

Employer Learning, Productivity and the Earnings Distribution: Evidence from Performance Measures¹

Lisa B. Kahn and Fabian Lange
Yale School of Management and McGill University²

August 19, 2013

¹We are grateful for helpful comments from Joe Altonji, David Autor, John Galbraith, Steffen Habermalz, Paul Oyer, Imran Rasul, Chris Taber, Michael Waldman, four anonymous referees, and numerous seminar and conference participants. We thank Mike Gibbs and George Baker for providing the data. Doug Norton provided able research assistance.

²Lisa Kahn, Yale School of Management, 135 Prospect St, PO Box 208200, New Haven, CT 06520. Email: lisa.kahn@yale.edu. Fabian Lange, McGill University, Department of Economics, Leacock Building, 855 Sherbrooke Street West, Montreal, QC H3A 2T7, Canada.

Abstract

Pay distributions fan out with experience. The leading explanations for this pattern are that over time, either employers learn about worker productivity but productivity remains fixed or workers' productivities themselves evolve heterogeneously. We propose a dynamic specification that nests both employer learning and dynamic productivity heterogeneity. We estimate this 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 productivity, some of which have not yet been observed by the firm when it sets wages. We use our estimates to investigate how learning and dynamic productivity heterogeneity jointly contribute to the increase in pay dispersion with age. We find that both mechanisms are important for understanding wage dynamics. The dispersion of pay increases with experience primarily because productivity differences increase. Imperfect learning however means that wages differ significantly from individual productivity all along the life-cycle because firms continuously struggle to learn about a moving target in worker productivity. Our estimates allow us to calculate the degree to which imperfect learning introduces a wedge between the private and social incentives to invest in human capital. We find that these disincentives exist throughout 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 might have large effects on investments especially among older workers.

1 Introduction

Observationally identical workers often earn vastly different wages, so much so that after controlling for education, experience, and demographics more than two-thirds of the variation in wages remain unaccounted for. Furthermore, this unexplained variation in wages increases with age. One explanation for this increase is that worker productivity evolves heterogeneously over the life-cycle. An alternative explanation is that wages only gradually diverge as employers learn to distinguish between skilled and unskilled workers. Employer learning and dynamic productivity heterogeneity (hereafter EL and DPH, respectively) represent two of the leading hypotheses for why the unexplained variance in wages increases with age. However, there is little to no evidence on how these forces interact in shaping careers and wage profiles.

Understanding the role of EL and DPH is crucial for many important questions in labor economics. For example, in models with incomplete information such as the learning model, the agent bearing the cost of a human capital investment does not see the full benefit. Models of employer learning thus can result in inefficient investment behavior. How large are these inefficiencies? How are they distributed over the life-cycle? Answers to these questions require estimates of how EL and DPH interact over the life-cycle.¹

In this paper, we develop a new methodology exploiting information commonly collected in personnel data sets to identify and estimate models that incorporate both EL and DPH. In this, we go beyond the common approach in the literature of testing pure versions of either EL or DPH, while assuming away any role for the other.² In our model employers constantly learn about a worker’s productivity, but this productivity varies over the life-cycle. We present and estimate a tractable specification to determine how EL and DPH interact in wage dynamics and how much they contribute to pay dispersion over the life-cycle.³ Based on these estimates, we can quantitatively assess how the disincentive to invest due to incomplete information varies over the life-cycle.

Distinguishing between EL and DPH using traditional data sources is intrinsically difficult. Typically, such data contain only wages, but not any independent measures of productivity. This forces researchers who want to estimate productivity dynamics to assume that employers are perfectly informed about workers’ skills so wages equal productivity.⁴

¹Such estimates are also crucial for many other aspects of labor economics. For example, they inform on the sources and size of earnings risk over the life-cycle and are important for understanding the incentives to engage in signaling through education.

²Hereafter, we refer to the “pure EL” model to mean that employers learn about worker productivity but productivity is itself fixed, while in the “pure DPH” model productivity evolves heterogeneously throughout the life-cycle but firms are perfectly informed about worker productivity.

³The literature on earnings dispersion is too large to review here; see the Neal and Rosen (2000) survey for a useful starting point.

⁴A rich literature (eg. Abowd and Card (1989), Baker (1997), Guvenen (2007), Hause (1980), and MaCurdy (1982), among many others) related to our work analyzes the covariance structure of wages, often within the context of the human capital framework based on Becker (1964), Mincer (1958), and Ben-Porath (1967). Implicit or explicit is an assumption that wages equal productivity.

Farber and Gibbons (1996) broke new ground in exploiting an independent measure of productivity (the AFQT, an aptitude test score) that is arguably not observed by firms to test for EL. Their finding that AFQT increasingly correlates with wages over the life-cycle suggests a substantial role for employer learning.⁵ An important drawback of this literature is its assumption that researchers are better informed than employers about worker skills. Employers are assumed not to collect the AFQT (or equivalent measures) even though the information in these measures is valuable to them. An equally important drawback is that the AFQT was collected only once at the outset of workers' careers. Consequently, the EL models analyzed in the literature cannot allow for individual heterogeneity in productivity dynamics over the life-cycle. Rather the scope of these studies is limited to understanding how employers learn about productivity differences that exist at young ages.

The key innovation of our paper is to use a panel of repeated performance measures and wages to relax the restrictive assumptions of both the pure EL and DPH models. 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 The panel structure allows us to observe performance ratings that were collected prior to, contemporaneous to, and after the current period. The latter provide 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, the repeat performance ratings obtained at various points over the life-cycle allow us to estimate dynamic specifications of productivity and 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 EL and DPH. For example, the pure EL model predicts that pay correlates more with past than with future performance measures because firms rely on past, but not future, performance measures to set current pay. In contrast, an implication of the pure DPH model is that pay correlates similarly with past and future performance evaluations.

We find evidence for employer learning in that we observe that wages are indeed more highly correlated with past rather than future performance ratings. However, we observe this pattern even among experienced workers. In contrast, the pure EL model implies that firms become increasingly well informed about more experienced workers and therefore

⁵The AFQT is a composite score derived from a battery of tests administered to the respondents of the NLSY79, prior to their labor market entry. Farber and Gibbons (1996), Altonji and Pierret (2001), Lange (2007), Arcidiacono, Bayer, and Hizmo (2010), Habermalz (2011), among others, exploit this measure to study employer learning.

⁶These 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 (2012) use the data to test the Waldman (1984) promotion-as-signal hypothesis.

update less on new signals. Our full model can rationalize this continued learning by allowing for heterogeneity in the evolution of worker productivity that is difficult to predict by firms. Consequently, firms continue to update their expectations about the worker's productivity even for experienced workers: they try to hit a moving target.

These findings have important implications for the questions raised above. We find that the majority of the observed growth in the dispersion of wage residuals reflects heterogeneous innovations in productivity. However, wages and productivity are not perfectly aligned as firms make substantial errors in wage setting, even at high levels of experience. We also find that individuals' incentives to invest in their human capital are affected by imperfect information through their careers.⁷ This effect looms larger for older workers since they have less time to capture the social returns of their investments. In prior work (Lange 2007), one of us argued that firms learn rapidly about differences in worker productivity present at the beginning of workers' careers, suggesting that younger workers are most affected by imperfect information. Our finding instead suggests the opposite: the incentives to invest in skills are more severely misaligned for older workers, rather than younger, workers. This reinterpretation of the traditional employer learning model represents a significant contribution to our understanding of workers' careers and pay evolution over the life-cycle.

The remainder of this paper is structured as follows. Section 2 introduces our main model, shows how this model nests the pure EL and DPH models, and discusses the identification of these two models. Section 3 describes the data and estimation method. Section 4 reports the results and evaluates the fit of the model. In Section 5, we discuss what these estimates imply for how EL and DPH contribute to wage dynamics over the life-cycle and we show how imperfect learning affects the incentives to invest into human capital. Section 6 discusses alternative assumptions on how to interpret performance ratings, the effect of selective attrition, and how relaxing the spot market and other assumptions might affect our results.⁸ A more general formulation of the model, and a formal identification argument of the two basic constituent models are relegated to the appendices.

⁷A theoretical literature (see Chang 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. We make a similar point for symmetric learning models. This is novel, likely because the literature has so far only estimate learning models under the assumption of constant productivity.

⁸In Section 6, we also discuss pay for performance as a competing explanation. As we explain there, a direct link of pay with contemporaneous performance is not consistent with the data. On the other hand, deferred incentive schemes, such as tournament models of promotions (Lazear and Rosen 1981) or performance based raises, are difficult to identify separately from employer learning. Fitting a richer model of productivity evolution, employer learning, and job assignment in the spirit of Gibbons and Waldman (1999, 2006) would be of obvious interest here. We have abstracted away from such an exercise to retain tractability. See Smeets, Waldman and Warzynski (2013) for a first step which qualitatively assesses a model of productivity, employer learning, and one dimension of job assignment (the span of control) in personnel data on Danish employees of a large multinational firm. Pastorino (2013) estimates a model with learning, productivity, and task assignment to explore the role that experimentation by placing workers in different tasks plays in the learning process.

2 A Model of Learning and Productivity

EL and DPH models represent distinct points of view about how wages evolve over the life-cycle. We provide a parsimonious formulation that nests both. To fix ideas, we first develop this nested specification by assuming that we have access to ideal data: a panel containing pay without measurement error and a continuously distributed objective correlate of productivity. We explain how we deal with the realities of the data in section 3.

