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 2024

We thank Peter Arcidiacono, Oriana Bandiera, Jonathan Heathcote, Erik Hurst, Mike Keane, Kjetil Storesletten, Chris Taber, Petra Todd, and Gabriel Ulyssea as well as participants at various seminars and conferences for their comments and suggestions. Braz Camargo gratefully acknowledges financial support from CNPq, Fabian Lange from the Canada Research Chairs Program, and Elena Pastorino from the Stanford Institute for Economic Policy Research (SIEPR) and the National Science Foundation (NSF). This paper is dedicated to the memory of Eddie Lazear, whose encouragement to pursue this project, mentorship, and generosity we will never forget. This paper is dedicated to the memory of Eddie Lazear, whose encouragement to pursue this project, mentorship, and generosity we will never forget. 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 2024
JEL No. D8, D86, J24, J3, J31, J33, J41, J44

ABSTRACT

Performance pay, through which firms provide workers with incentives for performance, is known to be an important source of the variation in wages across workers. Little is known, though, about the impact of performance pay on wages over the life cycle or about the sources of its variability with workers' labor market experience. In this paper, we fill this gap by accounting for the possibility that incentives for performance also arise implicitly, as often argued, from workers' desire to prove themselves whenever their productivity is uncertain and to accumulate human capital when employed. We propose a framework that integrates and extends well-known models of dynamic moral hazard and of information and human capital acquisition in the labor market. This framework allows us to analytically decompose performance pay into distinct terms that capture the basic forces we nest, and is identified under intuitive conditions. We parameterize the model using data from foundational papers in personnel economics and find that the most important determinants of performance pay are workers' desire to insure against the wage risk due to the uncertainty about their productivity, which explains the relatively low level of performance pay, and their incentive to acquire human capital through learning-by-doing. The contemporaneous risk-incentive trade-off that much of the literature on performance incentives has emphasized is instead less important. Our estimates imply that performance pay is central to the dynamics of wages over the life cycle because of its direct impact on the variability of wages and its indirect impact on the process of human capital acquisition with experience.

Braz Camargo
Sao Paulo School of Economics
Rua Itapeva 474
Sao Paulo, SP 01332-000
Brazil
braz.camargo@fgv.br

Fabian Lange
Department of Economics
McGill University
855 Sherbrooke Street West
Montreal, QC H3A, 2T7
and NBER
fabian.lange@mcgill.ca

Elena Pastorino
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
and Hoover Institution
and also NBER
epastori@stanford.edu

1 Introduction

To align workers' incentives to firms' objectives, firms often link compensation to performance on the job through bonuses, commissions, and piece rates, a practice that has become more prevalent in recent decades (Lemieux et al. [2009]). For most workers, though, performance pay amounts to only a small fraction of pay, which raises the question of whether incentives for performance are important for the typical worker in practice. For instance, in the PSID, variable pay accounts for less than 5% of workers' pay and does not represent a major component of pay at any point over the life cycle (Frederiksen et al. [2017]).

One reason why performance pay is small is that workers may already face strong implicit incentives for performance. In particular, they may be motivated to work hard to convince potential employers, who are likely to be uncertain about their talent, that their productivity is high. These *implicit* incentives for effort on the job can then substitute for the *explicit* incentives from performance pay. This well-known *career-concerns* argument provides a common explanation not only for why performance pay often makes up only a small portion of pay but also for how performance pay varies over the life cycle. That is, as workers' experience accumulates and their productivity becomes better known, the implicit incentives for performance from career concerns weaken. To compensate for them, explicit incentives from performance pay should become increasingly more important (Holmström [1999] and Gibbons and Murphy [1992]). As we document, though, the *opposite* pattern is common in the data; namely, relative to total pay, performance pay eventually tends to decrease with experience. Thus, the questions of whether performance incentives matter and, if so, why they are so small for the typical worker are very much open ones.

In this paper, we start from the premise that workers face other powerful implicit incentives for effort on the job, as they are also motivated to acquire new skills when employed. Namely, since the effort to produce output can substitute for the effort to invest in human capital, as in models of learning-or-doing or on-the-job training (Ben-Porath [1967]), or complement it, as in models of learning-by-doing (Heckman et al. [1998]), performance pay influences how much human capital workers accumulate with experience by affecting their incentives to produce output. We argue that the implicit incentives for performance induced by career concerns and the opportunity to acquire human capital are key for understanding both why performance pay is relatively low and how it varies over the life cycle.

Specifically, we find that uncertainty about workers' productivity, which induces *correlated* wage risk over time, is primarily responsible for the low level of performance pay. But although small, performance pay crucially shapes life-cycle wages through its indirect impact on workers' process of human capital acquisition in the labor market.

We formalize these intuitions by proposing a tractable model for the multiple incentives for workers' effort on the job that arise from performance pay, career concerns, and the opportunity to accumulate human capital during employment. Our model thus offers a unified framework to investigate how these forces together determine the level and dynamics of wages and 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 contribution of the mechanisms we nest, we can shed light on the forces governing how performance pay evolves with experience and so on the specific environments in which alternative life-cycle profiles of performance pay emerge. Conversely, we show that performance pay provides rich information that helps identify the determinants of fixed and variable pay that we integrate.

Second, our model explains why the ratio of performance pay to total pay tends to follow a hump-shaped pattern over the life cycle, contrary to the prediction of leading models of performance incentives. Human capital considerations are key to this feature of the data. When human capital is acquired through learning-by-doing so that effort to produce output also increases human capital, explicit incentives from performance pay naturally *complement* the implicit ones from human capital acquisition: they support greater effort and so human capital accumulation early on, when human capital is most valuable. As accumulating human capital becomes less valuable over time, explicit incentives for performance correspondingly weaken later on, and thus performance pay eventually decreases relative to total pay. Third, using our parameterized model, we show that performance pay, through its direct effect on total pay and its indirect effect on human capital acquisition, plays a critical role for the growth and dispersion of wages over the life cycle.

Our framework integrates and extends existing models of learning and performance incentives by building on the notion that workers can exercise effort on *simple* tasks that are contractable and *complex* tasks that are non-contractable, which both contribute to a worker's output and human capital—the latter accumulates through learning-or-doing and learning-by-doing. We think of different jobs as distinct bundles of simple and complex tasks. Whereas a worker's ability is unobserved to all and effort on complex tasks and human capital are observed only by a worker, effort on simple tasks as well as output or performance are publicly observed—output is then a noisy signal of ability, effort on complex tasks, and human capital. We assume that firms compete for workers by offering contracts that allow for variable pay contingent on a worker's output. Hence, workers face two types of performance incentives: explicit incentives from wages being linked to performance and implicit ones from their desire to influence the market's perception of their ability and human capital. We can then investigate not only how wages and their components evolve with experience but also how the allocation of effort across activities that are more or less difficult to contract varies over time. Our simple micro-foundation of the notion of a job and the resulting task assignment process can account for the joint dynamics of wages, their structure, and the types of activities that workers perform in firms as their careers progress.

Our main results are as follows. Through our model, we can decompose performance pay into distinct terms that capture *i*) the trade-off between risk and incentives typical of settings of moral hazard; *ii*) the career-concerns incentives for performance generated by the uncertainty about workers' ability; *iii*) the insurance firms provide against the wage risk due to this uncertainty; and *iv*) the incentives for performance arising from the opportunity to acquire human capital. We can then determine the primitive conditions that lead to different life-cycle patterns of performance pay relative to total pay. We find that a human capital motive for effort rationalizes the hump-shaped life-cycle profile of performance pay that characterizes 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). In the absence of human capital considerations, performance pay would rise with experience.

Evaluating the incentive power of wage contracts and, more generally, distinct mechanisms for the variability of wages across individuals and over time involves a difficult measurement exercise, since the underlying sources of the variation in wages are unobserved and mediated by firm and worker behavior. We do so 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 parameterize our model by matching how mean wages, their variance, and the ratio of variable to total pay evolve with experience in the BGH data under alternative restrictions on the model's parameters, so as to reduce our model to special cases studied in the literature. This way, we can measure the strength of the incentives we focus on in typical data; examine how they shape performance pay and overall wages over the life cycle; and, by comparing our model to existing ones, assess the extent to which integrating known frameworks offers novel insights about the impact of performance incentives on wages.

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 contributes to both output and the accumulation of human capital—as we find it to be of the learning-by-doing type. Indeed, according to our baseline parameterization, performance pay accounts for more than 30% of wage growth, once the cumulative impact of effort on human capital accumulation is taken into account. Performance pay is also crucial for wage dispersion: it accounts for a large portion of the variability of wages over the first 10 to 20 years of labor market experience. To the best of our knowledge, these estimates of the role of performance pay for life-cycle wages are new to the literature.

Interestingly, we find that the insurance against the wage risk due to the uncertainty about ability, which wage contracts provide through low performance pay, is the primary force depressing performance pay, rather than the contemporaneous risk-incentive trade-off that much of the literature on performance incentives has emphasized. 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 who are paid according to performance-pay contracts effectively hold a portfolio of state-contingent claims to output whose value comoves with their perceived ability. Such a portfolio pays out *more* in good times—when output and thus signals about ability are high—and *less* in bad times—when output and thus signals about ability are low—thereby compounding the risk that workers already face because their output fluctuates over time. Like any other risk-averse investor, though, workers would prefer, and are willing to pay a premium for, assets that diversify their risk. Accordingly, workers demand contracts that reduce the risk generated by the variability of their output due to the variability of the beliefs about their ability as learning about it occurs. Performance pay then tends to be small to partially shield workers against the risk in lifetime wages induced by the uncertainty about their ability.[1]

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

Since uncertainty about ability increases the variability of wages, a natural conjecture—and a common reading of learning models—is that the variance of wages would be lower in the absence of such uncertainty. This indeed would be the case if wage contracts did *not* respond to changes in uncertainty. But since wage contracts are endogenous, performance pay increases when uncertainty declines, as workers demand less insurance against the implied wage risk. Higher performance pay, in turn, amplifies any residual productivity risk, leading, on balance, to much greater wage dispersion. 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 in wages induced by performance pay. Such an exercise illustrates the importance of accounting for the endogeneity of the wage structure when assessing the role of different sources of wage dispersion, which has implications for the broader debate on inequality. For instance, in the settings we consider, reducing the dispersion in worker productivity early in the life cycle—say, through more schooling—may induce firms to offer wages more sensitive to performance. Then, workers who are *more* homogeneous in terms of their initial skills might end up experiencing *more* wage inequality.

Our paper relates to multiple strands of literature, including work on *i*) measuring the impact of uncertainty and learning about ability for wages and job mobility (Arcidiacono et al. [2010], Kahn and Lange [2014], Aryal et al. [2022], Keane et al. [2017], and Pastorino [2024]); *ii*) assessing the importance of human capital acquisition in the labor market for wage growth (Heckman et al. [1998], Gladden and Taber [2009], and Taber and Vejlin [2020]); *iii*) exploring the role of incentive pay for wage inequality and productivity (Lemieux et al. [2009] and Bloom and Van Reenen [2010]); and *iv*) estimating the importance of implicit non-monetary incentives for performance and their interaction with performance pay (Bandiera et al. [2005], Bandiera et al. [2009], and Bandiera et al. [2010]). Much work has also emphasized the role for wages of persistent unobserved worker heterogeneity (Geweke and Keane [2000], Meghir and Pistaferri [2004], Low et al. [2010], and Adda and Dustmann [2023]), which is at the heart of the incentive mechanisms we study.

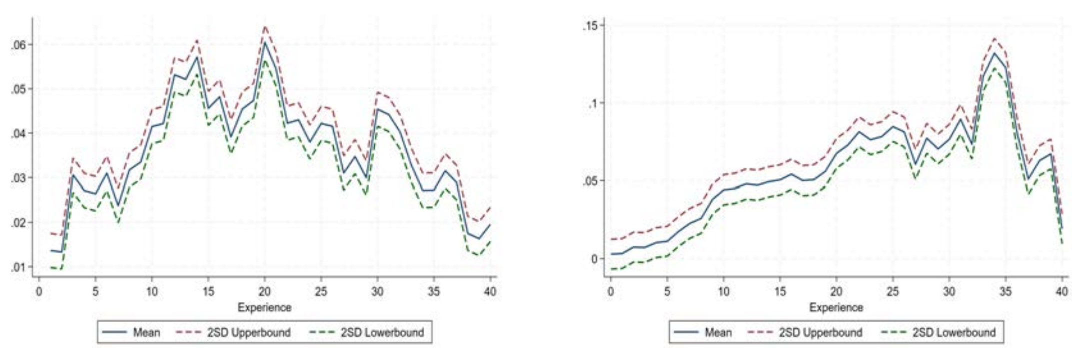
The paper proceeds as follows. We discuss the evidence on performance pay in Section 2, introduce the model in Section 3, and characterize equilibrium in Sections 4 and 5. We show how the model is identified in Section 6 and present our empirical exercises in Section 7. Section 8 concludes. The online appendix contains all omitted details.

2 Performance Pay over the Life Cycle

We start by providing evidence on the experience profile of the ratio of variable (performance) pay to total pay using proprietary data from the personnel records of two firms studied in three influential papers in the literature on careers, namely, Baker et al. [1994a,b] (BGH) and Gibbs and Hendricks [2004] (GH) as well as public data from the PSID, the NLSY79, and the NLSY97, described in the online appendix. All these data sets contain information on workers' fixed f_{it} and variable pay v_{it}—we focus on bonus pay for consistency across them and because it is the most important component of variable pay—which determine worker i's *wage* $w_{it} = f_{it} + v_{it}$ in period t. Note that $v_{it} = b_t y_{it}$ when variable pay is proportional to output; also, if firm entry into the labor market is free, then average wages equal average output—we will maintain both of these assumptions. In this case, the ratio $\mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$ equals the piece rate b_t, which measures the sensitivity of pay to performance. We estimate this ratio for men aged 21 to 65 from each dataset and plot it in Figures 1 and A.1 against experience; see the online appendix for details.

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

(a) BGH (b) GH



As Figures 1 and A.1 show, using these data spanning multiple years, workers, and firms, we document that the importance of performance pay relative to total pay eventually *declines* with experience, contrary to the prediction of existing models of career concerns and performance incentives (see Gibbons and Murphy [1992]). The model we present next is able to account for this pattern together with workers' wages and tasks over the life cycle.

3 A Model of Learning, Human Capital, and Incentives

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. 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 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.[2]

Production. The output technology, common to all firms, is such that worker i's output is

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

in period t, 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 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. That the coefficient ξ_θ multiplying θ_{it} in (1) is one is without loss since ξ_θ can be incorporated into θ_{it}. 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

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

[2] The labor market considered can be interpreted as one of many segmented by location, occupation, or industry, each defined by a distribution of worker productivity and common output, learning, and human capital technologies. What is important for our results is that these markets are sufficiently separate that employment opportunities in other markets are irrelevant for workers' and firms' decisions in a given market.

