Employer Learning, Productivity and the Earnings Distribution: Evidence from Performance Measures*

Lisa B. Kahn and Fabian Lange†
Yale University

February 28th, 2011

Abstract

Two ubiquitous empirical regularities in pay distributions are that the variance of wages increases with experience and innovations in wage residuals have a large, unpredictable component. The leading explanations for these patterns are that over time, either firms learn about worker productivity but productivity remains fixed or workers' productivities themselves evolve heterogeneously. In this paper, we seek to disentangle these two models and place magnitudes on their relative importance. We derive a dynamic model of learning and productivity that nests both models and then estimate our model on a 20-year panel of pay and performance measures from a single, large firm. The advantage of these data is that they provide us with repeat measures of correlates of productivity that are in part not observed by the firm when it sets wages. Our estimates show that wages differ significantly from individual productivity all along the life-cycle and both heterogenous productivity changes and employer learning are important for understanding the wage dynamics. We then use our estimates to calculate the degree to which imperfect learning introduces a wedge between the private and social incentives to invest into human capital. We find that these disincentives exist all over the life-cycle but increase rapidly after about 15 years of experience. Thus, in contrast to the existing literature on employer learning, we find that imperfect learning is highly relevant for older workers.

*We are grateful for helpful comments from Joe Altonji, David Autor, Steffen Habermalz, Paul Oyer, Chris Taber, Michael Waldman, and seminar participants at Collegio Alberto Turin, Columbia University, the 9th IZA/SOLE Transatlantic Meetings, IZA - Bonn, the 2010 SOLE meetings, University of California - Berkeley, the 2010 NBER Summer Institute, Maastricht University, Princeton University, University of Alicante, University of British Columbia, University of California - San Diego, University of Koeln, University of Rochester, University of Southern California, University of Wisconsin, the Utah Winter Business Economics Conference 2010 and Yale University. We thank Mike Gibbs and George Baker for providing the data. Doug Norton provided able research assistance.

†Lisa Kahn, Yale School of Management, 135 Prospect St, PO Box 208200, New Haven, CT 06510. Email: lisa.kahn@yale.edu. Fabian Lange, Yale Department of Economics, 37 Hillhouse Ave, New Haven, CT 06511. Email: fabian.lange@yale.edu.

1 Introduction

Observationally identical workers often earn vastly different wages. Controls for education, experience, and demographic characteristics typically remove only 20 to 30% of the variation in wages. Furthermore, the variance in wage residuals tends to increase with age. Why wages vary so much across observationally similar workers and why this variation increases with age are central questions of labor economics. One set of answers to these questions stresses that workers experience unforeseen productivity shocks and that productivity changes heterogeneously over the life-cycle in ways not captured by standard controls in Mincer earnings regressions. An alternative set of answers emphasizes that worker skills are not easily observed by employers; instead employers need to learn about the skills and abilities of workers. Central to these explanations is a process by which information about differences between workers is slowly revealed to the labor market. Wages diverge as potential employers learn to distinguish individual skills.

Both hypotheses can account for two fundamental empirical regularities regarding wage residuals: the variance of wage residuals increases with experience and innovations in wage residuals have a large unpredictable component.[1] Estimating the relative contributions of heterogeneous productivity and employer learning to pay changes over the life cycle is the task of this paper.[2] We develop a new methodology exploiting information such as that commonly collected in personnel data sets to identify models that incorporate both explanations. We derive a dynamic model

[1] These findings are intuitive. In learning models, wages equal expected productivity conditional on the information available at any age. The variance of conditional expectations increases as the conditioning set increases, implying the same for the variance of wage residuals. Furthermore, because past wages are included in the firm's information set, wage growth will be uncorrelated over time. Unrestricted productivity models can match the observed patterns in wages by simply assuming a stochastic dynamic process on productivity that mirrors the observed stochastic properties of the wages.

[2] The paper assumes that individuals productivity includes a deterministic component that evolves similarly over the life-cycle for all workers. This component captures the empirical regularities captured by typical Mincer earnings equations. Our interest lies not in this deterministic component of wages, but rather in the variation around this deterministic profile.

where firms must learn about the skills of their workers and worker productivity itself varies stochastically over time. Firms set wages equal to expected productivity, hence wages vary over the life-cycle because of both firm updating and productivity evolution. This nested model provides a well specified alternative against which we can test the alternative pure models that restrict either firm learning or heterogenous productivity dynamics to play no role in wage dynamics. We find that both learning and heterogenous productivity are important for explaining the dynamics in wages and estimate the parameters of the nested model.[3] To illustrate how relevant learning might be along the life-cycle, we use our estimates to provide empirical evidence on how disincentives to invest in human capital due to incomplete information change with age.[4]

Distinguishing between the alternative hypotheses of heterogeneous productivity and employer learning is intrinsically difficult because it is rarely possible to directly observe the productivity of individual workers. The growing empirical literature on employer learning (Farber and Gibbons (1996), Altonji and Pierret (2001), Lange (2007), Schönberg (2007), Arcidiacono, Bayer, and Hizmo (2010), and others) exploits a correlate of productivity, measured prior to labor market entry, that is available to researchers but (they argue) not to firms. In practice, this literature relies almost exclusively on the AFQT score, a composite score derived from a battery of tests administered to the respondents of the NLSY79. The fact that wages increasingly correlate with the AFQT score over the life-cycle is seen as evidence for employer

[3] Our model is designed to empirically evaluate the relative importance of the employer learning model and the hypothesis that productivity evolves stochastically. In order to arrive at a tractable, estimable specifications we need to abstract from an important mechanism that relates individuals careers to employer learning: task assignment based on information learned by employers over time (see eg. Gibbons and Waldman (1999, 2006)). In our model, individuals are endowed with a single skill that evolves over time and that employers learn about. Employers however do not assign individuals to different tasks based on what they have learned. Such assignment can lead to interesting feedback mechanisms between learning on the part of employers and how individual productivity evolves over the life-cycle. This paper however abstracts from these mechanisms in order to arrive at a simple estimable model. This can only be justified by the fact that we believe that this empirical model does produce interesting insights into the dynamics of individuals'careers.

[4] To our knowledge, we are the first to provide evidence on this question.

learning. However, a major drawback of this literature is that the AFQT score was collected prior to the labor market entry of these workers. Therefore models examined in this literature cannot allow productivity to vary heterogeneously over the life cycle. Another drawback is that one needs to assume that employers did not use the AFQT score when setting wages, even though knowledge of the test score is valuable and that it may have been possible to collect.[5]

In this paper, we provide new evidence on whether employer learning or changes in worker productivity drive individual wage dynamics over the life-cycle. To do this, we use a 20-year unbalanced panel data set of all managerial employees in one firm, previously analyzed in Baker, Gibbs and Holmstrom (1994a and 1994b, BGHa and BGHb hereafter).[6] For our purposes, these data have the crucial advantage that they contain both annual pay of workers as well as performance ratings in the form of subjective managerial assessments. The panel structure allows us to observe performance ratings that were collected prior to, contemporaneous to, and after the current period. The latter provides us with information about worker productivity that the firm was not able to exploit when setting wages. We can thus dispense with the ad-hoc assumption on the information available to employers that was previously required in this literature. Further, these repeat performance ratings obtained at various points over the life-cycle allow us to estimate dynamic specifications of productivity and

[5] A related, prior literature analyzes the second moments of wage residuals to understand the roles information revelation and hetergeneous productivity play in wage dynamics (e.g., Abowd and Card 1989, Hause 1980, MaCurdy 1982 and Baker 1997). The observation that log wage residuals have a large unpredictable component is seen as evidence against human capital models. By contrast, observing that wage changes are correlated over time is seen as evidence for systematic differences in human capital accumulation. A major obstacle in this literature is that it focuses exclusively on wages and does not use other information on worker productivity. Therefore this literature cannot determine whether changes in wages are due to changes in productivity itself or the information held by employers.

[6] These landmark studies provided early empirical evidence on the internal organization and pay dynamics of the firm. Their findings have inspired the well known contributions by Gibbons and Waldman (1999 and 2006) who reconcile most of the BGH findings by combining simple models of job (and later task) assignment, human-capital acquisition and learning. In addition, Gibbs (1995) describes the empirical relationship between pay, promotions and performance and DeVaro and Waldman (2007) use the data to test the Waldman (1984) promotion-as-signal hypothesis.

learning that go beyond those currently estimated in the literature.

We show that the correlations of pay with performance, measured at various lags and leads, are particularly informative for distinguishing between employer learning and dynamic productivity models. For example, a pure learning model predicts that these correlations of pay with past performance measures exceed the correlations of pay with future performance measures. This is because firms rely on past, but not future, performance measures to set current pay. Over time, as firms' priors become more precise and they update less on new signals, this difference in the correlations of pay with past and future performance measures should decline. In contrast, an implication of the full information pure productivity model is that wages correlate similarly with past and future performance evaluations.

In isolation, neither model can fully reproduce the moments of the data. We find evidence for employer learning in that we observe that wages are more highly correlated with past rather than future performance ratings. However, we observe this pattern even for workers at high experience levels, contradicting the pure learning model. When estimating the full model, these facts lead us to conclude that the firm does learn about worker ability *and* that productivity evolves over time. Somewhat surprisingly, we find that the initial variance in productivity is quite small and employers seem to be well informed about the skills of workers at the outset of their careers.[7] Over time, productivity evolves substantially, thanks to both a predictable and a random walk component. Therefore the firm must continuously learn about a moving target, even at high experience levels. Thus, the majority of the observed growth in dispersion in wage residuals reflects increasing heterogeneity in individual productivity. However, imperfect employer learning means that it requires a number of years for productivity differences to be priced into wages.

[7] This finding is however consistent with Arcidiacono et al. who show that firms have more precise initial expectations about college graduates than high school graduates. Our sample of managers, reflecting highly skilled workers, should be more similar to the college graduates sample.

These findings have important implications about individuals' incentives to invest in their human capital. If labor markets find it difficult to distinguish productive and unproductive workers, then workers find it less valuable to invest in their human capital.[8] In prior work (Lange (2007)), one of us argued that productivity of young workers is rapidly revealed to the labor market implying that individuals will capture most of the benefits of early investments in human capital. In this paper, we allow productivity to vary heterogeneously all along the life-cycle and then estimate how fast employers learn about these changes. The size of the gap between the social returns to investing in human capital and the private returns depends on how rapidly employers learn about worker producitivity and how long the horizon facing individuals is. If learning is relatively rapid, any human capital investment made by younger workers will be priced into wages after a few years. Younger workers then can enjoy the fruits of their investments for the remainder of their career. For older workers, the period of learning when their investment are imperfectly priced into wages looms larger and they will capture smaller and smaller proportions of the social returns of their human capital investments.

We find that the fraction of the social returns to human capital investments going to workers declines steadily over the life-cycle. Workers in their twenties and thirties capture about three quarters of the productivity return. However, after about 15 years of experience, the share of the returns to investment going to individuals declines rapidly. It declines to about 65% after 20 years of experience, 40% after 30 years, and 25% after 35 years of experience. Therefore the incentives to invest are much more severely misaligned for workers in middle and old age than for younger workers. The prior literature on employer learning has focused on the consequence

[8] A theoretical literature (see Change and Wang 1996, Katz and Ziderman 1990 and Waldman 1990) posits that when firms learn asymmetrically about worker ability, workers could underinvest in general skills. This point has not been applied to symmetric employer learning models (the case we consider in this paper), likely because estimating these models has required the assumption that worker productivity cannot evolve heterogeneously.

of learning for young workers. Given the observed speed of learning (Lange 2007), this literature suggests that the consequences of incomplete information for human capital investments for this age group are limited. Our study suggests that incentives are more severely misaligned for human capital investments of older workers. We believe this reinterpretation of the traditional employer learning model – that the consequences of imperfect learning are more severe among older rather than younger workers – represents a significant contribution to the empirical literature on employer learning.

The remainder of this paper is structured as follows. Section 2 introduces the model of learning and productivity, shows how this model nests the pure learning and pure productivity models, and discusses the identification of these two models. Section 3 describes the data. Section 4 reports the estimation method and results and evaluates the fit of the model. In Section 5, we discuss what these estimates imply for how learning and productivity contribute to wage dynamics over the life-cycle and we show how imperfect learning affects the incentives to invest into human capital. Section 6 briefly touches upon alternative models of wage dynamics that might explain the data. Section 7 concludes the paper. A more general formulation of the model, a formal identification argument of the two basic constituent models, and a discussion of attrition are relegated to the appendices.

2 A Model of Learning and Productivity

In this section, we introduce the model that we use to organize the discussion and empirical evidence.[9] We have chosen a parsimonious specification that nests two of the main models of wage dynamics, the pure employer learning model and the pure productivity model. Each model represents a distinct view point about how wages

[9] The models we analyze in this paper are special cases of a more general class of models of learning and productivity that can be analyzed using the tools developed in this paper. We present this more general class of models in the appendix.

evolve over the life-cycle. The pure learning model goes back to the specification analyzed by Farber and Gibbons (1996) (see also Altonji and Pierret (2001) and Lange (2007)). This model assumes that individual heterogeneity in productivity is fixed across the life-cycle and that wage dynamics are driven entirely by learning on the part of employers. The pure productivity model instead assumes that employers are perfectly informed about worker productivity. Wage dynamics in this model reflect variation in productivity over the life-cycle. The nested model allows for both: productivity varies heterogeneously over the life-cycle and employers are assumed to constantly update their information about individual productivity.

2.1 The Nested Model

A number of assumptions apply not only to the models we analyze in this section, but also to the more general model developed in appendix I. We assume that labor markets are spot markets and that information is symmetric across all employers.[10] This implies that workers are paid their expected productivity in each period. Furthermore, we assume that firms know the structure of the economy and they update their expectations in a Bayesian manner.

More specific to the models analyzed in this paper are the assumptions on the productivity process, the information structure, and the measurement of wages and performance that we will detail now.

