t}^* for all t, since $e_{2t}^* = \xi_2 b_t^* + \xi_2 R_{CC,t}^* + R_{HK,t}^*$. Hence, e_{10}^* is identified from $\mathbb{E}[w_{i0}]$ and e_{20}^* up to m_θ, as $\mathbb{E}[w_{i0}] = m_\theta + \xi_1 e_{10}^* + \xi_2 e_{20}^*$. In turn, γ_1 is identified from e_{10}^* and λ, as $e_{10}^* = \xi_1 + \gamma_1 \sum_{\tau=0}^T \delta^\tau \lambda^{\tau-1}$, and so $e_{1t}^* = \xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$ is identified for all t. Now, human capital k_t^* in $1 \leq t \leq T$ is identified from $\mathbb{E}[w_{it}]$, e_{1t}^*, and e_{2t}^* up to m_θ, since $\mathbb{E}[w_{it}] = m_\theta + k_t^* + \xi_1 e_{1t}^* + \xi_2 e_{2t}^*$ for all such t. We can then identify the terms $\{\beta_t\}_{t=0}^{T-1}$ from $\{k_t^*\}_{t=1}^T$, $\{e_{1t}^*\}_{t=0}^{T-1}$, $\{e_{2t}^*\}_{t=0}^{T-1}$, λ, γ_1, and γ_2, as $k_{t+1}^* = \lambda k_t^* + \gamma_1 e_{1t}^* + \gamma_2 e_{2t}^* + \beta_t$ for all such t.

We now consider the case of unobserved heterogeneity. Suppose there exist $J \geq 1$ types of workers who differ in their distributions of initial ability, shocks to output, and shocks to ability; their degree of risk aversion; and their human capital process. The model parameters for each type j are observable to model agents but not to the econometrician. Denote the probability that a worker is type j by π_j, and let $\sigma_{j\theta}^2$, $\sigma_{j\varepsilon}^2$, $\sigma_{j\zeta}^2$, r_j, $1-\lambda_j$, γ_{j1}, γ_{j2}, and β_{jt} be, respectively, the variance of the initial distribution of ability, the variance of the output shocks, the variance of the ability shocks, the risk aversion parameter, the depreciation rate of human capital, the rate of human capital accumulation in the simple task, the rate of human capital accumulation in complex task, and the period-t drift term in the human capital process for type-j workers. Proposition 1 holds for each worker type. Let e^*_{j1t}, e^*_{j2t}, and b^*_{jt} be, respectively, the effort in the simple task, effort in the complex task, and piece rate in period t for type-j workers, and let k^*_{jt} be the human capital of such workers in t. By (3), the wage of worker i of type j with ability θ_{ijt} in t is $w_{ijt} = f_{ijt} + v_{ijt}$, where $f_{ijt} = (1 - b^*_{jt})\mathbb{E}[\theta_{ijt} + k^*_{jt} + \xi_1 e^*_{j1t} + \xi_2 e^*_{j2t} | I_{it}]$ and $v_{ijt} = b^*_{jt}(\theta_{ijt} + k^*_{jt} + \xi_1 e^*_{j1t} + \xi_2 e^*_{j2t} + \varepsilon_{ijt})$ are its fixed and variable components. Thus, w_{ijt} is normally distributed by (5) and the distribution of wages in each period is a finite mixture of normal distributions. As such mixtures are identifiable (Teicher [1963]), both the mixture weights $\{\pi_j\}_{j \in J}$ and the component distributions are identified in each period, and so are their component means $\{\mathbb{E}_j[w_{ijt}]\}_{j \in J}$. Since v_{ijt} is normally distributed as well, the distribution of the variable component of wages in each period is also a finite mixture of normal distributions with the same component weights as the corresponding mixture distribution of wages. Hence, for each worker type j and period t, mean variable wages $\mathbb{E}_j[v_{ijt}]$ are identified as well so that the piece rate of type-j workers in t is identified as $b^*_{jt} = \mathbb{E}_j[v_{ijt}]/\mathbb{E}_j[w_{ijt}]$.[48] The rest of the argument is as in the proof of Proposition 4 for each type j.

Proposition A.1. *Suppose that each worker is one of $J \geq 1$ types. For each worker type j, the piece rates $\{b^*_{jt}\}_{t=0}^{T}$ and the variance parameters $(\sigma^2_{j\theta}, \sigma^2_{j\varepsilon}, \sigma^2_{j\zeta})$ are identified from a panel of wages and their variable components. Once piece rates and $(\sigma^2_{j\theta}, \sigma^2_{j\varepsilon}, \sigma^2_{j\zeta})$ are identified, the risk aversion parameter r_j, the rate of human capital accumulation in the complex task γ_{j2}, and the depreciation rate $1 - \lambda_j$ are identified from piece rates. Once piece rates and $(\sigma^2_{j\theta}, \sigma^2_{j\varepsilon}, \sigma^2_{j\zeta}, r_j, \gamma_{j2}, \lambda_j)$ are identified, the rate of human capital accumulation in the simple task γ_{j1} and the drift terms $\{\beta_{jt}\}_{t=0}^{T-1}$ are identified from mean wages up to $m_{j\theta}$.*

Proposition A.1 immediately extends to the case in which wages and their fixed and variable components are measured with error, provided that 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 accommodates alternative settings in which work-

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

ers of higher ability may be more or less efficient at acquiring new skills. This more general setup thus relaxes the impact of our functional-form assumptions by leading to a flexible dependence of wages on ability, uncertainty about it, human capital, risk, and workers' risk attitudes.

A.5 Extension: Wage Markdowns

We first extend our model to the case in which workers capture a fraction $\alpha \in (0, 1]$ of the surplus from their matches with firms; our baseline model corresponds to $\alpha = 1$. We omit most of the details in what follows, as derivations for this more general model follow very closely derivations for the baseline model.

A.5.1 Setup

The setup is the same as the baseline model except that now workers capture a fraction $\alpha \in (0, 1]$ of the surplus from their matches with firms. Consider worker i in period t. The expected value of the match between the worker and a firm is $\mathbb{E}[y_{it}|I_{it}]$. Thus, if Π_{it} is the expected flow profit of the firm that employs i in t, then $\Pi_{it} = (1-\alpha)\mathbb{E}[y_{it}|I_{it}]$. On the other hand, since $w_{it} = c_{it} + b_{it}y_{it}$, we have that $\Pi_{it} = \mathbb{E}[y_{it} - w_{it}|I_{it}] = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}] - c_{it}$. Thus, $c_{it} = (\alpha - b_{it})\mathbb{E}[y_{it}|I_{it}]$, and so $w_{it} = (\alpha - b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it}y_{it}$.

A.5.2 Equilibrium Characterization

The process of learning about ability is as in the baseline model. Thus, posterior beliefs about a worker's ability are normally distributed with mean and variance that evolve according to the laws of motion in (4), and the evolution of workers' reputation is as in (5).

As in the baseline model, the equilibrium is unique, symmetric, and such that effort choices and piece rates depend only on time. If workers face a sequence $\{(e_{1t}, b_t)\}_{t=0}^{T}$ of employment contracts such that efforts in the simple task and piece rates depend only on time, then effort in the complex task in period t is $e_{2t} = \xi_2 b_t + \xi_2 R_{CC,t} + R_{HK,t}$, where $R_{HK,t}$ has the same expression as in the baseline model and now $R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^{\tau}(\alpha - b_{t+\tau})(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k})(1-\mu_t)$. The intuition for this result is simple. The derivation of $R_{HK,t}$ does not depend on the surplus-sharing rule, so its expression does not change. The expression for $R_{CC,t}$ follows from the fact that the fixed component of a worker's wage in period $t + \tau$ with $0 \le \tau \le T - t$ is now a fraction $\alpha - b_{t+\tau}$ of the worker's expected output in $t + \tau$.

The derivation of equilibrium employment contracts follows the same steps as those in the baseline case. Given that workers now capture only a fraction α of their expected output, we have that $\partial\mathbb{E}[W_t|I_t]/\partial b = \alpha(\xi_2^2 + \xi_2\gamma_2\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1})$ and $\partial\mathbb{E}[W_t|I_t]/\partial e_{1t} = \alpha(\xi_1 + \gamma_1\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1})$. One can adapt the argument in the baseline case to show that $\mathrm{Cov}[w_t, w_{t+\tau}|I_t] = \alpha b\sigma_t^2$ for all $0 \leq t \leq T$ and $1 \leq \tau \leq T-t$, from which it follows that $\partial\mathrm{Var}[W_t|I_t]/\partial b = 2b(\sigma_t^2 + \sigma_\varepsilon^2) + 2\alpha H_t^*$. Finally, for the same reason as in the baseline case, $\partial\mathrm{Var}[W_t|I_t]/\partial e_1 = 0$. Therefore, the period-$t$ employment contract is (e_{1t}^*, b_t^*) with $e_{1t}^* = \alpha(\xi_1 + \gamma_1\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1})$ and

$$b_t^* = b_t^0\left\{\alpha\left[1 + (\gamma_2/\xi_2)\sum\nolimits_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\right] - (1/\xi_2)R_{HK,t}^* - R_{CC,t}^* - (r\alpha/\xi_2^2)H_t^*\right\},$$

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

A.5.3 Identification

The share α is pinned down by the ratio of firm wages to revenues. Since now $\mathbb{E}[w_{it}] = \alpha\mathbb{E}[y_{it}]$, it follows that $b_t^* = \mathbb{E}[v_{it}]/\mathbb{E}[y_{it}] = \alpha\mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$. Thus, piece rates are identified from α and a panel of wages and their variable components. To identify the variance parameters $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$, note that the wage residual in period t is now $r_{it} = (\alpha - b_t^*)\mathbb{E}[\theta_{it}|I_{it}] + b_t^*(\theta_{it} + \varepsilon_{it})$. The same steps as those in the derivation of the second moments of the wage distributions in the baseline model show that $\mathrm{Var}[w_{it}] = \alpha^2(\sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2) + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ and $\mathrm{Cov}[w_{it}, w_{it+s}] = \alpha^2(\sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2) + \alpha b_t^*\sigma_t^2$. The rest of the identification argument is the same as in the baseline model.

A.6 Extension: General Cost Function

Now we consider the case in which $c(e_1, e_2) = (\rho_1 e_1^2 + 2\eta e_1 e_2 + \rho_2 e_2^2)/2$ with $\rho_1, \rho_2 > 0$. By redefining e_1 as $e_1/\sqrt{\rho_1}$ and e_2 as $e_2/\sqrt{\rho_2}$, this case is equivalent to the one in which $c(e_1, e_2) = (e_1^2 + 2\widehat{\eta}e_1 e_2 + e_2^2)/2$ with $\widehat{\eta} = \eta/\sqrt{\rho_1\rho_2}$, the rates at which effort in the simple and complex tasks increase output are $\xi_1\sqrt{\rho_1}$ and $\xi_2\sqrt{\rho_2}$, respectively, and the rates of human capital accumulation in the simple and complex tasks are $\gamma_1\sqrt{\rho_1}$ and $\gamma_2\sqrt{\rho_2}$, respectively. Thus, we set $\rho_1 = \rho_2 = 1$ in what follows. We also assume that $\eta^2 < 1$.[49]

[49] When $\eta^2 \geq 1$, the complementarity or substitutability between tasks is strong enough that a change in the effort in one task changes the marginal cost of effort in the other task by more than it changes the marginal cost in the task itself, making the worker's problem ill-behaved.

Learning about Ability and Effort in the Complex Task. The process of learning about ability is the same as in the baseline model ($\eta = 0$). Likewise, the equilibrium is unique, symmetric, and such that effort choices and piece rates depend only on time. Suppose workers face a sequence $\{(e_{1t}, b_t)\}_{t=0}^{T}$ of employment contracts in which effort choices in the simple task and piece rates depend only on time and consider a worker's choice of effort in the complex task in period t, e_{2t}, when the worker's future effort choices in this task depend only on time. Since now the marginal cost of effort e_2 in the complex task when effort in the simple task is e_1 is $e_2+\eta e_1$, the necessary and sufficient first-order condition for the optimal choice of e_{2t} is $e_{2t} + \eta e_{1t} = \xi_2 b_t + \xi_2 R_{CC,t} + R_{HK,t}$, where $R_{CC,t}$ and $R_{HK,t}$ are still given by (7).

Equilibrium Employment Contracts. As in the baseline model, we use a backward induction argument to derive the equilibrium employment contracts and show that they are symmetric across workers and such that efforts in the simple task and piece rates depend only on time. Here, we only discuss the induction step in the derivation of the equilibrium employment contracts. It is straightforward to adapt the induction step to derive the equilibrium employment contracts in the last period and show that efforts in the simple task and pieces rates are the same for all workers and (trivially) depend only on T.