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 human capital, and β_t is a time-varying constant.[3] We can set $\xi_k = 1$ since we can absorb it into γ_1, γ_2, and β_t, and redefine human capital accordingly.[4] 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 standard learning-by-doing models, and the case in which the effort to acquire human capital *substitutes* the effort to produce output in both tasks ($\gamma_1, \gamma_2 < 0$), as in models of learning-or-doing à la Ben-Porath [1967]. In the first 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 second case, the investments in t are $\overline{e}_t - e_{i1t}$ and $\overline{e}_t - e_{i2t}$ with corresponding rates of human capital accumulation $|\gamma_1|$ and $|\gamma_2|$, where $\overline{e}_t$ is a worker's endowment of time or efficiency units in t and we absorb $(\gamma_1 + \gamma_2)\overline{e}_t$ into β_t. Cases in which the effort to acquire human capital complements the effort to produce output for one task and substitutes it for the other task are also possible. Throughout, we refer to γ_1 and γ_2 as the rates of human capital accumulation.

Worker Preferences. The lifetime utility from period t on of a worker who receives the wages $\{w_{t+\tau}\}_{\tau=0}^{T-t}$ 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 $-\exp\{-r\sum_{\tau=0}^{T-t}\delta^\tau[w_{t+\tau} - c(e_{1t+\tau}, e_{2t+\tau})]\}$, where $r > 0$ and δ are a worker's coefficient of (absolute) 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).[5] In what follows, we assume that $\rho_1 = \rho_2 = 1$ and $\eta = 0$, and consider the general case in the online appendix. There, we show that equilibrium piece rates do not depend on η and we can renormalize the model parameters to set $\rho_1 = \rho_2 = 1$.[6]

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 period-t effort in the simple task, e_{i1t}, and wage schedule to incentivize effort in the complex task, $w_{it} = c_{it} + b_{it}y_{it}$, where c_{it} is the fixed component of worker i's wage in t and b_{it} is worker i's

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

[4]Indeed, just rewrite (1) and (2) with $\widehat{k}_{it} = \xi_k k_{it}$, $\widehat{\gamma}_1 = \xi_k\gamma_1$, $\widehat{\gamma}_2 = \xi_k\gamma_2$, and $\widehat{\beta}_t = \xi_k\beta_t$.

[5]We assume quadratic costs for simplicity. Our equilibrium characterization extends to more general cost functions, and so do our identification results, provided that the cost function is known.

[6]The value of η does not affect equilibrium piece rates since effort in the simple task is contractable, so the provision of incentives for effort in the complex task is not influenced by effort in the simple task. In the online appendix, we also consider a version of our model in which both tasks feature non-contractable effort.

piece rate in t.[7] Three reasons lead us to consider wage schedules that are linear in output. 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 together with their risk neutrality 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 worker i's expected output in t given I_{it}. So,

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

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, $\mathbb{E}[\theta_{it}|I_{it}]$. We discuss how the latter term is determined in the next section.

3.3 Discussion

We conclude by discussing some features of our model and dimensions along which it can be extended; see the online appendix for details. We first note that our model can be interpreted as the log version of a model in which the output and human capital technologies are of the standard Cobb-Douglas form, namely, *i*) the period-t output of a worker with unobserved 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 $Y_t = \Theta_t K_t E_{1t}^{\xi_1} E_{2t}^{\xi_2} \Omega_t$, where ξ_1 and ξ_2 are the parameters in (1) and Ω_t is a mean-one noise term; and *ii*) the human capital acquired in period $t+1$ by a

[7] 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.

worker who in period t has human capital K_t and exerts efforts E_{1t} in the simple task and E_{2t} in the complex task is $B_t K_t^{\lambda} E_{1t}^{\gamma_1} E_{2t}^{\gamma_2}$, where B_t is a positive time-varying constant, and λ, γ_1, and γ_2 are the parameters in (2).

As in Gibbons and Murphy [1992], we assume that 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 to 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 certainty-equivalent 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 by linking a worker's job to the content or responsibilities it entails, which are contractable to different degrees and can change with a worker's experience. For instance, in the BGH data we use, as discussed in Section 7.6, workers progress over time to more complex jobs for which general management duties—such as general administration or planning—requiring workers to perform activities difficult to contract become increasingly more important, whereas simpler activities easier to contract—such as creating or selling products—correspondingly decrease in importance.

Our model also 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 time expended to produce output.[8] When effort is not a choice variable, the model specializes to one of human capital acquisition with experience, which is "passive" if $\gamma_1 > 0$ (learning-by-doing) and "active" if $\gamma_1 < 0$ (learning-or-doing). 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 simpli-

[8]The version of our model featuring just the task requiring non-contractable effort, known ability, and no noise in output is equivalent. In this case, piece rates are one so workers are paid their output.

fies 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].[9]

Finally, 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 as to allow for a wage markdown of workers' output; see the online 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 $y_{it} = p_{it} + \theta_{it} + k_{it} + \xi_1 e_{i1t} + \xi_2 e_{i2t} + \varepsilon_{it}$—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 for simplicity, as we use data from one firm in our empirical exercises.

4 Learning and Effort in the Complex Task

In this section, we present key results for our equilibrium characterization. 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 characterize how it depends on career concerns and human capital acquisition incentives.

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; we omit the dependence of effort choices and human capital on i for ease of notation. 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

[9] 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.

(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 $m_{it} = \mathbb{E}[\theta_{it}|I_{it}]$ and variance $\sigma_{it}^2 = \mathrm{Var}[\theta_{it}|I_{it}]$, with $m_{i0} = m_\theta$ and $\sigma_{i0}^2 = \sigma_\theta^2$. We refer to m_{it} as worker i's *reputation* in t. By standard results,

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

The recursions for m_{it} and σ_{it}^2 in (4) describe how a worker's reputation and the variance of posterior beliefs about a worker's ability evolve over time. These expressions are valid even when a worker's effort in the complex task deviates from the equilibrium path as any output realization is possible regardless of a worker's effort choices, so firms cannot use realized output to infer a worker's unobservable effort in the complex task. Note that σ_{it}^2 evolves independently of z_{it} and so is common to all workers in t. Thus, we can suppress the subscript i and simply denote this variance by σ_t^2. By iterating on the law of motion for m_{it} in (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 period $t+\tau$ with $1 \leq \tau \leq T-t$ given reputation m_{it} in period t is

$$m_{it+\tau} = \left(\prod\nolimits_{k=0}^{\tau-1} \mu_{t+k}\right) m_{it} + \sum\nolimits_{s=0}^{\tau-1} \left(\prod\nolimits_{k=1}^{\tau-1-s} \mu_{t+\tau-k}\right) (1-\mu_{t+s}) z_{it+s}. \tag{5}$$

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. We present here the workers' problem and decompose the returns to effort in the complex task into terms capturing the incentives from workers' desire to affect their reputation—the career-concerns incentive—and to acquire human capital.

Worker Problem. Suppose 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 $0 \leq \tau \leq T-t$. The worker chooses e_{2t} to maximize $U_{it}(e_{2t}) = \mathbb{E}[-\exp\{-r[W_{it} - c(e_{1t}, e_{2t})]\}|h_i^t]$, where $W_{it} = \sum_{\tau=0}^{T-t} \delta^\tau w_{it+\tau}$. Note that 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 choice of e_{2t} that maximizes $U_{it}(e_{2t})$ is independent of h_i^t; it is also independent of e_{1t}. Since signals about ability are normally distributed, it follows from (3) and (5) that the 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 $\mathbb{E}[W_{it}|h_i^t] - r\text{Var}[W_{it}|h_i^t]/2 - e_{2t}^2/2$.[10]

First-Order Conditions for Effort. Notice that $\partial\mathbb{E}[w_{it}|h_i^t]/\partial e_{2t} = \xi_2 b_t$ by (1) and (3). Worker i's choice of effort in the complex task also affects the worker's future wages through its impact on the worker's future reputation—which affects the *fixed* component of future pay—and on the worker's future human capital—which affects both the *fixed* and *variable* components of future pay. Since $\mathbb{E}[W_{it}|h_i^t] = \sum_{\tau=0}^{T-t}\delta^\tau\mathbb{E}[w_{it+\tau}|h_i^t]$ and, as we show in the online appendix, effort in the complex task does not affect the variance of future pay, the first-order condition for worker i's effort in the complex task is

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

The right side of (6), which describes the marginal benefit of effort in the complex task in t, consists 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 $t < T$.[11]

Let $\widehat{\mu}_{t,\tau} = (\prod_{k=1}^{\tau-1}\mu_{t+\tau-k})(1-\mu_t)$ and define the terms $R_{CC,t}$ and $R_{HK,t}$ to be such that

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

In the online 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 $R_{HK,t}$ respectively describe the dynamic marginal benefit of effort in the complex task arising from its effect on a worker's reputation and human capital, which provides implicit incentives for effort through variation in both the fixed and variable components of pay.[12] To understand $\xi_2 R_{CC,t}$, note that at the margin, higher e_{2t} increases the expected period-t signal about worker ability by ξ_2. By (5), this increases a worker's

[10] 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.

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

[12] Note from (8) that effort choices in the complex task depend only on time and are identical across workers—the mean-variance representation of worker preferences implies that a worker's trade-off between wages and effort in the complex task does not depend on a worker's history. This fact plays a key role in the proof that the equilibrium is symmetric and such that effort choices and piece rates depend only on time.

expected reputation in period $t+\tau$, 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 the present-discounted value of all these marginal increases. Similarly, to understand $R_{HK,t}$, note that at the margin, higher e_{2t} changes worker i's output $t+\tau$ by $\gamma_2 \lambda^{\tau-1}$, which amounts to the change in the worker's 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 the worker's ability in $t+\tau$ by $\gamma_2\lambda^{\tau-1}$, which, by the same argument used to derive $R_{CC,t}$, increases the present-discounted value of the fixed component of the wages from $t+\tau$ on by $\gamma_2\lambda^{\tau-1}R_{CC,t+\tau}$. The term $R_{HK,t}$ is the present-discounted value of all these marginal changes.

5 Equilibrium Characterization and Properties

We first characterize equilibrium and then examine the life-cycle pattern of piece rates, effort choices, and job assignments implied by it.

5.1 Equilibrium Characterization

Recall that $\mu_t = \sigma_\varepsilon^2/(\sigma_t^2+\sigma_\varepsilon^2)$ and $\prod_{k=1}^{0} a_k = 1$ for any sequence $\{a_k\}$. Let $\{\sigma_t^2\}_{t\geq 0}$ be such that $\sigma_0 = \sigma_\theta^2$ and $\sigma_{t+1}^2 = \mu_t\sigma_t^2 + \sigma_\zeta^2$. By the results in Section 4, the variance σ_t^2 describes the uncertainty about a worker's ability in period t. Since output signals do not perfectly reveal ability, this uncertainty persists throughout a worker's career and converges to a non-negative value σ_∞^2, which is positive if $\sigma_\zeta^2 > 0$. In particular, the variance σ_t^2 monotonically decreases to σ_∞^2 if $\sigma_\theta^2 > \sigma_\infty^2$ and monotonically increases to σ_∞^2 if $\sigma_\theta^2 < \sigma_\infty^2$.[13]

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}^*, e_{2t}^*, and b_t^* be, respectively, the equilibrium efforts in the simple and complex tasks and the equilibrium piece rate in t. For each t, let $b_t^0 = 1/[1+(r/\xi_2^2)(\sigma_t^2+\sigma_\varepsilon^2)]$, $R_{CC,t}^*$ 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*

$$b_t^* = b_t^0\left(1+\frac{\gamma_2}{\xi_2}\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - \frac{1}{\xi_2}R_{HK,t}^* - R_{CC,t}^* - \frac{r}{\xi_2^2}H_t^*\right). \tag{9}$$

We discussed the expression for e_{2t}^* in Section 4. In any period t, effort in the complex

[13] One can show that $\sigma_\infty^2 = [\sigma_\zeta^2 + (\sigma_\zeta^4 + 4\sigma_\zeta^2\sigma_\varepsilon^2)^{1/2}]/2$; see Holmström [1999] for a proof of this result and of the properties of σ_t^2. Kahn and Lange [2014] refer to $\sigma_\zeta^2 > 0$ as "learning about a moving target" and find evidence of it from the correlation between performance ratings and wages in the BGH data.

task equates the marginal cost of effort to its marginal private benefit, which features a static and a dynamic component. The latter arises from the impact of effort on a worker's future reputation and human capital. By contrast, in any period t, effort in the simple task equates the marginal cost of effort to its marginal social (output) benefit, $\xi_1 + \gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$.[14]

To understand the expression for equilibrium piece rates, first note that b_t^0 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. Now, by the expression for e_{2t}^*, 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, equilibrium piece rates deviate from first-best piece rates since risk-averse workers are unwilling to bear all output risk.[15] In a static setting, this distortion results in piece rates being adjusted by the factor $b_t^0 < 1$. 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 evolves with a worker's realized output. The insurance term $(r/\xi_2^2)H_t^*$ mitigates this risk by reducing the correlation between a worker's performance and pay when $t < T$.[16]

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

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

This decomposition of piece rates helps illustrate how the economic forces at play in 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 to workers 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].

[14]Note that e_{1t}^* is positive for all t when $\gamma_1 \geq 0$. When $\gamma_1 < 0$, we have that e_{1t}^* is positive for all t if $\gamma_1 > \xi_1(\lambda - 1/\delta)$ and e_{10}^* is positive. By the expressions for $R_{CC,t}^*$ and $R_{HK,t}^*$, it follows that e_{2t}^* is positive for all t if $\gamma_2 > 0$ and piece rates are between zero and one. Moreover, if piece rates are strictly positive and bounded above by one, then e_{2t}^* is positive even when $\gamma_2 < 0$ as long as $|\gamma_2|/\xi_2$ is small.

[15]Indeed, $b_T^* = 1$ if $r = 0$. This, in turn, implies that $R_{CC,T-1}^* = 0$ and $R_{HK,T-1}^* = (\gamma_2/\xi_2)\delta$ so that $b_{T-1}^* = 1$. 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. In this case, implicit incentives for effort arise solely from human capital considerations.

[16]That the insurance term $(r/\xi_2^2)H_t^*$ depends only on the uncertainty about ability σ_t^2 and not on total output risk $\sigma_t^2 + \sigma_\varepsilon^2$ follows from the fact that the life-cycle wage risk due to uncertainty and learning about ability is due to the correlation between current and future wages through worker ability. This correlation depends only on the uncertainty about ability, as output shocks—the other source of output risk—is idiosyncratic.

The second term lowers the explicit incentives for effort provided by piece rates in light of the implicit reputational incentives arising from the uncertainty about ability. As discussed above, the third term in (10) lowers piece rates to provide workers with insurance against the life-cycle wage risk due to the process of learning about ability.

The last term in (10), which is novel, captures the contribution of human capital acquisition and consists of two further terms. The first term is proportional to $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}$, which is the present-discounted change in lifetime output resulting from the change in a worker's human capital following 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 arising from the prospect of human capital acquisition, which *substitute* for explicit incentives. This last term in (10) corrects piece rates to better align the private marginal returns to effort in the complex task with the corresponding social marginal 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 are not present, which lead to counterfactual implications for piece rates, and then turn to the general case, which is consistent with the data.