Productivity Evolution

A scalar $\widetilde{Q}_{it}$ summarizes worker productivity. Productivity varies with observed

[10] A large literature deviates from the assumptions of spot markets and symmetric information. For example, Gibbons and Katz (1991), Kahn (2009a),Schönberg (2007) and DeVaro and Waldman (2007) provide evidence, in a variety of settings, that employers learn asymmetrically. Further BGH (1994b), Beaudry and DiNardo (1991), Kahn (2010), and Oreopoulos et al. (2006) show that pay is in part dependent on past labor market conditions. We are enormously sympathetic to this literature, especially since one of us has contributed to it. However, it would be intractible to include features of these models in our paper. What is important for us is that despite evidence of the existence of these market imperfections, there is also substantial evidence that market forces constrain firms in setting pay policies. For example, BGH (1994b) find that the firm analyzed here does not fully shelter pay from market fluctuations.

characteristics (x_i) and experience t. Thus, we let $\widetilde{Q}_{it} = Q(x,t) * Q_{i,t}$. Here $Q(x,t) = E\left[\widetilde{Q}_{it}|x,t\right]$ captures systematic variation in productivity over the life-cycle and is necessary to explain the strong regularities in log wages with experience and schooling that characterize all labor market data. $Q_{i,t}$ is the idiosyncratic component of individual productivity.

The difference equation (1) provides a simple of representation of how the individual component of log productivity $q_{it} = \log(Q_{it})$ evolves with experience:[11]

$$q_{it} = q_{it-1} + \kappa_i + \varepsilon^r_{it} \tag{1}$$

We assume $\kappa_i \sim N(0, \sigma^2_\kappa)$ and $\varepsilon^r_{it} \sim N(0, \sigma^2_r)$ and that the ε^r_{it} are uncorrelated over time and with κ_i. We initialize this difference equation in period 0 with a draw of q_{i0}, drawn from a normal distribution $N(0, \sigma^2_q)$. This draw is independent of κ_i.[12]

According to equation (1), log productivity q_{it} evolves with three sources of heterogeneity. The heterogeneity in q_{i0} captures differences in initial ability. The heterogeneity in the drift parameter κ_i models persistent differences in the intensity with which individuals accumulate human capital over the life-cycle.[13] Finally, ε^r_{it} captures innovations in individual productivity that are not predictable. The i.i.d. assumption on the ε^r_{it} implies that the variation in these innovations does not decline with experience and that individual productivity diverges even for relatively experienced workers.

There are various possibilities for why worker productivity might evolve randomly over time. Some workers might experience bad health. Others find some of their skills to become obsolete due to technological change. Another possibility still is

[11] By construction q_{it} is mean zero and uncorrelated with the controls x. From now on, we will suppress the dependence on x. We generally follow the notational convention that upper case and lower case letters refer to variables measured in levels and logs, respectively.

[12] We adopt the convention that period 0 is a period prior to the first period the individual spends in the labor market.

[13] Persistent differences in intensity would arise, for example, if individuals differ in either their preferences or ability to invest (Becker (1964), Ben-Porath (1967)).

that individuals are asked to perform different tasks as they acquire more experience. If productivity on past tasks does not perfectly predict productivity on future tasks, then worker productivity would indeed be subject to unpredictable variation as individuals gain experience (Gibbons and Waldman 2006).

Information Structure

The flow of information to employers is modeled using three different signals. Any information firms have about worker productivity at the beginning of their career is embodied in an initial signal z_{i0}. As individuals spend time in the labor market, firms observe two signals in each time period: $\{p_{it}, z_{it}\}_{t=1}^{T}$. The signals z_{i0} and $\{z_{it}\}_{t=1}^{T}$ are not observed in the data available to researchers. The only signal that is (partially) contained in our data is p_{it}.[14] We assume that all three signals are normally distributed around q_i and therefore have $z_{i0} = q_i + \varepsilon_{i0}$, $p_{it} = q_i + \varepsilon_{it}^{p}$, $z_{it} = q_i + \varepsilon_{it}^{z}$ where $\varepsilon_{i0}\tilde{}N\left(0, \sigma_0^2\right)$, $\varepsilon_{it}^{p} \sim N\left(0, \sigma_p^2\right)$, and $\varepsilon_{it}^{z}\tilde{}N(0, \sigma_z^2)$. The normality assumptions allow us to analyze the learning process using the tools of Kalman filtering and ensure great parsimony for the model. Without loss of generality, we impose that $cov\left(\varepsilon_{it}^{z}, \varepsilon_{it}^{p}\right) = 0$.[15]

Based on the spot market assumption made above, wages will equal expected productivity conditional on all signals firms have observed up to that point.

Measurement Issues

Two measurement issues arise when we try to map the above model onto the particular data we consider. First, and quite standard, we allow for measurement error in wages:

$$W_{i,t} = W_{i,t}^{*}\Omega_{i,t} \tag{2}$$

where W_{it} is the observed wage, W_{it}^{*} is the wage measured without error and Ω_{it}

[14] In the data section, we describe more precisely the information we have on p_{it}.

[15] The information in correlated normal signals is identical to the information contained in orthogonalized signals. The correlations between p_{it} and wages implied by a model with correlated signals and those implied by a model with orthogonal signals are therefore identical.

represents the measurement error. Taking logs we get

$$w_{it} = w_{it}^{*} + \omega_{it} \tag{3}$$

We assume that ω_{it} is classical measurement error with $\omega_{it} \tilde{} N(0, \sigma_{\omega}^{2})$.

The second issue arises from the fact that our observed productivity signals, p_{it}, are subjective managerial performance evaluations (described in more detail below). As we estimated the model, we found that these performance ratings were very highly correlated across short time horizons. We believe this pattern arises from temporary stickiness in performance evaluations and does not reflect true productivity evolution. Such persistence could occur, for example, if workers are temporarily matched with the same manager for several periods who may then give similar ratings. Or, managers may be reluctant to give ratings that deviate too far from past performance, if they anticipate the unpleasantness of dealing with worker complaints or needing to provide extra justification. We model this effect by assuming that the ε_{it}^{p} evolve according to equation (4) :

$$\varepsilon_{it+1}^{p} = \rho \varepsilon_{it}^{p} + u_{it+1} \tag{4}$$

where the initial noise is $\varepsilon_{i1}^{p} = 0$ and $u_{it} \tilde{} N(0, \sigma_{u}^{2})$. The parameter ρ governs the degree of persistence in manager ratings and will be estimated. Other than this, we assume that signals reflect new information, i.e., the signal errors $(\varepsilon_{i0}, \varepsilon_{it}^{z}, u_{it})$ are uncorrelated across time.[16]

[16] A different modeling assumption would be to put the auto-regressive component, ρ, directly into the productivity evolution equation. This would yield some auto-correlation in performance measures. However, this assumption violates several of the observed patterns in our data, which we describe below. Specifically, because p_{it} contains noise terms, ε_{it}^{p}, the AR-1 process in observed performance would exhibit less persistence than the AR-1 process in true productivity. In order to generate the relatively large auto-correlations between p_{it} and p_{it-1}(we show below that these are on the order of 0.6), we would need the signal noise in ε_{it}^{p} to be very small. But, if the ε_{it}^{p} were very precise, then we would necessarily require wages and performance signals to be very highly correlated, contradicting the findings in the data.

Summary

The model described above is governed by only 8 parameters: $\left(\sigma_q^2, \sigma_r^2, \sigma_0^2, \sigma_u^2,\ \sigma_\omega^2, \sigma_\kappa^2, \rho, \sigma_z^2\right)$. Because of this parsimony, it becomes transparent what features of the data drive the parameter estimates. At the same time the model is sufficiently complex to nest the two interpretations of wage dynamics that are the object of our inquiry: employer learning and productivity dynamics. By imposing the appropriate restrictions on these parameters, we can estimate either the pure learning or the pure productivity model. The restriction $\sigma_\kappa^2 = \sigma_r^2 = 0$ eliminates any heterogenous dynamics in productivity and delivers the pure learning model. By contrast, the restriction $\sigma_0^2 = \sigma_z^2 = 0$ implies that the firm is perfectly informed at any stage of the life-cycle and thus delivers the pure productivity model.

2.2 Implications and Identification

We now derive several intuitive implications from the model which illustrate how one can empirically distinguish between the employer learning and productivity models. In appendix II, we provide a more formal discussion of how the parameters in the model can be identified using the second moments of wages and performance ratings.

First, it is worth pointing out that wage data alone does not allow one to reject models with unrestricted productivity processes under full information. It is alwasy possible to rationalize wage data within a full information model by assuming that the productivity process follows the same process governing the wage data. For example, observing that log wages follow a random walk has been taken as evidence of employer learning (Farber and Gibbons 1996). However, this pattern would also be obtained under a full information model if productivity itself evolves as a random walk. Therefore, identifying joint models of learning and productivity dynamics using wage data alone will either require functional form restrictions that one is willing to impose on the productivity process or it will require an additional source of information on productivity. Access to productivity correlates such as performance ratings helps resolve this identification problem.

Especially helpful is the co-variation in pay with performance across experience. To illustrate, we consider what the pure learning model[17] implies for how pay covaries with past performance measures as opposed to future performance measures. To simplify assume, for now, that performance ratings are uncorrelated over time ($\rho = 0$) and that wages are measured without error ($\sigma_\omega^2 = 0$). Then, an individual's wage will be given by the following expression:

$$
\begin{aligned}
w_{it} &= E\left[q_i | I^t\right] = \chi_t + (1 - K_{t-1}) * E\left[q_i | z_{i0}\right] + K_{t-1} \frac{1}{t-1} \sum_{j=1}^{t-1} \phi_{ij} && (5) \\
\text{where} \quad & \phi_{it} = (1 - \phi)\, p_{it} + \phi z_{it} && (6) \\
K_t &= \frac{t\sigma_q^2}{t\sigma_q^2 + \sigma_\phi^2} && (7)
\end{aligned}
$$

This expression contains both a component χ_t that is common across individuals and a component that depends on the signals the firm obtains.[18] In each period, we combine the two signals z_{it} and p_{it} into a single scalar ϕ_{it} that represents a sufficient statistic for the information obtained in period t. The weight ϕ depends on how much variance there is in both signals respectively. [19]

From equations (5)-(7), it is easy to derive the covariances between pay and performance measures across time:

$$
cov(w_{it}, p_{i\tau}) = \left\{ \begin{array}{cc} K_{t-1}(\sigma_q^2 + \frac{1-\phi}{t-1}\sigma_p^2) & \tau < t \\ K_{t-1}\sigma_q^2 & \tau \geq t \end{array} \right\} \qquad (8)
$$

[17] We thus impose that $\sigma_\kappa^2 = \sigma_r^2 = 0$.

[18] The time effects χ_t capture both the common variation in log productivity over time and also how the variance of the prediction error varies with experience. A convenient feature of the normal learning model is that the variance of the prediction error does not depend on the observed signals and is instead common across all individuals with the same level of experience.

[19] The exact expressions for ϕ and σ_ϕ^2, the variance of the scalar signal σ_ϕ^2, are known, but not of particular interest at this point.

Equation (8) encapsulates three of the features implied by the pure learning model that are particular noteworthy.

First, for $\tau > t$, the $cov(w_{it}, p_{i\tau})$ increases with experience t because K_{t-1}, the weight placed on the stream of performance measures, grows. Intuitively, as the firm learns, the wage becomes increasingly correlated with underlying productivity and therefore will also correlate more with any signal of productivity, i.e., future performance ratings. Second, $cov(w_{it}, p_{i\tau})$ is larger for performance measures that occurred before the wage was set $(\tau < t)$, than for performance measures that were not yet observed when the wage was set $(\tau \geq t)$. This is because current pay incorporates the realizations of ε^p from previously observed performance measures, but not from future performance measures. Therefore, under the learning model, the relationship between $cov(w_{it}, p_{i\tau})$ and τ will be a step function with a step at $\tau = t$. The size of the step can be obtained by differencing the two expressions in equation (8) and is equal to $K_{t-1}\frac{1-\phi}{t-1}\sigma_p^2$. This yields the third prediction: the size of the step decreases in t. Intuitively, firms' expectations are based on substantially more productivity ratings when t is large and they therefore put less weight on any given signal p_{it} when setting wages.[20]

Thus, the learning model implies a discontinuity at the present when we compare how pay in any period correlates with past and future performance ratings. For learning models, the distinction between the past and the future is fundamental, because it separates observed and unobserved information, generating the discontinuity in correlations. By contrast, the pure productivity model treats the past and the future symmetrically, since the firm has full knowledge of productivity when setting pay. It therefore could not generate the step function described above. This asymmetry illustrates that having performance and wage data available provides a source

[20] While correlations of wages with future performance rise as workers gain experience, this does not happen for correlations with past performance. In a learning model, though wages increasingly correlate with true productivity, that effect is offset by the fact that firms use any given productivity measure less for older workers since their expectations have become more precise.

of identification that allows distinguishing learning models from productivity models and is not functional form dependent.

3 Data

3.1 General description

This paper analyzes data first used by BGHa and BGHb in their canonical studies of the internal organization of the firm. The data consist of personnel records for all managerial employees of a medium-sized, US-based firm in the service sector from 1969-1988. We have annual pay and performance measures, as well as some demographics and a constructed measure of job level (see BGHa for more detail). The original sample contains 16,133 employees. Of these, we restrict attention to the 9,391 employees with non-missing education who can be observed with a wage or performance measure between the ages of 25 and 54 and at least one more wage or performance measure.[21]

Because we have data on only one firm, we may suffer from several selection problems. We are concerned that attrition from the sample is non-random, since nonrandom turnover could bias our results. In appendix III, we estimate a selection corrected version of our model that corrects for attrition based on observables and find that our results are unchanged when we estimate this version of the model.