Let $0 \le t < T$ and suppose the equilibrium efforts and piece rates from period $t+1$ on depend only on time. For each $1 \le \tau \le T - t$, let $b^*_{t+\tau}$ be the equilibrium piece rate in period $t + \tau$, and define $R^*_{CC,t}$ and $R^*_{HK,t}$ as in the baseline model. Then, a worker's effort in the complex task in period t when the period-t employment contract is (e_1, b) is $e_2 = -\eta e_1 + \xi_2 b + \xi_2 R^*_{CC,t} + R^*_{HK,t}$. A worker's equilibrium employment contract in t is the pair (e_1, b) maximizing $V_t = \mathbb{E}[W_t|I_t] - (r/2)\text{Var}[W_t|I_t] - (e_1^2 + 2\eta e_1 e_2 + e_2^2)/2$, where W_t and I_t are as before. We determine the pair (e_1, b) maximizing V_t in what follows. As in the baseline model, this pair is independent of I_t and so the same for all workers. The expression for $\partial\mathbb{E}[W_t|I_t]/\partial b$ is the same as in the baseline model as it is still the case that $\partial e_2/\partial b = \xi_2$. Now note that

$$\begin{aligned}\frac{\partial \mathbb{E}[W_t|I_t]}{\partial e_1} &= \xi_1 + \gamma_1 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} + \frac{\partial e_2}{\partial e_1}\left[\xi_2 + \gamma_2 \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\right] \\ &= \xi_1 - \eta\xi_2 + (\gamma_1 - \eta\gamma_2) \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1};\end{aligned}$$

unlike in the baseline model, effort in the simple task now affects effort in the complex task

$(\partial e_2/\partial e_1 = -\eta)$. As effort in either task does not affect the variance of current or future wages, the partial derivatives $\partial \text{Var}[W_t|I_t]/\partial b$ and $\partial \text{Var}[W_t|I_t]/\partial e_1$ remain the same. Finally, given that $dc(e_1, e_2)/db = (\eta e_1 + e_2)\partial e_2/\partial b = \xi_2(\eta e_1 + e_2)$ and $dc(e_1, e_2)/de_1 = e_1 + \eta(e_2 + e_1\partial e_2/\partial e_1) + e_2\partial e_2/\partial e_1 = (1-\eta^2)e_1$, the necessary and sufficient conditions for the problem of maximizing V_t are then given by $\xi_1 - \eta\xi_2 + (\gamma_1 - \eta\gamma_2)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - (1-\eta^2)e_1 = 0$ and

$$\xi_2^2 + \gamma_2\xi_2\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - rb(\sigma_t^2 + \sigma_\varepsilon^2) - rH_t^* - \xi_2\left(\xi_2 b + \xi_2 R_{CC,t}^* + R_{HK,t}^*\right) = 0.$$

The unique solution to the above first-order conditions is the pair $(e_1, b) = (e_{1t}^*, b_t^*)$ with $e_{1t}^* = (1-\eta^2)^{-1}[\xi_1 - \eta\xi_2 + (\gamma_1 - \eta\gamma_2)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}]$ and b_t^* given by the equilibrium piece rate in the baseline model; the pair (e_{1t}^*, b_t^*) is the equilibrium employment contract in t. To understand why η does not affect equilibrium piece rates, it is useful to recall why in multi-tasking models such as in Holmström and Milgrom [1991] the degree of substitutability (or complementarity) between effort choices matters for equilibrium piece rates. In such models, all effort choices are non-contractable and so must be incentivized by output-contingent contracts. Thus, increasing incentives for effort in one task affects the power of contracts to incentivize effort in other tasks. In our model, since effort in one of the tasks is contractable, the provision of incentives for effort in the task with non-contractable effort is not affected by the level of effort in the task with contractable effort. In the supplementary appendix, we present an alternative multi-tasking model in which both tasks feature non-contractable effort and the degree of substitutability between effort choices matters for incentive provision.

A.7 Extension: Cobb-Douglas Technology

We now show how our model can be viewed as the log version of a model in which the output and human capital technologies are of the standard Cobb-Douglas form.

A.7.1 Setup

We begin by describing the setup. To keep the exposition brief, we just detail what changes in the setup relative to the baseline model.

Production. The output of worker i in period t is $Y_{it} = \Theta_i K_{it} E_{i1t}^{\xi_1} E_{i2t}^{\xi_2}\Omega_{it}$, where Θ_i is the worker's unobserved ability, which we assume is time-invariant for simplicity, E_{i1t} and E_{i2t} are the worker's effort in the simple and complex task, respectively, K_{it} is the worker's human capital, Ω_{it} is an idiosyncratic noise term, and ξ_1 and ξ_2 are positive constants. The ability Θ_i and the noise terms Ω_{it} are drawn from log-normal distributions with parameters $(m_\theta, \sigma_\theta^2)$ and $(0, \sigma_\varepsilon^2)$, respectively.

Human Capital. The human capital of workers evolves over time according to the law of motion $K_{it+1} = B_t K_{it}^\lambda E_{i1t}^{\gamma_1} E_{i2t}^{\gamma_2}$, where B_t is a positive constant, $1 - \lambda \in [0, 1]$, γ_1 and γ_2 are constants, and $K_{i0} \equiv K_0$ is the worker's initial stock of human capital.

Preferences. The lifetime utility of a worker who, from period t on, receives the wages $\{W_{t+\tau}\}_{\tau=0}^{T-1}$ and exerts the efforts $\{E_{1t+\tau}\}_{\tau=0}^{T-t}$ and $\{E_{2t+\tau}\}_{\tau=0}^{T-t}$ in the simple and complex tasks, respectively, is $\sum_{\tau=0}^{T-t} \delta^\tau \left(\ln(W_{t+\tau}) - \ln(E_{1t+\tau})^2/2 - \ln(E_{2t+\tau})^2/2\right)$.[50]

Contracts and Equilibrium. An employment contract for worker i in period t is a pair (E_{i1t}, W_{it}), where W_{it} is the worker's wage schedule in t. We assume that $W_{it} = C_{it} Y_{it}^{b_{it}}$ with $C_{it} \in \mathbb{R}_+$ and $b_{it} \in \mathbb{R}$. Since $b_{it} = (Y_{it}/W_{it}) dW_{it}/dY_{it}$, the elasticity of wages with respect to output, we can thus interpret b_{it} as a piece rate. As in the baseline model, in equilibrium firms make zero expected profits in every period. Hence, if (E_{i1t}, W_{it}) is worker i's equilibrium contract in period t when the public information about the worker is I_{it}, then $\mathbb{E}[Y_{it}|I_{it}] = \mathbb{E}[W_{it}|I_{it}] = C_{it}\mathbb{E}[Y_{it}^{b_{it}}|I_{it}]$, and so $\ln(C_{it}) = \ln(\mathbb{E}[Y_{it}|I_{it}]) - \ln(\mathbb{E}[Y_{it}^{b_{it}}|I_{it}])$. We determine the implications of this fact for (log) wages in what follows.

A.7.2 Equilibrium Characterization

We now characterize the equilibrium.

Learning about Ability. Let $y_{it} = \ln(Y_{it})$, $\theta_i = \ln(\Theta_i)$, $k_{it} = \ln(K_{it})$, $e_{i1t} = \ln(E_{i1t})$, $e_{i2t} = \ln(E_{i2t})$, and $\varepsilon_{it} = \ln(\omega_{it})$. Moreover, let I_{it} be the public information about worker i in period t. Given that $y_{it} = \theta_i + k_{it} + \xi_1 e_{i1t} + \xi_2 e_{i2t} + \varepsilon_{it}$ and ε_{it} is normally distributed with mean 0 and variance σ_ε^2, it follows from the argument in the main text that $\theta_i|I_{it}$ is normally distributed with mean m_{it} and variance σ_t^2, where $m_{it} = \mathbb{E}[\theta_{it}|I_{it}]$ and σ_t^2 have the same expressions as in the baseline model when $\sigma_\zeta^2 = 0$.

[50] Note that unlike in the baseline model, where it is a present-discounted sum of wage payments, here W_t is a wage. Our analysis extends to the case in which the payoff to a worker from wages $\{W_{t+\tau}\}_{\tau=0}^{T-1}$ and efforts $\{E_{1t+\tau}\}_{\tau=0}^{T-t}$ and $\{E_{2t+\tau}\}_{\tau=0}^{T-t}$ is $-\exp\left\{-r\left[\sum_{\tau=0}^{T-t} \delta^\tau \left(\ln(W_{t+\tau}) - \ln(E_{1t+\tau})^2/2 - \ln(E_{2t+\tau})^2/2\right)\right]\right\}$ with $r > 0$. The parameter r is not the workers' coefficient of risk aversion, though.

Effort in the Complex Task. We first derive the implications of free entry of firms for (log) wages. Since in any period t the posterior belief about θ_i is normally distributed with mean $\mathbb{E}[\theta_{it}|I_{it}]$ and variance σ_t^2, we have that $\mathbb{E}\{\exp[a(\theta_i + \varepsilon_{it})]|I_{it}\} = \exp\left(a\mathbb{E}[\theta_i|I_{it}]+a^2(\sigma_t^2 + \sigma_\varepsilon^2)/2\right)$ for all $a \in \mathbb{R}$, and so $\mathbb{E}[Y_{it}^a|I_{it}] = \mathbb{E}[\exp(ay_{it})|I_{it}] = \exp\left(a\mathbb{E}[y_{it}|I_{it}] + a^2(\sigma_t^2 + \sigma_\varepsilon^2)/2\right)$. Hence, $\ln(C_{it}) = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}] + (1 - b_{it}^2)(\sigma_t^2 + \sigma_\varepsilon^2)/2$, and so

$$w_{it} = \ln(W_{it}) = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it}y_{it} + (1 - b_{it}^2)(\sigma_t^2 + \sigma_\varepsilon^2)/2. \tag{18}$$

Thus, as in the baseline model, employment contracts are summarized by the pair (e_{i1t}, b_{it}). Note, however, that the expression for $\ln(W_{it})$ differs from the expression for w_{it} in (3) by the term $(1 - b_{it}^2)(\sigma_t^2 + \sigma_\varepsilon^2)/2$. We determine the implications of this below.

As in the baseline model, the equilibrium is unique, symmetric, and is such that (log) effort choices and piece rates depend only on time. Assume workers face a sequence $\{(e_{1t}, b_t)\}_{t=0}^T$ of employment contracts in which efforts in the simple task and piece rates depend on time. Consider worker i's period-t choice of effort in the complex task, e_{2t}, when the worker's future effort choices in this task depend only on time. The worker chooses e_{2t} to maximize $\sum_{\tau=0}^{T-t} \delta^\tau \mathbb{E}[w_{it+\tau}|h_i^t] - e_{2t}^2/2$, where h_i^t is the worker's history in period t and $w_{it+\tau}$ is given by (18) with $b_{it+\tau} \equiv b_{t+\tau}$ for all $0 \leq \tau \leq T - t$. Since the terms $(1 - b_{t+\tau}^2)(\sigma_t^2 + \sigma_\varepsilon^2)$ in $w_{it+\tau}$ are deterministic and the law of motion for (log) human capital is the same as the law of motion in (2) with $\beta_t = \ln(B_t)$, it follows that the optimal choice of e_{2t} is the same as in the baseline model.

Equilibrium Employment Contracts. We again proceed by backward induction to compute the equilibrium employment contracts. As in the baseline model, a worker's choice of (log) effort in the complex task in period T is $e_2 = \xi_2 b$ if the worker's piece rate is b. Free entry of firms implies that a worker's employment contract in T is the pair (e_1, b) maximizing $V_T = \mathbb{E}[w_T|I_T] - (e_1^2 + e_2^2)/2$, where I_T is as before.[51] Since, by (18), $\mathbb{E}[w_T|I_T] = \mathbb{E}[y_T|I_T] + (1-b^2)(\sigma_T^2 + \sigma_\varepsilon^2)/2 \propto \xi_1 e_1 + \xi_2^2 b + (1 - b^2)(\sigma_T^2 + \sigma_\varepsilon^2)/2$, it follows that the pair maximizing V_T is $(e_1, b) = (e_{1T}^*, b_T^*)$ with $e_{1T}^* = \xi_1$ and $b_T^* = 1/[1 + (1/\xi_2^2)(\sigma_T^2 + \sigma_\varepsilon^2)]$.