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 coefficient of risk aversion $r' = r/\xi_2^2$. Then, (9) becomes $b^*_t = b^0_t(1 - R^*_{CC,t} - rH^*_t)$. This setup generalizes the model in Gibbons and Murphy [1992]—under the assumption of quadratic effort costs—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 enough to one.[17]

[17] Also as in Gibbons and Murphy [1992], piece rates are smaller than one. Indeed, $b^*_t < b^0_t$ 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.

Since equilibrium piece rates do not depend on 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] shows that in this case, firms compensate for the decline in implicit incentives for effort by increasing the strength of explicit incentives over time. In the online 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.

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 varies 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, whereas $R_{CC,t}^*$ 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 $e_{2t}^* - e_{1t}^*$ increases strictly with t whenever $\sigma_\theta^2 \geq \sigma_\infty^2$.[18]

[18] In Section 7.6, 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 $\ln(1 + e) \approx e$.

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

[19] 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.

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 $\xi_2 + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_2)\sigma_\varepsilon^2 b^*_t$, the socially optimal level of effort in this task net of a term proportional to piece rates.[20] 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 than or greater than γ_1; namely, 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—task complexity increases with experience and $\gamma_1 < \gamma_2$—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. 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 $\gamma_2 > \lambda\xi_2[1 + (r/\xi_2^2)\sigma_\varepsilon^2]$.[21] By contrast, when the importance of human capital for the complex task is small, 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. Proposition 2 below summarizes this discussion.[22]

[20] 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.

[21] 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.

[22] 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 online appendix that this is true when human capital depreciation is small.

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 in the complex task complements the effort to acquire human capital.

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, when the initial uncertainty 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—piece rates between zero and one ensure that both the career-concerns and human capital motives contribute negatively to piece rates, so that they are bounded above by $b_t^0[1 + (\gamma_2/\xi_2)\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1} - (r/\xi_2^2)\sigma_\theta^2\sum_{\tau=1}^{T}\delta^\tau]$. Hence, 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 large enough.*

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^* arises. 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.

6 Empirical Content of the Model

Our model is identified from the first and second moments of the distribution of wages and the ratio of variable to total pay over the life cycle up to the level normalization of mean ability m_θ.[23] In establishing this result, we treat the discount factor δ and the sensitivity of output to effort in the simple and complex tasks, ξ_1 and ξ_2, as known. We discuss in Section 7.6 how these restrictions can be relaxed. Since we can absorb k_0 into m_θ, we set $k_0 = 0$.[24]

Proposition 4. *The piece rates $\{b_t^*\}_{t=0}^T$ and variance parameters $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ are identified from a panel of wages and their components. 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. Finally, 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 average wages up to m_θ.*

We divide our identification argument into three sequential steps. First, we show how piece rates and the variance of ability, output shocks, and ability shocks—namely, σ_θ^2, σ_ε^2,

[23] See Margiotta and Miller [2000], Gayle and Miller [2009, 2015], and Golan et al. [2015] on the identification of moral-hazard models of 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.

[24] 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.

and σ_ζ^2—are identified. Because wages are linear in output, the ratio of average variable pay to average total pay identifies piece rates at each level of experience, as argued in Section 2. Once piece rates are known, the covariance structure of the distribution of wages identifies σ_θ^2, σ_ε^2, and σ_ζ^2. Next, we show how r, λ, and γ_2 are identified. The last-period piece rate identifies the risk aversion parameter, since once the variance of output, which is determined by the variance of workers' ability and output shocks, is recovered, its only unknown parameter is r. Piece rates in previous periods reflect the depreciation rate of human capital $1 - \lambda$ and the rate γ_2 at which effort in the complex task changes human capital, so they pin down those parameters. Lastly, as for γ_1 and $\{\beta_t\}_{t=0}^{T-1}$, the average wage in the first period identifies effort in the simple task in this period (up to m_θ), which in turn pins down γ_1. Then, effort in the simple task in all subsequent periods is identified so that average wages in all such periods can be used to back out the evolution of a worker's human capital stock, from which $\{\beta_t\}_{t=0}^{T-1}$ can be recovered. Intuitively, any difference between a worker's stock of human capital in t and undepreciated stock of human capital from $t-1$ that is not explained by effort in the two tasks in $t-1$ is due to the drift term β_{t-1}.

We conclude this section with three remarks. First, unlike in the instrumental-variable approaches common in the literature (Gibbons et al. [2005]), our argument does not require any exogenous variation. Second, since the parameters of the learning process are identified independently of those of the human capital process, the recovery of the learning process is robust to the specification of the human capital process.[25] Third, in the online appendix, we show that a more general version of the model, in which parameters unobservably differ across workers, is also identified, even when wages are measured with error.

7 The Role of Learning, Human Capital, and Incentives

In this section, we analyze how the incentives provided by performance pay, learning about ability, and human capital acquisition and the risk implied by them together shape the life-cycle profile of wages and their structure. Since our model admits a variety of life-cycle patterns for average wages, their variance, and piece rates, we propose in Section 7.1 three alternative parameterizations. For each of them, we restrict the model to feature only non-contractable effort so as to focus on the interplay between implicit and explicit incen-

[25] With performance information, which is often available (Frederiksen et al. [2017]), these results extend to the case in which human capital evolves nonparametrically with effort. Details are available upon request.

tives for performance. The first parameterization assumes that piece rates are exogenous, whereas the second one treats them as endogenous. Although our model is very parsimonious and highly overidentified, both of these parameterized versions of it fit the targeted moments very well. We also explore a third parameterization that 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. The drawback of this parameterization is that its implications for piece rates are clearly rejected by the data. Nevertheless, we think of it as useful in demonstrating how central the learning process—specifically, the speed of learning about ability—is to the evolution of wages with experience. The three parameterizations clearly illustrate how accounting for performance pay is key to accounting for the variance of wages over the life cycle.

In Section 7.2, we show how our integrated model compares with 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 section demonstrates the ability of our model to combine existing frameworks to offer a novel and more comprehensive account of the wage process.

We then explore the implications of the three parameterizations described for lifetime wage risk and the incentives for effort (Section 7.3), decompose piece rates into their primitive components (Section 7.4), and examine the contribution of performance pay to life-cycle wage growth and dispersion (Section 7.5). This analysis illustrates how uncertainty about worker productivity is a powerful force depressing piece rates and how performance pay is nonetheless central to the dynamics of wages.

In Section 7.6, we explore a final parameterization with both contractable and non-contractable effort to examine how the allocation of effort towards activities that are easy (simple tasks) and activities that are difficult (complex tasks) to contract varies over time in response to the incentives we focus on. 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. Importantly, this exercise shows that even once we incorporate a standard dimension of labor supply—contractable effort—performance incentives and the effort they sustain still play a key role for the growth and dispersion of wages throughout the life cycle.

All our parameterizations are disciplined by the experience profiles of average wages, the variance of wages, and piece rates from the BGH data introduced in Section 2. Numerous papers—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, jobs, and worker performance in these data.[26] All this work supports the notion that learning about ability matters for careers at the BGH firm, which is a key assumption of our model. Crucially, wage profiles in the BGH data are comparable with those documented based on more representative data sets. For instance, the log wages of male college-educated workers increase by 0.67 log points (95%) during the first 30 years of labor market experience in the BGH data, which is consistent with the wage growth of about 1 log point documented by Elsby and Shapiro [2012] from U.S. census data between 1960 and 2000. Rubinstein and Weiss [2006] find similar estimates using the PSID between 1968 and 1997 and the NLSY79.

7.1 Three Model Parameterizations

As mentioned, our first three parameterizations assume that all effort is non-contractable and thus feature $\xi_1 = \gamma_1 = 0$. Our first parameterization targets the life-cycle profile of the variance of wages and average wages from the BGH data using the three learning parameters $(\sigma_\theta^2, \sigma_\varepsilon^2, \sigma_\zeta^2)$ and the two parameters (γ_2, λ) governing human capital acquisition—its accumulation rate in the complex task and its depreciation rate, respectively. This first parameterization simply imposes that piece rates $\{b_t^*\}$ are given by their empirical counterparts in the BGH data. That is, we deliberately choose not to impose the restrictions on piece rates imposed by optimal contracting. By Section 6, we know that without performance pay, we lack crucial information to identify the risk aversion parameter r/ξ_2^2.

The implied parameter values are reported in column 1 of Table 1. Figure 2 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) across 40 years of experience. It turns out that this version of the model fits the data best when $\sigma_\zeta = 0$ 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 derives from the fact that this variance in any period t, $\text{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$, incorporates piece rates—namely, the square of the period piece rate b_t^* multiplied by the

[26] We thank Michael Gibbs for sharing these data. 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 Waldman [2012] for the similarity of patterns across the BGH firm and many others.

variance of output $\sigma_t^2 + \sigma_\varepsilon^2$, as shown in Lemma A.2. The large variance of output shocks σ_ε^2 implies that the hump-shaped pattern of piece rates in the data leads to a hump-shaped pattern for the variance of wages. (A high value for σ_θ^2 would have a similar effect on the variance of wages late in working life but would counterfactually imply a high variance early on.) Our model can thus match both the overall *level* of the variance of wages and its *decline* late in working life. This evidence validates that performance pay is central to understanding the variance of wages over the life cycle.

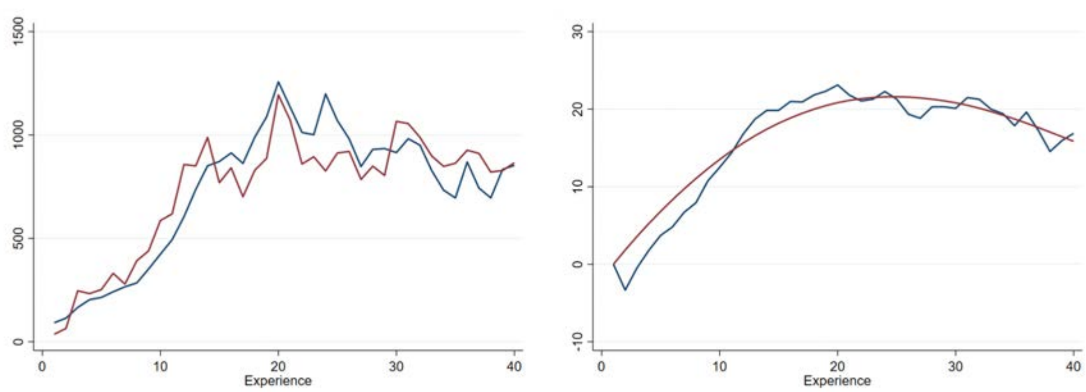
Table 1: Parameter Estimates Based on BGH Data

Parameters	Exogenous Piece Rates (1)	Endogenous Piece Rates (2)	Faster Learning ($K_0 = 0.2$ and $K_\infty = 0.05$) (3)
σ_θ: std. dev. ability (1,000 of 1988 \$)	49.1	49.2	28.9
σ_ζ: std. dev. ability shock (1,000 of 1988 \$)	0.0	0.0	2.97
σ_ε: std. dev. output shock (1,000 of 1988 \$)	439.4	522.9	57.8
γ_2: human capital acc. rate (complex task)	0.939	0.804	0.461
λ: human capital depr. rate	0.967	0.991	0.974
r/ξ_2^2: effective risk aversion	N/A	0.00024	N/A

Note: The model in column 1 imposes the observed piece rates. The model in column 2 frees all parameters and endogenizes piece rates. The model in column 3 imposes a speed of learning in $t = 0$ of $K_0 = 0.2$ and in the limit as $t \to \infty$ of $K_\infty = 0.05$ as additional parameter restrictions. All models feature only non-contractable effort ($\xi_1 = \gamma_1 = 0$), $\xi_2 = 1$, 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.

Figure 2: Fit of Model with Exogenous Piece Rates

(a) Variance of Wages (b) Wage Growth



Our second parameterization endogenizes piece rates so that we can recover the parameter r/ξ_2^2 that captures the curvature of workers' utility with respect to consumption and effort governing the trade-off between risk and incentives. Endogenizing piece rates imposes $T = 40$ additional constraints given by (9), which describes how piece rates vary as a function of the model parameters and experience. Panel c of Figure 3 shows that our full model (red line) successfully fits the life-cycle profile of performance pay in the BGH data (blue line), although it allows for only one additional parameter, r/ξ_2^2, relative to the model with exogenous piece rates. The model still matches the variance of wages and their growth over the life cycle very well. Except for r/ξ_2^2, parameter estimates are very similar to those obtained with exogenous piece rates. Thus, the model with endogenous piece rates is able to reproduce the data closely so endogenizing piece rates does not compromise the fit of our model to the remaining moments.

Figure 3: Fit of Model with Endogenous Piece Rates

It is instructive to consider the features of the data behind the large estimate of the variance σ_ε^2 of output shocks. For this result, it is not sufficient that piece rates at the end of working life (T) are low, since both the variance of output shocks and workers' degree of risk aversion can account for low piece rates. However, when σ_ε^2 is small and output signals are informative about ability so that learning about it is rapid, the life-cycle wage risk due to the uncertainty about ability would rapidly decrease over the first half of workers' careers and hence piece rates would drastically vary over the life cycle, a pattern that is highly counterfactual. As we discuss below, our third parameterization, which forces a lower value for σ_ε^2, leads to this counterfactual prediction. Instead, a large value for σ_ε^2

is consistent with both the observed level and range of piece rates over the life cycle—together with the decline of piece rates late in working life, it also helps account for the hump-shaped pattern of the variance of wages, as argued.

The estimated rates of human capital accumulation and depreciation (γ_2, λ) are reported in columns 1 and 2. That γ_2 is positive implies that human capital is acquired through learning-by-doing—we obtain a positive value since piece rates are declining late in working life, as established in Section 5.2. The estimated values of 0.939 for γ_2 and 0.967 for 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. Over the life cycle, the resulting increase in the present-discounted value of output amounts to 11.2 dollars. The estimates for the model with endogenous piece rates in column 1 have similar implications: they entail large output (social) returns to effort through human capital accumulation.