Summary statistics are reported in table 1. The majority of managers are white males with at least a college degree. Annual salary averages almost $54,000 in cpi-adjusted 1988 dollars and measures base pay.[22]

[21] Age 25 might be considered slightly old to begin the processes of employer learning and post-school skill accumulation for most education groups. However, our sample consists of workers who have already been promoted to the level of manager. As we have no way of learning about their labor market experiences before they enter this sample, we start at the earliest age which still yields a decent sample size. This is also why we extend the analysis to age 54. From now on, we adopt the convention that age 25 is the first year of experience.

[22] We have information on bonus pay for some years (1981-1988) but do not include it in the

The performance ratings range from 1 to 4, with higher rating reflecting better performance.[23] From table 1, we see the average rating is a little over a 3 and the distribution is top heavy, with more than 75% of workers receiving one of the top two ratings.[24]

Table 1: Summary Statistics

Figure 1 plots log pay and performance residuals by age, with solid and dashed lines, respectively.[25] The solid line shows that the earnings are rising with age, but at a decreasing rate, reflecting typical life-cycle patterns. The dashed line reveals, somewhat surprisingly, that average performance falls with age. This is unexpected if we think part of the explanation for the rising age-earnings profile is that workers are accumulating more skills. Medoff and Abraham (1980, 1981) find similar patterns in their data: wage-experience profiles often deviate substantially from experience profiles observed for subjective performance measures. Gibbons and Waldman (1999) argue that this finding can be explained if employees of the same experience level are rated relative to each other. This interpretation can reconcile the patterns from subjective performance measures with the finding that objective productivity measures typically have similar experience profiles as wages do (Waldman and Avolio 1986). It also explains why studies (see Jacob and Lefgren 2008 and Bommer et al. 1995) that have access to both objective and subjective performance measures find that these

analysis to maintain consistency in our data across years. In these years, 22% of workers receive a bonus and, conditional on receiving a bonus, the amount is on average 12% of base salary. We have separately estimated the model with the bonus and the salary data using the 1981-1988 period only. The results are consistent with those presented here but less precise.

[23] We inverted and recoded the original measures, which ranged from 1 to 5, combining the worst two ratings since almost nobody receives the worst.

[24] This distribution of performance ratings is similar to those found in Medoff and Abraham (1980 and 1981) and Murphy (1991) in their studies of performance ratings across various industries and firms. Gibbs (1995) shows that these performance ratings do contain meaningful information. For example, high performance ratings are correlated with higher raises and bonuses, and increase the probability of promotions.

[25] Both variables are residualized on the following set of variables, all interacted with education group (high school, some college, exactly college, advanced degree): gender, race and year dummies and gender- and race-specific time trends.

performance measures are significantly positively correlated.

Figure 1: Log Wages and Performance by Age

In our analysis, we follow the common practice in the literature to treat the performance measures as relative. That is, we interpret observed performance, $\tilde{p}_{it}$, as arising from a latent signal on individual productivity, p_{it}, according to the mapping in equation (9)

$$\tilde{p}_{it} = \sum_{k=1}^{K-1} \mathbf{1}(p_{it} \geq c_{kt}) \tag{9}$$

A worker is assigned the ranking $\tilde{p}_{it} = k$ if his or her latent productivity signals falls between the two thresholds, c_{k-1t} and c_{kt}. We allow these thresholds to differ across age groups, thus incorporating the assumption that ratings are relative to individuals of the same age.[26] The structure assumed in section 2 yields that the latent signal, p_{it}, is normally distributed. We can therefore estimate correlations of p_{it} with other normally distributed variables (such as log wage residuals and lagged performance) using maximum likelihood methods. Of course, since the observed performance ratings are categorical, we cannot identify the variance of p_{it}.

3.2 Moments for estimation

Our model outlined above generates implications about the second moments of wages and performance across different experience levels. Here we present the empirical analogs which we use to estimate our model. In principle, we could match correlations in wages and performance ratings across all 30 age levels, 25-54. Instead, we simplify the estimation and exposition by constructing a set of 68 moments, that we

[26] Age may not capture the exact reference group for a worker. We could easily include demographics, such as race, gender and education, in forming these groups, though we have not done so here. However, our results are robust to allowing performance to be relative to other workers in one's entry cohort or job level.

think are particularly informative for distinguishing learning and productivity models. These moments are displayed in figures 2a and 2b with 95% bootstrapped confidence intervals.[27][28] The information contained in figures 2a and 2b is also represented with standard errors in table 2.

Figures 2a and 2b: Moments and 95% CI

Table 2

Panel A in figure 2a shows the variance in log wage residuals for six 5-year experience groups[29] ranging from 1-5 to 26-30 years. The variance in pay around the age profile is substantial and increases almost linearly with age. It is only after 25 years of experience that the growth in the variance of pay slows.[30] Understanding this variation and its increase over the life-cycle is the primary task of this paper.

Panels B and C in figure 2a show auto-correlations in performance and pay residuals, respectively, for up to 6 lags and for two experience groups: experience 1-15 shown with solid dots and 16-30 with hollow dots. For both pay and performance,

[27] In constructing these moments, we first residualize all pay and performance measures by the following variables all interacted with education group: gender, race age and year dummies, gender- and race-specific time trends as well as gender and race interacted with a quadratic in age. We then take average correlations and variances across the specified set of experience years weighted by the number of individuals for which we observe that moment.

[28] We have investigated to what extend these patterns are similar if we slice the data by education group and by gender. Regardless how we cut the data, the second moments of wages and the second moments of performance measures are consistently similar to those reported for the aggregate sample, with some minor deviations. The correlations between pay and performance measures are also consistent with those reported here for most subgroups. The one major exception is when we consider the less educated. Among these, the evidence for an asymmetry due to pay and performance is less pronounced especially for younger workers. Given the evidence in Arcidiacono et al. (2010) on differential learning by education, we find this deviation from the observed patterns for less educated workers of interest and hope it will attract further research. We are happy to provide Figure 2 seperately by gender and education upon request.

[29] We measure experience as potential experience: schooling minus age - 6.

[30] It is worth noting that these variances are quite a bit lower than one would observe in a cross-section (for example, the variance in log earnings residuals is 0.04 in the first experience bucket). This is because we are already restricting attention to workers in the same firm and occupation.

the more experienced group exhibits higher auto-correlations which fall the further away in time the observation was. In panel B, the performance auto-correlations are more highly correlated at short horizons. As discussed above, we fit this stickiness in performance ratings by allowing for an autoregressive component in the signal noise.

Panel D in figure 2a shows correlations in pay changes for up to 9 lags and for the same two experience groups. As has been observed in MaCurdy (1982), Baker (1997) and many other papers that investigate the 2nd moment properties of log wages, the autocorrelation in wage growth identifies permanent heterogeneity in productivity growth (when $w_{it} = q_{it}$, as in the pure productivity model). In contrast, a pure learning model could not yield this implication because each wage innovation reflects new information obtained by the firm in that period.[31] Here we clearly have evidence consistent with productivity evolution since all correlations in pay changes are sizeable and statistically distinguishable from zero.[32]

In Panel D, we also see that the wage growth correlations decline sharply over the first few periods and then stabilize after the 3rd lag and remain fairly constant through the 9th lag. We believe this decline may be evidence for stickiness in wage growth. Given our spot market assumption and the current structure of our productivity process we cannot fit this decline and we will only fit the 4th through 9th lag in our estimation.[33]

Lastly, we focus on figure 2b, which gives correlations of current pay with past, current and future performance measures for up to 6 lags and leads. These correlations are again shown for the two experience groups. We pay particular attention to these moments throughout the paper because we believe they represent the major innovation to the previous literature . In section 2, we argued that these correlations

[31] Farber and Gibbons (1996) propose testing the pure learning model using exactly this absence of autocorrelation in wage growth.

[32] BGHb also obtained this result and took it as evidence of hetergeneous growth in productivity.

[33] We fit up to 9 lags here because we wanted to gain a better sense of the decay process past the first 3 lags. These long run correlations are of particular interest because they cannot be generated by any temporary correlations in wage growth.

are informative about employer learning. In particular, the pure learning model yields three testable implications: correlations of wages with future performance measures rise with experience; correlations of wages with past performance measures decline with experience; and the relationship between $cov(w_{it}, p_{i\tau})$ and τ will be a step function. A corollary of these three implications is that the size of the step should decline with experience.

Figure 2b provides evidence consistent with two of the predictions. Correlations for future performance measures are larger for the higher experience group, suggesting firm expectations approach true worker productivity over time. Also,there is an asymmetry in correlations of wages with past relative to future performance evaluations. As presented in table 3, the differences in the correlation of pay with future and past performance measures is stastically significant, espescially for older workers. For young workers and the first three leads and lags, the correlation of pay with lagged performance are between 0.015-0.04 larger than those with future performance at similar lag/lead length. These differences are statistically signficant at the 5% level for the first two leads and lags and at the 10% for the third. Contrary to the third prediction of the pure learning model, the step size does not appear to fall with experience. For older workers, the correlations of pay with past performance are on average 0.06 larger than those with future performance. For these older workers, the differences between correlations at similar leads and lags are significant at all conventional levels and for all leads and lags.

Table 3

Thus we see reduced form evidence consistent with both heterogenous productivity growth and employer learning. However, firms continue to exhibit patterns of learning even for workers at high experience levels, suggesting that the pure learning model alone will not be able to fit the data.

4 Estimation

In Section 2 we developed a model of learning and productivity that represents a special case of the more general model described in Appendix I. In Appendix I, we also show how one can use linear state space methods to derive the moments of these more general models. Applying these methods to our specific case, we obtain the implied second moment matrices for wages and performance ratings. These second moment matrices allow us to estimate the parameters of our model using a method of moments estimator.

Table 4 displays our parameter estimates for the three models which we obtain via method of moments with equal weights on all moments. Standard errors, obtained by bootstrapping with 500 repetitions, are shown in parentheses.[34] Figure 3 summarizes the fit of the model for all 3 models.

Table 4: Parameter Estimates.

Figure 3: Correlations of Pay and Performance

We now discuss the fit of each model. As we have mentioned, we pay particular attention to how pay and performance are correlated at various lags and leads.

The Pure Learning Model

Panel B of figure 3 and figure 4 summarize the results of the pure learning model, contrasting the empirical moments with the predictions based on the estimated parameters for the pure learning model (restricting $\sigma_\kappa^2 = 0$ and $\sigma_r^2 = 0$). The predicted moments are shown using solid lines for younger workers and dashed lines for older

[34]The exact bootstrapping procedure is as follows. We draw the sample randomly, with replacement and generate the bootstrapped moments. We then estimate the parameters to match these moments, taking as starting values the true parameters values shown in table 2. We do not search across starting values to find the global minimum for each of the 500 samples. However, in each bootstrap, we go through four optimization routines (alternating between Newton-Rapson and the simplex method), which should ensure we have found the global minimum.

workers.

We find that the learning model does succeed in a number of ways. Using a small set of parameters, it matches the variance of wages across experience levels. It also matches the approximate levels of the auto-correlations in wages by experience, though not the decline across lags. It matches the decay across lags in the auto-correlations of performance measures, thanks to the parameter ρ, but not the differences across experience. The model, by construction, predicts that wages follow a random walk and therefore the learning model is not able to match any of the long-run positive correlations in pay growth that we observe in the data and report in panel D of figure 4.

Figure 4: Results for the pure learning model.

However, as is evident in Figure 3, panel B, the pure learning model does not fit the correlations between pay and performance ratings that we believe to be the most important new empirical evidence we add to the literature. The data show that the correlations of pay and performance are generally increasing with experience, resulting in a sizeable asymmetry between correlations of wages with past and future performance measures even at high experience levels. In contrast, the fitted learning model predicts a cross-over pattern. For young workers, firms rely heavily on past performance measures, since current expectations are imprecise. This should result in wages that are more highly correlated with past performance for younger, relative to older, workers. The model predicts the reverse for the correlation of current wage with future performance. Because firm expectations become more precise, wages of older workers should approach true worker productivity and become increasingly correlated with future performance.

This failure reflects general features of pure learning models and, in our view, is not a result of any particular distributional assumptions. Overall, we therefore find significant evidence against the pure learning model.

The Pure Productivity model

Figure 3, panel C and figure 5 show the fit of the pure productivity model.

Figure 5: Results for the pure productivity model

Along a number of dimensions, the pure productivity model does better than the pure learning model. First, because the variance of the heterogeneous growth term κ_i reported in Table 4 is non-zero, the pure productivity model generates long run correlations in wage changes that are positive, though smaller in magnitude than the observed moments. The pure productivity model also fits both the auto-correlations in pay and performance, better than the learning model did. Allowing productivity to vary yields stronger declines in auto-correlations across lags and experience groups that the learning model could not predict. However, this model does poorly in fitting the experience profile of variance of log pay. Growth rate heterogeneity implies that the variance rises in the square of experience, producing the convex pattern fitted by the model.

Turning to our main set of moments (figure 3 panel C), the evidence regarding the pure productivity model is mixed. A success for the model is that it manages to fit the approximate levels of correlations across experience groups. Intuitively, these correlations increase with experience because, as the variance in productivity increases with experience, the common component in performance ratings and wages becomes more important, relative to the noise in the performance ratings.

However, we find that within experience, the pure productivity model predicts that the correlations of current pay is larger for performance measures that are collected later in an individuals career. This is because current pay (which equals current productivity) is correlated with the systematic growth component κ_i and κ_i have a larger on performance further into the future. Therefore the current wage is more highly correlated with wages that are further in time. This results in the upwards slope of the lines in Figure 3, Panel C which represent the predicted moments from the pure productivity model. As is clear from this panel, the empirical moments do not show this upward slope. Instead, the empirical moments show the asymmetry around current pay and they show lower correlations of current pay with future rather than past performance. This asymmetry is clearly not matched by the pure productivity model.

The Nested Model

Finally, we consider how the parameter estimates and fit of the nested model compare with those of the pure learning and productivity models. Results from the nested model are shown in figure 6 and panel D of figure 3.