Now let $0 \leq t < T$ and suppose that equilibrium piece rates and (log) efforts from period $t + 1$ on depend only on time. For each $1 \leq \tau \leq T - t$, let $b_{t+\tau}^*$ be the equilibrium piece rate in $t + \tau$ and once again define $R_{CC,t}^*$ and $R_{HK,t}^*$ as in (7) with $b_{t+\tau} = b_{t+\tau}^*$ for each $1 \leq$

[51] The variance of (log) wages does not show up in V_T given our preference specification.

$\tau \leq T-t$. Then, a worker's period-t effort in the complex task as a function of the piece rate b in t is $e_2 = \xi_2 b + \xi_2 R^*_{CC,t} + R^*_{HK,t}$ and a worker's equilibrium employment contract in t is the pair (e_1, b) maximizing $V_t = \sum_{\tau=0}^{T-t} \delta^{\tau-1}\mathbb{E}[w_{t+\tau}|I_t] - (e_1^2 + e_2^2)/2$, where I_t is as before. Since $\mathbb{E}[w_t|I_t] = \mathbb{E}[y_t|I_t] + (1-b^2)(\sigma_t^2+\sigma_\varepsilon^2)/2$ and, for all $1 \leq \tau \leq T-t$, $\mathbb{E}[w_{t+\tau}|I_t] = \mathbb{E}[y_{t+\tau}|I_t] + w^0_{t+\tau}$, where $w^0_{t+\tau}$ is a constant term, it follows that

$$\sum_{\tau=0}^{T-t} \delta^\tau \frac{\partial \mathbb{E}[w_{t+\tau}|I_t]}{\partial b} = \xi_2^2 + \xi_2\gamma_2 \sum_{\tau=0}^{T-t} \delta^\tau \lambda^{\tau-1} - b(\sigma_t^2 + \sigma_\varepsilon^2).$$

and that $\sum_{\tau=0}^{T-t} \partial\mathbb{E}[w_{t+\tau}|I_t]/\partial e_1 = \xi_1 + \gamma_1 \sum_{\tau=0}^{T-t} \delta^\tau \lambda^{\tau-1}$. Thus, the pair (e_1, b) maximizing V_t is $(e_1, b) = (e^*_{1t}, b^*_t)$, where $e^*_{1t} = \xi_1 + \gamma_1 \sum_{\tau=0}^{T-t} \delta^\tau \lambda^{\tau-1}$ and b^*_t is the period-t piece rate in the baseline model when the workers' (effective) coefficient of risk aversion is $1/\xi_2^2$.[52]

A.8 Extension: More General Human Capital Process

We now discuss the case in which the law of motion for human capital depends nonparametrically on effort. For simplicity, we assume only the complex task is present. The model is that of the main text with $\xi_1 = \gamma_1 = 0$ except that the law of motion for human capital is $k_{it+1} = \lambda k_{it} + F_t(e_{i2t})$, where $F_t : \mathbb{R} \to \mathbb{R}$ is thrice differentiable and weakly concave with $\sup_{e\in\mathbb{R}} F'_t(e) < \infty$, $F'''_t(e)$ nonpositive and nondecreasing, and $\inf_{e\in\mathbb{R}} F''_t(e) > -\infty$. For ease of notation, in what follows we denote e_{i2t} simply by e_{it}. This case reduces to the case in the main text with only the complex task when $F_t(e) = \gamma_2 e$ for all $t \geq 0$.

A.8.1 Equilibrium Characterization

Given that the main focus here is on extending the identification argument to a more general human capital process, we state the equilibrium characterization of the model considered in this section without proof; see the supplementary appendix for a proof.

Proposition A.2. *Suppose that*

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

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, in the unique equilibrium, piece rates and effort choices in the complex task are the same for all workers and

[52] With the more general preference specification of footnote 50, one can show that the equilibrium is as in the baseline model when workers' coefficient of risk aversion is $(1+r)/\xi_2^2$.

depend only on time. Moreover, piece rates are in the interval $(0,1)$*. Let* e_t^* *be the equilibrium effort in the complex task in period* t *and* b_t^* *be the equilibrium piece rate in the same period. Moreover, let* $R^*_{CC,t}$ *be given by* (7) *with* $b_t = b_t^*$, $R^*_{HK,t} = F_t'(e_t^*)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(b^*_{t+\tau}+R^*_{CC,t+\tau})$, $H_t^* = \sigma_t^2\sum_{\tau=1}^{T-t}\delta^\tau$*, and* $r_t^* = r\{1-[F_t''(e_t^*)/F_t'(e_t^*)]R^*_{HK,t}\}$*. Then* $e_t^* = b_t^* + R^*_{CC,t} + R^*_{HK,t}$ *and*

$$b_t^* = \frac{1}{1+r_t^*(\sigma_t^2+\sigma_\varepsilon^2)}\left[1+F_t'(e_t^*)\sum\nolimits_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - R^*_{HK,t} - R^*_{CC,t} - r_t^*H_t^*\right].$$

The restrictions on the rates $F_t'(e)$ of human capital accumulation are natural if piece rates are to be in the $(0,1)$ interval. These rates cannot be too positive, otherwise piece rates are greater than one. Likewise, these rates have to be sufficiently positive, otherwise the learning-about-ability motive dominates, and we know from Gibbons and Murphy [1992] that this can lead to negative piece rates. Additionally, workers cannot be too risk averse, otherwise the demand for insurance against the lifetime wage risk due to the uncertainty about ability overwhelms all other factors determining equilibrium piece rates, leading to negative piece rates. Finally, since human capital depreciation erodes the impact of human capital accumulation, it must be small.

A.8.2 Identification

We now show that this version of our model is also identified provided that information on workers' performance is available in addition to information on wages, as is the case in the firm-level data sets we examine and in many others commonly used (Frederiksen et al. [2017]). We first consider the simpler case in which the available performance measure is a noisy measure of only a worker's effort. Let then $p_{it} = e_{it} + \eta_{it}$ be the performance measure of worker i in period t observed by the econometrician, where η_{it} is a continuously distributed noise term independent across workers and over time with cumulative distribution function $G(\eta)$ and known mean.[53]

To transparently convey the role of performance information for the recovery of the primitives of interest in this more general case, we keep the discussion here purposely informal. The identification argument proceeds as follows. Note from (3) that $\mathbb{E}[w_{it}] = m_\theta + k_t^* + e_t^*$. Given

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

that $\mathbb{E}[p_{it}] = e_t^* + \mathbb{E}[\eta_{it}]$ and $\mathbb{E}[\eta_{it}]$ is known, both e_t^* and k_t^* in each t are identified from average wages and average performance in t up to m_θ. Observe next that $\mathbb{E}[w_{iT}] - \lambda\mathbb{E}[w_{iT-1}] = k_T^* + e_T^* - \lambda(k_{T-1}^* + e_{T-1}^*) + (1-\lambda)m_\theta$. Thus, λ is identified from the vector $(k_{T-1}^*, e_{T-1}^*, k_T^*, e_T^*)$ up to m_θ. Now, since $k_{t+1}^* = \lambda k_t^* + F_t(e_t^*)$, we can identify $(F_0(e_0^*), \ldots, F_{T-1}(e_{T-1}^*))$ from λ and the sequence of equilibrium efforts and stocks of human capital from 0 to T. Hence, if the functions $F_t(e)$ do not depend on experience or, alternatively, if they do and any of the parameters σ_θ^2, σ_ε^2, σ_ζ^2, r, and λ vary across observable groups of workers so that different choices of effort are induced among different groups of workers in each t, then these functions are identified from $(F_0(e_0^*), \ldots, F_{T-1}(e_{T-1}^*))$ and $(e_0^*, \ldots, e_{T-1}^*)$.[54] The identification of piece rates and the remaining parameters follows by the same argument as in the proof of Proposition 4.

An analogous argument applies when the econometrician observes only a discrete version of p_{it} if $G(\eta)$ is known. Assume that for each t, there exist thresholds $\overline{p}_{1t} < \ldots < \overline{p}_{Kt}$ and that the econometrician observes the discrete performance measure p_{it}^o given by

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

This is a plausible representation of performance scales that are common in firms; see, for instance, the evidence from Baker et al. [1994a]. Since $\mathbb{P}\{p_{it}^o = K\} = 1 - \mathbb{P}\{p_{it} \leq \overline{p}_{Kt}\}$ and $\mathbb{P}\{p_{it}^o = k\} = \mathbb{P}\{p_{it} \leq \overline{p}_{k+1t}\} - \mathbb{P}\{p_{it} \leq \overline{p}_{kt}\}$ for all k, the probabilities $\mathbb{P}\{p_{it} \leq \overline{p}_{1t}\}$ to $\mathbb{P}\{p_{it} \leq \overline{p}_{Kt}\}$ are identified from the probabilities $\mathbb{P}\{p_{it}^o = 1\}$ to $\mathbb{P}\{p_{it}^o = K\}$—that is, from the distribution of the discrete performance measure p_{it}^o in period t. Now observe that since $\mathbb{E}[w_{it}] = m_\theta + k_t^* + e_t^*$ and $\mathbb{P}\{p_{it} \leq \overline{p}_{kt}\} = \mathbb{P}\{\eta_{it} \leq \overline{p}_{kt} - e_t^*\} = G(\overline{p}_{kt} - e_t^*)$ for each k with G strictly increasing and so invertible, we have a linear system of $K+1$ equations

$$k_t^* + e_t^* = \mathbb{E}[w_{it}] - m_\theta$$

$$\overline{p}_{kt} - e_t^* = G^{-1}(\mathbb{P}\{p_{it} \leq \overline{p}_{kt}\}) \text{ for } k \in \{0, \ldots, K\}$$

[54] Our argument holds regardless of the length of the time interval between periods. Thus, when the functions $F_t(e)$ are independent of t, wage and performance information available at higher frequency, and so for more periods overall, would allow us to identify the common function $F(e)$ at a greater number of support points. When the functions $F_t(e)$ depend on t, it is easy to see from the first-order conditions for effort that variation in σ_θ^2, σ_ε^2, σ_ζ^2, r, or λ among workers, say, with different age at entry into the firm or who entered in different periods, would induce variation in effort in each t that would allow us to identify $F_t(e)$ at every possible equilibrium choice of effort in t.

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

The identification argument so far has relied on a specific functional form for the performance measure p_{it}. We now extend this argument to the more general case in which $p_{it} = f_t(e_{it}, k_{it}) + \eta_{it}$, where $f_t : \mathbb{R}^2 \mapsto \mathbb{R}$ is a known continuously differentiable function nondecreasing in each of its arguments such 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 $e \mapsto f_t(e, \alpha - e)$ is surjective for all $\alpha \in \mathbb{R}$. These assumptions trivially hold in the case just discussed.[55] We only consider the case in which the econometrician observes p_{it}, as it will be clear that we can extend the analysis to the case in which the econometrician observes the discrete version p_{it}^o by following the approach discussed above. For each t, we have the system of equations

$$\begin{aligned} e_t^* + k_t^* &= \mathbb{E}[w_{it}] - m_\theta \\ f_t(e_t^*, k_t^*) &= \mathbb{E}[p_{it}] - \mathbb{E}[\eta_{it}] \end{aligned}, \tag{19}$$

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 (19) has a unique solution. Indeed, using the first equation in (19) to solve for k_t^*, the second equation in (19) becomes

$$f_t(e_t^*, \mathbb{E}[w_{it}] - m_\theta - e_t^*) = \mathbb{E}[p_{it}] - \mathbb{E}[\eta_{it}]. \tag{20}$$

Since $e \mapsto f_t(e, \alpha - e)$ is surjective for all $\alpha \in \mathbb{R}$, equation (20) has a solution regardless of m_θ, $\mathbb{E}[w_{it}]$, $\mathbb{E}[p_{it}]$, and $\mathbb{E}[\eta_{it}]$. Now let $h(e) = f_t(e, \mathbb{E}[w_{it}] - m_\theta - e)$. Since $\partial f_t(e,k)/\partial e \neq \partial f_t(e,k)/\partial k$ for all $(e,k) \in \mathbb{R}^2$ implies that $h'(e) \neq 0$ for all $e \in \mathbb{R}$, the solution to (20) is unique. Thus,

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

workers' effort and stock of human capital in each period t are identified from the mean wage and the mean performance measure in t up to m_θ. The rest of the identification argument is as before.