The wage (private) returns to workers, however, are much smaller due to the risk that workers face. Recall from (7) that the marginal return to effort due to human capital acquisition, $R_{HK,t} = \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}(b_{t+\tau} + R_{CC,t+\tau})$, captures the impact of an increase in current effort on future human capital, which affects a worker's presented-discounted value of future variable pay, $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} b_{t+\tau}$, and fixed pay, $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} R_{CC,t+\tau}$, as higher human capital increases (expected) output leading on average to higher beliefs about ability. Note that the term $b_{t+\tau} + R_{CC,t+\tau}$ would be equal to one each period if workers were risk neutral, and social and private returns to human capital would coincide. But since workers are risk averse, ability (σ_t^2) and output (σ_ε^2) risk dampens the benefit to workers of investing in human capital in terms of future wages—the largest value of $b_{t+\tau}$ is about 0.05 and, as shown in panels a and b of Figure 8, 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 wage risk is sizable over the life cycle is not an artifact of the large estimate of σ_ε^2 under this parameterization. Rather, it is due to a large portion of this risk being correlated over time because it arises from the uncertainty about workers' ability. We return to this point in Section 7.3. See Low et al. [2010] for very similar results on the role of persistent individual productivity risk for the wage process.[27]

So far we have shown that our model captures well the life-cycle evolution of the vari-

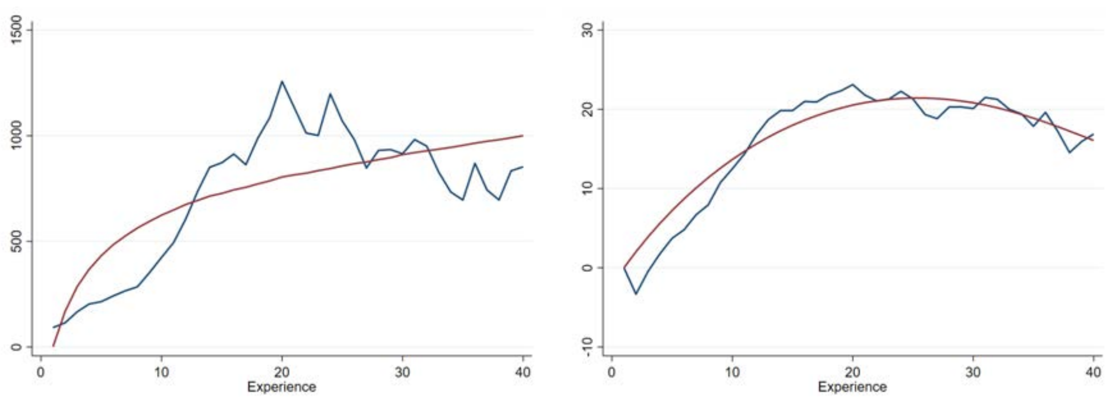
[27] Matching a shorter life-cycle horizon ($T = 30$) or allowing for worker exit from the labor market considered (which leads to a higher effective discount rate of future wages $\widehat{\delta} = \delta s$, where s is the exogenous rate of separation) would imply much lower values for σ_ε^2 and γ_2, due to the lower insurance workers demand, and a much higher speed of learning than under our baseline. Details are available upon request.

ance of wages, average wages, and piece rates with only six parameters. Some of our estimates differ from estimates in the literature on learning, though. In particular, Altonji and Pierret [1997], Lange [2007], Arcidiacono et al. [2010], and Aryal et al. [2022] suggest that firms are able to learn rapidly about unobserved worker productivity. In addition, Kahn and Lange [2014] find evidence that worker productivity changes over the life cycle so that firms continue to learn about this "moving target". Our parameter estimates, which are more in line with those in Pastorino [2024], instead imply that firms learn slowly about worker productivity. In light of these observations, we consider next a parameterization that forces a faster learning speed about the ability of both young and old workers. In particular, 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 reestimate our model imposing both the exogenous piece rates taken from the data and the two restrictions $K_0 = \sigma_\theta^2/(\sigma_\theta^2 + \sigma_\varepsilon^2) = 0.2$ and $K_\infty = \sigma_\infty^2/(\sigma_\infty^2 + \sigma_\varepsilon^2) = 0.05$. For this exercise, we target the life-cycle profiles of the variance of wages and average wages.

Figure 4 displays how this version of the model fits the data. Naturally, the fit worsens when we impose the restrictions described on the speed of learning. Yet, the model fits the variance of wages and the growth of wages over the life cycle quite well. As panel a of Figure 4 shows, however, this parameterization does not capture the decline in the variance of wages over the second-half of the life cycle. That is, it does not produce the observed hump shape in the variance of wages. Indeed, rapid learning implies that the variance of wages tends to *increase* over time. Column 3 of Table 1 reports the resulting estimates, which imply a smaller rate of accumulation of human capital, γ_2. Now, it turns out that imposing these additional restrictions while allowing piece rates to be endogenously determined as in (9) leads to piece rates vastly different from the observed ones—that is, negative when workers are young, very large in magnitude, and rapidly increasing with experience. Intuitively, since 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, thus permanently lowering their future wages.They are partially insured against this risk by firms through negative piece rates when young.[28] But as the remaining working life shortens and posterior beliefs about ability become more precise, lifetime risk decreases and hence piece rates rapidly increase—all features at odds with the data.

Figure 4: Fit of Model with Faster Learning

(a) Variance of Wages (b) Wage Growth



We have proposed three alternative parameterizations that emphasize different aspects of the data and incorporate priors on values of the model's parameters that reflect a number of estimates in the literature. As such, they offer a useful contrast that underscores the challenges of matching multiple key features of the dynamics of wages and the ability of existing frameworks to address them. Below, we will use these parameterizations to provide substantive answers to the questions about the magnitude of different sources of wage risk for workers, the determinants of the returns to effort and of piece rates, and the importance of performance pay. Before we do so, we turn to consider four prominent models in the literature that are nested by our model, which help isolate the mechanisms we integrate. By examining them one by one, we can explore the role of these mechanisms in more detail and illustrate why none of these models in isolation can account for the data.

7.2 Comparison with Leading Models in Labor Economics

Our model nests several models central to labor and personnel economics. We now turn to estimate four of them in order to illustrate the features of the data that *cannot* be matched by these nested models, which helps validate our integrated approach. This exercise also builds intuition for the features of the data that account for the estimates of the parameters

[28]Balancing negative piece rates would otherwise require extremely high output returns to human capital acquisition through γ_2, which would lead to counterfactual estimates for wage growth.

of our model. Table 2 lists the four 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 then reestimate the free parameters of each nested model using the moments it is designed to match and examine how each one fits the data.

Table 2: Description of Nested Models

Model (1)	Economic Content (2)	Restriction (3)	Untargeted Moments (4)
Human Capital (HK)	Full Information, 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 Perf. Pay (PP)	Learning, PP, no HK	$\gamma_2 = 0$	Wage growth

Human Capital Model. Consider the pure human capital model with accumulation rate γ_2 and depreciation rate λ; leading references are Ben-Porath [1967] and Becker [1962]. Contrasting panel a of Figure 5 with panel b of Figure 3 reveals that this model fits wage growth over the life cycle better than our baseline model. A depreciation rate $1-\lambda$ of roughly 4% results in the decline in wage growth late in a worker's career, whereas an accumulation rate γ_2 of 0.471 generates the rapid growth in wages early on. This result is not surprising because the parameters (γ_2, λ) are free to match the growth of wages over the life cycle, whereas in our model, they are also constrained to match the life-cycle profile of the variance of wages and piece rates. In its basic form, the human capital model has no predictions about the variance of wages—conditional on acquired human capital—or their structure, so it is silent about the level and life-cycle variability of performance pay.[29]

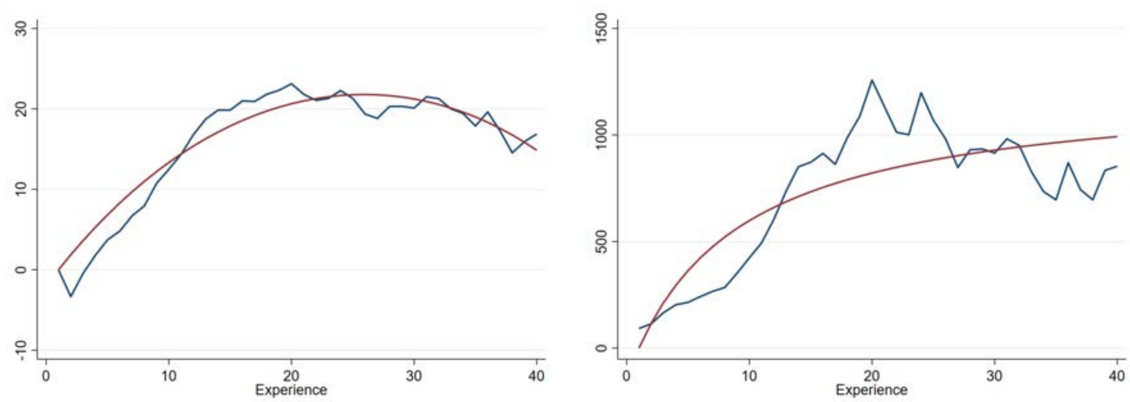
Learning Model. Consider the standard model of learning about ability without any effort choice, human capital acquisition, or contracting on performance. Farber and Gibbons [1996] propose a tractable log-linear formulation of it to account for the variance of wages over the life cycle. This learning model allows for heterogeneity in ability but has no implications for wage growth and no prediction for either the level of performance pay or its variation with workers' experience (see Table 2). Panel b of Figure 5 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 experience profile for it that

[29] Augmenting this model with heterogeneity in worker productivity would lead to variation in wages across workers if wages reflect expected output.

is at odds with the hump-shaped profile observed in the BGH data. These data suggest 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. As argued, our model improves substantially on this pattern because piece rates exhibit a hump-shaped pattern with experience and the variance of wages inherits it.

Figure 5: Fit of Nested Models

(a) Human Capital Model: Wage Growth (b) Learning Model: Variance of Wages



Career-Concerns Model. A process of learning about ability also governs the dynamics of the model of implicit incentives for performance posited 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 performance pay. Holmström [1999] shows that in the absence of explicit incentives for performance, effort cannot by sustained over time. 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, however, in equilibrium all workers choose the same effort level in any period and so the variance of wages is entirely determined 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 5.

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

Figure 6: Fit of Nested Career-Concerns Model of Gibbons and Murphy [1992]

(a) Variance of Wages

(b) Piece Rates

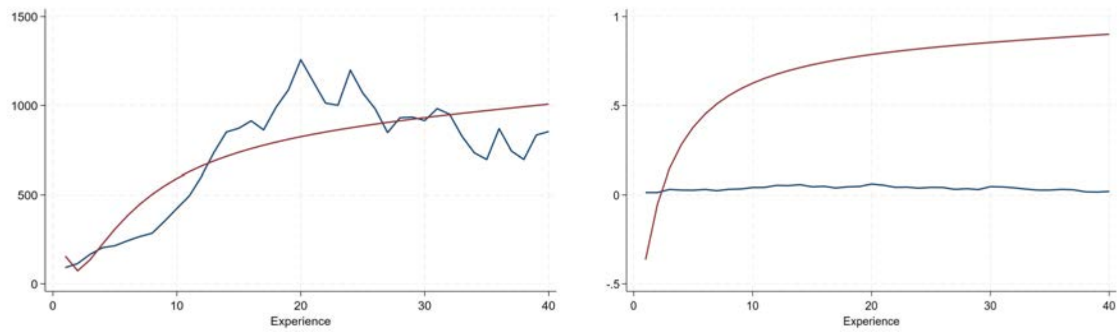


Figure 6 shows the fit of this model (red lines) to the two sets of targeted moments (blue lines). This standard career-concerns model with explicit contracting is unable to simultaneously reproduce the evolution of the variance of wages and piece rates over the life cycle. In particular, the implications of this model for the level and experience profile of piece rates are starkly at odds with the data. In the absence of a human capital motive, piece rates are predicted to start at a negative level (due to the career-concerns and insurance effects by (10)) and increase rapidly over time as ability is progressively revealed and the wage risk due to the uncertainty about it decreases. The estimate of the standard deviation of output shocks σ_ε is 29.4 thousand dollars—much smaller than the estimates for the full model in columns 1 or 2 of Table 1—so ability is effectively known after just a few years. Both workers' career concerns and desire to insure against the risk of low realizations of ability become less and less important as learning rapidly takes place, which leads workers and firms to agree to high piece rates since the performance incentive problem becomes easier to solve. As is common to learning models, workers are characterized by large differences in ability relative to σ_ε. We estimate the standard deviation of ability σ_θ to be 17 thousand dollars. This heterogeneity in ability, the rapid market learning about it, and the variability of beliefs about ability as information about it is acquired, taken together, impose substantial wage risk on workers. Such risk and firms' desire to partially insure workers against it are crucial determinants of the magnitude and variation of performance pay over the life cycle.We examine next the nature of this risk through the lens of our model.

7.3 Life-Cycle Wage Risk and Strength of Incentives

We now discuss the implications of our model for workers' wage risk and returns to effort.

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, namely, $b_t^* y_t$. 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 signal about ability z_t extracted from y_t net of expected output $\mathbb{E}[y_t|I_t]$ and $\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)$ is the weight that the updating process places on new signals by (4). Since beliefs follow a martingale process, such a change in beliefs persists in expectation over time. Thus, realized output changes the expected PDV of wages in t by $[\sigma_t^2/(\sigma_t^2+\sigma_\varepsilon^2)]\sum_{\tau=1}^{T-t}\delta^\tau\{y_t - \mathbb{E}[y_t|h_t]\}$. The standard deviations of these two components of the effect of output on future 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, which affects performance pay, as captured by the first measure, and the information about workers' ability and so their future reputation, as captured by the second measure.

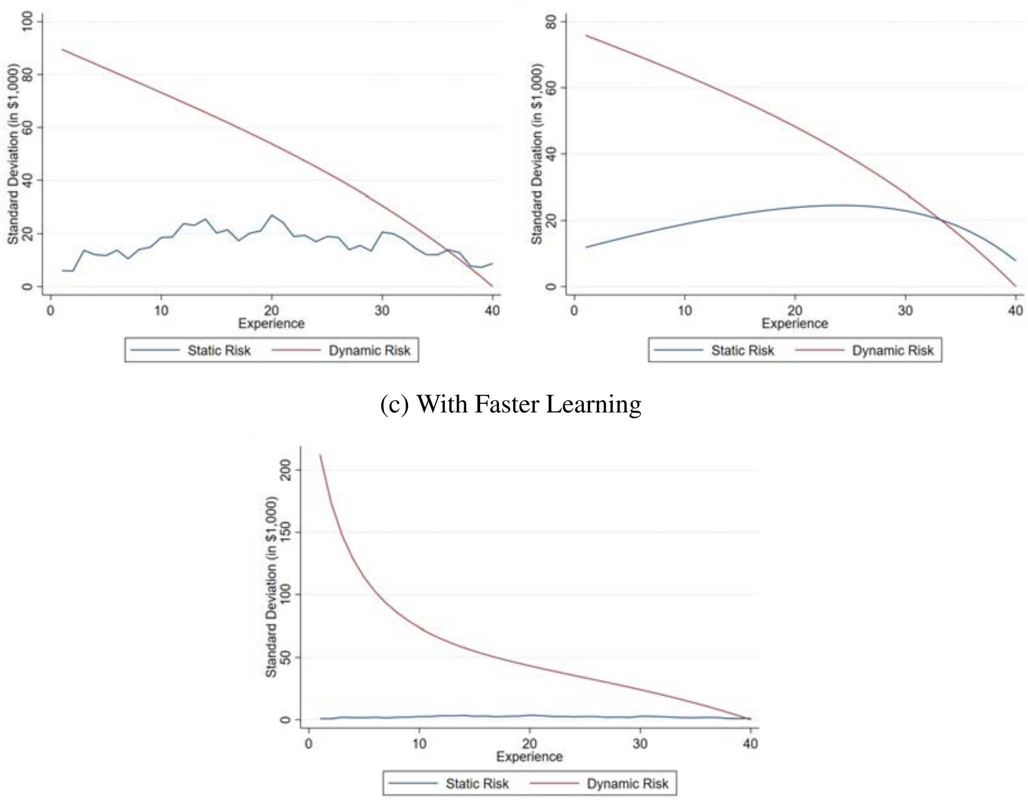
Figure 7 shows these two measures for our three parameterizations in Table 1. For all of them, the wage risk induced by learning about ability is substantially larger than the wage risk implied by contemporaneous performance pay, at least early in the life cycle. For the two models in panels a and b that do not impose a fast speed of learning, 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 roughly follows the shape of piece rates. When we impose a fast speed of learning in our third parameterization in column 3 of the table, we estimate a substantially *higher* degree of risk associated with learning about ability, as panel c shows. Specifically, the dynamic risk component early in a worker's career amounts to more than 200 thousand dollars. Young workers face such risk because firms update their beliefs about workers' ability rapidly early on, which leads to a large 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 over time given that ability is learned quickly. Under this parameterization, ability is governed by a random walk process and thus learning continues throughout the life cycle. Indeed, sizable dynamic risk persists even at 20 and 30 years of experience and it is substantially higher than the risk from contemporaneous performance pay.