Figure 6: Results for the combined model

Overall, our estimates emphasize that it is important to account for both learning and productivity growth in explaining the data. The greatest failure of the nested model lies in its inability to fit the concavity in the variance of log wages across experience. It is successful though in fitting the correlations between pay and performance, both the levels across experience and the asymmetry across lags, though it admittedly has trouble fitting the decline after about four leads into the future. In addition, because of imperfect information, productivity innovations are not immediately incorporated into pay. Therefore, the model is also able to fit higher correlations in pay growth, resulting from a larger κ_i.

In fact, the nested model attributes a larger role to persistent differences in productivity growth, κ_i, and less of a role to random innovations in productivity, ε_i^r. An increase in κ of one standard deviation corresponds to 45% extra productivity growth over our time horizon in the nested model, and 35% in the pure productivity model. Over the same time period, a standard deviation of the sum of random walk components is about 12% and 25% for the nested and pure productivity models, respectively.

Turning to the estimates of learning parameters, we find that the variance in all of the signals is much greater for the pure learning model than the nested model. The variance of the initial signal (σ_0^2) and of the dynamic signals (σ_z^2, σ_u^2) is substantially larger for the pure learning than for the nested model. The pure learning model requires more signal noise to match the evidence for learning even at higher experience levels. This additional noise enables the pure learning model to fit the increase in the variance of wages even at high experience levels and it allows it to fit approximate correlations between pay and performance for both young and old workers. The nested model instead allows for much less signal noise. The variance of log wages continues to increase because productivity itself evolves and the evidence of learning even for the old workers results from the fact that firms need to learn about a moving target: learning about κ_i is quite small and furthermore there are always new innovations ε_{it}^r that the firm needs to learn about.

Statistically, we reject the pure learning and pure productivity models in favor of the nested model. We reject the restrictions of the pure productivity model $(\sigma_0^2 = \sigma_z^2 = 0)$ against the unrestricted model at a 97.5% significance level (the χ^2 statistic with two degrees of freedom is 7.51). The restrictions of the pure learning model $(\sigma_\kappa^2 = \sigma_r^2 = 0)$ are rejected at any reasonable significance level with a χ^2 of 487. Overall, we thus find support for a model that combines elements of learning with heterogenous changes in productivity over the life-cycle.

5 Interpretation and Discussion

In this section, we interpret the estimates of the nested model. We begin by discussing what they imply for the overall variation in productivity and wages over the life-cycle and in particular what they imply about the size of the expectation error made by firms over the life-cycle. In particular, we are interested in how far productivity and wages can deviate from each other because firms are imperfectly informed. We then turn to the question of how incentives to engage in productivity enhancing activities are impacted by imperfect labor market learning.

5.1 Productivity and Wage Variance of the Life-Cycle

The estimated parameters of the nested model allow us to derive the variances in productivity, wages, and in the expectation error over the life-cycle. Figure 7 plots these variances as a function of experience.

Figure 7: Variances in productivity, wages and expectation error, by experience

The top line shows the variance in log productivity with the variance of log wages just below. Even at 30 years of experience, the variances of wage and productivity are quite similar (0.174 and 0.154, respectively). Clearly, the shape and magnitude of the variance of log wages derive from the shape and variance of productivity. Thus, to understand why wages diverge between individuals over the life-cycle means first and foremost understanding why productivity evolves heterogeneously.

The difference between the variance in wages wages and productivity is accounted for by the variance of the firm's error in expectations. During the first years in the labor market, this variance declines as firms learn about initial productivity, q_{i0}, and the persistent component of productivity growth, κ_i. Subsequently, the variance stabilizes at a fairly constant level around 0.022, reflecting that firms must continue to learn about the constantly accruing random innovations in productivity.

While it might seem that the variance of the expectation error is small and thus imperfect learning is of small consequence, we would disagree. The implied standard deviation for the expectation error is about 0.15, which means that the average expectation error of the firm is about 10% of annual productivity for most of the life-cycle. Firms make sizeable errors when estimating individual productivity and face substantial incentives to learn about how productive their workers are. The observed size of the expectation error and the fact that these errors persist late into individual careers make it plausible that worker turnover and human resource policies are substantially shaped by employer learning.

5.2 Incomplete Learning and the Returns to Investment

The estimated model of productivity and learning allows us to answer a simple, yet fundamental question: if individual productivity at experience t increases by 1%, then what fraction of the present discounted value of this increase accrues to the individual? If this fraction is less than one, then the incentives to privately invest in human capital fall short of the full social returns. In this case, investments that are difficult to observe on the part of employers - such as health investments or efforts to keep up with technological change and/or prevent depreciation of existing skills will be below socially optimal levels.

Following a change in productivity there will first be a period during which this change is only partially priced into wages. Eventually, after employers learn, wages will fully reflect individual productivity and only then will individuals fully benefit from any changes in their skills. However, as workers age, the period over which individuals' wages fully incorporate the productivity change shortens, resulting in a smaller fraction of any productivity change accrueing to older individuals. The size of the share of the return to human capital investments going to workers and how rapidly it declines depends on how fast firms learn as well as the discount rate individuals face.

In order to estimate the share of a productivity increase that accrues to individuals, we use our parameter estimates for the nested model as, a range of discount rates (3 to 10%) and assume individuals work for 40 years. For a one unit permanent increase in labor productivity, we ask how much the present discounted value of earnings changes relative to the present discounted value of productivity. Table 5 reports these estimates for workers experiencing the productivity shock at different points along the life-cycle. These estimates, while admittedly rough, provide an indication of how important learning and incomplete information can be for understanding investment patterns throughout a career.

Table 5: The Wedge Between Social and Private Returns to Productivity Investments.

Regardless of the discount rate considered, we find that the share of a productivity increase going to workers is greatest if the increase occured prior to entering the labor market. This is because firms receive fairly precise signals about initial productivity differences (σ_0^2 is relative small). During the first 15 years of individuals' careers, between 60 and 80% of the social returns to productivity changes are captured by individuals, depending on the discount rate. However, as individuals approach the half-way mark of their careers their share of the return declines fairly rapidly. If we consider a discount rate of 5%, then we observe that during the first 10 years about 75% of the returns are captured by individual workers. This percentage declines to about 65% after 20 years of experience, 40% after 30 years of experience and only about 25% after 35 years of experience.

These estimates therefore suggest that incomplete learning on the part of employers can generate gaps between the private and the social returns of human capital investments that are relatively small for young workers. In that sense, we reach a similar conclusion to Lange (2007).[35] Lange finds that initial expectation errors about productivity differences existing at the beginning of individual careers decline by about 50% in the first 3 years and that only 25% remains after 8 years. The parameter estimates obtained in this paper imply that expectation errors about productivity differences existing at the beginning of individual careers decline by about one third within 3 years and about 70% within 8 years. Thus, our estimates regarding the speed of learning about initial productivity differences are strikingly consistent with those of Lange, despite the differences in methodology.[36] Similar to Lange, we therefore conclude that signaling about existing productivity differences is not likely to be the main motivation for obtaining additional schooling degrees.

However, in contrast to the static model in Lange (2007), our estimates here suggest that the importance of imperfect information increases with age and that incentives might be most severely misaligned during old age. Older individuals are likely to refrain from efficient human capital incentives, because they can not count on wages to accurately reflect the productivity returns of their investments. Our estimates suggest that the focus of models of incomplete information and employer learning should not be place exclusively on young workers, but rather that employer learning models also have important implications for behavior of older workers. As evident from Table 5, incomplete learning generates the largest gaps between the

[35] Lange (2007) builds on the empirical strategy proposed first by Farber and Gibbons (1996) and developed by Altonji and Pierret (2001), using data on the AFQT from the NLSY 1979, to estimate how quickly firms learn about heterogeneity in worker productivity. He argues that this speed of employer learning is crucial for understanding how relevant signaling motives are in schooling decisions, because if firms learn rapidly about worker productivity, then workers have little reason to signal their productivity by taking costly actions such as acquiring schooling.

[36] Firms learn about 2 productivity states, κ_i and q_{it}. This imparts some complicated dynamics into the speed of learning, which does not allow us to summarize the speed of learning in a single parameter, as in Lange (2007). The dynamics in fact generate overshooting, such that initial productivity differences in q_{0i} will have a more than one-for-one impact on log wages for part of the individuals life-cycle.

private and social returns to investing into human capital late in individuals careers.

6 Alternative Theories linking Pay and Performance

We have concluded above that productivity evolves heterogeneously throughout the life-cycle and firms continue to learn about this moving target. The main piece of evidence for this is the fact that past performance correlates more highly with pay than does future performance, even at high experience levels. Here we discuss briefly whether varying precision of the measures, direct pay for performance, or tournament models might generate the same patterns in the data.

It is possible that performance evaluations become more precise as workers age and the firm learns how to evaluate them. This would explain why firms still update on worker productivity, for a time, at high experience levels. However, it would not explain why firms continue to update at all points along the life-cycle – recall, the difference between the correlation of pay with lags of performance, compared to leads of performance is always positive and statistically significant for the older group. Even if the variance in the productivity signal falls with age, firms should eventually stop updating, and we do not see that.

Alternatively, one might be worried about direct pay for performance. If firms incentivize individuals by linking their wages to current performance, then we should observe performance measures to correlate highly with current pay. We clearly do not observe this pattern in the correlations between pay and performance presented in Figure 2b. We see instead that all past performance measures have roughly the same correlation with current pay (around 0.28 for young workers and almost 0.40 for old workers). This is inconsistent with a direct pay for performance scheme.

If firms directly incorporate past performance into pay, we would see larger correlations for lagged performance and wages, relative to those of future performance.

Further, as workers age and are promoted, the scopes of their jobs might broaden and firms might want to strengthen the incentive. If this were the case, we should see a spike at one lag of performance (or possibly the past few performance measures). We might also see a larger spike for the high experience group. However, we do not see these patterns.[37]

However, consider a deferred form of incentive pay where firms operate tournaments to determine promotions and pay raises (a la Lazear and Rosen 1981). Such tournaments can lead current pay to correlate more highly with past performance measures (those being used to determine tournament winners) rather than future performance measures. Such a model can therefore generate asymmetries in the correlations between pay and performance of the type we observe in our data, even if firms know everything about worker characteristics. To rule out that such deferred incentive pay generates the observed patterns, we would need more information on the structure of pay setting and promotions. Lacking such information in this dataset, we are forced to simply note this identification problem with the hope that in the future, better and more comprehensive human resource data will permit progress in distinguishing alternative explanations from the productivity and learning based model analyzed in this paper

7 Conclusion

In this paper, we provide new evidence on employer learning and productivity evolution by exploiting performance evaluations, along with pay data, from a panel of workers in a single firm. We derive a nested model and show how we can uncover both the learning and productivity parameters by matching moments in the data. We find

[37] It is worth pointing out that had we incorporated bonuses into our pay day, this might be different. We have not done so because we cannot get consistent bonus measures throughout the sample. However, it also means that our current measure of pay probably does not include direct incentives.

that problems of accurately predicting productivity are important for employers and that average expectation errors are large at all stages of individuals careers. However, the learning process is not the primary driver of wage dynamics. Instead, our model suggests that heterogeneous variation in productivity drives most of the observed increase in the variance of wages over the life-cycle. We believe these findings represent a significant reinterpretation of the employer learning literature.

An important caveat to our conclusion is that we are only able to study one firm and further, only one occupation (broadly defined). Our finding that firms have quite precise expectations over worker ability at the beginning of the worker's career could be explained by the fact that these workers have *already* been promoted to manager. Thus the market probably had opportunities to learn about these workers before they entered our sample. In the future, we hope to analyze other data sets containing pay and performance measures to establish the generalizability of these findings.

Seemingly contradictory to most models of human capital accumulation (Becker 1964, Ben-Porath 1967), we find that a significant component of productivity evolves unpredictably throughout the life cycle. One explanation for this finding is that workers are assigned to different tasks throughout the life cycle and performance on past tasks does not predict performance on future tasks. This interpretation suggests that firms shift workers into job levels and tasks with little ability to predict worker success there.

We believe that this paper contributes to the literature on employer learning in two ways, methodologically and substantively. First, we provide and implement an approach for estimating models of employer learning and dynamic productivity that can be implemented when data contain multiple signals of worker productivity at various points along the life-cycle. We hope that this approach will prove useful for analyzing the growing set of firm level data-sets comprising personnel records that are appearing in the literature. Second, we show that employer learning continues throughout the life-cycle and we provide evidence against the implication of the existing models on employer learning (Farber and Gibbons 1996; Altonji and Pierret 2001; Lange, 2007) that incomplete information and employer learning are most important early in the life-cycle. To the contrary, in our context, incomplete learning will generate the largest distortions in individual behaviors late in their careers.

References

[1] Abowd, J. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2): 411-455.

[2] Altonji, J.G. and C.R. Pierret (2001): "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 113:79-119.

[3] Arcidiacono, P., P. Bayer, and A. Hizmo (2010) "Beyond Signaling and Human Capital: Education and the Revelation of Ability" forthcoming, *AEJ: Applied Microeconomics.*

[4] Baker, G., M. Gibbs and B. Holmstrom (1994a): "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, CIX: 921-955.

[5] Baker, G., M. Gibbs and B. Holmstrom (1994b): "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, CIX: 881-919.

[6] Baker, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15: 338-375.

[7] Beaudry, P. and J. DiNardo (1991): "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Microdata," *Journal of Political Economy*, XCIX: 665-688.

[8] Becker, G. (1964): *Human Capital*, New York, NY: NBER.

[9] Ben-Porath, Y.(1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, LXXV: 352-365.

[10] Bommer, W., J. Johnson, G. Rich, P. Podsakoff and S. Mackenzie (1995): "On the Interchangeability of Objective and Subjective Measures of Employee Performance: A Meta-Analysis," *Personnel Psychology*, Vol 48: 587-605.

[11] Chang, C. and Y. Wang (196): "Human Capital Investment Under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14, pp. 555-570.

[12] DeVaro, J. and M. Waldman (2007): "The Signaling Role of Promotions: Further Theory and Empirical Evidence," Cornell University, mimeo.