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

Proposition A.3. *Suppose that for each worker i and period t, the econometrician observes the performance measure $p_{it} = f_t(e_{it}, k_{it}) + \eta_{it}$, where $f_t : \mathbb{R}^2 \to \mathbb{R}$ is a known continuously differentiable function nondecreasing in each of its arguments and η_{it} is a continuously distributed noise term independent across workers and over time with known mean. For each period t, the choice of effort in the complex task e_t^* and workers' stock of human capital k_t^* are identified from the mean wage and the mean performance measure in t up to m_θ if, and only if, $\partial f_t(e,k)/\partial e \neq \partial f_t(e,k)/\partial k$ for all $(e,k) \in \mathbb{R}^2$ and $e \mapsto f_t(e, \alpha - e)$ is surjective for all $\alpha \in \mathbb{R}$.*

The availability of the performance measure discussed begs the question of why firms would not offer contracts in which they condition wages not only on output but also on this measure. As

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

argued by Holmström [1979], firms should do so as long as a worker's output is not a sufficient statistic for this additional 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, for instance, the discussion in Baker [1992]. A common explanation for this feature of contracts observed in practice is that although they are observable, performance measures often are not verifiable or are manipulable by workers. When workers' ability is uncertain, although firms cannot or may not wish to explicitly link wages to all performance measures they observe, they can still use them to form expectations about workers' ability. This influences offered contracts even if contracts do not explicitly depend on all 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.

A.9 Extension: Heterogeneous Workers

In the last extension we consider, we allow workers to be either heterogeneous in their ability to perform the complex task or heterogeneous in how their ability contributes to output. Since the analysis of both cases is very similar, we focus on the first case, briefly discussing the second case in our remarks at the end. The case in which workers are heterogeneous in their ability to perform the simple task also follows immediately.

A.9.1 Setup and Equilibrium

Workers are heterogeneous in the rate at which their effort in the complex task affects output but homogeneous in all other model parameters. There are $J \geq 1$ such types of workers. Let $\pi_j \in (0,1)$ be the fraction of workers of type $j \in \{1,\ldots,J\}$ and ξ_{j2} be the rate at which effort in the complex task affects output for type-j workers, with $0 \leq \xi_{12} < \xi_{22} < \cdots < \xi_{J2}$. The rates $\{\xi_{j2}\}_{j=1}^{J}$ are observable to agents in the model but not to the econometrician. As workers are homogeneous with respect to ξ_1, equilibrium efforts in the simple task are the same for all workers and given by the equilibrium efforts in the baseline model. For type-j workers, equilibrium piece rates and effort choices in the complex task are as in the baseline model with $\xi_2 = \xi_{j2}$.

Let e^*_{1t} be the period-t effort in the simple task and e^*_{j2t} and b^*_{jt} be, respectively, the period-t effort in the complex task and period-t piece rate for workers of type j. By the same argument as in the baseline model, it follows that $e^*_{j2t} = \xi_{j2} + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_{j2})[(\sigma^2_t + \sigma^2_\varepsilon) b^*_{jt} + H^*_t]$, where σ^2_t and H^*_t are as in the baseline model. Now let $s^*_{jt} = (1 + e^*_{j2t})/(1 + e^*_{1t})$ be the task complexity of the job that workers of type j perform in period t. When piece rates are small, as we observe in the data, s^*_{jt} is approximately equal to $(1 + \xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1})^{-1}[1 + \xi_{j2} + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_{j2}) H^*_t]$, which strictly increases with j; that is, at every experience, workers with higher productivity in the complex task are assigned to higher-complexity jobs.

A.9.2 Identification

Assume that δ and ξ_1 are known and suppose that in addition to information on total and variable pay, we have information on a worker's job as defined by its complexity; see Section 6.5. Let w_{ijt} and v_{ijt} be, respectively, the total and variable pay of worker i of type j in period t. Since w_{ijt} and v_{ijt} are normally distributed for each type j, the distributions of total and variable pay in each period are finite mixtures of normal distributions. By the same argument as in Appendix A.4, the mixture weights $\{\pi_j\}_{j=1}^J$ and the mean total and variable pay, $\mathbb{E}[w_{ijt}]$ and $\mathbb{E}[v_{ijt}]$, are identified for each type j and period t. Therefore, as in the baseline model, the piece rate of type-j workers in t is identified as $b^*_{jt} = \mathbb{E}[v_{ijt}]/\mathbb{E}[w_{ijt}]$, and we can identify the parameters $(\sigma^2_\theta, \sigma^2_\varepsilon, \sigma^2_\zeta)$, and so the variances σ^2_t for all t, from the piece rates and the second moments of the distributions of wages of a given type of workers in different time periods.[57] Since $s^*_{jT} = (1 + \xi_{j2} b^*_{jT})/(1 + \xi_1)$ increases strictly with j, and so type-J workers, and only them, occupy the highest-level job in period T, it follows that ξ_{J2} is identified from b^*_{JT} and s^*_{JT}. Then, r can be recovered from σ^2_T, σ^2_ε, and $b^*_{jT} = 1/[1 + (r/\xi^2_{2J})(\sigma^2_T + \sigma^2_\varepsilon)]$. From this, it follows that for each $1 \leq j \leq J$, the parameter ξ_{j2} is identified from r, σ^2_T, σ^2_ε, and $b^*_{jT} = 1/[1 + (r/\xi^2_{jT})(\sigma^2_T + \sigma^2_\varepsilon)]$. The rest of the identification argument for each type j of workers is the same as in the baseline model.

A.9.3 Remarks

We can also extend our analysis to a setup in which workers are heterogeneous in the rate ξ_θ at which their ability θ affects output—our baseline model is such that $\xi_\theta = 1$ for all workers. Since we can redefine worker ability to absorb the rate ξ_θ into it, this setup is equivalent to one in which workers are heterogeneous in the uncertainty σ^2_θ about their ability. Intuitively, a higher ξ_θ means

[57] By this argument, it would be straightforward to allow these parameters to vary across types of workers.

that a worker's performance is more informative about a worker's ability, which is equivalent to a higher σ_θ^2. The equilibrium in this version of the model is such that workers are heterogeneous in their piece rates, and thus in their effort in the complex task. One can show that when piece rates are small, effort in the complex task strictly increases with σ_θ^2, as a higher σ_θ^2 translates into a smaller σ_t^2 for all t. Thus, as in the above, workers with higher productivity in the complex task are assigned to higher-complexity jobs. Identification of this setup proceeds along the same lines as above.

A.10 Additional Results and Figures

We present here additional results and the omitted figures referenced in Section 6.

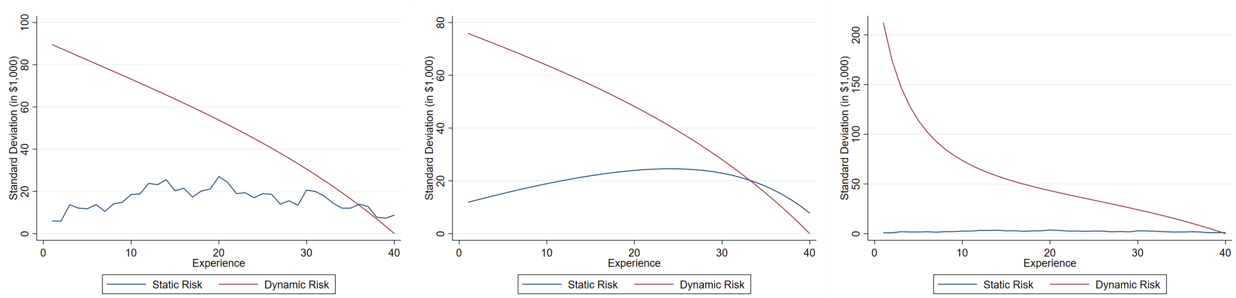
Life-Cycle Wage Risk. The output y_t realized in any period t has two effects on a worker's present-discounted value (PDV) of wages. First, it determines performance pay in t, $b_t^* y_t$. Second, it leads firms and workers to revise the mean of their beliefs about a worker's ability by $[\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)]\{y_t-\mathbb{E}[y_t|I_t]\}$, where $y_t-\mathbb{E}[y_t|I_t]$ is the innovation in the information about ability—that is, the signal z_t about ability extracted from realized output y_t net of expected ability $\mathbb{E}[\theta_{it}|I_t]$—and $\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)$ is the weight placed on new output signals by the updating rule in (4). Since beliefs follow a martingale process, such a change in beliefs persists in expectation over time. Thus, $[\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)]\sum_{\tau=1}^{T-t}\delta^\tau\{y_t-\mathbb{E}[y_t|h_t]\}$ is a measure of how this belief revision affects the expected PDV of wages in t. The standard deviations of these two effects of output on wages are $b_t^*\sqrt{(\sigma_t^2+\sigma_\varepsilon^2)}$ (*static risk*) and $[\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)](\sum_{\tau=1}^{T-t}\delta^\tau)\sqrt{(\sigma_t^2+\sigma_\varepsilon^2)}$ (*dynamic risk*), respectively. They reflect the variability of the expected PDV of wages in t due to the variability of output in t—which influences performance pay as captured by the first measure—and to the implied variability of a worker's future reputation—as captured by the second measure.

Figure A.3 shows the values of these two measures for our three parameterizations in Table 3. For all of them, the dynamic wage risk induced by learning about ability is notably larger than the static wage risk induced by performance pay, especially early in the life cycle. For the two parameterizations in panels a and b that do not impose a fast speed of learning, this dynamic wage risk at $t=0$ is close to 90 thousand dollars (per year) with exogenous piece rates and close to 80 thousand dollars with endogenous piece rates. Under both parameterizations, dynamic wage risk declines fairly linearly over the life cycle. By contrast, static wage risk never exceeds 25 thousand dollars (per year) and, since it is proportional to piece rates, roughly follows their shape.

When we impose a fast speed of learning as per our third parameterization in column 3 of Table 3, we estimate a substantially *higher* degree of risk associated with learning about ability, as panel c shows—dynamic risk early in a worker's career exceeds 200 thousand dollars. Young workers face such large risk because firms rapidly update their beliefs about workers' ability early on, which leads to a high variability in beliefs and so wages. The convex shape of the profile of dynamic risk arises because beliefs quickly become more precise and less volatile given that ability is learned fast. Under this parameterization, ability is governed by a random walk process thus learning continues throughout the life cycle. Indeed, a sizable dynamic risk persists even at 20 and 30 years of experience, and it is much larger than the static risk from performance pay.

Figure A.3: Sources of Dispersion in Lifetime Wages

(a) With Exogenous Piece Rates (b) With Endogenous Piece Rates (c) With Faster Learning



For all three parameterizations, the dynamic wage risk due to learning about ability is much larger than static one due to performance pay. When learning is slow, the difference in magnitude between the two types of risk is not as stark. But because uncertainty about ability is persistent, dynamic wage risk remains sizable over most of the life cycle. With fast learning, highly volatile beliefs lead to an even larger degree of dynamic wage risk over the first half of the life cycle.

Additional Figures. We collect here figures omitted from the main text.

Figure A.4: Results for Parameterization with Faster Learning

(a) Wage Growth Decomposition (b) Perf. Pay and Wage Growth (c) Wage Variance

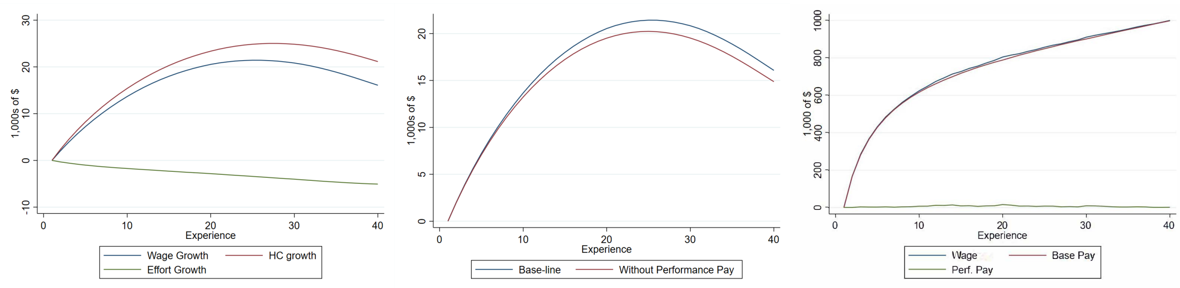


Figure A.5: Variance of Wages and Piece Rates without Uncertainty about Ability

(a) Variance (Fixed P. Rates) (b) Variance (Endogenous P. Rates) (c) Endogenous P. Rates

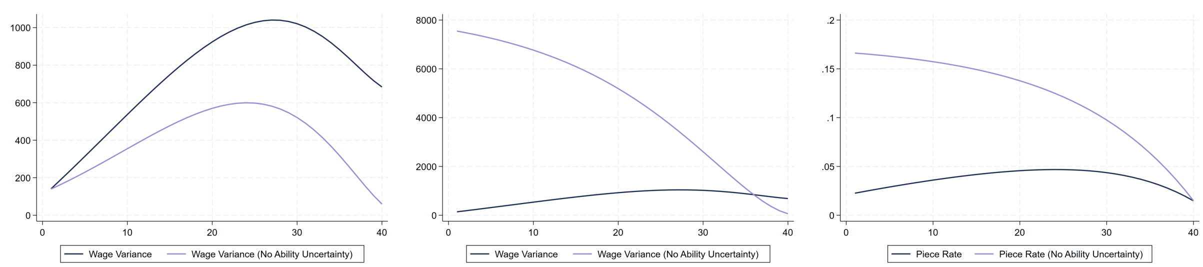
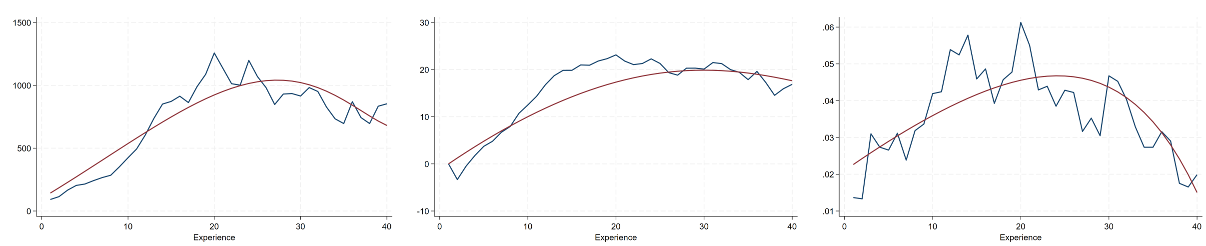


Figure A.6: Fit of Model with Endogenous Piece Rates and Simple and Complex Tasks

(a) Variance of Wages (b) Wage Growth (c) Piece Rates



References

Baker, G. P. (1992): "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 100 (3), 598–614.