Figure 7: Sources of Dispersion in Lifetime Wages

(a) With Exogenous Piece Rates

(b) With Endogenous Piece Rates

(c) With Faster Learning



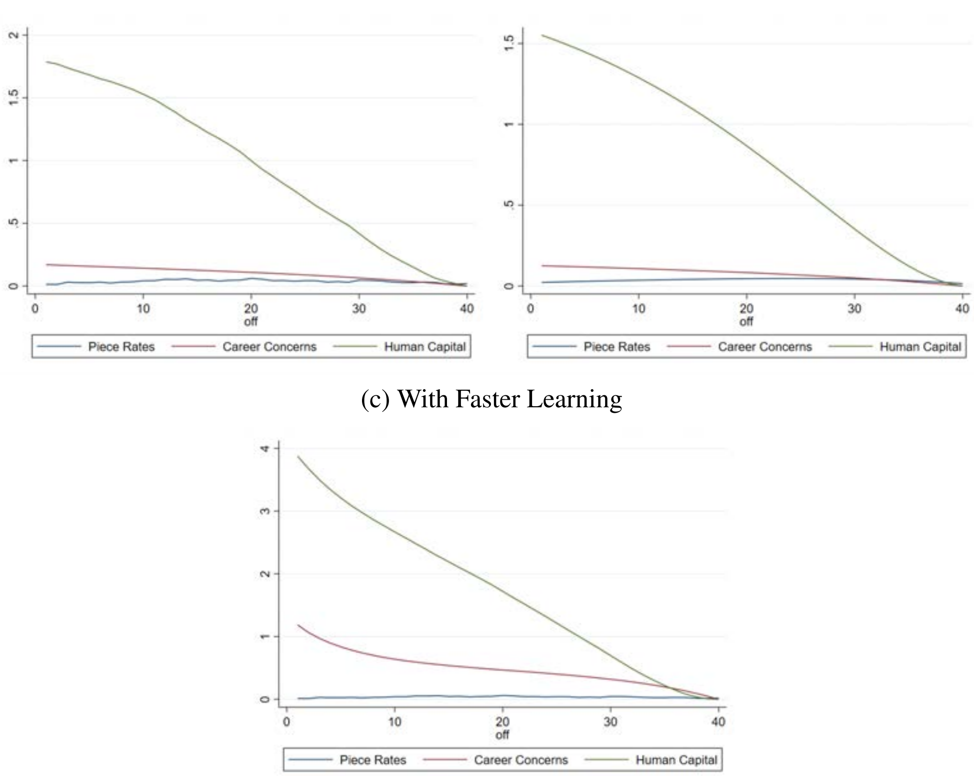
To summarize, for all three parameterizations, life-cycle wage risk due to learning about ability is much larger than contemporaneous wage risk due to performance pay. When learning is slow, the difference in magnitude between the two types of risk is not quite as stark. As a large noise in output signals results in a gradual learning process, the persistent wage risk arising from the uncertainty about ability is nonetheless sizeable. When learning is fast, highly volatile beliefs lead to an even larger degree of wage risk over the first half of the life cycle. That dynamic wage risk greatly exceeds static wage risk is then robust across very different versions of our model, whose common feature is that uncertainty about ability is a key source of the dynamics of wages over the life cycle.

Returns to Effort in the Complex Task. By (6) and (7), we can decompose the marginal returns to this effort into its determinants. In Figure 8, we plot them for the three parameterizations in Table 1. As the figure shows, across all three of them, the implicit dynamic returns from acquiring new human capital exceed those from both career concerns and performance pay by a large margin over most of a worker's career. Contrasting panels a and b with c reveals that this finding is robust to very different values for the speed of learning about ability, although the parameterization imposing fast learning implies significantly *higher* returns to 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—for two reasons. First, effort raises output in expectation and so induces firms to favorably update their beliefs about a worker's ability, which increases a worker's reputation and thus expected future wages. A higher learning speed amplifies the impact of effort on a worker's future reputation, since it increases the weight $\sigma_t^2/(\sigma_t^2 + \sigma_\varepsilon^2)$ on new information in the updating of beliefs about ability by (4), thereby increasing the marginal benefit of effort.

Figure 8: Returns to Effort in the Complex Task

(a) With Exogenous Piece Rates

(b) With Endogenous Piece Rates



(c) With Faster Learning

Second, when learning is faster, increments to human capital are more rapidly capitalized into wages—in fact, any increase in the $R_{CC,t}$ term also increases $R_{HK,t}$—which raises the benefit of higher effort. Intuitively, faster learning is associated with a smaller noise in output signals and so results in a rapidly declining variance of posterior beliefs about ability, which leads to higher piece rates—b_t^0 increases and H_t^* decreases in (10).

Higher future piece rates in turn increase the marginal benefit of effort from accumulating human capital by the first component of $R_{HK,t}$, $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} b_{t+\tau}$. Faster learning also strengthens career concerns due to the impact of effort on a worker's future reputation, as just argued, which in turn raises the marginal benefit of effort from accumulating human capital by the second component of $R_{HK,t}$, $\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} R_{CC,t+\tau}$. Hence, greater incentives for effort from learning also imply greater incentives for effort from human capital acquisition. As we show next, that dynamic incentives from human capital matter more than those from career concerns is a central feature of piece rates as well.

7.4 Decomposing Piece Rates

We have demonstrated that lifetime wage risk due to the uncertainty about ability greatly exceeds the risk due to performance pay and that the dominant component of the returns to effort are the returns from human capital. Decomposing piece rates using (10) into distinct terms representing the incentives that workers face reinforces these findings.

Figure 9 shows the contribution of each such term across 40 years of experience for our baseline parameterization in column 2 of Table 1, which features endogenous piece rates—we place the four terms in two panels because of their different scale. 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, since the variance of output shocks is large (σ_ε^2) and uncertainty about ability (σ_θ^2) is relatively small. The human capital, $(b_t^0/\xi_2)(\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - R_{HK,t}^*)$, and insurance (or hedge), $b_t^0 (r/\xi_2^2) H_t^*$, 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 in the data.

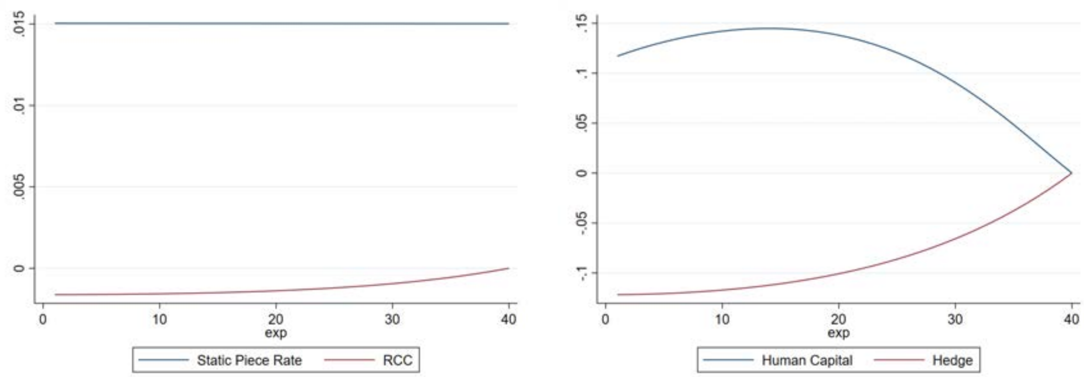
The fact that dynamic wage risk outweighs static wage risk as discussed in Section 7.3 explains why the insurance term $H_t^* = -\sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$ is so large (and negative). Because of the magnitude of life-cycle wage risk, workers have a strong desire to insure against it. Without markets providing insurance against low output realizations and so the revelation of low ability, firms partially offer such insurance in the form of low piece rates.

This insurance effect can be interpreted using basic principles in portfolio theory. A worker in any period t can be thought of as holding two assets. The first asset is performance pay, whose payout occurs at the end of t—a *performance asset*. The variation in its payout with output depends on the sign and size of the piece rate in t. When piece rates are positive, the value of this performance asset increases (respectively, decreases) with positive (respectively, negative) output realizations. The second asset is expected ability—an *ability asset*—whose value increases with output, which is a signal about its worth. With a high degree of uncertainty about ability and a long remaining career, the risk from holding the ability asset is large. The term H_t^* is an insurance premium built into piece rates in response to workers' preference for assets that correlate little (ideally, negatively) with the ability asset, which makes workers favor contracts with small or even negative piece rates.

Figure 9: Decomposition of Piece Rates

(a) Static Piece Rate and Career Concerns

(b) Human Capital and Insurance



Finally, we find that the human capital term is sizable over the life cycle. Hence, human capital has an important effect not only on the implicit incentives for effort from higher future wages but also on the explicit incentives for effort via piece rates. This result follows, conceptually, from the human capital process being of the learning-by-doing type and, practically, from the high estimated productivity of effort in terms of human capital production and the low estimated rate of human capital depreciation. Thus, although piece rates are small, given the size of the static piece-rate and career-concerns terms, acquired human capital is key to their determination—acquired human capital offsets the large estimated insurance term H_t^*. As argued in Section 5.2, it is also critical to accounting for the observed decline in piece rates over the second half of the life cycle, a pattern at odds with the predictions of standard models of career concerns and performance pay.

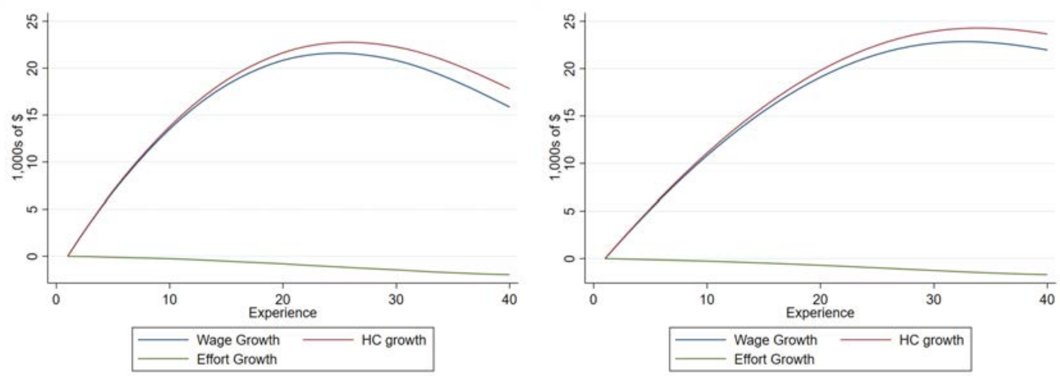
7.5 The Importance of Performance Pay

We have demonstrated how workers' aversion to the wage risk induced by the process of learning about their ability is an important force governing performance pay, which leads to small observed piece rates. The question then arises as to whether performance pay can be simply abstracted from when studying the wage process. We now argue how such an approach would miss an important source of the dynamics of wages with experience.

Wage Growth. In our model, average wages change over time as effort in the complex task and human capital vary, since $\mathbb{E}[w_{it}] - \mathbb{E}[w_{i1}] = (e^*_{i2t} - e^*_{i21}) + (k^*_{i2t} - k^*_{i21})$. In Figure 10 and panel a of Figure A.2, we show how changes in effort $(e^*_{i2t} - e^*_{i21})$ and human capital $(k^*_{i2t} - k^*_{i21})$ contribute to the life-cycle profile of wages under our three parameterizations.

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

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



Clearly, the accumulation of human capital drives wage growth over the life cycle as effort tends to decline, although its decline is moderate compared to the increase in human capital.[30] Since effort 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.2 only traces out the *direct* effect of effort on wage growth. According to our model, though, effort is instrumental to accumulating human capital: 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 the setup of Holmström [1999]. Figure 11 and panel b of Figure A.2 show the wage profiles obtained with (red lines) and without (blue lines) this restriction for our three parameterizations. Since our third parameterization assumes that piece rates are exogenous and, once they are endogenized, has implications for them at odds with the data, we find it useful to focus on our first two parameterizations for this exercise.

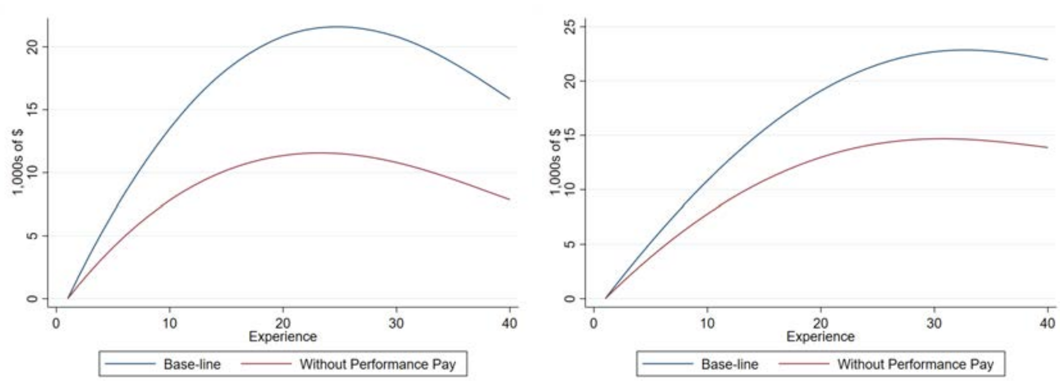
Without performance pay, firms lack an important instrument to reward performance and thereby encourage workers to exert effort. Relative to the baseline, much less effort is

[30] Under our third parameterization with a rapid speed of learning (panel a of Figure A.2), learning primarily governs effort so effort declines more rapidly than under the other two, since ability is learned quickly.