2.1 The Nested Model

Throughout, we assume that labor markets are spot markets and that information is symmetric across employers.⁹ This implies that workers are paid their expected product each period. Firms know the structure of the economy and update expectations in a Bayesian manner. These assumptions keep the model tractable. They are also standard in the previous literatures on employer learning and productivity evolution. By invoking these same assumptions we ensure that our results can be compared to these literatures. In Section 6 we discuss informally the implications of relaxing some of these assumptions.

We next impose a specific productivity process and information structure on our model. Appendix I shows how to relax these specific assumptions.

Productivity Evolution

A scalar \tilde{Q}_{it} summarizes worker productivity which evolves with observed characteristics x_i and experience t according to $\tilde{Q}_{it} = Q(x_i, t) * Q_{it}$. Here $Q(x_i, t) = E[\tilde{Q}_{it}|x_i, t]$ 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_{it} is the idiosyncratic, time-varying component of individual productivity. Denote $\tilde{q}_{it} = \log(\tilde{Q}_{it}) = \chi_{it} + q_{it}$, where χ_{it} is common to individuals with the same observable characteristics and $q_{it} = \log(Q_{it})$ represents the idiosyncratic component of productivity.

The difference equation (1) provides a simple representation of how q_{it} evolves with experience:

$$q_{it} = q_{it-1} + \kappa_i + \varepsilon_{it}^r \quad (1)$$

We assume $\kappa_i \sim N(0, \sigma_\kappa^2)$ and $\varepsilon_{it}^r \sim N(0, \sigma_r^2)$ and that the ε_{it}^r are uncorrelated over time and with κ_i . We initialize this difference equation in period 0 with a draw of q_{i0} from a normal distribution $N(0, \sigma_q^2)$.¹⁰ This draw is independent of κ_i . By construction, q_{it} is mean zero.

⁹A large literature deviates from the assumptions of spot markets and symmetric information. For example, Gibbons and Katz (1991), Kahn (2013), Schönberg (2007) and DeVaro and Waldman (2012) provide evidence, in a variety of settings, that employers learn asymmetrically. Further BGH (1994b), Beaudry and DiNardo (1991), Kahn (2010), and Oreopoulos et al. (2012) 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 intractable to include features of these models in our paper.

¹⁰We adopt the convention that period 0 is a period prior to the first period the individual spends in the labor market.

Equation (1) allows for three sources of heterogeneity in the evolution of log productivity q_{it} . Individual differences in q_{i0} reflect differences in initial ability. Differences in the drift parameter κ_i allow for persistent differences in the intensity with which individuals accumulate human capital over the life-cycle.¹¹ ε_{it}^r captures unpredictable innovations in individual productivity that can stem from various sources, such as task evolution, health shocks, or technological change rendering skills obsolete. These innovations are persistent because we assume that ε_{it}^r follows a random walk. And, since ε_{it}^r are i.i.d, the variation in these innovations does not decline with experience. The productivity process (1) implies that productivity continues to diverge even among experienced workers.

Information Structure

We use three different types of signals to model how employers learn. At the onset, firms receive an initial signal z_{i0} . In each subsequent period, employers observe two signals: $\{p_{it}, z_{it}\}_{t=1}^T$. 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} . The signal structure is:

$$\begin{aligned} p_{it} &= q_{it} + \varepsilon_{it}^p \\ z_{i0} &= q_{i0} + \varepsilon_{i0} \\ z_{it} &= q_{it} + \varepsilon_{it}^z \end{aligned} \tag{2}$$

where $(\varepsilon_{i0}, \varepsilon_{it}^p, \varepsilon_{it}^z)$ are independently distributed, mean zero, normal random variables with variances $(\sigma_0^2, \sigma_p^2, \sigma_z^2)$. The normality assumptions allow us to analyze the learning process using the tools of Kalman filtering and ensure great parsimony for the model.¹²

In Appendix I, we show how one can use linear state space methods to derive second moments of wages and performance in a more general class of models. Applying these methods to our specific case, we obtain the implied second moment matrices for wages and performance ratings which depend only on 6 parameters: $(\sigma_q^2, \sigma_r^2, \sigma_\kappa^2, \sigma_0^2, \sigma_p^2, \sigma_z^2)$. We will later estimate these parameters by matching empirical moments in the data. The parsimony of the model makes it fairly clear how the moments and parameters map onto each other. At the same time the model is sufficiently complex to nest EL and DPH. The restriction $\sigma_\kappa^2 = \sigma_r^2 = 0$ eliminates any heterogeneous dynamics in productivity and results in the pure EL model. By contrast, the restriction $\sigma_0^2 = \sigma_z^2 = 0$ removes any noise in the signals observed by the firm (but not the employer) and thus delivers the pure DPH model.

¹¹Persistent differences in intensity would arise, for example, if individuals differ in either their preferences or ability to invest (Becker 1964, Ben-Porath 1967).

¹²The assumption that $cov(\varepsilon_{it}^p, \varepsilon_{it}^z) = 0$ is without loss of generality since 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 either correlated or orthogonal signals are therefore identical.

2.2 Implications and Identification

Our goal is to separately identify how productivity evolves from how employers' expectations about productivity, which are reflected in wages, evolve. Clearly, this requires more than just observing wages. One approach to separate learning from productivity is to impose strong functional form assumptions on the productivity process. The alternative approach, pursued in this paper, is to obtain additional information about the underlying productivity process. We rely on a productivity correlate observed at multiple times over a worker's life-cycle to provide this additional information.

In the remainder of this section, we develop intuition about the identification of the model by contrasting the pure EL model with the pure DPH model. In the pure EL model, information is imperfect and productivity is constant over the life-cycle; in the pure DPH model, firms have perfect information and productivity evolves stochastically. We discuss each model in isolation not because we believe that either describes the world well; our empirical analysis below indeed shows that combining heterogeneous productivity dynamics with employer learning substantially improves the fit of the data. Rather, we discuss the two pure models in detail to clearly contrast the empirical implications of both forces.

Appendix II contains a more formal discussion of how to identify the parameters in the model using the second moments of wages and performance ratings.

The Pure Employer Learning Model

It has long been appreciated that wage changes in pure EL models result only from new information and are therefore serially uncorrelated. It is also well known that the variance in pay increases with experience at a decreasing rate. As firms learn to distinguish among workers, pay becomes more and more dispersed. However, eventually learning and the increase in the pay variance slows down.

Central to our analysis and novel to the literature are implications of the pure EL model for how wages covary with performance measures at various leads and lags. Given the restrictions of the pure EL model, $\sigma_\kappa^2 = \sigma_r^2 = 0$, wages are given by:

$$w_{it} = E[q_i | I^t] = \chi_t + (1 - K_{t-1}) * E[q_i | z_{i0}] + K_{t-1} \frac{1}{t-1} \sum_{j=1}^{t-1} \phi_{ij} \quad (3)$$

$$\text{where } \phi_{it} = (1 - \phi) p_{it} + \phi z_{it} \quad (4)$$

$$K_t = \frac{t\sigma_q^2}{t\sigma_q^2 + \sigma_\phi^2} \quad (5)$$

Recall, χ_t captures the variation in expected log productivity with age that is common across individuals.¹³ The remaining parts of equation (3) show how wages depend on the

¹³This variation is due to changes in average productivity with experience itself and to changes in the variance of the expectation error in productivity, which enter due to the non-linearity of the logarithmic

signals observed by the firm.¹⁴

From equations (3)-(5), it is easy to derive the covariances between pay and the performance measures observed in the data:

$$\text{cov}(w_{it}, p_{i\tau}) = \begin{cases} K_{t-1}(\sigma_q^2 + \frac{1-\phi}{t-1}\sigma_p^2) & \tau < t \\ K_{t-1}\sigma_q^2 & \tau \geq t \end{cases} \quad (6)$$

Inspecting equation (6) we observe that $K_{t-1}\sigma_q^2$ appears in the covariance between wages and both leading ($\tau \geq t$) and lagging ($\tau < t$) performance measures. $K_{t-1}\sigma_q^2$ reflects the joint dependence of both performance measures and wages on productivity q_i . It increases in experience t since the expectation error in wages declines with experience making wages and productivity more closely aligned. The additional term in the covariance between wages and *lagged* performance captures the fact that past performance measures are used to form expectations and thus to set wages. Therefore the signal noise in past performance measures directly enters wages. This raises the covariance between wages and lagged performance measures.

Reflecting this intuition, equation (6) generates two additional implications of the pure EL model for the covariances of wages and performance measures. First, the $\text{cov}(w_{it}, p_{i\tau})$ for $\tau < t$ exceeds that for $\tau \geq t$; the $\text{cov}(w_{it}, p_{i\tau})$ will be a step function of τ with a negative discontinuity at $\tau = t$. This is because current pay incorporates past, but not future, realizations of p_{it} . The second prediction is that the size of the step decreases in t . Differencing the two expressions in equation (6) we see that the step size is equal to $K_{t-1}\frac{1-\phi}{t-1}\sigma_p^2$, which decreases with experience t .