Cunha, F., J. J. Heckman, and S. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78 (3), 883–931.

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

Gibbs, M., and W. Hendricks (2004): "Do Formal Salary Systems Really Matter?" *ILR Review*, 58 (1), 71–93.

Gordon, W. B. (1972): "On the Diffeomorphisms of Euclidian Space," *American Mathematical Monthly*, 79 (7), 755–759.

Holmström, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10 (1), 74–91.

Holmström, B., and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 7, 24–52.

Lemieux, T., W.B. MacLeod, and D. Parent (2009): "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 124 (1), 1–49.

Teicher, H. (1963): "Identifiability of Finite Mixtures," *Annals of Mathematical Statistics*, 34 (4), 1265–1269.

B Supplementary Appendix

We first describe the alternative multi-tasking model in which both tasks feature non-contractable effort (Section B.1) and then describe the model with observable productivity shocks (Section B.2). We conclude by deriving the equilibrium of the version of our model with a more general human capital process (Section B.3) and discussing the extension in which an observable but unverifiable performance measure is available (Section B.4).

B.1 Extension: Alternative Multi-Tasking Model

We now consider a version of our model in which effort in both tasks is non-contractable, so that both effort choices must be incentivized by output-contingent contracts. In this case, the degree of substitutability between effort choices will matter for the incentive power of contracts. The model we consider here also differs from the baseline model— namely, the model in the main text—in that it allows a worker's ability and human capital to contribute differently to each task.

B.1.1 Setup

Each task has its own output and firms care about a worker's total output. The output of worker i in task $\ell \in \{1, 2\}$ in period t is $y_{i\ell t} = \xi_{\ell\theta}\theta_i + \xi_{\ell k}k_{it} + \xi_{\ell e}e_{i\ell t} + \varepsilon_{i\ell t}$, where θ_i is the worker's time-invariant unobserved ability, k_{it} is the worker's human capital, $e_{i\ell t}$ is the worker's effort in task ℓ, and $\varepsilon_{i\ell t}$ is an idiosyncratic noise term. The parameter $\xi_{\ell\theta}$ captures the contribution of ability to output in task ℓ, the parameter $\xi_{\ell k}$ captures the contribution of human capital to output in task ℓ, and the parameter $\xi_{\ell e}$ captures the contribution of effort to output in task ℓ. Worker i's ability is draw from normal distribution with mean m_θ and variance σ^2_θ and $\varepsilon_{i\ell t}$ is normally distributed with mean zero and variance $\sigma^2_{\ell\varepsilon}$. The law of motion for workers' stock of human capital is the same as in the baseline model, and so are worker preferences. In particular, the cost of the effort pair (e_1, e_2) is $c(e_1, e_2) = (e_1^2/2 + \eta e_1 e_2 + e_2^2/2)$ with $\eta^2 < 1$. Now, an employment contract for worker i in period t consists of a wage schedule $w_{it} = c_{it} + b_{i1t}y_{i1t} + b_{i2t}y_{i2t}$, where c_{it} is the fixed component of worker i's wage in t and $b_{i\ell t}$ is worker i's piece rate for task ℓ in t. We again consider pure-strategy perfect Bayesian equilibria. Free entry of firms together with their risk neutrality implies that $c_{it} = (1 - b_{i1t})\mathbb{E}[y_{i1t}|I_{it}] + (1 - b_{i2t})\mathbb{E}[y_{i2t}|I_{it}]$, where I_{it} is the public portion of worker i's history in t. Thus, employment contracts can be described by a pair of piece rates, one for each task.

Remarks. The assumption of constant worker ability is done for simplicity. In what follows, we also assume that $\xi_{1e} = \xi_{2e} = 1$; we can easily extend our analysis to the case in which, as in the baseline model, the contribution of effort to output differs across tasks. We assume that worker ability is common across tasks but can matter differently for each task; we can extend the model to allow for task-specific abilities. Human capital is also common across tasks and it can also matter differently for each task. A more general model allowing for task-specific human capital is possible. Such an extension is straightforward and does not affect the substance of our results. We can also extend the model to the case in which output shocks are correlated across tasks.

B.1.2 Equilibrium Characterization

We now characterize the equilibrium.

Learning about Ability. Consider worker i in period t, whose equilibrium effort choices and human capital in t are e^*_{1t}, e^*_{2t}, and k^*_t, respectively; as in the main text, we omit the dependence of effort choices and human capital on i for ease of notation. Let $z_{i\ell t} = (y_{i\ell t} - \xi_{\ell k} k^*_t - e^*_{\ell t})/\xi_{\ell\theta}$ be the part of the worker's period-t output in task $\ell \in \{1, 2\}$ that is not explained by the worker's human capital and effort in ℓ. Then, $z_{i\ell t} = \theta + \widetilde{\varepsilon}_{i\ell t}$ with $\widetilde{\varepsilon}_{i\ell t} = \varepsilon_{i\ell t}/\xi_{\ell\theta}$ is the signal about the worker's ability extracted from the worker's output in task ℓ. As in the baseline model, it then follows that posterior beliefs about a worker's ability in any period are normally distributed and so fully described by their conditional mean m_{it}, namely, the worker's reputation, and variance σ^2_{it}. Now let $\sigma^2_\varepsilon = \sigma^2_{1\varepsilon}\sigma^2_{2\varepsilon}/(\xi^2_{2\theta}\sigma^2_{1\varepsilon} + \xi^2_{1\theta}\sigma^2_{2\varepsilon})$, $\omega_1 = \xi^2_{1\theta}\sigma^2_{2\varepsilon}/(\xi^2_{2\theta}\sigma^2_{1\varepsilon} + \xi^2_{1\theta}\sigma^2_{2\varepsilon})$, $\omega_2 = \xi^2_{2\theta}\sigma^2_{1\varepsilon}/(\xi^2_{2\theta}\sigma^2_{1\varepsilon} + \xi^2_{1\theta}\sigma^2_{2\varepsilon})$, and $z_{it} = \omega_1 z_{i1t} + \omega_2 z_{i2t}$. One can show that the laws of motion for m_{it} and σ^2_{it} are[58]

$$m_{it+1} = \sigma^2_\varepsilon m_{it}/(\sigma^2_{it} + \sigma^2_\varepsilon) + \sigma^2_{it} z_{it}/(\sigma^2_{it} + \sigma^2_\varepsilon) \;\text{ and }\; \sigma^2_{it+1} = \sigma^2_{it}\sigma^2_\varepsilon/(\sigma^2_{it} + \sigma^2_\varepsilon).$$

Note that if $\xi_{1\theta} = 0$, and ability does not matter for output in task 1, then $\omega_1 = 0$, $\omega_2 = 1$, and $\sigma^2_\varepsilon = \sigma^2_{2\varepsilon}/\xi^2_{2\theta}$, the variance of $\widetilde{\varepsilon}_{i2t}$. In this case, the above formulas reduce to the ones in (4)

[58] Since noise terms are independent across tasks, we can break the belief-updating process in any period into two parts. First, agents update their beliefs about a worker's ability θ by using z_{i1t}, then they update their beliefs about θ using z_{i2t}. We obtain the above formulas by applying the formulas used in the baseline case.

if $\xi_{2\theta} = 1$. This result is expected, as in this case firms can learn about a worker's ability only through the worker's performance in task 2 and the rate at which ability contributes to output in task 2 is one. Similar results hold if $\xi_{2\theta} = 0$ and $\xi_{1\theta} = 1$. Also note that σ_ε^2 strictly decreases with both $\xi_{1\theta}$ and $\xi_{2\theta}$. Intuitively, increasing the importance of ability for either task makes workers' performance more informative about ability.

As in the baseline model, since σ_{it}^2 evolves independently of z_{it}, and so is common for all workers in t, we can suppress the subscript i and denote this variance by σ_t^2. For each $0 \leq t \leq T$ and $0 \leq \tau \leq T - t$, let $\Sigma_{t+\tau} = \sigma_t^2/(\tau\sigma_t^2 + \sigma_\varepsilon^2)$. Iterating on the law of motion for m_{it}, we obtain that worker i's reputation in period $t + \tau$ given reputation m_{it} in t is

$$m_{it+\tau} = \sigma_\varepsilon^2 m_{it}/(\tau\sigma_t^2 + \sigma_\varepsilon^2) + \Sigma_{t+\tau} \sum_{s=0}^{\tau-1} z_{it+s}.$$

Effort Choices. As in the baseline model, the equilibrium is unique, symmetric, and has the property that effort choices and employment contracts depend only on time. Suppose workers face a sequence $\{(b_{1t}, b_{2t}\}_{t=0}^{T}$ of employment contracts in which pieces depend only on time and consider a worker's choices of effort in tasks 1 and 2 in period t, e_{1t} and e_{2t}, when the worker's future effort choices in both tasks depend only on time. Define the terms $R_{CC,\ell t}$ and $R_{HK,\ell t}$, with $\ell \in \{1, 2\}$, to be such that

$$R_{CC,\ell t} = \sum_{\tau=1}^{T-t} \delta^\tau \left[\xi_{1\theta}(1 - b_{1t+\tau}) + \xi_{2\theta}(1 - b_{2t+\tau})\right] (\omega_\ell/\xi_{\ell\theta})\Sigma_{t+\tau};$$
$$R_{HK,\ell t} = \gamma_\ell \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \left[\xi_{1k}(b_{1t+\tau} + R_{CC,1t+\tau}) + \xi_{2k}(b_{2t+\tau} + R_{CC,2t+\tau})\right],$$

where γ_ℓ is the rate of human capital accumulation in task ℓ. The necessary and sufficient first-order conditions for effort are

$$e_{1t} + \eta e_{2t} = b_{1t} + R_{CC,1t} + R_{HK,1t};$$
$$e_{2t} + \eta e_{1t} = b_{2t} + R_{CC,2t} + R_{HK,2t}.$$

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

To understand the term $R_{CC,\ell t}$, note that, at the margin, higher $e_{\ell t}$ increases the expected period-t signal about a worker's ability resulting from the worker's performance in task ℓ by $1/\xi_{\ell\theta}$. From the law of motion for a worker's reputation given above, at the margin, a higher signal about ability resulting from performance in task ℓ increases a worker's expected reputation in period $t+\tau$, with $1 \leq \tau \leq T-t$, by $\omega_\ell \Sigma_{t+\tau}$. Thus, at the margin, higher $e_{\ell t}$ increases a worker's expected reputation in $t+\tau$ by $(\omega_\ell/\xi_{\ell\theta})\Sigma_{t+\tau}$. Now note that the signal about ability in one task influences future fixed pay in both tasks and that the importance of this signal for task ℓ is proportional to the importance of ability for performance in ℓ as measured by $\xi_{\ell\theta}$. To understand the term $R_{HK,\ell t}$, note that effort in task ℓ changes human capital at rate γ_ℓ and that the importance of human capital for ℓ is proportional to $\xi_{\ell k}$—as in the baseline model, higher human capital affects both the variable component of a worker's future wages and the future signals about the worker's ability.