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 line) relative to our baseline (blue line), 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. Critically, although performance pay is small relative to total pay, it has a substantial impact on life-cycle wage growth through its indirect effect on workers' process of human capital accumulation.[31]

Figure 11: Sensitivity of Wage Growth to Eliminating Performance Pay

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



Wage Inequality. To measure how much performance pay contributes to wage dispersion, we begin by decomposing the variance of wages into the variance of their fixed and variable components at the estimated parameter values—recall that the variance of performance pay is $(b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2)$ in t. Figure 12 shows this decomposition for our first two parameterizations. According to our baseline parameterization in column 2 of Table 1, since shocks to output (σ_ε^2) are large, the variance of performance pay is large relative to the variance of fixed (or 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.[32] 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. (Effectively, this panel also offers a decomposition of the variance of wages in the data, since fixed pay correlates little with performance pay and so its variance can be backed out as the vertical difference between the blue and green lines.) Thus, despite performance pay

[31] As discussed, with a rapid speed of learning, learning about ability has a large impact on effort. Correspondingly, performance pay has a relatively small effect on wage growth; see panel b of Figure A.2.

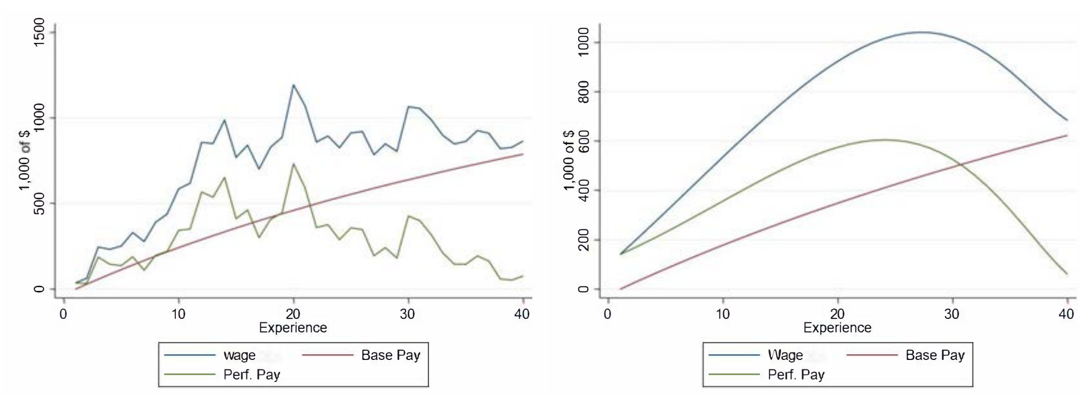
[32] Relatedly, Lemieux et al. [2009] find in the PSID that the increased prevalence of performance pay from the 1970s to the 1990s accounts for about 21% of the increase in the variance of (log) wages over this period.

accounting for only a small fraction of total pay at any time, it is responsible for a large portion of the variability of wages over the life cycle.[33]

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 panel a of Figure A.3 illustrates, its indirect impact on the level of piece rates b_t^*. In that panel, 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, 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 contracts in the absence of uncertainty about ability. Indeed, when we take into account firms' incentives to adjust piece rates, the variance of wages becomes six times larger (lavender line) than that in the baseline model (blue line); see panel b of A.3. As panel c shows, this increase is due to piece rates becoming much higher in response—up to three times as high (lavender line) as those in the baseline (blue line). Higher piece rates amplify the residual productivity risk faced by workers, leading to much higher wage dispersion over most of the life cycle. Hence, reducing differences in ability among workers *ex ante* induces firms to offer contracts with a higher sensitivity of pay to performance *ex post*, which more than compensates for the lower dispersion in ability giving rise, on balance, to much more variable wages. So, lower dispersion in ability can actually lead to much *higher* wage dispersion.

[33] As panel c of Figure A.2 shows, imposing a fast speed of learning, as we do in our third parameterization, and thus a low variance of output shocks 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.

7.6 Task Complexity over the Life Cycle

So far we have focused only on incentives for effort in the complex task.We now consider the general case of our model in which workers perform both simple and complex tasks, devoting possibly different amounts of effort to each, respectively e_{i1t} and e_{i2t} in any period t. By interpreting these two efforts as proxies of the task content of a job, the ratio $(1 + e_{i2t})/(1 + e_{i1t})$ can then be viewed as capturing the *task complexity* of worker i's job in t.

In the BGH data, an occupation or *job* is defined at the very granular level of the occupation's job title (there are 276 in the dataset); its nature, and so the complexity of the tasks it involves, can be inferred from the description of its cost center—the organizational unit a job belongs to. BGH construct the firm's job hierarchy based on 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, jobs at higher levels of the job 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, which correspond to 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%, respectively, whereas general management descriptions such as general administration or planning increase to about 30%—in practice, job levels 5 and 6 are the highest ones managers reach in our sample.[34] Based on these job descriptions and the definition of a complex task as related to "managing large groups, coordinating across business units, and strategic planning" as BGH suggest, we can then measure the degree to which a job engages a worker in non-contractable effort by the proportion of general management or overhead activities it entails and in contractable effort by the proportion of specific and easier to monitor activities, involving direct contact with customers or creating and selling products, it entails.

[34] At job levels 7 or 8, all activities are general management or planning.

We then estimate the version of our model augmented with effort in the simple task to match the variance of wages, average wages, and piece rates over the life cycle, with the additional parameter γ_1 for the rate of human capital accumulation in the simple task—recall that effort in the simple task affects a worker's output, human capital, and so wage by (1), (2), and (3), respectively. We find that this version of our model better fits life-cycle wage growth; see panel b of Figure A.4. The estimates of the parameters that are common to the baseline parameterization in column 2 of Table 1 are fairly similar. We note that the estimate of γ_1 is 0.14 and that of γ_2, around 0.75, is now only slightly lower, which confirms that effort in the complex task and performance pay still matter for the growth and dispersion of wages over the life cycle. 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, although none of these statistics has been targeted. Qualitatively, this richer model implies that as experience accumulates, workers eventually engage in more complex and harder to contract activities—our estimated measure of task complexity tends to increase with workers' tenure in the data—in line with what BGH document.

We conclude that our model replicates well the intuitive feature that workers progress over time to jobs that involve more complex tasks. Importantly, this exercise illustrates that the life-cycle profile of workers' tasks in our data validates not only our notion of jobs and the resulting assignment process but also the learning, human capital, and incentive mechanisms of our model for the wage process at the BGH firm.[35]

8 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 components. This framework is flexible in that it both nests a number of leading models of wage growth and dispersion and can be extended in several directions. For instance, it is easy to allow for a much richer stochastic process for worker productivity and for differences among workers in the efficiency of their ability and effort, which give

[35]We can extend our model to allow workers to be heterogeneous in their efficiency in performing the complex task, namely, in the rate ξ_2 at which their effort in the complex task increases output, and in how their ability contributes to output, namely, in the rate ξ_θ at which their ability θ increases output. Both cases lead to heterogeneous task assignment paths among workers. See the online appendix for details.

rise to highly heterogeneous effort paths and task assignment profiles, as we show in the online appendix. We find that two motives—namely, workers' demand for insurance against the substantial wage risk arising from the uncertainty about their productivity, and their desire to invest in human capital—have sizable effects of opposite sign on the experience profile of performance pay relative to total pay—negative and positive ones, respectively. 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 pay, our analysis illustrates its centrality to the dynamics of wages, which is due to its direct impact on the variability of wages and, most importantly, its indirect impact on the process of human capital acquisition with experience. We hope our results offer a first step toward richer models of incentives that can shed light on the sources of wage inequality and its persistence over time.

References

J. Adda and C. Dustmann. Sources of wage growth. *Journal of Political Economy*, 131(2): 456–503, 2023.

J. G. Altonji and C. Pierret. Employer learning and the signaling value of education. In *Industrial Relations, Incentives, and Employment*, Isao Ohashi and Toshiaki Tachibanaki, Eds., London:Macmillan Press Ltd., 159–195, 1997.

P. Arcidiacono, P. Bayer, and A. Hizmo. Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, 2(4): 76–104, 2010.

G. Aryal, M. Bhuller, and F. Lange. Signaling and employer learning with instruments. *American Economic Review*, 112(5):1669–1702, 2022.

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

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

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

O. Bandiera, I. Barankay, and I. Rasul. Social preferences and the response to incentives: Evidence from personnel data. *Quarterly Journal of Economics*, 120(3):917–962, 2005.

O. Bandiera, I. Barankay, and I. Rasul. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094, 2009.

O. Bandiera, I. Barankay, and I. Rasul. Social incentives in the workplace. *Review of Economic Studies*, 77(2):417–458, 2010.

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

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

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

J. DeVaro and M. Waldman. The signaling role of promotions: Further theory and empirical evidence. *Journal of Labor Economics*, 30(1):91–147, 2012.

E. Ekinci, A. Kauhanen, and M. Waldman. Bonuses and promotion tournaments: Theory and evidence. *Economic Journal*, forthcoming, 2018.

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

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

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

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

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

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

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

R. Gibbons, L. F. Katz, T. Lemieux, and D. Parent. Comparative advantage, learning, and sectoral wage determination. *Journal of Labor Economics*, 23(4):681–723, 2005.

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

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

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

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

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

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

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

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

M. P. Keane, A. Ching, and T. Erdem. Empirical models of learning dynamics: A survey of recent developments. In *Handbook of Marketing Decision Models*, Berend Wierenga and Ralf van der Lans, Eds., New York City:Springer, 223–257, 2017.

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

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

H. Low, C. Meghir, and L. Pistaferri. Wage risk and employment risk over the life cycle. *American Economic Review*, 100(4):1432–67, 2010.

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

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

E. Pastorino. Careers in firms: The role of learning and human capital. *Journal of Political Economy*, 132(6):1994–2073, 2024.

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

C. Taber and R. Vejlin. Estimation of a roy/search/compensating differential model of the labor market. *Econometrica*, 88(3):1031–1069, 2020.

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

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

A Online Appendix

In this 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 several extensions of our framework: *i*) the general cost-function case (Section A.5); *ii*) the alternative multi-tasking model in which both tasks feature non-contractable effort (Section A.6); *iii*) the model with Cobb-Douglas technology (Section A.7); *iv*) the model with wage markdowns (Section A.8); *v*) the model with observable productivity shocks (Section A.9); and *vi*) the model with heterogeneous workers, either in their ability at the complex task or in how their ability increases output (Section A.10). We conclude by reporting the figures omitted from Section 7 (Section A.11).

A.1 Data Samples

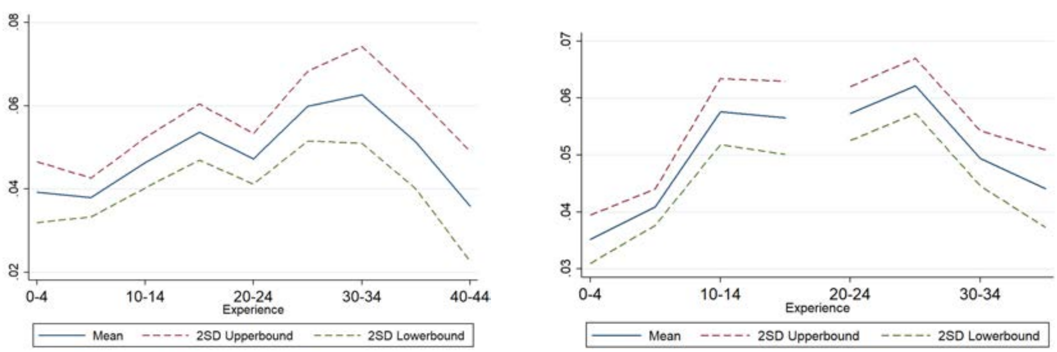
We provide here details about the data we use in our analysis.

Public Data: PSID, NLSY79, and NLSY97. 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 et al. [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 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.

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, and find very similar results—we note that only bonus pay is recorded for both the 1979 and the 1997 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 from the 2002 to the 2016 waves. The survey is administered biyearly over this period and 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 these two data sets, in the NLSY79, we can observe only individuals with 20 to 39 years of labor market 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 pay is $3,800 and $2,000, with a standard deviation of $12,000 and $4,000. We note that 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 one 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 NLSY97 and the NLSY79, which is quite similar to the one we document in the PSID—the break in panel b is due to the different cohorts tracked by the NLSY79 and the NLSY97.

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

(a) PSID (b) NLSY97 and NLSY79



Proprietary Data: BGH and GH Data. We also use data from two large U.S. firms studied in previous work and described in detail by Frederiksen et al. [2017]. As the identities of these firms cannot be disclosed, we refer to them by the names of the authors who first analyzed their data and so refer to them as the Baker-Gibbs-Holmström (BGH) firm and the Gibbs-Hendricks (GH) firm. For both firms, we have information only on white-collar workers—managers (supervisory workers) in the case of the BGH data. The BGH firm operates in a service industry, and the data from it cover the period from 1969 to 1988. Our analysis, however, is limited to the period between 1981 and 1988 because bonus pay, which is the only form of variable pay that managers receive, is not available before 1981. The BGH data contain 36,695 person-year observations and 9,800 unique individuals. As the data cover only supervisory workers, the average salary is fairly high: $55,000 (in 1988 dollars), with a standard deviation of $31,500. On average, bonus pay accounts for almost $2,000, with a standard deviation of about $7,600. Base salary makes up the remaining $53,000, with a standard deviation of $27,700. The GH data cover the years from 1989 to 1993. We cannot reveal the industry that the firm belongs to. For the GH firm, we have information about 15,648 individuals for a total of 47,715 person-year observations. As these data pertain to all white-collar employees of the firm, the average salary is intuitively lower than in the BGH data and close to $40,000 (in 1988 dollars), with a standard deviation of \$28,000. Bonus pay on average accounts for almost \$2,000, with a standard deviation of about \$9,300. The two panels of Figure 1 report the experience profile of the sensitivity of pay to performance in the BGH and GH data, respectively, for managers between 21 to 65 years of age. In both firms, performance pay is hump-shaped relative to total pay, peaking after about 20 years of experience in the BGH data and 30 years in the GH data. Analogous patterns emerge if we focus on college-educated or non-college-educated workers.