Figure 1, panel A illustrates both predictions by plotting a simulated set of $\text{cov}(w_{it}, p_{i\tau})$ for $\tau \in (t-6, t+6)$ for a younger (experience 7) and an older worker (experience 20).¹⁵

Figure 1: Simulated Correlations of Pay and Performance

Intuitively, firms incorporate past performance when setting current pay, but cannot incorporate performance measures that have not yet been realized. This results in higher correlations of pay with past performance measures than with future performance measures, or a “step” in the $\text{cov}(w_{it}, p_{i\tau})$. This distinction between the past and the future is fundamental to learning models because it separates observed and unobserved information. The size of this step provides information about the amount of learning that takes place at

function. 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.

¹⁴In 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_{it} depends on how much variance there is in both signals respectively. The exact expressions for ϕ and σ_ϕ^2 , the variance of the scalar signal ϕ_{it} are known, but not of particular interest at this point.

¹⁵We use our eventual estimates of $(\sigma_q^2, \sigma_0^2, \sigma_p^2, \sigma_z^2)$ from the pure EL model (estimated in section 4 and reported in table 4, column 1) to simulate data and generate these covariances.

different experience levels. It is therefore very influential in identifying the role of learning over the life-cycle.

The step is smaller for older workers because the correlation of pay with future performance measures is higher for this group, while that with past performance measures is unchanged. Intuitively, as workers age, firm expectations become more precise so wages and productivity are more highly correlated. This force works to increase the correlations of pay with past and future performance measures. However, for older workers firms rely on many more signals than for younger workers and they place less weight on any given signal when setting pay. Consequently, among older workers there is an offsetting tendency that lowers the correlations between pay and past (but not future) performance measures. On balance, the correlations of pay with future performance measures increase, the correlations with past performance measures are constant, and the difference between the correlations with past and future performance measures (the “step”) decline in experience.

For future reference, we restate the primary implications of the pure EL model that are relevant for distinguishing it from a pure DPH model.

- (EL 1) Wage changes are serially uncorrelated.
- (EL 2) The variance in pay increases with experience at a decreasing rate.
- (EL 3) The covariance of pay with past performance is larger than that with future performance: the $cov(w_{it}, p_{i\tau})$ is a step function with a discontinuity at $t = \tau$.
- (EL 4) The size of the step in (EL 3) declines with experience.

The Pure Dynamic Productivity Heterogeneity Model

We will now discuss the pure DPH model obtained by setting $\sigma_0^2 = \sigma_z^2 = 0$. With these restrictions, log wages and the performance measures observable to the researcher are:¹⁶

$$\begin{aligned} w_{it} &= q_{it} \\ p_{it} &= q_{it} + \varepsilon_{it}^p \end{aligned} \tag{7}$$

In contrast to the pure EL model, wages do not typically follow a random walk. Instead, wages and productivity have the same stochastic dynamic properties. Our specification (equation (1)), for instance, implies that the covariance in wage growth at different experience levels is σ_κ^2 , the variance of the heterogeneous trend in productivity.¹⁷ Equation (1) also implies that the variance in pay rises in experience at an increasing rate.

¹⁶Under perfect information, employers have no incentive to collect the measures p_{it} . Thus, the observation that firms collect p_{it} can be taken as evidence against the pure productivity model.

¹⁷MaCurdy (1982), Baker (1997), Abowd and Card (1989), Guvenen (2007) and many others use auto-correlation in wage growth to test for permanent heterogeneity in productivity growth. The findings in this literature on this question vary.

These two implications (serial correlation in pay growth and that the variance of log pay increases convexly with experience) are somewhat specific to the production process we imposed. If $\sigma_\kappa^2 = 0$, then the variance of log pay would increase at a constant rate and log pay would follow a random walk. A more fundamental distinction between the pure DPH model and the pure EL model can be drawn by considering the covariance of pay with different leads and lags of the performance signals. Since the signal noise ε_{it}^p is orthogonal to q_{it} , we have the following expression for the covariance between performance measures at τ and pay at t :

$$\text{cov}(w_{it}, p_{i\tau}) = \text{cov}(q_{it}, q_{i\tau}) \quad (8)$$

It follows immediately from equation (8) that at $t = \tau$, $\text{cov}(w_{it}, p_{it})$ increases in the variance of q_{it} and consequently with the variance of w_{it} . Thus, as long as the variance of pay increases (as is generally observed over the life-cycle), the covariance of pay with performance measures should also increase. Furthermore, equation (8) implies a fundamental smoothness in the $\text{cov}(w_{it}, p_{i\tau})$ at $t = \tau$; there will be no discontinuity. This is because t and τ are interchangeable in equation (8). Thus, we have that $\text{cov}(w_{it}, p_{i\tau}) = \text{cov}(w_{i\tau}, p_{it})$.¹⁸

Additional implications follow from combining (8) with our specific productivity process (1). In particular, the $\text{cov}(w_{it}, p_{it+k})$ increases in k . To see this, assume for the moment that $k > 0$. The covariance between $p_{i,t+k}$ and w_{it} is $\text{var}(q_{it}) + k * \text{cov}(q_{it}, \kappa_i)$. Since κ_i also enters into q_{it} , we obtain that $\text{cov}(w_{it}, p_{it+k})$ increases linearly in k . A similar argument applies for $k < 0$.

We illustrate these implications using simulated data in Panel B of figure 1.¹⁹ Notice the covariances are increasing in experience and in k . In a full information world the error in past performance measures is irrelevant for wage setting. In contrast, if there is incomplete information, then the firm will not be able to separate the error in the past performance measure from the signal. It will therefore set pay partially based on this error. It cannot do this for the performance measures that will be observed in the near future. This generates the discontinuity at $t = \tau$ in the pure learning model that is not present in the pure productivity model.

We now summarize the primary implications of the pure DPH model that allow distinguishing it from a pure EL model.

- (DPH 1) Wage changes are serially correlated.
- (DPH 2) The variance in pay increases in experience at an increasing rate.
- (DPH 3) The covariance of pay with performance, $\text{cov}(w_{it}, p_{i\tau})$, is increasing in experience, t , and in τ .

¹⁸Sufficient conditions for a lack of a discontinuity in the full information model are: p_{it} is correlated with q_{it} , wages equal expected productivity, and wages and current performance measures are related only through the correlation between productivity and expected productivity.

¹⁹To simulate the data for these covariances, we use estimates of $(\sigma_q^2, \sigma_r^2, \sigma_\kappa^2, \sigma_p^2)$ from the pure DPH model reported in table 4, column 2.

- (DPH 4) There is no discontinuity in the correlation of wages with past and future performance measures at $t = \tau$.

The above discussion illustrates the basic predictions that allow us to distinguish between EL and DPH using our data. In addition to testing each model in its pure form, we can use the nested model to study how employer learning and productivity evolution interact in generating observed dynamics of wages.

3 Data and Measurement Issues

In this section, we describe the data and we show how we adapt the above model in the face of the existing measurement issues. We summarize the key moments in the data and explain how we use these to estimate our model via general method of moments (GMM).

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 9,626 employees with non-missing education who can be observed with at least one wage or performance measure between the ages of 25 and 54 and at least one more wage or performance measure. We adopt the convention that age 25 is the first year of experience.²⁰

Table 1 reports summary statistics. The majority of managers are white males with at least a college degree. Average annual salary is \$54,000 in 1988 dollars and measures base pay.²¹

Table 1: Summary Statistics

Models of EL or DPH are about deviations of pay and performance from their average profiles. We thus residualize both log pay and performance on observable characteristics.²²

²⁰Age 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. Since we don't observe them before they enter this sample, we start at the earliest age which still yields a decent sample size. As a robustness check we have estimated the model separately for each education category thus effectively using a potential experience measure. The results are reported below.

²¹We follow the same restrictions on salary as those in BGH. We have information on bonus pay for some years (1981-1988) but do not include it in the analysis to maintain consistency in our data across years. In these years, 25% of workers in our sample 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.

²²Specifically, we residualize on dummies for age, race, gender and year, all interacted with education group (high school, some college, exactly college, advanced degree). In order to allow for different career

Since our data come from one firm only, the variation in pay will be lower than in the population. Appendix figure A1 compares the life-cycle variation in wage residuals in our data with the variation in the Current Population Survey over the same time period.²³ The average residual wage variance in the CPS is about 2.5 times that in BGH (0.23 compared to 0.09), while the experience profile is a bit steeper in BGH. This is important to be mindful of in interpreting our results.

3.2 Subjective Performance Ratings (and other measurement issues)

So far we have treated the observed performance measures p_{it} as noisy signals of productivity $p_{it} = q_{it} + \varepsilon_{it}^p$ where ε_{it}^p is normally distributed white noise. We will use the short-hand “objective performance measures” to describe performance correlates that have this structure. Unfortunately, the subjective, managerial performance assessments at our disposal do not conform with these assumptions. An obvious difference is that the ratings are ordinal rather than continuous random variables. They range from 1 to 4, with higher ratings reflecting better performance. From table 1, we see that 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

It is straightforward to accommodate the discrete nature of the performance ratings. For this purpose denote the observed performance rating as \tilde{p}_{it} . We assume that this discrete random variable is generated by a latent signal on individual productivity, p_{it} , which satisfies the assumptions made in the previous section. From the joint distribution of compensation w_{it} and observed discrete measure $\tilde{p}_{i\tau}$ for any t and τ we can identify the correlation between $p_{i\tau}$ with w_{it} and with $p_{i\tau \neq t}$ using maximum likelihood methods described in more detail below.