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

[14] Gibbons, R. and L. Katz (1991): "Layoffs and Lemons," *Journal of Labor Economics*, 9:351-380.

[15] Gibbons, R. and M. Waldman (1999): "A Theory of Wage and Promotion Dynamics inside Firms," *Quarterly Journal of Economics*, 114:1321-1358.

[16] Gibbons, R. and M. Waldman (2006): "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," *Journal of Labor Economics*, 24: 59-107.

[17] Gibbs, M. (1995): "Incentive compensation in a coporate hierarchy," *Journal of Accounting and Economics,* 19: 247-277.

[18] Gibbs, M. and W. Hendricks (2004): "Do Formal Salary Systems Really Matter?" *Industrial and Labor Relations Review*, October.

[19] Hamilton, T (1994): *Time Series Analysis*. Princeton University Press: Princeton, NJ.

[20] Hause, J. (1980): "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica*, 48: 1013-29.

[21] Jacob, B. and L. Lefgren (2008): "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics*, Vol. 26(1): 101-136.

[22] Kahn, L. (2009): "Asymmetric Information between Employers," Yale University, mimeo.

[23] Kahn, L. (2010): "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy," *Labour Economics*, Vol. 17(2): 303-316.

[24] Katz, E. and A. Ziderman (1990): "Investing in General Training: The Role of Information and Labour Mobility," *Economic Journal*, 100, pp. 1147-1158.

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

[26] Lazear, E. and S. Rosen (1981): "Rank-order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5): 841-864.

[27] MaCurdy, T.(1982) "The Use of Time Series Processes to Model the Error Structure of Earnings in Longitudinal Data Analysis," *Journal of Econometrics*, 18: 83-114.

[28] Medhoff, J. and K. Abraham (1980): "Experience, Performance and Earnings," *Quarterly Journal of Economics*, 95: 703-736.

[29] Medhoff, J. and K. Abraham (1981): "Are those paid more really more productive? The case of experience," *Journal of Human Resources*, 16: 186-216.

[30] Murphy, K. J. (1991): "Merck & Co., Inc. (A), (B), & (C)," Harvard Business School Press, Boston, MA.

[31] Oreopoulos, P., T. von Wachter and A. Heisz (2009),"The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates," Columbia University, mimeo.

[32] Schönberg, U. (2007): "Testing for Asymmetric Employer Learning" *Journal of Labor Economics,* 25, pp. 651-692.

[33] Waldman, M. (1984): "Job Assignment, Signalling and Efficiency," *RAND Journal of Economics*, 15: 255-267.

[34] Waldman, M. (1990): "Up-or-Out Contracts: A Signaling Perspective," *Journal of Labor Economics*, 8, pp. 230-250.

[35] Waldman, D. and B. Avolio, "A Meta-Analysis of Age Differences in Job Performance", Journal of Applied Psychology, LXXI: 33-38.

I A More General Class of Models

In section 2, we have presented a model with particular productivity and learning structures. In this section, we show a more general class of models of learning about worker productivity, drawing from Hamilton (1994). We will show how to derive the second moment matrices of productivity signals and wages in this larger class of models. To estimate the parameters of these models, one naturally will fit the predicted and the observed second moment matrices of productivity signals and wages.

I.1 The Productivity Process

In period 0 (before production starts), individuals are endowed with a $(n_q x1)-$vector of productivity parameters $\theta_{i,0}$ with $E\left[\theta_{i,0}\right]=0$ and $E\left[\theta_{i,0}\theta_{i,0}^{\prime}\right]=P_0$. In subsequent periods, productivity evolves according to a stochastic process represented by the stochastic difference equation:

$$\begin{aligned} \theta_{i,t+1} &= \Phi\theta_{i,t}+\varepsilon_{i,t+1}^{\theta} \\ &\varepsilon_{i,t+1}^{\theta}\,\tilde{}\,N(0,R_\theta) \end{aligned} \tag{10}$$

This implies that the productivity states in period 1, the first period of actual production are $\theta_{i,1}=\Phi\theta_{i,0}+\varepsilon_{i,1}^{\theta}$.

I.2 Prediction in the Initial Period

Before any production takes place, firms draw a signal about θ_{i0}. This signal is summarized by an initial $(n_z x1)$ vector of signals $z_{i,0}$. This vector is not observed in the data, but represents the information available to firms at the beginning of an individual's career.

$$\begin{aligned} z_{i,0} &= H_0^{\prime}\theta_{i,0}+\varepsilon_{i,0}^{z} \\ &\varepsilon_{i,0}^{z}\,\tilde{}\,N(0,R_{z,0}) \end{aligned} \tag{11}$$

The dimensions of $\left(H_0,\varepsilon_{i,0}^{z},R_{z,0},P_0\right)$ are implicitly defined to conform to $z_{i,0}$ and $\theta_{i,0}$.

Based on the signal vector $z_{i,0}$ firms predict the state $\theta_{i,0}$:

$$\begin{aligned} \widehat{\theta}_{i,0|0} &= P_0H_0\left(H_0{}^{\prime}P_0H_0+R_{z,0}\right)^{-1}z_{i,0} \\ &= K_z z_{i,0} \end{aligned} \tag{12}$$

Firms set wages based on this predicted state $\widehat{\theta}_{i,0|0}$ taking into account that productivity will evolve between the pre-period and period 1 according to equation (10). Firms best guess about productivity in period 1 is:

$$\begin{aligned}\widehat{\theta}_{i1|0} &= \Phi\widehat{\theta}_{i0|0} \\ &= \Phi K_z z_{i,0}\end{aligned}$$

and the posterior variance of the expectation error is:

$$P_{1|0} = \Phi\left(P_0 - K_z H_0' P_0\right)\Phi' + R_\theta$$

I.3 The Recursion

At the end of each period $t > 0$, a new $(n_x x1)-$signal vector x_{it} is drawn by the firm.

$$\begin{aligned}x_{i,t} &= H_x'\theta_{i,t} + \varepsilon_{i,t}^x \\ &\varepsilon_{i,t}^x \tilde{\ } N(0, R_x)\end{aligned} \tag{13}$$

Based on this signal, the expected posterior of θ_{it} conditional on x_{it} is:

$$\begin{aligned}\widehat{\theta}_{it|t} &= \widehat{\theta}_{i,t|t-1} + P_{t|t-1}H_x\left(H_x{}'P_{t|t-1}H_x + R_x\right)^{-1}\left(x_{i,t} - H_x'\widehat{\theta}_{it|t-1}\right) \\ &= \widehat{\theta}_{it|t-1} + K_t\left(x_{it} - H_x'\widehat{\theta}_{it|t-1}\right) \\ &= \left(1 - K_t H_x'\right)\widehat{\theta}_{it|t-1} + K_t x_{it}\end{aligned} \tag{14}$$

Again, when firms form expectation they account for the evolution in productivity described in equation (10). Therefore firms best guess about productivity in period

$t+1$ is:

$$\begin{aligned}
\widehat{\theta}_{it+1|t} &= \Phi\widehat{\theta}_{it|t} && (15)\\
&= \Phi\left(1-K_t H_x'\right)\widehat{\theta}_{it|t-1} + \Phi K_t x_{it}
\end{aligned}$$

The variance of the expectation error then evolves according to

$$P_{t+1|t} = \Phi\left(P_{t|t-1} - K_t H_x' P_{t|t-1}\right)\Phi' + R_\theta \tag{16}$$

This defines the complete prediction problem of the firm. The parameters are $(P_0, R_{z,0}, R_x, R_\theta, H_x, H_0, \Phi)$.

I.4 Wages

So far, we have described how the vector of individual productivity states θ_{it} and the expectation of this state evolves over time. One component of the individual productivity state is q_{it}, the idiosyncratic component of log productivity. We now show how log wages are related to log productivity. Because we assume that labor markets are frictionless spot markets and all information is common, we have that wages W_{it}^* equal expected productivity: $W_{it}^* = E\left[Q(x,t)\,Q_{it}|I^t\right] = E\left[Q(x,t)\exp\left(q_{it}\right)|I^t\right].$ Here $Q(x,t)$ is a productivity profile common to all individuals and Q_{it} represents individual productivity and I^t represents the information set available at time t. We assume also that wages are measured with multiplicative measurement error Ω_{it}.

We have made a number of normality assumptions. One advantage of these assumptions is that expected log productivity $\widehat{q}_{it}$ is normally distributed in each period.

We can therefore write:

$$\begin{aligned} W_{it} &= Q(x,t)\, E\left[Q_{i,t}|I_{it}\right]\Omega_{it} \\ &= Q(x,t)\, E\left[\exp\left(q_{i,t}\right)|I_{it}\right]\Omega_{it} = Q(x,t)\exp\left(\widehat{q}_{it}+\frac{1}{2}v(t)\right)\Omega_{it} \end{aligned}$$

where $v(t)$ is the variance of the expectation of log productivity. Taking logs, we obtain

$$\begin{aligned} w_{it} &= \left(q(x,t)+\frac{1}{2}v(t)\right)+\widehat{q}_{it}+\omega_{it} \\ &= h(x,t)+\widehat{q}_{it}+\omega_{it} \end{aligned} \tag{17}$$

where ω_{it} is the noise in the measurement error with variance σ_ω^2. We assume that ω_{it} is uncorrelated with all other variables in the model.

We residualize wages to remove the common age profile $h(x,t)$ and denote the residual as r_{it}.

I.5 Link to Observable Data: A State-Space Specification

The next task is to derive the second moments that the model implies for observable quantities (r_{it}, p_{it}). We note that our problem takes the form of a linear state-space specifications. The states that describe individuals are the individual productivity states θ_{it} as well as the expectations firms hold $\widehat{\theta}_{it}$. We stack these two vectors and denote the state vector by $\xi_{it}=\begin{pmatrix}\widehat{\theta}_{it} & \theta_{it}\end{pmatrix}'$. The states evolve in a linear stochastic way and the observed data is linearly related to the states. We denote the observed data as $y_{it}=\begin{pmatrix}r_{it} & p_{it}\end{pmatrix}'$.

The linear state space model consists of three parts. First, we need to specify how the state evolves. This is done in equation (18). Second, we need to specify how the states map into observed variables. This measurement equation is given by

(19). Finally, we need to specify the distribution of the initial state ξ_{i1}, the forcing variables v_{it}, and the unobservable noise in the measurement equation e_{it}.

$$\xi_{it+1} = F_t \xi_{it} + v_{it+1} \tag{18}$$

$$y_{it} = M\xi_{it} + e_{it} \tag{19}$$

$$\xi_{i1} = \begin{pmatrix} \Phi K_z z_{i,0} \\ \theta_{i1} \end{pmatrix}$$

The matrix M has as many rows as there are observable objects. The vector e_{it} contains the noise in the measurement equations. The matrix F_t is given by

$$F_t = \begin{pmatrix} \Phi \left(1 - K_t H_x'\right) & \Phi K_t H_x{}' \\ 0 & \Phi \end{pmatrix}$$

and the innovation v_{it+1} to the state vector is defined as:

$$v_{it+1} = \begin{pmatrix} \Phi K_t \varepsilon_{it}^x \\ \varepsilon_{it}^\theta \end{pmatrix}$$

The (K_z, K_t) $-$matrices were implicitly defined in equations (12) and (14) above.

I.6 The 2nd Moment Matrix of Observables

We can now derive the variance-covariance matrix for the observables y_{it} and $y_{i\tau}$. Without loss of generality, we can limit ourselves to $\tau \geq t$.

Because e_{it} contains only measurement error, we can write the second moment matrices of the observables as follows:

$$E\left[y_{it} y_{i\tau \geq t}^{'}\right] = ME\left[\xi_{it}\xi_{i\tau}'\right]M' + E\left[e_{it}e_{i\tau}'\right] \tag{20}$$

The M are deterministic and we therefore just have 2 components $E\left[\xi_{it}\xi_{i\tau}'\right]$, and $E\left[e_{it}e_{i\tau}'\right]$ that need to be determined as functions of the parameters of the model. The matrix $E\left[e_{it}e_{i\tau}'\right]$ is 0 for $\tau \neq t$ and is directly given from the is variance-covariance matrix of measurement error within t. We therefore simply need to determine how $E\left[\xi_{it}\xi_{i\tau}'\right]$ is related to the parameters.

Tedious, but straightforward algebra yields

$$E\left[\xi_{it}\xi_{i\tau}'\right] = \sum_{j=2}^{j=t}\left\{\left(\prod_{l=j}^{l=t-1} F_l\right) E\left[v_{i,j}v_{i,j}^{'}\right]\left(\prod_{l=j}^{l=\tau-1} F_l\right)'\right\} + \left(\prod_{l=1}^{l=t-1} F_l\right) E\left[\xi_{i1}\xi_{i1}'\right]\left(\prod_{l=1}^{l=\tau-1} F_l\right)' \tag{21}$$

where

$$E\left[\xi_{i1}\xi_{i1}'\right] = \begin{pmatrix} \Phi K_z\left(H_0'P_0H_0 + R_z\right)K_z'\Phi' & \Phi K_z H_0' P_0 \Phi' \\ \Phi P_0 H_0 K_z' \Phi' & \Phi P_0 \Phi' + R_\theta \end{pmatrix} \tag{22}$$

and

$$E\left[v_{i,j}v_{i,j}^{'}\right] = E\begin{pmatrix} \Phi K_{j-1} R_x K_{j-1}' \Phi' & 0 \\ 0 & R_\theta \end{pmatrix} \tag{23}$$

We have thus shown how to generate $E\left[y_t y_\tau\right]$ as functions of the parameters $(P_0, R_{z,0}, R_x, R_\theta, H_x, H_0, \Phi)$ and the measurement matrix for any dynamic specification of productivity that follows equation (10) and any normal learning model that follows equations (11) and (13).

I.7 The Nested Model as a Member of the General Linear State Space Models

In this Appendix, we have described how the second moment of observable variables is linked to the parameters of a general linear learning model. The nested model encountered in Section 2 is a special case of such a linear learning model. We now show in the remainder of this appendix what the nested model implies for the parameter matrices of the learning model: $(P_0, R_{z,0}, R_x, R_\theta, H_x, H_0, \Phi)$ and M. This will allow us to implement equation (20) together with equations $(21), (22)$, and (23) to generate the covariance matrices of the wage residuals and performance ratings.