Solving the above system of equations for e_{1t} and e_{2t}, we obtain that

$$e_{1t} = (1-\eta^2)^{-1}\left[b_{1t} + R_{CC,1t} + R_{HK,1t} - \eta\left(b_2 + R_{CC,2t} + R_{HK,2t}\right)\right]; \tag{21}$$

$$e_{2t} = (1-\eta^2)^{-1}\left[b_{2t} + R_{CC,2t} + R_{HK,2t} - \eta\left(b_1 + R_{CC,1t} + R_{HK,1t}\right)\right]. \tag{22}$$

Note that $\partial e_{\ell t}/\partial b_{\ell t} = 1/(1-\eta^2) > 0$ and $\partial e_{\ell t}/\partial b_{-\ell t} = -\eta/(1-\eta^2)$ for $\ell \in \{1,2\}$, where we use the subscript $-\ell$ to denote the task other than task ℓ. Thus, an increase in a task's piece rate increases effort in the task. Whether such an increase increases or decreases effort in the other task depends on whether tasks are complements ($\eta < 0$) or substitutes ($\eta > 0$). If tasks are complements, then increasing the piece rate at one task increases effort at the other task. If, instead, tasks are substitutes, then increasing the piece rate for one task decreases effort in the other task.

Equilibrium Employment Contracts. We use a backward induction argument to derive the equilibrium employment contracts and show that they are symmetric across workers and such that piece rates in both tasks depend only on time. Here, we only discuss the induction step in the derivation of the equilibrium employment contracts. Since in the last period our multi-tasking model reduces to the static multi-tasking model of Holmström and Milgrom [1991], last-period employment contracts and effort choices are the same for all workers and (trivially) depend only on T.

Let $0 \leq t < T$ and suppose the equilibrium employment contracts and effort choices from period $t+1$ on depend only on time. For each $1 \leq \tau \leq T-t$ and ℓ, let $b^*_{\ell t+\tau}$ be the equilibrium piece rate for task ℓ in period $t+\tau$. Also, let $R^*_{CC,\ell t}$ and $R^*_{HK,\ell t}$ be respectively given by $R_{CC,\ell t}$ and $R_{HK,\ell t}$ with $b_{\ell t+\tau} = b^*_{\ell t+\tau}$ for all $1 \leq \tau \leq T-t$ and ℓ. A worker's effort in task ℓ in period t when the employment contract is (b_1, b_2) is defined implicitly by $e_\ell = -\eta e_{-\ell} + b_\ell + R^*_{CC,\ell t} + R^*_{HK,\ell t}$. If we let w_t is a worker's wage in t and $W_t = \sum_{\tau=0}^{T-t} \delta^\tau w_{t+\tau}$, then a worker's equilibrium employment contract in t is the pair (b_1, b_2) maximizing $V_t = \mathbb{E}[W_t|I_t] - (r/2)\text{Var}[W_t|I_t] - c(e_1, e_2)$, where I_t has the same definition as in the baseline model. We determine the pair (b_1, b_2) maximizing V_t in what follows. In the same way as in the baseline model, this pair is independent of I_t and so the same for all workers.

First note that since workers capture the entire value of their matches with firms, then

$$\frac{\partial \mathbb{E}[W_t|I_t]}{\partial b_\ell} = \sum\nolimits_{i=1,2} \frac{\partial e_i}{\partial b_\ell} \left[1 + \gamma_i(\xi_{1k} + \xi_{2k}) \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}\right].$$

Now note that

$$\text{Var}[W_t|I_t] = \sum\nolimits_{i=1,2} b_i^2(\xi_{i\theta}^2\sigma_t^2 + \sigma_{i\varepsilon}^2) + 2b_1b_2\xi_{1\theta}\xi_{2\theta}\sigma_t^2 + 2\sum\nolimits_{\tau=1}^{T-1} \delta^\tau \text{Cov}[w_t, w_{t+\tau}|I_t] + \text{Var}_0,$$

where Var_0 is independent of (b_1, b_2), and, as in the baseline case, $\text{Cov}[w_t, w_{t+\tau}|I_t]$ is linear in b_1 and b_2. Thus,

$$\frac{\partial \text{Var}[W_t|I_t]}{\partial b_\ell} = 2b_\ell(\xi_{\ell\theta}^2\sigma_t^2 + \sigma_{\ell\varepsilon}^2) + 2b_{-\ell}\xi_{1\theta}\xi_{2\theta}\sigma_t^2 + 2H_{\ell t}^*,$$

where $H_{\ell t}^* = \sum_{\tau=1}^{T-1} \delta^{\tau-1}\partial\text{Cov}[w_t, w_{t+\tau}|I_t]/\partial b_\ell$ is independent of b_1 and b_2. Since

$$\frac{\partial c(e_1, e_2)}{\partial b_\ell} = \left(b_\ell + R_{CC,\ell t}^* + R_{HK,\ell t}^*\right)\frac{\partial e_\ell}{\partial b_\ell} + \left(b_{-\ell} + R_{CC,-\ell t}^* + R_{HK,-\ell t}^*\right)\frac{\partial e_{-\ell}}{\partial b_\ell},$$

the necessary and sufficient first-order conditions for the problem of maximizing V_t are

$$\begin{aligned}
&\sum\nolimits_{\ell=1,2} \frac{\partial e_\ell}{\partial b_1}\left[1 + \gamma_\ell(\xi_{1k} + \xi_{2k})\sum\nolimits_{\tau=1}^{T-1} \delta^\tau\lambda^{\tau-1} - R_{HK,\ell t}^* - R_{CC,\ell t}^*\right] \\
&\qquad - b_1\left[\frac{\partial e_1}{\partial b_1} + r(\xi_{1\theta}^2\sigma_t^2 + \sigma_{1\varepsilon}^2)\right] - b_2\left(\frac{\partial e_2}{\partial b_1} + r\xi_{1\theta}^2\xi_{2\theta}^2\sigma_t^2\right) - rH_{1t}^* = 0; \\
&\sum\nolimits_{\ell=1,2} \frac{\partial e_\ell}{\partial b_2}\left[1 + \gamma_\ell(\xi_{1k} + \xi_{2k})\sum\nolimits_{\tau=1}^{T-1} \delta^\tau\lambda^{\tau-1} - R_{HK,\ell t}^* - R_{CC,\ell t}^*\right] \\
&\qquad - b_2\left[\frac{\partial e_2}{\partial b_2} + r(\xi_{2\theta}^2\sigma_t^2 + \sigma_{2\varepsilon}^2)\right] - b_1\left(\frac{\partial e_1}{\partial b_2} + r\xi_{1\theta}^2\xi_{2\theta}^2\sigma_t^2\right) - rH_{2t}^* = 0.
\end{aligned}$$

To finish, $b_{\ell t}^0 = 1/[1 + r(1-\eta^2)(\xi_{\ell\theta}^2\sigma_t^2 + \sigma_{\ell\varepsilon}^2)]$ and $\mathcal{W}_{\ell t} = 1 + \gamma_\ell(\xi_{1k} + \xi_{2k})\sum_{\tau=1}^{T-1} \delta^\tau\lambda^{\tau-1} - R_{HK,\ell t}^* - R_{CC,\ell t}^*$. Given that $\partial e_{-\ell}/\partial b_\ell = -\eta\partial e_\ell/\partial b_\ell$, we can rewrite the above first-order conditions as

$$\begin{aligned}
b_1 &= b_{1t}^0\left[\mathcal{W}_{1t} - \eta(\mathcal{W}_{2t} - b_2) - r(1-\eta^2)(H_{1t}^* + b_2\xi_{1\theta}\xi_{2\theta}\sigma_t^2)\right]; \\
b_2 &= b_{2t}^0\left[\mathcal{W}_{2t} - \eta(\mathcal{W}_{1t} - b_1) - r(1-\eta^2)(H_{2t}^* + b_1\xi_{1\theta}\xi_{2\theta}\sigma_t^2)\right].
\end{aligned}$$

This last system of equations admits a unique solution (b_{1t}^*, b_{2t}^*), which is independent of I_t and is the equilibrium employment contract in t. Note that the expression for $H_{\ell t}^*$ does not matter for the derivation of equilibrium piece rates. One can show that $H_{\ell t}^*$ is equal to

$$\xi_{\ell\theta}\sigma_t^2 \sum_{\tau=1}^{T-t} \delta^{\tau-1} \left[\xi_{1\theta} b_{1t+\tau}^* + \xi_{2\theta} b_{2t+\tau}^* + (1 - b_{1t+\tau}^* - b_{2t+\tau}^*) \frac{\tau(\omega_1\xi_{1\theta} + \omega_2\xi_{2\theta})\sigma_t^2 + \sigma_\varepsilon^2}{\tau\sigma_t^2 + \sigma_\varepsilon^2} \right].$$

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

By definition, $\mathcal{W}_{\ell t}$ is the wedge in period t between the marginal social benefit of effort in task ℓ and the marginal private benefit of effort in the same task. A piece rate for task ℓ in period t equal to $\mathcal{W}_{\ell t}$ would induce workers to exert the first-best level of effort in ℓ. As in the baseline case, the piece rate for task ℓ in period t is proportional to $\mathcal{W}_{\ell t}$ minus a term, $r(1-\eta^2)(H_{\ell t}^* + b_{-\ell t}^* \xi_{1\theta}^2 \xi_{2\theta}^2 \sigma_t^2)$, that reflects the insurance workers demand against the life-cycle wage risk due to uncertainty and learning about ability. Also as in the baseline model, the constants of proportionality b_{1t}^0 and b_{2t}^0 capture the standard risk-incentives trade-off. In contrast to the baseline model, the insurance component of the piece rate for task ℓ in t features an additional term that depends on the period-t piece rate for the other task. This is intuitive. Because ability is common across tasks, uncertainty about ability implies that an increase in the piece rate in a task increases the risk associated with (the contemporaneous) performance in the other task. Another difference from the baseline model is that the piece rate for task ℓ in period t features an additional term proportional to $-\eta(W_{-\ell t} - b_{-\ell t})$. This term captures both the interdependence in the human capital accumulation process across tasks—by exerting effort in one task, workers affect their productivity in both tasks—and the fact that providing incentives for effort in one task affects the incentives for effort in the other task.

B.2 Extension: Productivity Shocks

We now consider an extension of our model that allows for observable productivity shocks.

B.2.1 Environment and Equilibrium

The environment is as in the baseline model, except that $y_{it} = \eta_{it} + \theta_{it} + k_{it} + \xi_1 e_{i1t} + \xi_2 e_{i2t} + \varepsilon_{it}$, where η_{it} is an idiosyncratic productivity shock to worker i in period t that is observed after firms offer employment contracts to workers. We assume that η_{it} is normally distributed with mean zero and variance σ_η^2.[59] Let $\widehat{y}_{it} = y_{it} - \eta_{it}$ be worker i's output in period t net of the productivity shock η_{it}. By definition, $\widehat{y}_{it}$ is worker i's period-t output in the baseline model. Free entry of firms implies that $w_{it} = (1-b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it}y_{it} = (1-b_{it})\mathbb{E}[\widehat{y}_{it}|I_{it}] + b_{it}(\widehat{y}_{it} + \eta_{it})$, as $\mathbb{E}[\eta_{it}] = 0$. As productivity shocks are observed, they do not affect the process of learning about ability; they only increase the variance of output, and so wage risk. Thus, the equilibrium is as in the baseline model except that now that static period-t piece rate is $b_t^0 = 1/[1 + (r/\xi_2^2)(\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)]$.

B.2.2 Identification

As in the baseline model, piece rates are identified from the ratio of variable to total pay. The parameters $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2, \sigma_\eta^2)$ are identified from the second moments of the wage distributions as follows. Since the productivity shocks η_{it} are idiosyncratic, the covariances of the wage distributions are the same as in the baseline model. The same argument as in the baseline model shows that $\mathrm{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)$; the sum of variances $\sigma_\varepsilon^2 + \sigma_\eta^2$ plays the role of σ_ε^2 in the baseline model. Thus, σ_θ^2 is identified from b_0^* and $\mathrm{Cov}[w_{i0}, w_{i1}] = b_0^*\sigma_\theta^2$. In turn, $\sigma_\varepsilon^2 + \sigma_\eta^2$ is identified from b_0^*, σ_θ^2, and $\mathrm{Var}[w_{i0}] = (b_0^*)^2(\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)$. Next, σ_1^2 is identified from b_1^*, $\sigma_\varepsilon^2 + \sigma_\eta^2$, and $\mathrm{Cov}[w_{i1}, w_{i2}] - \mathrm{Var}[w_{i1}] = (b_1^*)^2(\sigma_1^2 + \sigma_\varepsilon^2 + \sigma_\eta^2) - b_1^*\sigma_1^2$, and so σ_ζ^2 is identified from b_1^*, σ_θ^2, σ_1^2, and $\mathrm{Cov}[w_{i1}, w_{i2}] = \sigma_\theta^2 + \sigma_\zeta^2 - \sigma_1^2 + b_1^*\sigma_1^2$. Finally, σ_ε^2 is identified from σ_θ^2, σ_ζ^2, and $\sigma_1^2 = \sigma_\theta^2\sigma_\varepsilon^2/(\sigma_\theta^2 + \sigma_\varepsilon^2) + \sigma_\zeta^2$ and thus σ_η^2 is identified from σ_ε^2 and $\sigma_\varepsilon^2 + \sigma_\eta^2$. The rest of the identification argument is as in the baseline model.