Besides accounting for the discrete nature of these ratings, we also need to address the fact that our performance ratings represent subjective assessments. If we maintain the assumptions embedded in eqs. (2) and also assume that workers are paid their expected marginal product, then we have that $E[p_{it}|t] = E[q_{it}|t] = E[w_{it}|t]$ so that the life-cycle of average ratings and wages should be equal to each other. This restriction is clearly rejected by the data. Figure 2 plots log pay and performance residuals by age. The solid line

trajectories and secular trends across race and gender, we also interact race and gender with a linear time trend and a quadratic in age.

²³To remain comparable with our data, we use survey years 1970-1989 and restrict attention to annual earnings (in the previous calendar year) of full-time, full-year, private sector workers age 26-55 with at least a high school degree. We reweight the CPS sample to match the age-race-gender-education distribution of the BGH data. We also drop those who earned less than \$2600 over the year since this would be less than the federal minimum wage over this time period. We then residualize log wage and salary income on the same control variables listed above.

²⁴We inverted and recoded the original measures, which ranged from 1 to 5, combining the worst two ratings since almost nobody receives the worst. Similar distributions of performance ratings are found in Medoff and Abraham (1980 and 1981), Murphy (1991), and Frederiksen, Lange, and Kriechele (2013) in their studies of performance ratings across various industries and firms.

shows that earnings are rising with age, but at a decreasing rate, reflecting typical life-cycle patterns. The dashed line reveals, somewhat surprisingly, that average performance ratings decline with age in our data.

Figure 2: Log Wages and Performance by Age

In pioneering work, Medoff and Abraham (1980, 1981) found a similar pattern in a different set of firms: the life-cycle profiles of compensation and subjective performance measures often deviate from each other. Frederiksen, Lange, and Kriechel (2013) examine these life-cycle profiles in some detail in data from various industries, countries, and time-periods. While compensation invariably has a familiar Mincerian shape, subjective ratings deviate substantially. In some firms they increase with experience, in others they decrease and sometimes they are even non-monotone in experience. The observation that the life-cycle profiles of wages and performance ratings deviate from each other makes it impossible to both assume that wages equal expected productivity and that subjective ratings are unbiased signals (in the sense that $E[p_{it}] = q_{it}$).

Findings from studies that have access to objective ratings (e.g. Waldman and Avolio 1986) suggest that productivity tends to have the shape familiar from Mincer earnings regressions. And, those studies that have data on both objective and subjective ratings (see Jacob and Lefgren 2008 and Bommer et al. 1995) find high correlations between both.

We thus face a situation where subjective ratings display different life-cycle profiles than compensation and objective performance correlates. In addition, the life-cycle profiles of subjective ratings vary significantly across firms. Finally, we know (Gibbs 1995, Frederiksen, Lange, and Kriechel 2013) that subjective ratings within narrowly defined demographic groups correlate with career outcomes such as compensation, promotions, and retention. Subjective performance ratings therefore contain information relevant for workers' compensation and career evolution, despite the differences in the life-cycle profiles of wages and performance ratings.

These diverse empirical findings can be reconciled by recognizing that the scales and frames of reference for the subjective ratings are likely to change across different stages of careers and with demographic characteristics. However, within demographics and career stages, subjective ratings do contain information about the relative performance of workers (see Gibbons and Waldman 1999). By interpreting the rankings as relative within narrowly defined experience, education, and demographic groups, we remove any information contained in variation across these narrowly defined peer groups. However, continue to exploit the variation in performance ratings within groups defined by demographics and possibly other characteristics. This approach can accommodate the variation in the ratings across experience and other observables that define the peer groups. And, it also accommodates the correlation of subjective ratings with compensation and other outcomes within peer groups.

In our analysis, we therefore 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)

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

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} , where we allow these thresholds to differ across age groups. In practice, we generate age-specific performance deciles on the residualized performance measures, thus incorporating the assumption that ratings are relative to individuals of the same age.²⁵ The structure imposed in section 2 implies that the latent signal, p_{it} , is normally distributed. We can therefore estimate correlations of the latent index p_{it} with other normally distributed variables (such as log wage residuals and lagged performance) using maximum likelihood methods.²⁶ As usual for categorical variables, we cannot identify the variance of p_{it} . For this reason, we focus from now on on correlations, rather than the covariances discussed in section 2.2. It is straightforward to show that identification arguments in section 2.2 also apply to correlations.

As we estimated the model, we found that the 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 noise in the performance measures, ε_{it}^p , evolve according to equation (10) :

$$\varepsilon_{it+1}^p = \rho \varepsilon_{it}^p + u_{it+1} \quad (10)$$

where the initial noise is $\varepsilon_{i1}^p = 0$ and $u_{it} \sim N(0, \sigma_u^2)$. 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.²⁷

²⁵We have experimented with different approaches in generating the reference group for a worker. Besides age, we have allowed performance to be relative to other workers in their entry cohort and also relative to the job level at worker attained. Our results are qualitatively and quantitatively robust to redefining the comparison groups in this manner.

²⁶Whenever we refer to “deciles”, we actually mean that the support is divided into 9 parts. This was made necessary by the specific requirements of the polychoric estimation command in Stata that we rely on.

²⁷In order to generate auto-correlation in performance measures we could also assume that the innovation in productivity follows an AR1. However, this assumption would force the nested model to be very similar to the full information DPH model. To see this note that p_{it} contains noise ε_{it}^p . Thus, the AR-1 process in

Finally, we also adapt the model to allow for measurement error in wages:

$$W_{i,t} = W_{i,t}^* \Omega_{i,t} \quad (11)$$

where $W_{i,t}$ is the observed wage, $W_{i,t}^*$ is the wage measured without error and $\Omega_{i,t}$ represents the measurement error. Taking logs we get

$$w_{it} = w_{it}^* + \omega_{it} \quad (12)$$

We assume that ω_{it} is classical measurement error with $\omega_{it} \sim N(0, \sigma_\omega^2)$. In practice, we residualize log wage on the same set of variables used to residualize the performance measures.

Thus, with the addition of measurement error in wages and auto-correlation in the signal noise, we now have 8 parameters governing our model: $(\sigma_q^2, \sigma_r^2, \sigma_\kappa^2, \sigma_0^2, \sigma_u^2, \rho, \sigma_z^2, \sigma_\omega^2)$. We next describe the empirical moments we use to estimate these parameters.²⁸

3.3 Moments for estimation

Our model 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 think are particularly informative for distinguishing learning and productivity models. These moments are shown in figures 3a and 3b and in table 2.^{29,30}

observed performance necessarily exhibits less persistence than the AR-1 process in true productivity. In order to generate the auto-correlations between p_{it} and p_{it-1} (on the order of 0.6), we would need the signal noise in ε_{it}^p to be very small. If however ε_{it}^p is very precise, then we are back to the full information DPH model, which we show to be rejected by various empirical findings described below.

²⁸Our key identification arguments regarding the “step” in the correlations of pay with lags and leads of performance carry through with the introduction of ρ and σ_ω^2 . First, classical measurement error in wages will not affect the covariances of wages with other performance measures or with wages in other periods. Second, for ρ , recall that in the pure EL model, $p_{it} = q_i + \varepsilon_{it}^p$. Allowing for $\rho \neq 0$, we have: $p_{it} = q_i + \rho \varepsilon_{it-1}^p + u_{it} = q_i + \varepsilon_{it-1}^p + (\rho - 1)\varepsilon_{it-1}^p + u_{it}$ (adding and subtracting ε_{it-1}^p). Thus $cov(w_{it}, p_{it}) = cov(w_{it}, q_i + \varepsilon_{it-1}^p + (\rho - 1)\varepsilon_{it-1}^p + u_{it}) = cov(w_{it}, p_{it-1}) + (\rho - 1)cov(w_{it}, \varepsilon_{it-1}^p)$, since u_{it} is iid noise. Since $(\rho - 1) < 0$ and $cov(w_{it}, \varepsilon_{it-1}^p) > 0$, we have that $cov(w_{it}, p_{it}) < cov(w_{it}, p_{it-1})$. Therefore we will still have a “step” at $t = \tau$ when $\rho > 0$, though it may be smaller. Furthermore, in the pure EL model, $cov(w_{it}, \varepsilon_{it-1}^p)$ will be declining in t ; as workers gain experience, firms place less weight on any given signal. Thus the step size will still be declining in t .

²⁹In constructing these moments, we take average correlations and variances across the specified set of experience years weighted by the number of individuals for which we observe that moment.

³⁰We have investigated to what extent these patterns are similar if we slice the data by education group. Regardless how we cut the data, the second moments of wages and performance measures are consistently similar to those reported for the aggregate sample, with some minor deviations. The one major exception is that the asymmetry in time for the correlations between pay and performance among the less educated 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. Versions of figure 2 estimated on subgroups in our data are available

Figures 3a and 3b: Moments and 95% CI

Table 2: Empirical Moments

Panel A in figure 3a shows the variance in log wage residuals for six 5-year experience groups ranging from 1-5 to 26-30 years. The variance in pay around the age profile increases almost linearly with age, slowing only slightly after about age 50. Understanding this variation and its increase over the life-cycle is the primary task of this paper. Note both the pure EL model and the pure DPH model predict increasing variances (EL2 and DPH2), but the former predicts a concave pattern, consistent with this figure, while the latter predicts a convex pattern.