Define first the individual productivity states as $\xi_{it} = (\widehat{\theta}_{it}, \theta_{it})'$ where:

$$\theta_{it} = \begin{pmatrix} q_{it} \\ \kappa_i \\ \varepsilon_{it}^p \end{pmatrix}$$

Note here that we let the individual chumminess term ε_{it}^p enter as an individual state.

The individual state evolves as

$$\begin{aligned} \theta_{it+1} &= \begin{pmatrix} q_{it+1} \\ \kappa_i \\ \varepsilon_{it+1}^p \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} q_{it} \\ \kappa_i \\ \varepsilon_{it}^p \end{pmatrix} + \begin{pmatrix} \varepsilon_{it+1}^r \\ 0 \\ u_{it+1} \end{pmatrix} \\ &= \Phi\theta_{it} + \varepsilon_{it}^\theta \end{aligned}$$

The vector v_{it+1} is therefore given by $v_{it+1} = \begin{pmatrix} \Phi K_t \varepsilon_{it}^x \\ \varepsilon_{it}^\theta \end{pmatrix}$.

Now, the measurement equation is $y_{it} = M\xi_{it} + e_{it}$. Thus, we need to define M and e_{it}. We assume that there is measurement error in r_{it} but that p_{it} is observed without error in our data. Thus:

$$e_{it} = \begin{pmatrix} \omega_{it} \\ 0 \end{pmatrix}$$

The measurement error variance is σ_ω^2 and thus $E[e_{it}e_{it}'] = \begin{pmatrix} \sigma_\omega^2 & 0 \\ 0 & 0 \end{pmatrix}$.

Next,

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Then

$$\begin{aligned}
P_0 &= \begin{pmatrix} \sigma_q^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
H_0 &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\
H_x &= \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \\
R_{z,0} &= \sigma_0^2 \\
R_x &= \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & 0 \end{pmatrix} \\
\Phi &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{pmatrix} \\
R_\theta &= \begin{pmatrix} \sigma_r^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}
\end{aligned}$$

This specialization of the general linear state space model represents the nested model we estimate in this paper.

II Identification

We now consider the identification of the pure learning and productivity model using second moments of wages and performance signals.[38] To simplify the discussion, we assume the length of individuals' careers is unbounded and that we can therefore observe these moments at arbitrarily high experience levels.

II.1 The Pure Learning Model - Identification

The pure employer-learning model allows only for learning and fixes the idiosyncratic component of worker productivity $q_{it} = q_i$ over the life-cycle. This amounts to assuming that there is no heterogeneity in the drift κ_i nor in the individual innovations ε_{it}^{r} and is achieved by setting $\sigma_{\kappa}^{2} = \sigma_{r}^{2} = 0$. There remain 6 parameters that need to be identified: $\left(\sigma_q^2, \sigma_0^2, \sigma_u^2, \sigma_\omega^2, \rho, \sigma_z^2\right)$.

The pure learning model implies that in the limit wages asymptote towards individual productivity. Therefore, we can identify the variance of productivity (σ_q^2) and the variance of the measurement error (σ_ω^2) using the variance and covariance of wages as experience grows. In particular, we obtain $\left(\sigma_\omega^2, \sigma_q^2\right)$ from $lim_{t\to\infty}\left(v\left(w_t\right)\right) = \sigma_q^2 + \sigma_\omega^2$ and $lim_{t\to\infty}\left(cov\left(w_t, w_{t+1}\right)\right) = \sigma_q^2$.

The auto-correlations of p_{it} with p_{it-k} at different lags k inform us about the parameters (ρ, σ_u^2) that govern the signal noise ε_{it}^{p}. As t grows, the distribution of p_{it} converges to an ergodic distribution which depends only on the parameters ρ and σ_u^2. In particular, we have that $\lim_{t\to\infty} v\left(p_{it}\right)) = \lim_{t\to\infty} v\left(q_{it} + \varepsilon_{it}^{p}\right) = \sigma_q^2 + \rho\frac{\sigma_u^2}{1-\rho^2}$ and that $cov\left(p_{it}, p_{it+k}\right) = cov\left(q_i + \varepsilon_{it}^{p}, q_i + \rho^k \varepsilon_{it}^{p} + \Sigma_{j=1}^{k} u_{it+j}\right) = \sigma_q^2 + \rho^k var\left(\varepsilon_{it}^{p}\right)$. Combining,

[38] As described in the data section of this paper, the performance ratings in our data are ordinal, which implies that we do not observe variances or covariances of performance ratings with other objects. Therefore, we show how auto-correlations in performance ratings and correlations with wages at different experience levels allows us to identify models of learnings and productivity.

we have that

$$\lim_{t\to\infty}\lim_{k\to\infty} cor(p_{it}, p_{it+k}) = \frac{\sigma_q^2}{\sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2}} \tag{24}$$

$$\lim_{t\to\infty} cor(p_{it}, p_{it+1}) = \frac{\sigma_q^2 + \rho\frac{\sigma_u^2}{1-\rho^2}}{\sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2}} \tag{25}$$

Since σ_q^2 is already identified, we get $\frac{\sigma_u^2}{1-\rho^2}$ from equation (24) and ρ from equation (25).

This leaves only two parameters (σ_z^2, σ_0^2) that need to to be identified. σ_0^2 determines how much information the has about workers as they begin their careers. We can identify this parameter using the variance of wages at $t = 0$, since $w_{0i} = E[q_i|z_{i0}]$ and $var\,(w_{0i}) = var\,(E[q_i|z_{i0}])$. Conditional on σ_q^2, this variance declines monotonically in σ_0^2 and we can therefore identify σ_0^2 using the variance of log wages for individuals beginning their careers.

The remaining parameter σ_z^2 governs (together with the already identified σ_u^2 and ρ) how much additional information becomes available in any period. Conditional on $(\sigma_0^2, \sigma_u^2, \rho)$, the variance of $w_{1i} = E[q_i|z_{0i}, p_{1i}, z_{1i}]$ declines monotonically in σ_z^2 (as the signal becomes less informative). Therefore we can identify σ_z^2 using $var\,(w_{1i})$, having already identified the other parameters of the learning model.

II.2 The Pure Productivity Model - Identification

The pure productivity model assumes that firms have full information about worker productivity and that wages equal productivity at all times. This assumption can be imposed by restricting the signal noise for the unobserved signals to 0: $\sigma_0^2 = \sigma_z^2 = 0$. There remain 6 parameters that need to be identified: $\left(\sigma_q^2, \sigma_r^2, \sigma_u^2, \sigma_\omega^2, \sigma_\kappa^2, \rho\right)$.

Because wages at all times equal expected productivity, we can write $\Delta w_{it} = w_{it+1} - w_{it} = \kappa_i + \varepsilon_{it+1}^r + \omega_{it+1} - \omega_{it}$. This implies that $cov\,(\Delta w_{it}, \Delta w_{it+2}) = \sigma_\kappa^2$,

$cov\left(\Delta w_{it}, \Delta w_{it+2}\right) = \sigma_\kappa^2 - \sigma_\omega^2$, and $var\left(\Delta w_{it}\right) = \sigma_\kappa^2 + \sigma_r^2 + 2 * \sigma_\omega^2$. This system is triangular and can easily be solved for the parameters $\left(\sigma_\kappa^2, \sigma_r^2, \sigma_\omega^2\right)$. Furthermore, we can identify σ_q^2 using $var\left(w_{i0}\right) = \sigma_q^2 + \sigma_\omega^2$.

The remaining parameters that need to be identified are the parameters $\left(\rho, \sigma_u^2\right)$ that govern the noise in the performance rating p_{it}. To identify these we rely on the correlations between wages and performance ratings:

$$corr\left(p_{it}, w_{it}\right) = \frac{var\left(q_{it}\right)}{\left(var\left(q_{it}\right) + var\left(\varepsilon_{it}^p\right)\right)^{1/2}\left(var\left(q_{it}\right) + \sigma_\omega^2\right)^{1/2}} \tag{26}$$

Since all the productivity parameters are identified, we can treat $var\left(q_{it}\right)$ and σ_ω^2 as known. Thus, eq (26) solves for the variance of the signal noise $var\left(\varepsilon_{it}^p\right)$ for arbitrary t:

$$\lim_{t\to\infty} var\left(\varepsilon_{it}^p\right) = \frac{\sigma_u^2}{1-\rho^2} \Rightarrow \sigma_u^2 = \left(1-\rho^2\right)\lim_{t\to\infty} var\left(\varepsilon_{it}^p\right) \tag{27}$$

Since we know the $var\left(\varepsilon_{it}^p\right)$ for arbitrary t, we can exploit equation (4) to get

$$\rho^2 = \frac{var\left(\varepsilon_{it+1}^p\right) - \lim_{t\to\infty} var\left(\varepsilon_{it}^p\right)}{var\left(\varepsilon_{it}^p\right) - \lim_{t\to\infty} var\left(\varepsilon_{it}^p\right)} \tag{28}$$

These last two equations therefore deliver the parameters ρ and σ_u^2. We have thus established the identification of both the pure learning and the pure productivity model. We will now turn to the estimation of these models.

III Attrition

To obtain the results reported in the main body of the paper we assumed that the individuals in our data are representative of the population of workers from which the firm draws its white-collar workforce. That is, we assume that wages or productivity of individuals with the same experience level do not depend on tenure at the firm. We can investigate this assumption in reduced form using wages and performance measures. Figure A-1 and Figure A-2 show by how much log salary and the performance of first-year workers of various ages differ from the incumbent workforce of the same age. We observe that wages of new entrants and the incumbent work-force are quite similar, but that performance is somewhat lower among new entrants compared with the existing work-force. Table A-1 illustrates the mechanism that gives rise to this relation. This table reports how the probability of exiting the firm depends on the log salary as well as on the performance ratings of individuals, after controlling for age.[39]

Table A-1: Probability of Exit from the Firm

The table illustrates that the performance ratings, but not the salary are statistically significant predictors of attrition from the sample. In particular, individuals with very low ratings are significantly more likely to attrit from the sample. The linear probability model indicates that being in the lowest decile of the performance distribution raises the attrition probability by about 5 percentage points compared to being in the second decile. Further moves up in the performance distribution have a much smaller effect on attrition from the sample. These point estimates therefore support the notion that leaving the firm is endogenous to performance rankings and that the effect of performance on attrition is particular strong for very low ratings. However, the R^2 of the linear probability model also suggests that the overall impact of turn-over based on performance ratings is small.

To be sure that our estimates reported in the paper are not spurious due to non-

[39]We control for performance ratings using dummies for each decile within the distribution of performance rankings. These deciles are populated even though the performance ratings themselves are only reported on a 5 point scale, because individuals ratings are regression adjusted for race, gender, and education in a flexible manner.

random attrition, we estimate an attrition corrected version of our model. For this purpose, we assume that individuals entering the firm are randomly drawn from the population, but that the separation from the firm is governed by the relationship captured by the Probit regression reported in column 1 of the Table A-1. That is, we assume that the probability of separating is given by

$$\Phi\left(\beta_w wage + \beta_{PD}' PD_i + \beta_a age\right) \tag{29}$$

where $\Phi(.)$ is the standard normal distribution, PD_i denotes a vector of performance deciles and the parameters $(\beta_w, \beta_{PD}', \beta_a)$ are obtained from column 1 of Table A-1.

We estimate the parameters of the learning and productivity model using a simulated method of moments. That is, we simulate a sample consisting of 1,000 workers entering the firm at each experience level for a total of 40,000 workers entering with experience levels 1-40. For a given point in the parameter space of the nested model of learning and productivity (described in section 2), we simulate a history of wages and performance ratings under the assumption that no worker attrits. We then apply the selection rule (29) to this sample and thus obtain a selected sample. Using this selected, simulated sample, we generate the same moments (variance of wages, performance and pay autocorrelation, pay-performance correlations at various leads and lags) that we use in Section 2 to estimate the parameters of the model. We can then estimate the parameters by minimizing the distance between the observed and simulated moments in the same manner as before.[40]

In Table A-2 we report the attrition corrected parameters.

Table A-2: Parameter Estimates from Attrition Corrected Model

[40] Note that we estimate the parameters of the selection rule first. This is possible, because we assume that conditional on performance, wages, and age, attrition is random. We can thus treat the observed wages and performance measures as exogenous in estimating the parameters of the attrition model. Because we estimate the parameters of the attrition rule separately, we do not expect the fit of the model to improve as we correct for attrition.

To facilitate comparison this table also shows the parameter estimates reported for the full model in table 4.[41] The estimated parameters are close. In fact, the fit of the attrition corrected and the uncorrected estimates is almost identical and none of our conclusions on the relative importance of learning or productivity evolution are sensitive to using the attrition corrected or the uncorrected estimates. Overall, we therefore believe that our results are robust to attrition based on observed performance or wages.

[41] The computational burden of implementing the attrition correction is significant. We therefore imposed two restrictions on the parameters to reduce the run-time. Because the main model provides little evidence for measurement error in wages, we restricted the variation of measurement error to 0. We also restricted the auto-regressive parameter in the performance ranking to 0.64. Even after imposeing these restrictions, estimating the attrition corrected parameters on our system requires about 1 week of computing time. We therefore refrained from bootstrapping the attrition corrected standard errors.

Table 1 Summary Statistics

Years		1969-1988
Data Description		Managers of a medium-sized US firm in the service sector
# Employees[1]		9391
# Employee-years		59485
% Male		76.2%
% White		89.4%
Age		39.02 (9.02)
Education		
% HS		16.9%
% Some College		18.8%
% College		36.6%
% Advanced		27.7%
Salary[2]		\$53,881 (25447) [n=54364]
Performance[3]		3.12 (0.72) [n=38933]
Performance Distribution		
	1	0.009
	2	0.177
	3	0.499
	4	0.315

Notes: Parentheses contain standard deviations.