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* $0 \le t \le T-1$,

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

Proof. The result is trivially true when $t = T-1$, as $R_{CC,T} = 0$. Fix $0 \le t \le T-2$, and let $u, v \in \mathbb{R}^{T-t-1}$ be such that $u = (a_t, \ldots, a_{T-t-1})$ and $v = \big(\delta^2(1-b_{t+2}), \ldots, \delta^{T-t}(1-b_T)\big)$. Moreover, let A be the square matrix of order $T-t-1$ such that $A_{ij} = 0$ if $i < j$ and $A_{ij} = \big(\prod_{k=1}^{i-j} \mu_{t+i+1-k}\big)(1-\mu_{t+j})$ if $i \ge j$. 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 $1 \le i \le T-t-1$; note the change of variables $j \mapsto j+i-1$. So, by (11),

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

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 that is available in the same period. 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, it follows that

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

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 $t+s$. Since $\partial \mathbb{E}[z_{it}|h_i^t]/\partial e_{2t} = \xi_2$ 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

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

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 rely on (8) to derive the equilibrium choices of effort in the complex task, provided that equilibrium efforts and piece rates depend only on time, which will be 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 is a consequence of (3) and the law of iterated expectations. Since $\text{Var}[w_T|I_T] = b^2(\sigma_T^2 + \sigma_\varepsilon^2)$ and $\mathbb{E}[y_T|I_T] \propto \xi_1 e_1 + \xi_2 e_2 = \xi_1 e_1 + \xi_2^2 b$, the pair maximizing V_T is $(e_1, b) = (e_{1T}^*, b_T^*)$ with $e_{1T}^* = \xi_1$ and $b_T^* = 1/[1 + (r/\xi_2^2)(\sigma_T^2 + \sigma_\varepsilon^2)]$. Observe 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 period t on as functions of b. A worker's equilibrium employment contract in t is the pair (e_1, b) that maximizes $V_t = \mathbb{E}[W_t|I_t] - r\text{Var}[W_t|I_t]/2 - (e_{1t}^2 + e_{2t}^2)/2$, 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

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

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, by a similar argument as above, $\partial \mathbb{E}[y_{t+\tau}|I_t]/\partial e_1 = \gamma_1 \lambda^{\tau-1}$ 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

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

where $H_t^* = \sigma_t^2 \sum_{\tau=1}^{T-t} \delta^\tau$. The second term in $\partial \text{Var}[W_t|I_t]/\partial b$ reflects the fact that output in t is correlated with future output through worker ability. By increasing b and so the correlation between a worker's wage and ability in t, firms also increase the correlation between a worker's wage in t and future wages, thereby increasing $\mathrm{Var}[W_t|I_t]$. Since $\mathrm{Var}[W_t|I_t]$ is independent of e_{2t}—this holds for the same reason that effort in the complex task in t does not affect $\mathrm{Var}[W_t|I_t]$—and $\partial e_{2t}/\partial b = \xi_2$, it then follows that the first-order conditions for the problem of maximizing V_t are

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

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. So,

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

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

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

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

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

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 $\sigma_\theta^2 \geq \sigma_\infty^2$ and σ_t^2 is nonincreasing with t. Since $H_{T-1}^* > 0$ and $R_{CC,T-1}^* = \delta(1 - b_T^*)(1 - \mu_{T-1}) > 0$, $b_{T-1}^* = b_{T-1}^0(1 - R_{CC,T-1}^* - rH_{T-1}^*) < b_{T-1}^0 \leq b_T^0 = b_T^*$. 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 is true for $t = T-1$. Therefore,

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

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

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

Now note that $1 - R_{CC,t}^* - rH_t^* - b\big[1 + r(\sigma_t^2 + \sigma_\varepsilon^2)\big] \leq 0$ if $b \geq b_t^*$. Since $R_{CC,s}^* > R_{CC,t}^*$, $H_s^* \geq H_t^*$, and $\sigma_s^2 \geq \sigma_t^2$, it then follows that $b \geq b_t^*$ implies that

$$1 - R_{CC,s}^* - rH_s^* - b\big[1 + r(\sigma_s^2 + \sigma_\varepsilon^2)\big] < 0.$$

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. This concludes the first part of the proof.

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

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

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 $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_t^* = b_t^\infty$ for all such t. First note that $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_T^* = b_T^\infty$ as $\sigma_{T_0}^2 < \sigma_T^2 < \sigma_\infty^2$. Now suppose, by induction, that there exists $T_0 < t \leq T$ such that $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_{t+\tau}^* = b_{t+\tau}^\infty$ for all $0 \leq \tau \leq T - t$; the induction hypothesis is true for $t = T$. The desired result holds if $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_s^* = b_s^\infty$ for $s = t - 1$. Given that $\sigma_{T_0}^2 \leq \sigma_{s+\tau}^2 < \sigma_\infty^2$ for all $0 \leq \tau \leq T - s$, it then follows that $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} \sigma_{s+\tau}^2 = \sigma_\infty^2$, and thus $\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} \mu_{s+\tau} = \mu_\infty$, for all such τ. This, in turn, implies that

$$\lim_{\sigma_{T_0}^2 \to \sigma_\infty^2} b_s^* = \frac{1}{1 + r(\sigma_\infty^2 + \sigma_\varepsilon^2)} \left[1 - \sum\nolimits_{\tau=1}^{T-s} \delta^\tau (1 - b_{s+\tau}^\infty)\mu_\infty^{\tau-1}(1 - \mu_\infty) - r\sigma_\infty^2 \sum\nolimits_{\tau=1}^{T-s} \delta^\tau\right]$$

by the induction hypothesis and the fact that the piece rate b_s^* is jointly continuous in $(b_{s+1}^*, \dots, b_T^*, \sigma_s^2, \mu_s, \dots, \mu_T)$. 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 $0\le t\le T-1$. 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 $\gamma_2<\widetilde{\gamma}_2=\lambda\xi_2[1+(r/\xi_2^2)\sigma^2_\varepsilon]$ (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 $\sigma^2_\zeta=0$. Since the equations for the equilibrium piece rates depend continuously on σ^2_ζ and σ^2_t eventually becomes small when σ^2_ζ is small, we can extend the argument to the case in which σ^2_ζ is positive but small. Fix $T_0>0$ and let $T>T_0$; we pin down T_0 below. Now consider the difference equation

$$b^{hc}_t=\frac{1}{1+r\sigma^2_\varepsilon}\left[1+\gamma_2\sum_{\tau=1}^{T-t}\delta^\tau\lambda^{\tau-1}(1-b^{hc}_{t+1})\right]$$

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

We now show that there exist $T_0\ge 0$ such that if $T>T_0$, then b^*_t is strictly increasing with t for all $T_0\le t\le T$ provided that $|\gamma_2|$ is sufficiently small. Fix $T_0\ge 0$ and let $T>T_0$. By Lemma 1, if $\gamma_2=0$, then piece rates are strictly decreasing with t for all $T_0\le t\le T$ 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 $T_0\le t\le T$. This concludes the proof of Proposition 2.

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

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

Therefore,

$$b^*_t=b^0_t\left(1+\frac{\gamma_2}{\xi_2}(1-\mu_t)^{-1}R^*_{CC,t}-R^*_{CC,t}-\frac{r}{\xi_2^2}H^*_t\right).$$

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

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

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. This completes the proof of the proposition.

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

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 $\mathbb{E}[r_{it}] \equiv 0$. Thus, $\mathrm{Var}[w_{it}] = \mathbb{E}[r_{it}^2]$ and $\mathrm{Cov}[w_{it}, w_{it+s}] = \mathbb{E}[r_{it}r_{it+s}]$. 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. We make use of this fact in what follows.

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

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

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, since $\theta_{it}|I_{it}$ is normally distributed with mean $\mathbb{E}[\theta_{it}|I_{it} = \iota_t]$ and variance σ_t^2 when $I_{it} = \iota_t$, the random variable $(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}])|I_{it}$ is normally distributed with mean zero and variance σ_t^2. Thus, $\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] = \mathbb{E}[\mathrm{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]]|I_{it}] = \sigma_t^2$, and so, as $\mathrm{Var}[\theta_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2$, 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 $\eta^s_{it} = \mathbb{E}[\theta_{it+s}|I_{t+s}] - \mathbb{E}[\theta_{it}|I_t]$. Since

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

we then have that

$$\begin{aligned}\text{Cov}[w_{it}, w_{it+s}] &= \text{Var}[\mathbb{E}[\theta_{it}|I_{it}]] + (1b^*_{t+s})\mathbb{E}\big[\mathbb{E}[\theta_{it}|I_{it}]\eta^s_{it}\big] + b^*_t b^*_{t+s}\text{Var}[\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}]] \\ &\quad + (1 - b^*_{t+s})b^*_t\mathbb{E}[(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}] + \varepsilon_{it})\eta^s_{it}] \\ &= \sigma^2_\theta + t\sigma^2_\zeta - \sigma^2_t + b^*_t b^*_{t+s}\sigma^2_t + (1 - b^*_{t+s})b^*_t\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta^s_{it}] \\ &\quad + (1 - b^*_t)(1 - b^*_{t+s})\mathbb{E}[\mathbb{E}[\theta_{it}|I_{it}]\eta^s_{it}].\end{aligned}$$

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

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

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

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

where Λ^s_t is a zero-mean random variable. As $\mathbb{E}[(\theta_{it} + \varepsilon_{it})(\theta_{it} + \varepsilon_{it} - \mathbb{E}[\theta_{it}|I_{it}])] = \sigma^2_t + \sigma^2_\varepsilon$ and $\mathbb{E}[\theta_{it}(\theta_{it} - \mathbb{E}[\theta_{it}|I_{it}])] = \sigma^2_t$, it then follows from $(\sigma^2_t + \sigma^2_\varepsilon)(1 - \mu_t) = \sigma^2_t$ that

$$\begin{aligned}&\mathbb{E}[(\theta_{it} + \varepsilon_{it})\eta^s_{it}] \\ &= (\sigma^2_t + \sigma^2_\varepsilon)(1 - \mu_t)\left(\prod\nolimits_{j=1}^{s-1}\mu_{t+s-j}\right) + \sigma^2_t\sum\nolimits_{k=1}^{s-1}\left(\prod\nolimits_{j=1}^{s-1-k}\mu_{t+s-j}\right)(1 - \mu_{t+k}) \\ &= \sigma^2_t\left\{\prod\nolimits_{j=1}^{s-1}\mu_{t+s-j} + \sum\nolimits_{k=1}^{s-1}\left(\prod\nolimits_{j=1}^{s-1-k}\mu_{t+s-j}\right)(1 - \mu_{t+k})\right\}.\end{aligned}$$

The desired result follows from the fact that the term in braces is one. □

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, $b^*_t = \mathbb{E}[v_{it}]/\mathbb{E}[w_{it}]$.[36] With piece rates recovered, the variances $(\sigma^2_\theta, \sigma^2_\varepsilon, \sigma^2_\zeta)$ are identified as follows. First, σ^2_θ and σ^2_ε are identified from b^*_0, $\text{Var}[w_{i0}]$, and $\text{Cov}[w_{i0}, w_{i1}]$. In turn, σ^2_ζ is identified from $\text{Var}[w_{i1}]$, b^*_1, σ^2_θ, and σ^2_ε, since $\sigma^2_1 = \sigma^2_\zeta + \sigma^2_\theta\sigma^2_\varepsilon/(\sigma^2_\theta + \sigma^2_\varepsilon)$.

Risk Aversion, Human Capital in Complex Task, and Depreciation. First note that if $\{b^*_t\}^T_{t=0}$, σ^2_θ, σ^2_ε, and σ^2_ζ are identified, so are σ^2_t, $R^*_{CC,t}$, and H^*_t for all t. Thus, r is identified from b^*_T, σ^2_T and σ^2_ε, as $b^*_T = 1/[1 + (r/\xi^2_2)(\sigma^2_T + \sigma^2_\varepsilon)]$, and so b^0_t is identified for all t from r, σ^2_t, and σ^2_ε, as $b^0_t = 1/[1 + (r/\xi^2_2)(\sigma^2_t + \sigma^2_\varepsilon)]$. In turn, γ_2 is identified from b^*_{T-1}, b^0_{T-1}, b^*_T, $R^*_{CC,T-1}$, and H^*_{T-1}, since $b^*_{T-1} = b^0_{T-1}[1 + (\gamma_2/\xi_2)\delta(1 - b^*_T) - R^*_{CC,T-1} - (r/\xi^2_2)H^*_{T-1}]$. To conclude, λ is identified from b^*_{T-2}, b^0_{T-2}, γ_2, b^*_{T-1}, $R^*_{CC,T-1}$, b^*_T, $R^*_{CC,T-2}$, and H^*_{T-2}, as $b^*_{T-2} = b^0_{T-2}\{1 + (\gamma_2/\xi_2)[\delta(1 - b^*_{T-1} - R^*_{CC,T-1}) + \delta^2\lambda(1 - b^*_T)] - R^*_{CC,T-2} - (r/\xi^2_2)H^*_{T-2}\}$.

[36] The fact that $\mathbb{E}[w_{it}] = \mathbb{E}[y_{it}]$ follows from (3) and the law of iterated expectations.

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 $b_{jt}^* = \mathbb{E}_j[v_{ijt}]/\mathbb{E}_j[w_{ijt}]$.[37] 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 $(\sigma_{j\theta}^2, \sigma_{j\varepsilon}^2, \sigma_{j\zeta}^2)$ are identified from a panel of wages and their variable components. Once piece rates and $(\sigma_{j\theta}^2, \sigma_{j\varepsilon}^2, \sigma_{j\zeta}^2)$ are identified, the risk aversion parameter r_j, the rate of human capital accumulation γ_{j2}, and the depreciation rate $1-\lambda_j$ are identified from piece rates. Finally, once piece rates and $(\sigma_{j\theta}^2, \sigma_{j\varepsilon}^2, \sigma_{j\zeta}^2, r_j, \gamma_{2j}, \lambda_j)$ are identified, the rate of human capital accumulation γ_{j1} and the drift terms $\{\beta_{jt}\}_{t=0}^{T-1}$ are identified from average 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. Note that through this latent-type formulation in which workers differ

[37] 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.

in their ability distribution and human capital process in an unrestricted way, the model accommodates alternative settings in which workers of higher ability may be more or less efficient at acquiring new skills. This more general setup thus relaxes the impact of our functional-form assumptions by leading to a flexible dependence of wages on ability, uncertainty about it, human capital, risk, and workers' risk attitudes.

A.5 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 $\eta^2 < 1$.[38]

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 $0 \leq t < T$ and suppose the equilibrium efforts and piece rates 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 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 $V_t = \mathbb{E}[W_t|I_t] - (r/2)\mathrm{Var}[W_t|I_t] - (e_1^2 + 2\eta e_1 e_2 + e_2^2)/2$, 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, since it is still the case that $\partial e_2/\partial b = \xi_2$. Now note that

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

[38] 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.

unlike in the baseline case, effort in the simple task now affects effort in the complex task ($\partial e_2/\partial e_1 = -\eta \neq 0$), and this has an impact on $\partial \mathbb{E}[W_t|I_t]/\partial e_1$. As effort in any task does not affect the variance of current or future wages, the fact that now effort in the complex task responds to changes in effort in the simple task does not affect the partial derivatives $\partial \text{Var}[W_t|I_t]/\partial b$ and $\partial \text{Var}[W_t|I_t]/\partial e_1$. Finally, since $dc(e_1, e_2)/db = (\eta e_1 + e_2)\partial e_2/\partial b = \xi_2(\eta e_1 + e_2)$ and $dc(e_1, e_2)/de_1 = e_1 + \eta(e_2 + e_1 \partial e_2/\partial e_1) + e_2 \partial e_2/\partial e_1 = (1 - \eta^2)e_1$, it follows that the necessary and sufficient conditions for the problem of maximizing V_t are $\xi_1 - \eta\xi_2 + (\gamma_1 - \eta\gamma_2)\sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (1 - \eta^2)e_1 = 0$ 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 $(e_1, b) = (e_{1t}^*, b_t^*)$ 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 (9), 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 next appendix, we extend our analysis to the case in which effort in both tasks in non-contractable.