Panels B and C in figure 3a show auto-correlations in performance and pay residuals, respectively, for up to 6 lags and for two experience groups: experience 1-15 with solid dots and 16-30 with hollow dots. For both pay and performance, the more experienced group exhibits higher auto-correlations which decline across lags. As discussed above, we allow for an auto-regressive component in the signal noise to match the lag-structure of the performance auto-correlations.

Panel D in figure 3a shows correlations in pay changes for up to 9 lags and for the same two experience groups. According to the EL model pay changes are serially uncorrelated (EL1) while the DPH model allows for positive correlations in pay changes. (DPH2). The sizable correlations in pay changes that are statistically distinguishable from zero therefore provide clear evidence against EL and in favor of DPH.

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 wages, which we can not account for given our spot market assumption. We will therefore only fit the 4th through 9th lag in wage growth when we estimate the model.³¹

Lastly, figure 3b presents correlations of current pay with past, current and future performance measures for up to 6 lags and leads, for the two experience groups. These correlations are the empirical analogues to the simulated covariances depicted in figure 1. We pay particular attention to these moments throughout the paper because we believe they represent the major innovation to the previous literature and are particularly informative for separately identifying EL and DPH models. To better understand the size of the step, we present these correlations and the difference between past and future for a given lag/lead

upon request.

³¹The only way to generate such high early correlations in wage growth within the context of our model would be if learning about productivity innovations is very rapid. However, such rapid learning is at odds with several patterns in the data, discussed below. Our model therefore fails along this dimension. A model which relaxes the assumption of spot markets will have better luck in fitting the joint patterns of slow learning and high early correlations in wage growth. See Section 6 for a discussion of the spot market assumption.

time in table 3.

The evidence in figure 3b and table 3 is not entirely consistent with either the pure EL or the pure DPH model. We do see higher correlations of pay with past performance measures than with future performance measures (or, a “step”), consistent with EL3 and inconsistent with DPH3 and DPH4. For young workers, the differences reported in table 3 are positive and statistically significant for the first three leads and lags. However, the step size tends to be larger for older workers, violating EL4. Finally, for both past and future performance measures, correlations are larger for older workers, partially consistent with DPH3.

Table 3: The Asymmetry in Correlations of Pay with Lags and Leads of Performance

Thus, the reduced form evidence is not fully consistent with either EL or DPH.

3.4 Estimation Methodology

Our model produces a mapping from the 8 parameters $(\sigma_q^2, \sigma_r^2, \sigma_\kappa^2, \sigma_0^2, \sigma_u^2, \rho, \sigma_z^2, \sigma_\omega^2)$ to the second moments of pay and performance. We estimate this model by matching the 68 moments described in section 3.3 via method of moments with equal weights on all moments.^{32,33} We obtain standard errors by bootstrapping with 500 repetitions.³⁴

As noted above, we estimate three versions of the model. First, we impose $\sigma_\kappa^2 = \sigma_r^2 = 0$, eliminating any heterogeneous dynamics in productivity. This yields the pure EL model. Second, we impose $\sigma_0^2 = \sigma_z^2 = 0$, implying the firm has full information (since there is no noise in the private signals the firm observes), obtaining the pure DPH model. Third, we estimate the model with no restrictions, combining EL and DPH in a single model.

³²We use the identity weighting matrix, rather than a two-step approach with optimal weights. Some moments are indeed estimated more precisely than others in the data (for example, table 2 shows that the precision of the estimated variance and auto-correlations in pay is an order of magnitude greater than that of the other moments). However, as explained in section 2.2, these are not the key moments for separately identifying learning from productivity models. Optimal weighting would emphasize these moments at the expense of the moments that we believe to be particularly informative for our investigation. We have instead chosen to weigh all moments equally.

³³We generate a candidate set of moments by simulating data at a point in the parameter space, then aggregating to the moments by experience group in the exact same manner as in the actual data. We then minimize a distance function (the sum of squared distances of the empirical and the candidate moments) by going through four optimization routines, alternating between Newton-Raphson and the simplex method. After each routine, we use the parameter estimate obtained from the previous routine as our starting value for the next step. To obtain the point estimates, we have also worked with a grid search in starting values to ensure that we have found global minima.

³⁴We randomly sample with replacement from the data to generate the bootstrapped moments. We then estimate the parameters, with the same method described above, to match these moments, taking as starting values the parameters values shown in table 4.

4 Estimation

Table 4 displays the parameter estimates for the three models described above and discuss the fit of the model using figures 4-7. These figures contrast the empirical moments (solid dots for young and hollow dots for older workers) with the predicted moments based on the estimated parameters for a given model (solid lines for young and dashed lines for older workers).

Table 4: Parameter Estimates

Figure 4: Correlations of Pay and Performance

The Pure Employer Learning Model

Panel B of figure 4 and figure 5 summarize the results of the pure EL model. Panel A of figure 5 shows that it roughly matches the variance of wages across experience levels (EL2). The learning model also roughly fits the auto-correlations in performance (panel B), even if it does not reproduce their differences across experience. Regarding the auto-correlation in wages (panel C), it matches the levels and differences across experience groups, but not the decline in the auto-correlations with lags.³⁵ By construction, the model predicts that wages follow a random walk (EL1) and is contradicted by the positive correlations in pay-growth observed in the data (panel D.)

Figure 5: Results for the pure EL model

As is evident in Figure 4, panel B, the pure EL 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. Though it can fit the higher correlations of pay with past than with future performance measures (EL3), it does also predict a much smaller step among older workers than is found in the data (EL4). Most importantly, it fails to reproduce the empirical fact that the correlations between pay and performance at all leads and lags are greater for experienced workers. A general feature of pure EL models is that pay is less highly correlated with past performance measures among the more experienced workers as firms rely less on each individual measure when setting pay. This failure is therefore not the result of particular distributional assumptions but rather reflects a more general failure of the pure learning model.

Overall, though the data is consistent with EL2 and EL3, it contradicts EL1 and EL4.

³⁵Because productivity is fixed, the learning model cannot explain performance auto-correlations that are rising in experience. Auto-correlations of wages increase in experience because firms revise their expectations less for older workers when new information arises and thus wages are more stable.

The Pure Dynamic Productivity Heterogeneity Model

Figure 4, panel C and figure 6 show how the pure DPH model fits the data. Along a number of dimensions, this model does better than the pure EL model. We find that the variance of heterogeneous growth term κ_i reported in Table 4 is non-zero generating positive correlations in pay changes (DPH1), even if these are smaller than those in the data (figure 6, panel D). The pure productivity model also fits the auto-correlations in performance (panel B) and wages (panel C) better than the learning model did. However, the model does poorly in fitting the variance of log pay across experience (panel A). Growth rate heterogeneity implies that the implied variance in wages rises in the square of experience (DPH2), producing the convex pattern in panel A that is not present in the data.

Figure 6: Results for the pure DPH model

Turning to our main set of moments (figure 4, panel C), the evidence regarding the pure DPH model is also mixed. Consistent with DPH3, the empirical correlations between pay and performance are increasing in experience. However, within experience, the model implies that the correlations with current pay are larger for performance measures collected in the future, resulting in the upward slope of the lines. Clearly, the empirical moments do not show this upward slope. Furthermore, the pure DPH model cannot rationalize the discontinuity at 0 (DPH4).

Thus, while the data are consistent with DPH1, they are only partially consistent with DPH3, and violate DPH2 and DPH4.

The Nested Model

Finally, we show results from the nested model in figure 7 and panel D of figure 4. Overall, combining EL with DPH helps substantially to fit the observed patterns in the data. Panel D of figure 4 shows that the nested model can fit the step in the correlations of pay across lags and leads of performance, for both young and older workers. Figure 7 shows that the nested model succeeds in fitting the auto-correlations for performance (panel B) and for pay (panel C), both across experience and across lags. It is also able to fit high correlations of pay changes (panel D). The nested model however fails to fit the concavity in the variance of log wages across experience (panel A) since the estimated heterogeneity in κ_i is large.³⁶ It is worth noting, though, that compared to the pure DPH model, the nested model rationalizes a larger role for κ_i , which helps fit the correlations of pay changes. This is because with imperfect information, innovations in κ_i take time to be priced into pay, thus reigning in

³⁶We find that the heterogeneity in κ contributes substantially to diverging productivity over the life-cycle. One standard deviation in κ_i corresponds to 45% of extra productivity growth over 30 years, while one standard deviation of the sum of random walk components over 30 years amounts to about 10-15% of extra productivity growth.

the convexity in the increase of pay variance with experience.

Figure 7: Results for the combined model

Turning to the estimates of learning parameters, we find that the variance in both the initial signal (σ_0^2) and the dynamic signals (σ_z^2, σ_u^2) are substantially smaller for the nested model than for the pure EL model. The latter requires more signal noise to match the evidence for learning even at higher experience levels. The nested model instead allows for much less signal noise. The variance of log wages continues to increase because productivity itself evolves and because firms need to learn about this moving target.