B.2.3 Remarks

We can extend our analysis to the case in which productivity shocks are serially correlated by assuming that they are governed by the following process: $\eta_{it} = \nu_{it}$, with $\nu_{i0} = \mu_{i0}$ and $\nu_{it+1} = \sqrt{\rho}\nu_{it} + \mu_{it+1}$ for all $t \geq 0$, where $\rho \in [0, 1]$ and μ_{it} is an idiosyncratic shock that is normally distributed with mean zero and variance σ_μ^2 for all $t \geq 0$. This case reduces to the case of idiosyncratic productivity shocks when $\rho = 0$; productivity shocks are permanent when $\rho = 1$ and mean-reverting otherwise. The equilibrium characterization is the same as above except that now the static period-t piece rate is $b_t^0 = 1/[1 + (r/\xi_2^2)(\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_{\eta t}^2)]$, where $\sigma_{\eta t}^2 = \mathrm{Var}[\eta_{it}] = (1 - \rho^{t+1})\sigma_\mu^2/(1-\rho)$. Because of the serial correlation of productivity shocks, this model admits a more general variance-covariance structure of wages.

[59] The assumption that η_{it} is mean zero is without loss as we can absorb $\mathbb{E}[\eta_{it}]$ into θ_{it}.

B.3 More General Human Capital Process: Equilibrium Characterization

We first derive effort choices in the complex task for workers facing a sequence of piece rates that depend only on time when workers' future effort choices in the complex task also depend only on time. Note that in the absence of the simple task, employment contracts reduce to piece rates. We then determine equilibrium piece rates and show that they are the same for all workers and depend only on time. Finally, we characterize the equilibrium.

B.3.1 First-Order Conditions for Effort in the Complex Task

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

$$e_t = b_t + R_{CC,t} + R_{HK,t}(e_t), \tag{23}$$

where $R_{CC,t}$ is given by (7) and

$$R_{HK,t}(e) = F_t'(e) \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b_{t+\tau} + R_{CC,t+\tau}\big). \tag{24}$$

Recall that we denote the effort of a worker in the complex task in period t simply by e_t (omitting the subscript i). The assumption that $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$ ensures that (23) always has a solution. This solution need not be an optimal choice of effort for the worker, though. Additional assumptions, which we will discuss below, are necessary for this to be the case.

Suppose that piece rates are $\{b_t\}_{t=0}^T$ and consider worker i's choice of effort in the complex task in period t when the worker's future behavior depends only on time. The argument in the main text—the particular form of the functions $\{F_t(e)\}_{t=0}^T$ does not matter—shows that the first-order condition for the worker's choice of effort in the complex task is

$$e_t = b_t + \sum\nolimits_{\tau=1}^{T-t} \delta^\tau \frac{\partial \mathbb{E}[w_{it+\tau}|h_i^t]}{\partial e_t}. \tag{25}$$

In what follows, we show that (25) reduces to (23).

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

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

for all $1 \leq \tau \leq T-t$. By the law of motion of human capital and the fact that behavior from period $t+1$ on depends only on time,

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

for all $1 \leq \tau \leq T-t$. Finally, note from (5) that

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

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

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

for all $1 \leq s \leq T-t$, we can rewrite (25) as

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

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

The first-order condition (23) is necessary for optimality. This condition is also sufficient for optimality when $F_t(e) = \gamma_2 e$ for all t. When the functions $\{F_t\}_{t=0}^T$ are nonlinear, (23) need not be sufficient for optimality, though. Since $F_t(e)$ is concave, (23) is sufficient for optimality if

$$\sum\nolimits_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}\left(b_{t+\tau} + R_{CC,t+\tau}\right) \geq 0. \tag{26}$$

Indeed, $R_{HK,t}(e)$ is nonincreasing with e if (26) holds, in which case the marginal benefit of effort is nonincreasing with effort. Condition (26) holds if piece rates are in the unit interval.

B.3.2 Equilibrium Piece Rates

We now solve for the last-period equilibrium piece rates and then proceed backwards to determine the equilibrium piece rates in previous periods. With this characterization of piece rates at hand, we use (23) to derive the equilibrium choices of effort in the complex task, provided that equilibrium efforts in the complex task and piece rates depend only on time, which is the case.

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

Piece Rates in Previous Periods. Let $t < T$, and suppose that *i*) equilibrium piece rates and effort choices in the complex task from period $t + 1$ on are the same for all workers and depend only on time; and *ii*) piece rates belong to the interval $(0, 1)$. This is true when $t = T - 1$. For each $1 \le \tau \le T - t$, let $b^*_{t+\tau}$ be the equilibrium piece rate in $t+\tau$ with $1 \le \tau \le T-t$ and define $R^*_{CC,t}$ as in (7) with $b_{t+\tau} = b^*_{t+\tau}$ for each τ. Moreover, define $R^*_{HK,t}(e)$ to be given by (24) with $b_{t+\tau} = b^*_{t+\tau}$ for all τ. Since $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big) \ge 0$ when $b^*_{t+\tau} \in (0, 1)$ for all $1 \le \tau \le T - t$, a worker's period-t choice of effort in the complex task as a function of the piece rate b in t is the unique solution to the necessary and sufficient first-order condition

$$e_t = b + R^*_{CC,t} + R^*_{HK,t}(e_t). \tag{27}$$

As in the baseline model, the fact that (27) does not depend on a worker's history implies that workers' equilibrium choices of effort in t are independent of their history, and so the same for all of them, if period-t piece rates are the same for all workers.

Equation (27) implicitly defines a worker's optimal choice of effort in period t as a function of the worker's piece rate in period t. In an abuse of notation, denote this function by $e_t = e_t(b)$. The implicit function theorem implies that e_t is continuously differentiable with

$$\frac{\partial e_t}{\partial b} = \frac{1}{1 - F''_t(e_t) \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big)} = \frac{1}{1 - \dfrac{F''_t(e_t)}{F'_t(e_t)} R_{HK,t}(e_t)}. \tag{28}$$

Given that $F'''_t(e) \le 0$ for all $e \in \mathbb{R}$ and $\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} \big(b^*_{t+\tau} + R^*_{CC,t+\tau}\big) \ge 0$, it follows from (28) that $\partial e_t / \partial b$ is positive, bounded above by one, and nonincreasing with b.

Once again, let $W_t = W_t(b)$ be a worker's present-discounted value of wages from period t on as function of b. An equilibrium piece rate for a worker is the value of b that maximizes $V_t = \mathbb{E}[W_t|I_t] - r\text{Var}[W_t|I_t]/2 - e_t^2/2$, where I_t is the public information about the worker in t. We determine the choice of b that maximizes V_t in what follows. As in the baseline model, this choice is independent of I_t and so the same for all workers in t.

First, note that

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

Now note that since the functions $\{F_t(e)\}_{t=0}^T$ do not matter for the derivation of $\text{Var}[W_t|I_t]$, the partial derivative $\partial\text{Var}[W_t|I_t]/\partial b$ is still given by (14). Thus, the necessary first-order condition for the problem of maximizing V_t is

$$\left[1 + F_t'(e_t)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - e_t\right] \frac{\partial e_t}{\partial b} - r\left[b(\sigma_t^2 + \sigma_\varepsilon^2) + H_t^*\right] = 0 \qquad (29)$$

with $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^{\tau-1}$. Below, we show that this condition is also sufficient for optimality. Using (27), we can rewrite (29) as

$$b = \frac{1}{1 + r_t(e_t)(\sigma_t^2 + \sigma_\varepsilon^2)} \left[1 + F_t'(e_t)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{CC,t}^* - R_{HK,t}^*(e_t) - r_t(e_t)H_t^*\right], \qquad (30)$$

where $r_t(e_t) = (\partial e_t/\partial b)^{-1} r$. The solutions to (30), if they exist, do not depend on I_t and so are the same for every worker.

In order to establish that (29) is sufficient for optimality, let

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

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

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

be the marginal cost to the worker of an increase in b. Given that e_t is nondecreasing with b and $\partial e_t/\partial b$ is nonincreasing with b, it follows that MB_t is nonincreasing with b. Now note that $F_t'''(e)$ nonpositive and nondecreasing implies that $F_t'''(e)e \geq F_t''(e)$ for all $e \in \mathbb{R}$.[60] It then follows from

[60] The desired inequality is immediate if $e \leq 0$. When $e > 0$, it follows from $F_t''(e) = F_t''(0) + \int_0^e F_t'''(s)ds$ that $F_t''(e) \leq \int_0^e F_t'''(s)ds \leq \int_0^e F_t'''(e)ds = eF_t'''(e)$. The first inequality holds since $F_t''(0) \leq 0$, the second inequality holds since $F_t'''(s) \leq F_t'''(e)$ for all $s \leq e$, and the last equality holds since $e > 0$.

(28) that

$$\frac{d}{db}\left(e_t\frac{\partial e_t}{\partial b}\right) = \left(\frac{\partial e_t}{\partial b}\right)^2\left[1+\frac{e_t F_t'''(e_t)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(b^*_{t+\tau}+R^*_{CC,t+\tau})}{1-F_t''(e_t)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(b^*_{t+\tau}+R^*_{CC,t+\tau})}\right]$$

$$\geq \left(\frac{\partial e_t}{\partial b}\right)^2\frac{1}{1-F_t''(e_t)\sum_{\tau=1}^{T-t}\delta^\tau(b^*_{t+\tau}+R^*_{CC,t+\tau})} > 0.$$

Thus, MC_t is strictly increasing with b, which establishes the sufficiency of (29).

We conclude this step by showing that (29), and so (30), has a unique solution b_t^*, which does not depend on I_t. First note that MB_t is bounded since $\sup_{e\in\mathbb{R}} F_t'(e) < \infty$ and that $\partial e_t/\partial b$ belongs to the unit interval. Given that $e_t\partial e_t/\partial b$ is strictly increasing with b, it then follows from the expression for MC_t that $\lim_{b\to-\infty} MC_t(b) = -\infty$ and $\lim_{b\to+\infty} MC_t(b) = +\infty$. Thus, (29) has a solution, which is unique given the properties of MB_t and MC_t established above. Note that $b_T^* = 1/[1+r(\sigma_T^2+\sigma_\varepsilon^2)]$, since $\partial e_T/\partial b = 1$.

Equilibrium Characterization. The above argument shows that if there exists $t < T$ such that from period $t+1$ on, equilibrium piece rates and effort choices in the complex task are the same for all workers and depend only on time, and equilibrium piece rates are in the unit interval, then equilibrium piece rates and effort choices from period t on are the same for all workers and depend only on time. We now show that if $\lambda = 1$, then equilibrium piece rates in period t are also in the interval $(0, 1)$ if r is sufficiently small; we discuss how we can relax the assumption that $\lambda = 1$ at the end. Since the last-period piece rates and effort choices in the complex task are the same for all workers and depend only on T, and the last-period piece rates are in the interval $(0, 1)$, it follows by induction that equilibrium piece rates and effort choices in the complex task are the same for all workers and depend only on time, and equilibrium piece rates are in the interval $(0, 1)$, provided that $\lambda = 1$ and r is sufficiently small. From this, it further follows that the equilibrium is characterized by Proposition A.2.