A.6 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.

A.6.1 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_θ^2 and $\varepsilon_{i\ell t}$ is normally distributed with mean zero and variance $\sigma_{\ell\varepsilon}^2$. 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; as in the baseline model, we consider short-term employment contracts with linear wage schedules. 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 that worker ability is constant over time 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.

A.6.2 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[39]

$$m_{it+1} = \frac{\sigma^2_\varepsilon}{\sigma^2_{it} + \sigma^2_\varepsilon} m_{it} + \frac{\sigma^2_{it}}{\sigma^2_{it} + \sigma^2_\varepsilon} z_{it} \quad \text{and} \quad \sigma^2_{it+1} = \frac{\sigma^2_{it}\sigma^2_\varepsilon}{\sigma^2_{it} + \sigma^2_\varepsilon}.$$

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) 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^2_ε 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^2_{it} 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^2_t. For each $0 \leq t \leq T$ and $0 \leq \tau \leq T - t$, let $\Sigma_{t+\tau} = \sigma^2_t/(\tau\sigma^2_t + \sigma^2_\varepsilon)$. 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

$$m_{it+\tau} = \frac{\sigma^2_\varepsilon}{\tau\sigma^2_t + \sigma^2_\varepsilon} m_{it} + \Sigma_{t+\tau} \sum_{s=0}^{\tau-1} z_{it+s}.$$

[39]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.

Effort Choices. As in the baseline model, the equilibrium is unique, symmetric, and such 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

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

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

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

These equations state that for each task, the marginal cost of effort in the 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 $1 \le \tau \le T-t$, 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

$$e_{1t} = \frac{1}{1-\eta^2}\left[b_{1t} + R_{CC,1t} + R_{HK,1t} - \eta\left(b_2 + R_{CC,2t} + R_{HK,2t}\right)\right]; \text{ and} \quad (18)$$
$$e_{2t} = \frac{1}{1-\eta^2}\left[b_{2t} + R_{CC,2t} + R_{HK,2t} - \eta\left(b_1 + R_{CC,1t} + R_{HK,1t}\right)\right]. \quad (19)$$

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, for ease of notation, 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 tasks are instead substitutes, then increasing the piece rate for one task decreases effort in the other task.

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 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 $t+\tau$. Moreover, 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. Therefore, a worker's effort in task ℓ in period t when the period-t 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 $V_t = \mathbb{E}[W_t|I_t] - (r/2)\mathrm{Var}[W_t|I_t] - c(e_1, e_2)$, 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

$$\begin{aligned} \mathrm{Var}[W_t|I_t] \;=\; & b_1^2(\xi_{1\theta}^2\sigma_t^2 + \sigma_{1\varepsilon}^2) + b_2^2(\xi_{2\theta}^2\sigma_t^2 + \sigma_{2\varepsilon}^2) + 2b_1b_2\xi_{1\theta}\xi_{2\theta}\sigma_t^2 \\ & + 2\sum\nolimits_{\tau=1}^{T-1} \delta^\tau \mathrm{Cov}[w_t, w_{t+\tau}|I_t] + \mathrm{Var}_0, \end{aligned}$$

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

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

where $H^*_{\ell t} = \sum_{\tau=1}^{T-1} \delta^{\tau-1} \partial \mathrm{Cov}[w_t, w_{t+\tau}|I_t]/\partial b_\ell$ 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, $b^0_{\ell t} = 1/[1 + r(1-\eta^2)(\xi^2_{\ell\theta}\sigma^2_t + \sigma^2_{\ell\varepsilon})]$ 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

$$b_1 = b^0_{1t}\left[\mathcal{W}_{1t} - \eta(\mathcal{W}_{2t} - b_2) - r(1-\eta^2)(H^*_{1t} + b_2\xi_{1\theta}\xi_{2\theta}\sigma^2_t)\right];$$

$$b_2 = b^0_{2t}\left[\mathcal{W}_{2t} - \eta(\mathcal{W}_{1t} - b_1) - r(1-\eta^2)(H^*_{2t} + b_1\xi_{1\theta}\xi_{2\theta}\sigma^2_t)\right].$$

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

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

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

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, $r(1-\eta^2)(H^*_{\ell t} + b^*_{-\ell t}\xi^2_{1\theta}\xi^2_{2\theta}\sigma^2_t)$, 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^0_{1t} and b^0_{2t} 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. When tasks are substitutes, this term tends to depress piece rates.

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 technology 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 $Y_{it} = \Theta_i K_{it} E^{\xi_1}_{i1t} E^{\xi_2}_{i2t} \Omega_{it}$, where Θ_i is the worker's unobserved ability, which we assume is time-invariant for simplicity, E_{i1t} is the worker's effort in the simple task, E_{i2t} is the worker's effort in the complex task, 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 $(m_\theta, \sigma^2_\theta)$ and $(0, \sigma^2_\varepsilon)$, 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 $\sum_{\tau=0}^{T-t} \delta^{\tau} \left(\ln(W_{t+\tau}) - \ln(E_{1t+\tau})^2/2 - \ln(E_{2t+\tau})^2/2\right)$.[40]

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 $b_{it} \in \mathbb{R}$. Note that $b_{it} = (Y_{it}/W_{it}) dW_{it}/dY_{it}$, the elasticity of wage payments with respect to output. Therefore, we can 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 $\mathbb{E}[Y_{it}|I_{it}] = \mathbb{E}[W_{it}|I_{it}] = C_{it}\mathbb{E}[Y_{it}^{b_{it}}|I_{it}]$, 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 below.

A.7.2 Equilibrium Characterization

We now characterize the equilibrium and show it is the same as in the baseline model when the workers' (effective) coefficient of risk aversion is $1/\xi_2^2$.

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 $\sigma_\zeta^2 = 0$.

Effort in the Complex Task. We first derive the consequences 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, $\mathbb{E}\{\exp[a(\theta_i + \varepsilon_{it})]|I_{it}\} = \exp\left(a\mathbb{E}[\theta_i|I_{it}] + a^2(\sigma_t^2 + \sigma_\varepsilon^2)/2\right)$ 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

$$w_{it} = \ln(W_{it}) = (1-b_{it})\mathbb{E}[y_{it}|I_{it}] + b_{it} y_{it} + (1-b_{it}^2)(\sigma_t^2 + \sigma_\varepsilon^2)/2. \tag{20}$$

Thus, as in the baseline model, employment contracts are completely described 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 has the property that (log) 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 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 (20) with $b_{it+\tau} \equiv b_{t+\tau}$ for all $0 \le \tau \le T-t$. Since

[40] 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 $-\exp\left\{-r\left[\sum_{\tau=0}^{T-t} \delta^{\tau}\left(\ln(W_{t+\tau}) - \ln(E_{1t+\tau})^2/2 - \ln(E_{2t+\tau})^2/2\right)\right]\right\}$, where $r > 0$. The parameter r is not the workers' coefficient of risk aversion, though.

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) that maximizes $V_T = \mathbb{E}[w_T|I_T] - (e_1^2+e_2^2)/2$, where I_T is as before.[41] Since, by (20), $\mathbb{E}[w_T|I_T] = \mathbb{E}[y_T|I_T] + (1-b^2)(\sigma_T^2+\sigma_\varepsilon^2)/2 \propto \xi_1 e_1 + \xi_2^2 b + (1-b^2)(\sigma_T^2+\sigma_\varepsilon^2)/2$, 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 $1 \leq \tau \leq T-t$. 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 $1 \leq \tau \leq T-t$, $\mathbb{E}[w_{t+\tau}|I_t] = \mathbb{E}[y_{t+\tau}|I_t] + w_{t+\tau}^0$, where $w_{t+\tau}^0$ 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 $1/\xi_2^2$.[42]

A.8 Extension: Wage Markdowns

We now 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.8.1 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 $\mathbb{E}[y_{it}|I_{it}]$. So, 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.8.2 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).

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

[42] With the more general preference specification of Footnote 40, one can show that the equilibrium is as in the baseline model when the workers' coefficient of risk aversion is $(1+r)/\xi_2^2$.

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 $R_{CC,t} = \sum_{\tau=1}^{T-t} \delta^\tau (\alpha - b_{t+\tau})(\prod_{k=1}^{\tau-1} \mu_{t+\tau-k})(1-\mu_t)$ and $R_{HK,t}$ has the same expression as in the baseline model. 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 $0 \leq \tau \leq T-t$ 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. Since workers now capture only a fraction α of their expected output, $\partial\mathbb{E}[W_t|I_t]/\partial b = \alpha(\xi_2^2 + \xi_2\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1})$ 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 $1 \leq \tau \leq T-t$, so $\partial\mathrm{Var}[W_t|I_t]/\partial b = 2b(\sigma_t^2 + \sigma_\varepsilon^2) + 2\alpha H_t^*$. Finally, for the same reason as in the baseline case, $\partial\mathrm{Var}[W_t|I_t]/\partial e_1 = 0$. From this, it follows that 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

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

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.8.3 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.9 Extension: Productivity Shocks

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

A.9.1 Environment and Equilibrium

The environment is the same 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.[43] 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 $b_t^0 = 1/[1 + (r/\xi_2^2)(\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)]$.

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

A.9.2 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 $\text{Var}[w_{it}] = \sigma_\theta^2 + t\sigma_\zeta^2 - \sigma_t^2 + (b_t^*)^2(\sigma_t^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)$; 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 $\text{Cov}[w_{i0}, w_{i1}] = b_0^*\sigma_\theta^2$. In turn, $\sigma_\varepsilon^2 + \sigma_\eta^2$ is identified from b_0^*, σ_θ^2, and $\text{Var}[w_{i0}] = (b_0^*)^2(\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_\eta^2)$. Next, σ_1^2 is identified from b_1^*, $\sigma_\varepsilon^2 + \sigma_\eta^2$, and $\text{Cov}[w_{i1}, w_{i2}] - \text{Var}[w_{i1}] = (b_1^*)^2(\sigma_1^2 + \sigma_\varepsilon^2 + \sigma_\eta^2) - b_1^*\sigma_1^2$, and so σ_ζ^2 is identified from b_1^*, σ_θ^2, σ_1^2, and $\text{Cov}[w_{i1}, w_{i2}] = \sigma_\theta^2 + \sigma_\zeta^2 - \sigma_1^2 + b_1^*\sigma_1^2$. 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.

A.9.3 Remarks

We considered the case of idiosyncratic productivity shock for simplicity. We can extend our analysis to the case in which productivity are serially correlated by assuming that they behave over time according to 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 $\sigma_{\eta t}^2 = \text{Var}[\eta_{it}] = (1 - \rho^{t+1})\sigma_\mu^2/(1 - \rho)$. Because of the serial correlation of productivity shocks, this model admits a more general variance-covariance structure of wages.

A.10 Extension: Heterogeneous Workers

In the last extension we consider, we allow workers to be either heterogeneous in their ability to perform at 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 in our remarks at the end.

A.10.1 Setup and Equilibrium

Workers are heterogeneous in the rate at which their effort in the complex task affects output, being 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 $e_{j2t}^* = \xi_{j2} + \gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_{j2})[(\sigma_t^2 + \sigma_\varepsilon^2)b_{jt}^* + H_t^*]$, where σ_t^2 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 $[1+\xi_{j2}+\gamma_2 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1} - (r/\xi_{j2}) H^*_t]/[1+\xi_1+\gamma_1 \sum_{\tau=1}^{T-t} \delta^\tau \lambda^{\tau-1}]$, 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.10.2 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 7.6. 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 a finite mixture 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 wage distributions of a given type of workers.[44] Since $s^*_{jT} = (1+\xi_{j2} b^*_{jT})/(1+\xi_1)$ increases strictly with j, and so the type-J workers and only them occupy the highest-level job in period T, it follows that ξ_{2T} is identified from b^*_{JT} and s^*_{JT}. Then, r is identified from σ^2_T, σ^2_ε, and $b^*_T = 1/[1+(r/\xi^2_{JT})(\sigma^2_T+\sigma^2_\varepsilon)]$. 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 worker is the same as in the baseline model.

A.10.3 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 that a worker's performance is more informative about their ability, which is equivalent to a higher σ^2_θ. The equilibrium in this model is such that workers are heterogeneous in their piece rates, and thus on their effort in the complex task. One can show that when piece rates are small, effort in the complex task strictly increases with σ^2_θ, as a higher σ^2_θ translates into a smaller σ^2_t for all t. Thus, as above, workers with higher productivity in the complex task are assigned to higher-complexity jobs at every experience. Identification of this setup proceeds along the same lines as above.

A.11 Omitted Quantitative Exercises

We present here figures omitted from the main text that are referenced in Section 7.

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

Figure A.2: Results for Parameterization with Faster Learning

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

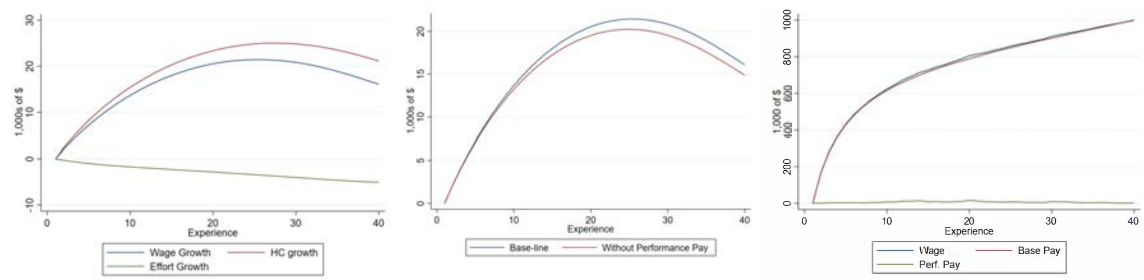


Figure A.3: Variance of Wages and Piece Rates without Uncertainty about Ability

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

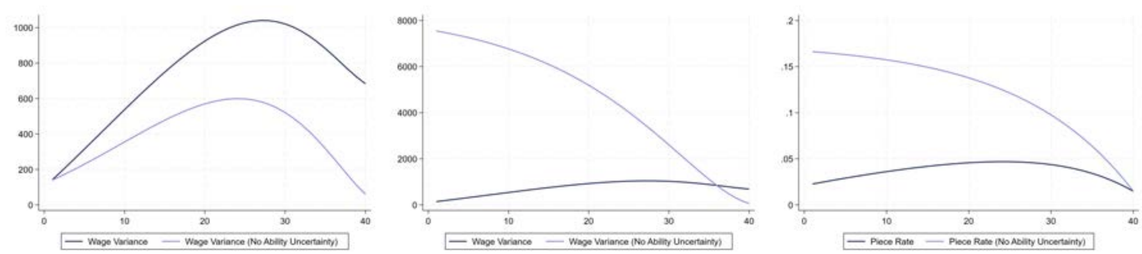


Figure A.4: Fit of Model with Endogenous Piece Rates and Multidimensional Effort

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