1. Sample includes all employees who have a pay or performance measure between the ages of 25 and 54 and at least one more pay or performance measure, with a non-missing education variable.

2. Salary is annual base pay, adjusted to 1988 dollars.

3. Performance is a categorical variable which we recode to be between 1 and 4, with 4 being the highest performance.

Table 2 The Second Moments of Wages and Experience

Variances in Wages by Experience

	Experience					
	1-5	6-10	11-15	16-20	21-25	26-30
	0.044	0.065	0.083	0.100	0.112	0.114
	(0.001)	(0.002)	(0.002)	(0.003)	(0.006)	(0.007)

Autocorrelation in Wages for lags 1-6

	Lags					
Experience	1	2	3	4	5	6
1-15	0.969	0.935	0.903	0.871	0.840	0.813
	(0.001)	(0.002)	(0.003)	(0.004)	(0.006)	(0.008)
16-30	0.990	0.975	0.958	0.940	0.921	0.903
	(0.000)	(0.001)	(0.002)	(0.003)	(0.004)	(0.005)

Autocorrelations in Performance for lags 1-6

	Lags					
Experience	1	2	3	4	5	6
1-15	0.568	0.413	0.315	0.207	0.155	0.154
	(0.008)	(0.011)	(0.014)	(0.016)	(0.018)	(0.026)
16-30	0.659	0.527	0.420	0.323	0.219	0.205
	(0.009)	(0.013)	(0.016)	(0.019)	(0.021)	(0.027)

Table 2, cont'd The Second Moments of Wages and Experience

Correlation of Performance of t with lags and leads in wages

	Lags						
	-6	-5	-4	-3	-2	-1	
Experience 1-15	0.205	0.232	0.266	0.287	0.290	0.281	
	(0.025)	(0.021)	(0.017)	(0.015)	(0.013)	(0.011)	
	Leads						
	0	1	2	3	4	5	6
	0.249	0.266	0.263	0.265	0.253	0.234	0.232
	(0.010)	(0.011)	(0.012)	(0.014)	(0.016)	(0.018)	(0.019)
	Lags						
	-6	-5	-4	-3	-2	-1	
Experience 16-30	0.371	0.379	0.392	0.395	0.393	0.384	
	(0.019)	(0.016)	(0.015)	(0.014)	(0.013)	(0013)	
	Leads						
	0	1	2	3	4	5	6
	0.361	0.36	0.349	0.329	0.309	0.291	0.269
	(0.013)	(0.013)	(0.015)	(0.017)	(0.019)	(0.022)	(0.024)

Autocorrelations in Wage Growths for lags 4-9

Lags					
4	5	6	7	8	9
0.086	0.07	0.077	0.06	0.06	0.081
(0.013)	(0.016)	(0.015)	(0.018)	(0.019)	(0.020)

The table displays the second moments of wages and performance measures that form the basis of the estimation described in the paper. The same correlations are displayed in figure 2a and 2b. The correlations involving performance measures are polychoric correlations. The correlations involving only wages are pearson correlations.

Table 3 The Asymmetry in Correlations between Pay and Performance

	Experience 1-15					
Lag / Lead	1	2	3	4	5	6
Lag	0.281	0.290	0.287	0.266	0.232	0.205
Lead	0.249	0.266	0.263	0.265	0.253	0.234
Difference	0.032	0.024	0.024	0.001	-0.021	-0.029
	(0.005)	(0.010)	(0.014)	(0.018)	(0.023)	(0.028)
	Experience 16-30					
Lag / Lead	1	2	3	4	5	6
Lag	0.384	0.393	0.395	0.392	0.379	0.371
Lead	0.361	0.36	0.349	0.329	0.309	0.291
Difference	0.023	0.033	0.046	0.063	0.070	0.080
	(0.004)	(0.009)	(0.014)	(0.017)	(0.021)	(0.026)

To illustrate the content of this table consider column 1 for younger workers. This column contains first the correlation of the current wage with the performance measure received in the same year (0.281). This performance measure is the first that was not used in setting the current wage. Below, the column contains the correlation of the current wage with the last performance measure received before the current wage was set (0.249). Finally the table contains the difference of these two correlations and their standard error (0.032 and 0.005). The second column performs the same comparision, but uses the second performance measure received prior and after the current wage was set.

Table 4 Parameter Estimates for 3 Models

	Employer Learning	Productivity	Combined
σ_q^2	0.118 (0.0057)	0.025 (0.0051)	0.037 (0.0072)
σ_r^2	-	0.0040 (0.00032)	0.00049 (0.00040)
σ_0^2	0.383 (0.061)	-	0.114 (0.071)
σ_u^2	0.650 (0.062)	0.405 (0.031)	0.488 (0.051)
σ_ω^2	0.0049 (0.00021)	0.00030 (0.00048)	2.83e-12 (4.95e-12)
σ_κ	-	0.00000027 (0.0000023)	0.00015 (0.000016)
ρ	0.645 (0.0084)	0.634 (0.0084)	0.640 (0.009)
σ_z^2	0.506 (0.131)	-	0.206 (0.075)

Reported are the parameter values for the pure employer learning model, the pure productivity model and combined model. The pure employer learning model and the pure productivity model are estimated imposing zero restrictions on the relevant parameters. Standard errors are obtained by bootstrapping with 500 repetitions.

Table 5 The Share of Returns to Investments Going to Individuals

	Discount Factor R			
Experience	0.9	0.92	0.95	0.97
0	0.67	0.71	0.78	0.84
5	0.60	0.66	0.75	0.82
10	0.61	0.67	0.75	0.81
15	0.60	0.64	0.71	0.76
20	0.56	0.60	0.64	0.68
25	0.49	0.51	0.55	0.57
30	0.39	0.40	0.42	0.43
35	0.25	0.25	0.26	0.26

The table displays the increase in the present discount value of life-time wages as a fraction of the increase in the present discounted value of remaining life-time production associated with a unit increase in worker productivity at experience level t. These ratios are shown for different experience levels and for the specified gross discount factors. The calculations are based on the parameter estimates for the combined model presented in Table 4. We assume that individuals careers last for 40 years.

Appendix Table A-1: Probability of Exit from the Firm

	(1)	(2)
	Probit	Linear Probability Model
Log wage	0.0154	0.0034
	(0.034)	(0.007)
Performance - 2nd Decile	-0.204***	-0.0498***
	(0.035)	(0.008)
Performance - 3rd Decile	-0.231***	-0.0552***
	(0.036)	(0.008)
Performance - 4th Decile	-0.212***	-0.0507***
	(0.036)	(0.008)
Performance - 5th Decile	-0.280***	-0.0650***
	(0.036)	(0.008)
Performance - 6th Decile	-0.350***	-0.0788***
	(0.037)	(0.008)
Performance - 7th Decile	-0.411***	-0.0895***
	(0.037)	(0.008)
Performance - 8th Decile	-0.335***	-0.0756***
	(0.038)	(0.008)
Performance - 9th Decile	-0.453***	-0.0964***
	(0.038)	(0.008)
Age	-0.00611***	-0.0013
	(0.0009)	(0.0002)
Constant	-0.657***	0.236***
	(0.043)	(0.009)
Observations	33,151	33,151
R-squared		0.008

Reported are the estimates results from a Probit and Linear Probability model of separating from the job.

Standard errors in parantheses. *** p<0.01, ** p<0.05, * p<0.1

Table A-2 Parameter Estimates from Attrition Corrected and Baseline Model

	Attrition Corrected	Baseline
σ_q^2	0.0420	0.037 (0.0072)
σ_r^2	3.7E-04	0.00049 (0.00040)
σ_0^2	0.0716	0.114 (0.071)
σ_u^2	0.5025	0.488 (0.051)
σ_ω^2	0 (fixed)	2.83e-12 (4.95e-12)
σ_κ	0.0125	0.00015 (0.000016)
ρ	0.64 (fixed)	0.640 (0.009)
σ_z^2	0.2464	0.206 (0.075)

Reported are the attrition corrected estimates of the nested model and as comparision the uncorrected estimates ("Baseline"). Standard errors for the attrition corrected estimates are not available due to the computational burden of estimating these parameters. Standard errors for the Baseline estimates are obtained from bootstrapping 500 times.

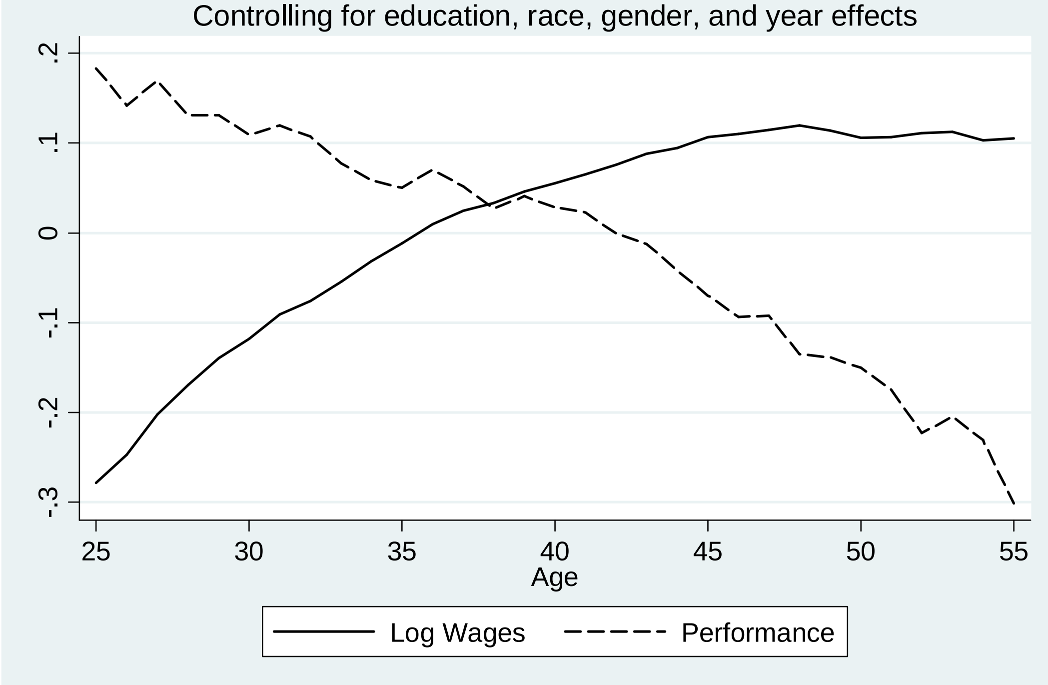

Figure 1: Log Wages and Performance, by Age

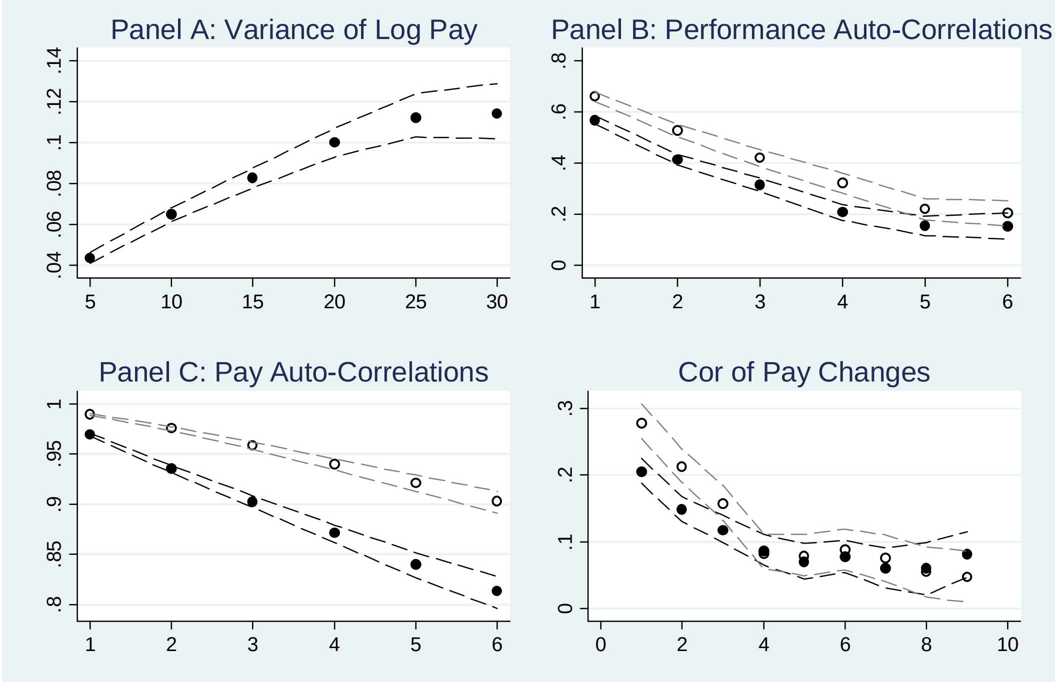

Figure 2a: Moments with 95% CI

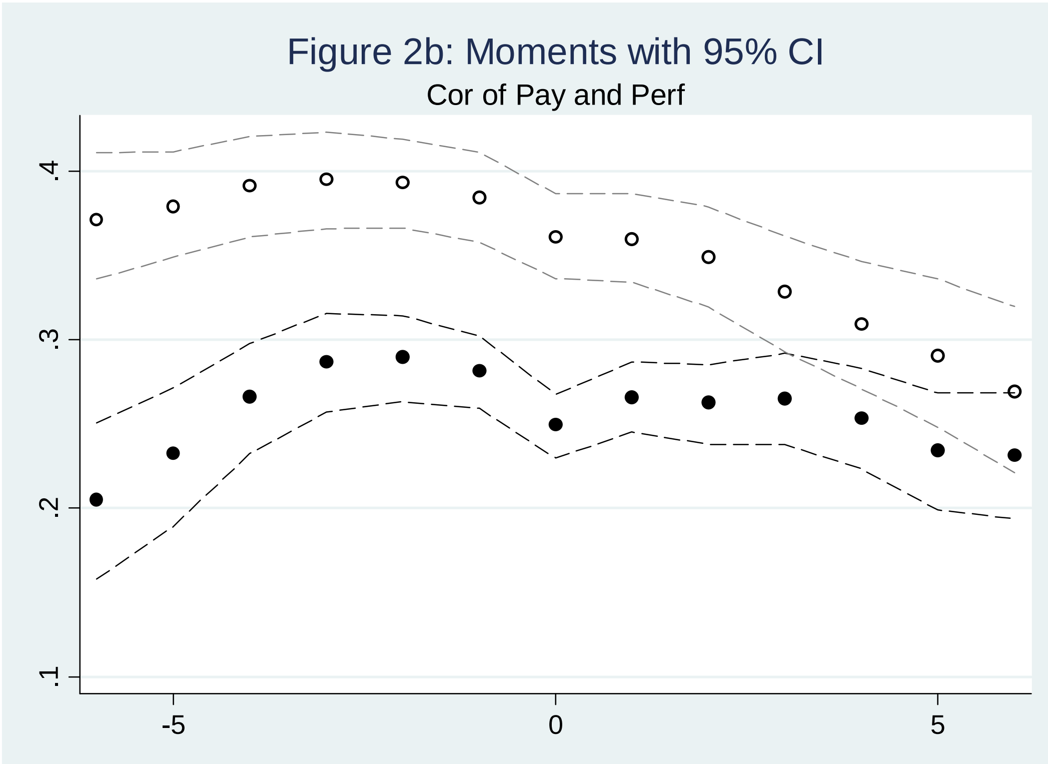
Figure 2b: Moments with 95% CI
Cor of Pay and Perf
.1
.2
.3
.4
-5
0
5