Testing the Three Models

Our discussion so far has focused on the qualitative fit of the models. Statistically, we clearly reject the pure EL and DPH models in favor of the nested model. Using a Wald test, we reject the restrictions of the pure DPH model ($\sigma_0^2 = \sigma_z^2 = 0$) against the unrestricted model at a 95% significance level (the χ^2 statistic with two degrees of freedom is 7.51). The restrictions of the pure EL model ($\sigma_\kappa^2 = \sigma_r^2 = 0$) are rejected at any reasonable significance level with a χ^2 of 487. These Wald statistics are derived using the weighting matrix of our objective function, which weights all moments equally.

We can also test the fit of the models using the squared distance of the fitted from the observed moments, using the sampling variation of the observed moments as the weighting matrix. The resulting χ^2 has 60 degrees of freedom for the nested model and 62 degrees of freedom for the pure EL or DPH model. Without a doubt, none of our models fits the data using this statistical criterion. The test statistic for the EL model is 64,926, that for the DPH model is 1,619 and the statistic for the nested model is 2,035.³⁷ In our view, it is not surprising that models with six or eight parameters will be rejected on the basis of fitting 68 moments obtained from a sample with almost 60,000 observations.

Overall, we interpret our estimates as supporting a model that combines elements of EL with DPH.

5 Interpretation

In this section, we interpret the estimates of the nested model. We discuss the implied variation in productivity and wages over the life-cycle and how far productivity and wages can deviate from each other at different ages 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.

³⁷To understand why the test statistic for the nested model exceeds that for the pure model, remember that this test statistic is not based on the criterion function for estimating the parameters. Therefore the test statistic for the nested model is not necessarily smaller than that of the restricted models.

5.1 Productivity and Wage Variance of the Life-Cycle

In this paper we strive to understand how EL and DPH each contribute to rising pay dispersion over the life-cycle. Individual pay is given by the sum of individual productivity and firms' expectation errors about worker productivity. To illustrate the role of productivity evolution and imperfect learning, figure 8 presents the variances of these two components implied by our estimates of the nested model as a function of experience. We should note that this decomposition of the variance in residual log wages needs to be taken with a grain of salt. The experience profile in the variance in log wages is a set of moments which we have particular difficulty pinning down.

Figure 8: 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 wages and productivity are quite similar (0.174 and 0.154, respectively). Clearly, the shape and magnitude of the variance of log wages over the life-cycle 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.

Expectation error accounts for the difference between wages and productivity. During the first few years in the labor market, the variance in the expectation error declines as firms learn about differences in initial productivity, q_{i0} , and the persistent component of productivity growth, κ_i . After a few years however, the variation in the expectation error stabilizes 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 that thus imperfect learning is of small consequence, we would disagree. The implied standard deviation for the expectation error is about 0.15 even late in the worker's career. This means that the average expectation error is about 10% of annual productivity for most of the life-cycle. Even for experienced workers, firms make sizable errors when estimating productivity and face substantial incentives to learn about how productive their workers are. It is plausible that worker turnover and human resource policies are substantially shaped by employer learning.

We conclude that the increase in the variance of wages is largely sustained by heterogeneous productivity evolution rather than by learning about productivity differences among younger workers. At the same time, firms continue to make sizable expectation errors about the productivity of even seasoned workers. Worker productivity is not an immutable constant. Instead, firms continue to learn about how the skills of workers evolve, even late in their careers: They try to hit a moving target.

5.2 Incomplete Learning and the Returns to Investment

We are now in a position to answer a simple, yet fundamental question: If individual productivity at experience t increases by 1%, 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.

For a time, any increase in productivity will only partially be priced into wages. Eventually, wages catch up with productivity and only then will individuals fully benefit from any changes in their skills. As workers age, the period during which wages fully reflect productivity shortens relative to the time that wages only partially reflect productivity, so that a smaller fraction of any productivity change accrues to older individuals. This effect occurs in addition to the well-known horizon effect in optimal investment decisions: that older workers will invest less because they have less time to reap the benefits of investments. 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 and the discount rate individuals face.

In Table 5 we present estimates of the fraction of a productivity increase that accrues to individuals at different points of the life-cycle. We base these estimates on the parameter estimates for the nested model as well as a range of discount rates varying from 3-10%. For all of these estimates we assume that individuals work for 40 years. For various points over the life-cycle, we report how much the present discounted value of earnings changes relative to the present discounted value of productivity in response to a permanent, one unit increase in productivity. These estimates, while admittedly rough, provide an indication of how important learning and incomplete information can be for understanding investment patterns along the life-cycle.

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

Regardless of the discount rate considered, we find that the share of any productivity increase going to workers is greatest prior to entering the labor market. This is because firms receive fairly precise signals about initial productivity differences (σ_0^2 is 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 rapidly. With a discount rate of 5%, we observe that during the first 10 years about 75% of the returns are captured by workers. This percentage declines to about 65% after 20 years, 40% after 30 years and only about 25% after 35 years of experience.

These estimates suggest that incomplete learning by employers can generate large gaps

between the private and the social returns of human capital investments for older workers. In contrast, these gaps are relatively small for young workers. For younger workers, our results are consistent with Lange (2007) who finds that initial expectation errors about productivity differences existing at the beginning of individual careers decline by about half in the first 3 years and 75% during the first 8 years.³⁸ Our parameter estimates imply that expectation errors about productivity differences existing at the beginning of individual careers decline by about one third within 3 years and 70% within 8 years. Thus, our estimates about the speed of learning about initial productivity differences are strikingly consistent with those of Lange, despite the differences in methodologies.³⁹ 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 suggest that incomplete learning can severely mis-align incentives late in individuals careers. As evident from Table 5, incomplete learning generates the largest gaps between the private and social returns to investing into human capital among the most experienced workers. Of course, these results are obtained in the context of a model with exogenous productivity, no private information, and the absence of strategic behavior on the part of workers and are therefore only suggestive. However, our estimates suggest that EL can have important implications for behavior of older workers. This is in sharp contrast to the existing literature, which has focused almost exclusively on young workers.

6 Discussion of Alternative Explanations

Much of our discussion above has focused on the finding that past performance correlates more highly with pay than does future performance, even at high experience levels and that the correlation of pay with performance continues to increase over the life-cycle. We rationalized these findings by concluding that productivity evolves heterogeneously throughout the life-cycle and firms continue to learn about this moving target. Here we consider alternative explanations for these findings. We focus on three separate categories: measurement issues, the spot markets assumption, and attrition.

³⁸Lange (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.

³⁹Firms 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_{i0} will have a more than one-for-one impact on log wages for part of the individuals life-cycle.

6.1 Measurement Issues

In order to remain parsimonious, yet still extract meaningful information from the subjective performance evaluations, we have placed some degree of discipline on their structure. In particular, we assume that the signal value in the performance evaluations is constant over the life-cycle, and that the performance ratings are correlates of current productivity, rather than statements on how workers performed relative to expectations. We discuss these two assumptions here starting with the latter.

We interpret performance ratings as correlates of overall productivity relative to a peer group. An alternative interpretation is that they measure whether the worker met, exceeded, or fell short of expectations during a given time-period. We did not adopt this interpretation because it is inconsistent with a few simple facts in the data. Specifically, we showed above that past performance measures and past wages are highly predictive of future performance measures. The ability to predict performance measures using past variables implies that these cannot simply reflect deviations from expected performance, because deviations from expected performance are necessarily uncorrelated with variables that are available when expectations are formed.

Another worry is that our model is mis-specified because the signal quality might vary over the life-cycle rather than remain constant. However, it is unlikely that allowing for such life-cycle variation in signal quality would overturn our main conclusion that EL and DPH are jointly present. Consider for instance the possibility that performance evaluations become more precise as workers age. This would lead firms to rely more heavily on performance evaluations obtained later in life. Nevertheless, this extension would not suffice to reconcile the pure EL model with the data. The data reveals fast learning even among young workers so that almost all information about workers in a pure EL model is revealed early on regardless of the quality of signals at older ages. Thus, even with high quality signals late in life, we would not be able to reconcile the pure EL model with the observed patterns in the data. Similarly, declines in the signal quality with age are likewise inconsistent with the pure EL model, since this would imply that learning would be absent among older workers. Again this is at odds with our data.

6.2 The spot market assumption

We assume throughout that workers are paid their expected product. This assumption bought us a tractable model and allowed us to compare our results to the prior literatures on EL and DPH which likewise assume spot markets. However, a number of realistic models violate this assumption. In this subsection, we discuss how incentives, and imperfect competition might alter our results. Due to the complexity of models that incorporate such assumptions, the discussion remains necessarily informal.

Incentives

Incentives in firms are provided in various forms. One possibility is that pay is directly linked to current performance measures. Such a direct link of pay with current performance will induce current pay to be more highly correlated with current performance measures than with those in the past or the future. Our data however does not show that the correlations of pay with current performance exceed those in the recent past. The data therefore does not support direct pay for performance.⁴⁰

Another possibility is that firms link pay raises to performance in an effort to elicit higher effort. Tournament models (Lazear and Rosen 1981) likewise provide incentives by linking future pay to performance measures via promotions. These forms of deferred compensation are difficult to empirically distinguish from EL using the pay-performance correlations that we have stressed in this paper. One reason is that linking pay to expected productivity in the way prescribed by the EL model does itself generate incentives for providing effort if effort and skills are difficult to distinguish (as in the career concerns model of Gibbons and Murphy 1992). Fundamentally, models of moral hazard and incentive pay as well as models of employer learning are both based on incomplete information on the part of firms, making it extremely challenging to cleanly separate and identify the role of both.