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 A.1 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

$$\sum_{s=1}^{\tau-1}\left(\prod_{k=1}^{\tau-1-s}\mu_{t+\tau-k}\right)(1-\mu_{t+s})+\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}=1, \tag{31}$$

we then have that

$$\sum_{\tau=1}^{T-t}\delta^\tau(1-b^*_{t+\tau}-R^*_{CC,t+\tau})=\sum_{\tau=1}^{T-t}\delta^\tau(1-b^*_{t+\tau})\prod_{k=1}^{\tau-1}\mu_{t+\tau-k}=\frac{\sigma_\varepsilon^2}{\sigma_t^2}R^*_{CC,t}, \tag{32}$$

where we used the definition of $R^*_{CC,t}$ and the fact that $\mu_t/(1-\mu_t)=\sigma_\varepsilon^2/\sigma_t^2$. Now observe that the right side of (30), and so b^*_t, is smaller than one if, and only if,

$$\begin{aligned}&F'_t(e_t)\sum_{\tau=1}^{T-t}\delta^\tau-R^*_{HK,t}-R^*_{CC,t}\\&\quad=F'_t(e_t)\sum_{\tau=1}^{T-t}\delta^\tau(1-b^*_{t+\tau}-R^*_{CC,t+\tau})-R^*_{CC,t}<\frac{r}{\partial e_t/\partial b}\left(\sigma_\varepsilon^2+\sigma_t^2\sum_{\tau=0}^{T-t}\delta^\tau\right). \end{aligned}\tag{33}$$

Since $\partial e_t/\partial b\leq 1$, (32) implies that

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

is sufficient for (33). The above inequality holds since *i*) $F'_t(e)<(\sigma_t^2/\sigma_\varepsilon^2)[1+r(\sigma_t^2+\sigma_\varepsilon^2)]$ for all $e\in\mathbb{R}$; and *ii*)

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

by (24) and the assumption that $b^*_{t+\tau}\in(0,1)$ for all $1\leq\tau\leq T-t$.

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 A.1, that

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

where the second equality follows from (31) and the inequality follows from the assumption that $b^*_{t+\tau}<1$ for all $1\leq\tau\leq T-t$. Thus, by (28)

$$\frac{rH^*_t}{\partial e_t/\partial b}<r\sigma_t^2\left[1-F''_t(\infty)\frac{\delta}{1-\delta}\right]\frac{\delta}{1-\delta}. \tag{34}$$

Now note that $F_t'(e) > \sigma_t^2/\sigma_\varepsilon^2$ for all $e \in \mathbb{R}$ and (32) together imply that

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

Hence, by (34), there exists $\overline{r} > 0$ such that

$$1+F_t'(e_t)\sum_{\tau=1}^{T-t}\delta^\tau(1-b_{t+\tau}^*-R_{CC,t+\tau}^*)-R_{CC,t}^*-\frac{rH_t^*}{\partial e_t/\partial b}>0 \tag{35}$$

if $r \in (0, \overline{r})$. This, in turn, implies that the right side of (30) is positive, and so is the piece rate b_t^*. To sum up, there exists $\overline{r} > 0$ such that $b_t^* \in (0,1)$ provided that $r \in (0, \overline{r})$. Since σ_t^2 is monotonically decreasing if $\sigma_\theta^2 > \sigma_\zeta^2$ and monotonically increasing if $\sigma_\theta^2 < \sigma_\zeta^2$, it follows that $\sigma_t^2 \leq \max\{\sigma_\theta^2, \sigma_\zeta^2\}$. Thus, by (34), we can take the upper bound $\overline{r}$ on the worker's risk aversion to be independent of t.

We conclude by discussing how we can relax the assumption that $\lambda = 1$. First, note that (27), (28), and (30) define the equilibrium piece rates continuously as a function of λ.[61] Thus, the maps $\lambda \mapsto \sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(1-b_{t+1}^*-R_{CC,t+\tau}^*)$ are continuous. From this, it follows that if we take λ sufficiently close to one, then the inequalities in (33) and (35) will continue to hold when $r \in (0, \overline{r})$, where $\overline{r}$ is the upper bound on r when $\lambda = 1$.

B.4 Equilibrium Contracts with Multiple Performance Measures

We conclude by extending our analysis of the model with the more general human capital process (Section A.8) to the case in which there exists an observable but unverifiable additional measure of workers' output. Since the argument here follows many of the steps of the derivations in the case with the more general human capital process, we keep the exposition brief. The environment is the same as in Section A.8 except that for each worker i and in every period t, firms now observe a noisy measure of workers' performance, p_{it}, in addition to output, y_{it}. Assume that

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

where γ_t^e and γ_t^k are known constants and η_{it} is an unobserved idiosyncratic shock to worker i's performance measure in t that is normally distributed with mean zero and variance σ_η^2 and is

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

independent of all other shocks. For ease of exposition, we 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 t. Since the performance measure is unverifiable, firms still offer linear one-period output-contingent contracts to workers. Thus, worker i's wage in period t is given by $w_{it} = (1 - b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it}y_{it}$, where b_{it} is the worker's piece rate in period t and I_{it} is the public information about the worker available in t. However, this case differs from the one without the performance measure in that I_{it} contains not only the worker's output realizations before t but also the realizations of the worker's performance measure before t.

Learning about Ability. We first discuss how the presence of the performance measure affects learning about workers' ability in equilibrium. Consider worker i in period t, and let e_{it}^* and k_{it}^* be, respectively, the worker's equilibrium effort in the complex task and stock of human capital in t; recall our convention of denoting effort in the complex task simply by e_{it}. Let $z_{it}^y = y_{it} - k_{it}^* - e_{it}^*$ and $z_{it}^p = p_{it} - e_{it}^*$ be, respectively, the part of worker i's output and performance measure in t that cannot be explained by the worker's effort in the complex task and stock of human capital in t. Since in equilibrium agents correctly anticipate a worker's effort in the complex task and stock of human capital at any point in time, the same argument as that in the main text shows that posterior beliefs about worker i's ability in period t are normally distributed with mean m_{it} and variance σ_{it}^2. In an abuse of notation, let $\sigma_{it+1/2}^2 = \sigma_{it}^2\sigma_\varepsilon^2/(\sigma_{it}^2 + \sigma_\varepsilon^2)$. By standard results, m_{it} and σ_{it}^2 evolve over time according to[62]

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

and

$$\sigma_{it+1}^2 = \frac{\sigma_{it+1/2}^2\sigma_\eta^2}{\sigma_{it+1/2}^2 + \sigma_\eta^2} + \sigma_\zeta^2.$$

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

$$z_{it} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}z_{it}^y + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}z_{it}^p. \tag{36}$$

[62]The equations for the evolution of m_{it} and σ_{it}^2 follow from a belief-updating process in which, in each period, agents first update their beliefs about a worker's ability based on the worker's output and then update their beliefs based on the realization of the worker's performance measure. The order in which agents use the information about a worker to update their beliefs about the worker's ability is clearly irrelevant.

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

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

Thus, the evolution of posterior means and variances follow the *same* laws of motion as those in the case without the additional performance measure, except that $\sigma_{\varepsilon\eta}^2$ plays the role of the variance of the noise in output and z_{it} given by (36) plays the role of the signal about worker i's ability in period t. When $\sigma_\eta^2 = \infty$ and the performance measure is uninformative, the laws of motion for m_{it} and σ_t^2 reduce to the laws of motion in the absence of the performance measure. If we let $\mu_t = \sigma_{\varepsilon\eta}^2/(\sigma_t^2 + \sigma_{\varepsilon\eta}^2)$, then the law of motion for a worker's reputation is still given by (5).

Dynamic Returns to Effort. We now consider the first-order conditions for worker effort in the complex task when piece rates and future worker behavior depend only on time. Since for any worker i, we have that $\partial\mathbb{E}[z_{it}|h_i^t]/\partial e_t = 1$ for any period t and any period-t private history h_i^t for the worker, it follows that the expressions for $R_{CC,t}$ and $R_{HK,t}(e_t)$ are the same as they are in the case with the more general human capital process without the performance measure, and so are the first-order conditions for worker effort in the complex task when piece rates and future worker behavior depend only on time.[63]

Equilibrium Piece Rates. Since the first-order conditions for effort in the complex task when piece rates and future worker behavior depend only on time are the same as they are in the case with the more general human capital process without the performance measure, the derivation of the equilibrium piece rates follows exactly the same steps as in Section B.3. The only step in which the presence of the performance measure can alter the derivation of equilibrium piece rates is the calculation of the derivative $\partial\text{Var}[W_t|I_t]/\partial b$, as the presence of the performance measure potentially affects the covariance of wage payments across periods. Note that I_t now describes past realizations of output and the performance measure.

We claim that $\partial\text{Var}[W_t|I_t]/\partial b$ has the same expression as in the case without the performance measure, so that the expression for equilibrium piece rates remains unchanged. Since it still is the

[63]More generally, we have that $\partial\mathbb{E}[z_{it}|h_i^t]/\partial e_t = (\sigma_\eta^2 + \sigma_\varepsilon^2)^{-1}(\sigma_\eta^2 + \gamma_t^e \sigma_\varepsilon^2)$, from which it follows that $R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (1 - b_{t+\tau})\widehat{\mu}_t(\sigma_\eta^2 + \sigma_\varepsilon^2)^{-1}(\sigma_\eta^2 + \gamma_t^y \sigma_\varepsilon^2)$, where $\widehat{\mu}_t = (\prod_{k=1}^{\tau-1} \mu_{t+\tau-k})(1 - \mu_t)$. The expression for $R_{HK,t}(e_t)$ remains the same. Since, as we will show, we can identify the variances $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\zeta^2)$ from a panel of wages with information on their fixed or variable components and $p_{it} = \widehat{f}(e_{it}, k_{it}) + \eta_{it}$, where $\widehat{f}_t(e, k) = \gamma_t^e e + \gamma_t^k k + m_\theta$ is known up to m_θ and satisfies the conditions for identification for the case with the more general human capital process if, and only if, $\gamma_t^e \neq \gamma_t^k$, we can adapt our identification argument to this more general case.

case that

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

where Var_0 does not depend on b, the desired result follows if $\mathrm{Cov}[w_t, w_{t+\tau}|I_t] = b\sigma_t^2$ for all $\tau \geq 1$. As in Appendix A.2, $\mathrm{Cov}[w_t, w_{t+\tau}|I_t] = b\,\mathrm{Cov}[y_t, w_{t+\tau}|I_t]$ and

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

for all $1 \leq \tau \leq T - t$, where $y_{t+\tau} = y_{t+\tau}(b)$ and $m_{t+\tau} = m_{t+\tau}(b)$ still respectively denote a worker's output and reputation in period $t + \tau$ as a function of the period-t piece rate. Hence, if $z_{t+s} = z_{t+s}(b)$ with $0 \leq s \leq T - t$ is once again the signal about a worker's ability in period $t + s$ as a function of b, then (5) implies that

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

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

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

Given that $\mathrm{Cov}[y_t, z_{t+s}^p|I_t] \equiv \sigma_t^2$, $\mathrm{Cov}[y_t, z_{t+s}^y|I_t] = \sigma_t^2 + \sigma_\varepsilon^2$ if $s = 0$, and $\mathrm{Cov}[y_t, z_{t+s}^y|I_t] = \sigma_t^2$ if $1 \leq s \leq T - t$, we then have that

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

Thus, $\mathrm{Cov}[w_t, w_{t+\tau}|I_t] = b\sigma_t^2$ has the desired expression since $\sigma_{\varepsilon\eta}^2(1 - \mu_t) = \sigma_t^2\mu_t$.

Identification. As in the main text, equilibrium piece rates are identified from a panel of wages and their variable components (Section 5.3). Since $\mathrm{Var}[w_{i0}] = (b_0^*)^2(\sigma_\theta^2 + \sigma_\varepsilon^2)$, $\mathrm{Cov}[w_{i0}, w_{i1}] = b_0^*\sigma_\theta^2$, and $\mathrm{Var}[p_{i0}] = \sigma_\theta^2 + \sigma_\eta^2$, the vector $(\sigma_\theta^2, \sigma_\eta^2, \sigma_\varepsilon^2)$ is identified from $\mathrm{Var}[w_{i0}]$, $\mathrm{Cov}[w_{i0}, w_{i1}]$, and $\mathrm{Var}[p_{i0}]$. In particular, we do not need to assume that the distribution of the shock term η_{it} is known in order to obtain identification. The variance σ_ζ^2 is then identified from $\text{Var}[w_{i1}]$ since $\text{Var}[w_{i1}] = \sigma_\theta^2 + \sigma_\zeta^2 - \sigma_1^2 + (b_1^*)^2(\sigma_1^2 + \sigma_\varepsilon^2)$ and σ_1^2 is known from $(\sigma_\theta^2, \sigma_\eta^2, \sigma_\varepsilon^2)$. Finally, given that $p_{it} = \widehat{f}(e_{it}, k_{it}) + \eta_{it}$, where $\widehat{f}_t(e, k) = e + m_\theta$ is known up to m_θ and satisfies the conditions for identification for the case with the more general human capital process, the rest of the argument proceeds as in Section A.8.

References

Holmström, B., and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 7, 24–52.