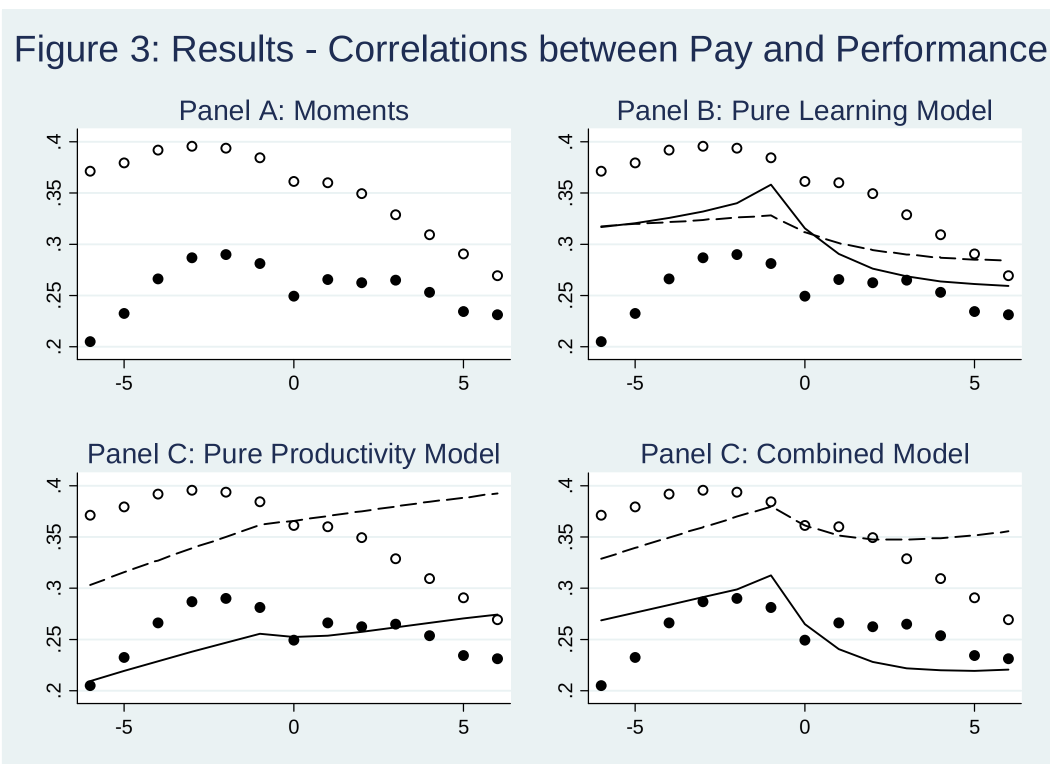
Figure 3: Results - Correlations between Pay and Performance
Panel A: Moments
Panel B: Pure Learning Model
Panel C: Pure Productivity Model
Panel C: Combined Model
.2
.25
.3
.35
.4
-5
0
5

Figure 4: Results - Pure Learning Model

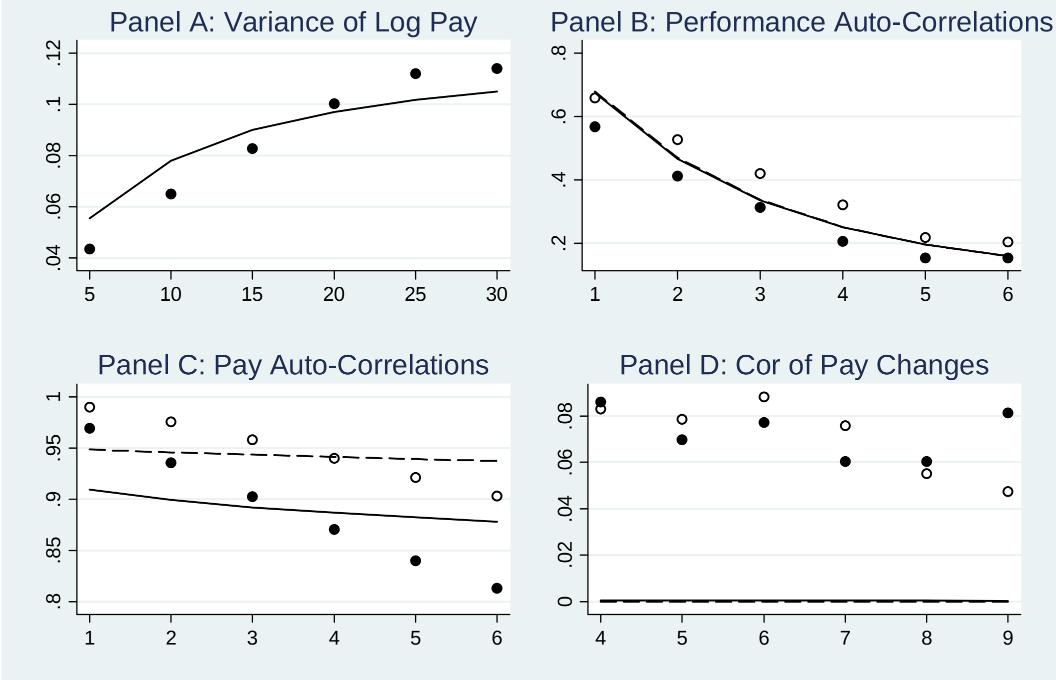

Figure 5: Results - Pure Productivity Model

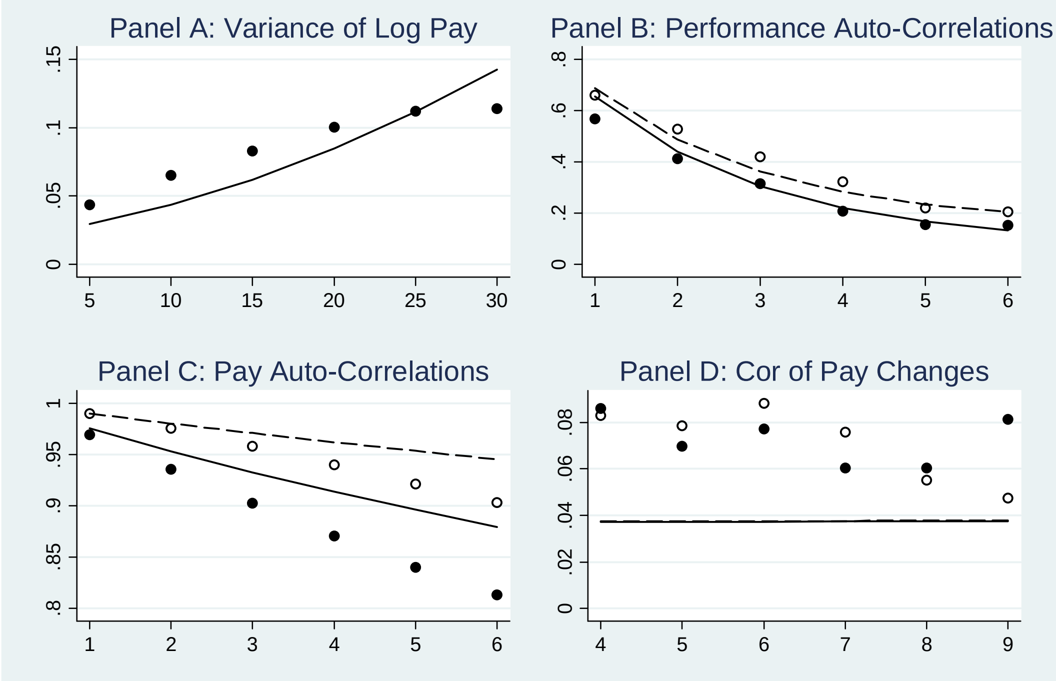

Figure 6: Results - Combined Model

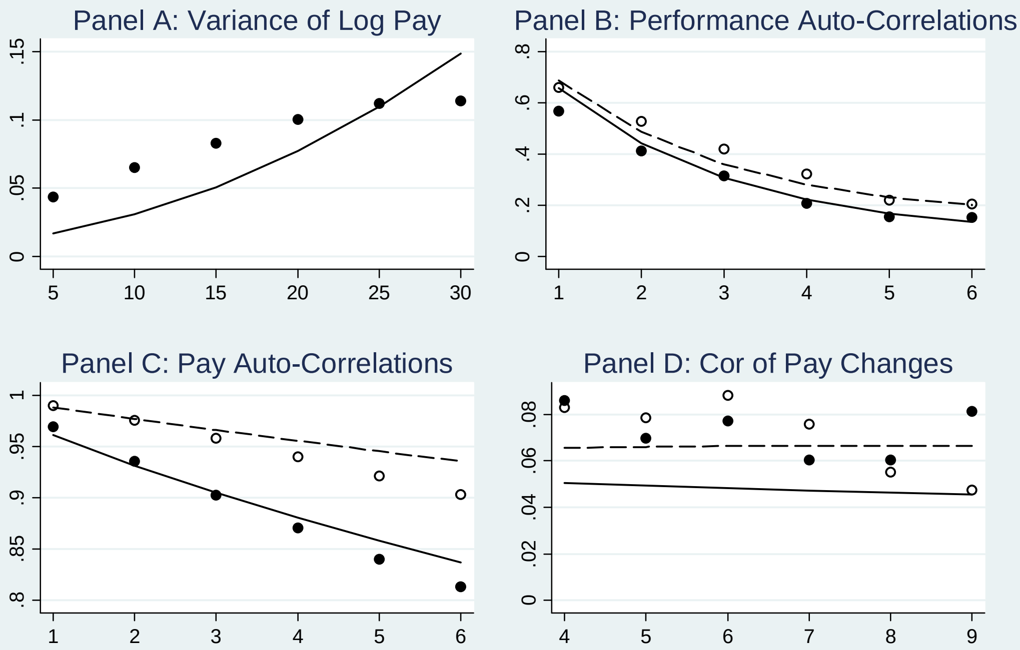

Figure 7: Productivity, Wage, and Error Variances

Full Model

Figure A-1: Log Salary in First Year as a Function of Entry Age

Controlling for age, education, race, gender, and year effects

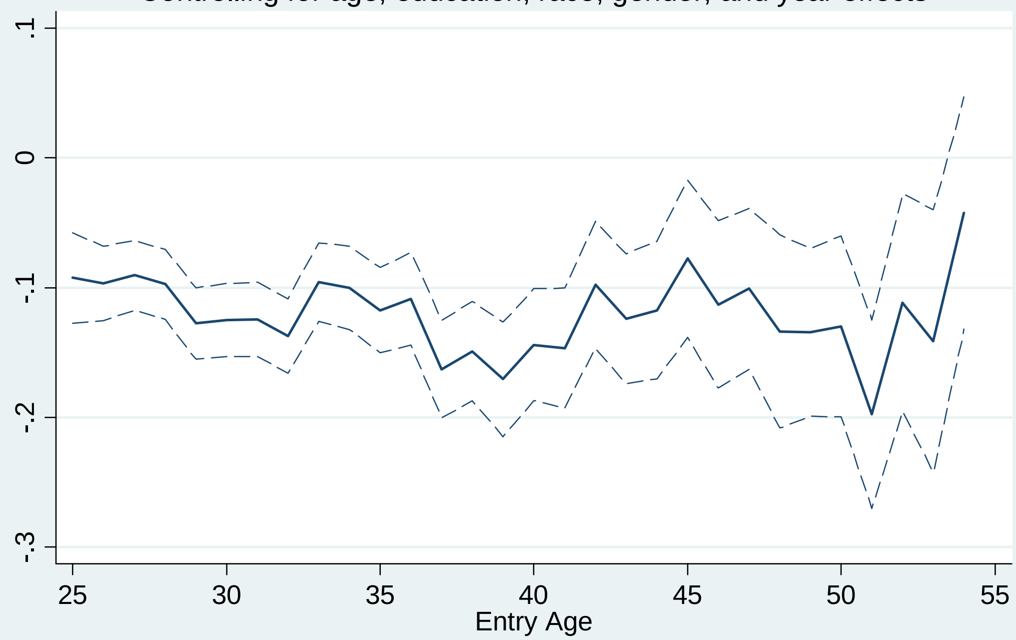

Figure A-2: Performance in First Year as a Function of Entry Age

Controlling for age, education, race, gender, and year effects