While we cannot rule out such deferred compensation models, we do note that our results are robust various ways of treating job levels.⁴¹ This somewhat allays concerns that a tournament model of promotions and associated pay rises drives our results. However, lacking more information on the structure of pay setting and promotions, 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 testing and separating models of incentive provision through deferred compensation from models of employer learning.

Imperfect Competition

In non-competitive labor markets, firms will try to extract informational rents from their employees. We conjecture that this will result in a less sharp discontinuity (“step”) in the correlations between wages and performance measures at various leads and lags because markets do not force firms to price new information into wages immediately. For example, after a firm receives positive information about a worker, it might raise wages only when the worker receives an outside offer. In models with search frictions outside offers arrive only infrequently and the discontinuity will therefore be less sharp. We would also expect the

⁴⁰Note that our pay data does not include bonuses. Frederiksen and Lange (2013) examine the bonus data available between 1981 and 1988 and find some support that bonuses might be used to set direct incentives. However, because bonuses are a small fraction of total compensation, they find little role for direct incentives in overall pay.

⁴¹We can residualize pay and performance on job level, make performance measure relative to other workers in the same level, and restrict the data to workers who do not change job level, all producing similar results.

discontinuity to be less sharp if the firm had contracts with limited commitment (a la Harris and Holmstrom 1982), since there also, innovations to productivity or new information would take time to be priced in.

These frictions might provide a rational for why wage growth is correlated over time - and correlated more highly at short leads. For example, models where incumbent firms hold an information advantage imply that wages only reflect new positive information about their employees after this information has been learned by outside firms. If this is the case, then wage growth in response to positive information should be correlated over time - and to be correlated more highly at short leads and lags (as we observe in the data). Furthermore, such frictions could result in pay being correlated somewhat more highly with performance at a few lags rather than with the most recent performance measure. Labor market friction therefore might rationalize the observed dip in the correlation patterns in figure 3b for $cor(w_{it}, p_{it-1})$.

6.3 Selective Attrition

Attrition is an obvious concern when analyzing data on workers at a single firm. Our exercise relies heavily on the correlations of pay with lags and leads of performance, thus we must necessarily restrict our attention to the subset of workers who survive at the firm for several years. If attrition is systematically related to observable or unobservable characteristics of the workforce, which is likely to be the case, then our results will be biased. The crucial question is then how severely selected the survivor sample is? We provide a detailed discussion of attrition in a web appendix in which we examine both how turnover is related to observables and how robust our estimates are to non-random selection of survivors. Here we briefly summarize our main findings. In the end, we conclude that any bias generated by selective attrition is likely to be very small.

First, as noted by BGH, quits at this firm are relatively rare, with roughly 10% annual attrition. In our sample, only 6% of new jobs end within a year and 19% end within two years. The corresponding population averages for full-time jobs reported by Farber (1994) over roughly the same time period are 50% and 66%, respectively. Second, as we demonstrate in the web appendix, the relationship between attrition and observable characteristics is relatively weak. Thus selective attrition will not generate substantial changes in the distribution of observables over time.

To illustrate this point, we compare the joint distribution of wage and performance deciles across two groups of workers: a base sample and the 54% of the base sample who will eventually survive for at least 10 years.⁴² We can thus compare the eventual survivors in this firm with the full set. The scatter in figure 9 contrasts the share in each overall wage decile-performance decile category in the base sample on the x-axis with the share in the

⁴²Our base sample includes all workers younger than 55 who entered the firm between 1970 and 1978, so they can potentially be observed for at least 10 years.

eventual survivor sample on the y-axis. Absent selective attrition and sampling variation, this scatter should line up on the 45-degree line. With selective attrition, we should expect the shares to move away from the 45 degree line. As figure 9 shows, the distributions are quite similar, with most of the points lining up on or near the 45-degree line.

Figure 9: Wage-Performance Distribution of Survivors and Attritors

In the web appendix, we also perform an attrition-corrected estimation which allows for selection on observables.⁴³ Our results are very similar when applying this correction. This and other evidence detailed in the appendix convinces us that selection on observables does not seem to be a significant problem in our data.

We have also explored the robustness of our estimates to attrition based on unobservables. We do this by simulating data from the parameter estimates of our nested model augmented with a variety of attrition models that might be of concern, including extreme forms. We then compare our estimating moments in these simulated samples to the true moments in the data. For details, we refer the reader to the web appendix, and simply note here that our set of estimating moments is quite robust to various forms of attrition.

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 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, heterogeneous variation in productivity drives most of the observed increase in the variance of wages over the life-cycle. 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). These workers have already been promoted to manager. Thus the market probably had opportunities to learn about these workers before they entered our sample, and these workers probably had an opportunity to accumulate skills heterogeneously. In the future, we hope to analyze other data sets containing pay and performance measures to establish how generalizable these findings are.

⁴³We use the parameters from the nested model to simulate a data set 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, we simulate a history of wages and performance ratings under the assumption that no worker attrites. We then apply a selection rule based on the empirical relationship between wages, performance, and attriting, and thus obtain a selected sample. Using this selected, simulated sample, we generate our estimating moments. We can then estimate the parameters by minimizing the distance between the observed and simulated moments in the same manner as before.

We believe that this paper contributes to the literature on the influences of worker's careers 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 contains 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 particularly important early in the life-cycle. To the contrary, our estimates imply that the 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" *AEJ: Applied Microeconomics*, 2(4): 76-104.
- [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, Chun, and Yijang Wang. 1996. "Human Capital Investment Under Asymmetric Information: The Pigovian Conjecture Revisited." *Journal of Labor Economics* 14(3): 555-570.
- [12] DeVaro, Jed, and Michael Waldman. 2012. "The Signaling Role of Promotions: Further Theory and Empirical Evidence." *Journal of Labor Economics* 30(1): 91-147.
- [13] Farber, H.S. (1994): "The Analysis of Inter-Firm Worker Mobility." *Journal of Labor Economics* 12(4): 554-593.
- [14] Farber, H.S. and R. Gibbons (1996): "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111:1007-1047.
- [15] Frederiksen, A., F. Lange, and B. Kriechel (2013): "Subjective performance evaluations and employee careers", mimeo.
- [16] Gibbons, R. and L. Katz (1991): "Layoffs and Lemons," *Journal of Labor Economics*, 9:351-380.
- [17] Gibbons, R. and M. Waldman (1999): "A Theory of Wage and Promotion Dynamics inside Firms," *Quarterly Journal of Economics*, 114:1321-1358.
- [18] Gibbons, R. and M. Waldman (2006): "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," *Journal of Labor Economics*, 24: 59-107.
- [19] Gibbons, R. and K. Murphy (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 100(3): 468-505.

- [20] Gibbs, M. (1995): "Incentive compensation in a corporate hierarchy," *Journal of Accounting and Economics*, 19: 247-277.
- [21] Gibbs, M. and W. Hendricks (2004): "Do Formal Salary Systems Really Matter?" *Industrial and Labor Relations Review*, October.
- [22] Guvenen, F. (2007): "Learning your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review*, 97(3): 687-712.
- [23] Habermalz, S. (2011): "Employer Learning and Rational Inattention", manuscript available at "<http://faculty.wcas.northwestern.edu/~sha562/files/SEDPaper.pdf>"
- [24] Hamilton, T (1994): *Time Series Analysis*. Princeton University Press: Princeton, NJ.
- [25] Harris, M. and B. Holmstrom (1982): "A Theory of Wage Dynamics," *Review of Economic Studies*, 49(3): 315-333.
- [26] Hause, J. (1980): "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica*, 48: 1013-29.
- [27] 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.
- [28] Kahn, L. (2010): "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy," *Labour Economics*, Vol. 17(2): 303-316.
- [29] Kahn, L. (2013): "Asymmetric Information between Employers," *American Economic Review: Applied*, 5(4) forthcoming.

- [30] Katz, Eliakim, and Adrian Ziderman. 1990. "Investment in General Training: The Role of Information and Labour Mobility." *Economic Journal* 100(403): 1147-1158.
- [31] Lange, F. (2007): "The Speed of Employer Learning." *Journal of Labor Economics*, 25: 1-35.
- [32] Lazear, E. and S. Rosen (1981): "Rank-order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5): 841-864.
- [33] 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.
- [34] Medoff, J. and K. Abraham (1980): "Experience, Performance and Earnings," *Quarterly Journal of Economics*, 95: 703-736.
- [35] Medoff, J. and K. Abraham (1981): "Are those paid more really more productive? The case of experience," *Journal of Human Resources*, 16: 186-216.
- [36] Mincer, Jacob. (1958). *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.
- [37] Murphy, K. J. (1991): "Merck & Co., Inc. (A), (B), & (C)," Harvard Business School Press, Boston, MA.
- [38] Neal, D. and S. Rosen (2000): "Theories of Labor Earnings," in the *Handbook of Income Distribution*, edited by A. B. Atkinson and F. Bourguignon, North Holland.
- [39] Oreopoulos, P., T. von Wachter and A. Heisz (2012), "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the

Market for College Graduates," *American Economic Journal: Applied* 4(1): 1-29.

[40] Pastorino, E. (2013): "Careers in Firms: Estimating a Model of Job Assignment, Learning, and Human Capital Acquisition", mimeo.

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

[42] Smeets, V., M. Waldman, and F. Warzynski (2013), "Performance, Career Dynamics, and Span of Control," mimeo.

[43] Waldman, D. and B. Avolio (1986), "A Meta-Analysis of Age Differences in Job Performance", *Journal of Applied Psychology*, LXXI: 33-38.

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

[45] Waldman, Michael. 1990. "Up-or-Out Contracts: A Signaling Perspective." *Journal of Labor Economics* 8(2): 230-250.

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 $(n_q \times 1)$ -vector of productivity parameters $\theta_{i,0}$ with $E[\theta_{i,0}] = 0$ and $E[\theta_{i,0}\theta'_{i,0}] = P_0$. In subsequent periods, productivity evolves according to a stochastic process represented by the stochastic difference equation:

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

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 $\theta_{i,0}$. This signal is summarized by an initial $(n_z \times 1)$ 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.

$$\begin{aligned}z_{i,0} &= H_0'\theta_{i,0} + \varepsilon_{i,0}^z \\ \varepsilon_{i,0}^z &\sim N(0, R_{z,0})\end{aligned}\tag{2}$$

The dimensions of $(H_0, \varepsilon_{i,0}^z, R_{z,0}, P_0)$ 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 $\theta_{i,0}$:

$$\begin{aligned}\widehat{\theta}_{i,0|0} &= P_0 H_0 (H_0' P_0 H_0 + R_{z,0})^{-1} z_{i,0} \\ &= K_z z_{i,0}\end{aligned}\tag{3}$$

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 (1). 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:

$$P_{1|0} = \Phi (P_0 - K_z H'_0 P_0) \Phi' + R_\theta$$

I.3 The Recursion

At the end of each period $t > 0$, a new $(n_x \times 1)$ -signal vector x_{it} is drawn by the firm.

$$\begin{aligned}x_{i,t} &= H'_x \theta_{i,t} + \varepsilon_{i,t}^x \\ \varepsilon_{i,t}^x &\sim N(0, R_x)\end{aligned}\tag{4}$$

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

$$\begin{aligned}\widehat{\theta}_{it|t} &= \widehat{\theta}_{i,t|t-1} + P_{t|t-1} H_x (H'_x P_{t|t-1} H_x + R_x)^{-1} (x_{i,t} - H'_x \widehat{\theta}_{i,t|t-1}) \\ &= \widehat{\theta}_{i,t|t-1} + K_t (x_{i,t} - H'_x \widehat{\theta}_{i,t|t-1}) \\ &= (1 - K_t H'_x) \widehat{\theta}_{i,t|t-1} + K_t x_{i,t}\end{aligned}\tag{5}$$

Again, when firms form expectations they account for the evolution in productivity described in equation (1). Therefore firms best guess about productivity in period

$t + 1$ is:

$$\begin{aligned}\widehat{\theta}_{it+1|t} &= \Phi \widehat{\theta}_{it|t} \\ &= \Phi (1 - K_t H'_x) \widehat{\theta}_{it|t-1} + \Phi K_t x_{it}\end{aligned}\tag{6}$$

The variance of the expectation error then evolves according to

$$P_{t+1|t} = \Phi (P_{t|t-1} - K_t H'_x P_{t|t-1}) \Phi' + R_\theta\tag{7}$$

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: $W_{it}^* = E [Q(x, t) Q_{it} | I^t] = E [Q(x, t) \exp(q_{it}) | I^t]$. 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:

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

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

$$\begin{aligned} w_{it} &= \left(q(x, t) + \frac{1}{2}v(t)\right) + \widehat{q}_{it} + \omega_{it} \\ &= h(x, t) + \widehat{q}_{it} + \omega_{it} \end{aligned} \tag{8}$$

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 specification. 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 (9). Second, we need to specify how the states map into observed variables. This measurement equation is given by (10).

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} .

$$\xi_{it+1} = F_t \xi_{it} + v_{it+1} \quad (9)$$

$$y_{it} = M \xi_{it} + e_{it} \quad (10)$$

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

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

$$F_t = \begin{pmatrix} \Phi (1 - K_t H'_x) & \Phi K_t H'_x \\ 0 & \Phi \end{pmatrix}$$

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 (3) and (5) 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:

$$E \left[y_{it} y'_{i\tau \geq t} \right] = M E \left[\xi_{it} \xi'_{i\tau} \right] M' + E \left[e_{it} e'_{i\tau} \right] \quad (11)$$

The M are deterministic and we therefore just have 2 components $E[\xi_{it}\xi'_{i\tau}]$, and $E[e_{it}e'_{i\tau}]$ that need to be determined as functions of the parameters of the model. The matrix $E[e_{it}e'_{i\tau}]$ 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 $E[\xi_{it}\xi'_{i\tau}]$ is related to the parameters.

Tedious, but straightforward algebra yields

$$E[\xi_{it}\xi'_{i\tau}] = \sum_{j=2}^{j=t} \left\{ \left(\prod_{l=j}^{l=t-1} F_l \right) E[v_{i,j}v'_{i,j}] \left(\prod_{l=j}^{l=\tau-1} F_l \right)' \right\} + \left(\prod_{l=1}^{l=t-1} F_l \right) E[\xi_{i1}\xi'_{i1}] \left(\prod_{l=1}^{l=\tau-1} F_l \right)' \quad (12)$$

where

$$E[\xi_{i1}\xi'_{i1}] = \begin{pmatrix} \Phi K_z (H'_0 P_0 H_0 + R_z) K'_z \Phi' & \Phi K_z H'_0 P_0 \Phi' \\ \Phi P_0 H_0 K'_z \Phi' & \Phi P_0 \Phi' + R_\theta \end{pmatrix} \quad (13)$$

and

$$E[v_{i,j}v'_{i,j}] = E \begin{pmatrix} \Phi K_{j-1} R_x K'_{j-1} \Phi' & 0 \\ 0 & R_\theta \end{pmatrix} \quad (14)$$

We have thus shown how to generate $E[y_t y_\tau]$ 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 (1) and any normal learning model that follows equations (2) and (4).

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 (11) together with equations (12), (13), and (14) to generate the covariance matrices of the wage residuals and performance ratings.

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

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

Note here that we let ε_{it}^p (the noise term in p_{it} , which is governed by the auto-correlation term, ρ and an iid error term) 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

$$P_0 = \begin{pmatrix} \sigma_q^2 & 0 & 0 \\ 0 & \sigma_\kappa^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$H_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

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

$$R_{z,0} = \sigma_0^2$$

$$R_x = \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Phi = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{pmatrix}$$

$$R_\theta = \begin{pmatrix} \sigma_r^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 \end{pmatrix}$$

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 EL and DPH models using second moments of wages and performance signals.¹ 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 Employer Learning Model - Identification

The pure EL 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: $(\sigma_q^2, \sigma_0^2, \sigma_u^2, \sigma_\omega^2, \rho, \sigma_z^2)$.

The pure EL 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 $(\sigma_\omega^2, \sigma_q^2)$ from $\lim_{t \rightarrow \infty} (v(w_t)) = \sigma_q^2 + \sigma_\omega^2$ and $\lim_{t \rightarrow \infty} (cov(w_t, w_{t+1})) = \sigma_q^2$.

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 $\lim_{t \rightarrow \infty} v(p_{it}) = \lim_{t \rightarrow \infty} v(q_{it} + \varepsilon_{it}^p) = \sigma_q^2 + \rho \frac{\sigma_u^2}{1-\rho^2}$ and that $cov(p_{it}, p_{it+k}) = cov(q_i + \varepsilon_{it}^p, q_i + \rho^k \varepsilon_{it}^p + \sum_{j=1}^k u_{it+j}) = \sigma_q^2 + \rho^k var(\varepsilon_{it}^p)$. Combining,

¹As 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

$$\lim_{t \rightarrow \infty} \lim_{k \rightarrow \infty} \text{cor}(p_{it}, p_{it+k}) = \frac{\sigma_q^2}{\sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2}} \quad (15)$$

$$\lim_{t \rightarrow \infty} \text{cor}(p_{it}, p_{it+1}) = \frac{\sigma_q^2 + \rho \frac{\sigma_u^2}{1-\rho^2}}{\sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2}} \quad (16)$$

Since σ_q^2 is already identified, we get $\frac{\sigma_u^2}{1-\rho^2}$ from equation (15) and ρ from equation (16).

This leaves only two parameters (σ_z^2, σ_0^2) that need to be identified. σ_0^2 determines how much information employers have 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 $\text{var}(w_{0i}) = \text{var}(E[q_i|z_{i0}])$. 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 $\text{var}(w_{1i})$, having already identified the other parameters of the learning model.

II.2 The Pure Dynamic Productivity Heterogeneity Model - Identification

The pure DPH 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: $(\sigma_q^2, \sigma_r^2, \sigma_u^2, \sigma_\omega^2, \sigma_\kappa^2, \rho)$.

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

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

$$corr(p_{it}, w_{it}) = \frac{var(q_{it})}{(var(q_{it}) + var(\varepsilon_{it}^p))^{1/2} (var(q_{it}) + \sigma_\omega^2)^{1/2}} \quad (17)$$

Since all the productivity parameters are identified, we can treat $var(q_{it})$ and σ_ω^2 as known. Thus, eq (17) solves for the variance of the signal noise $var(\varepsilon_{it}^p)$ for arbitrary t :

$$\lim_{t \rightarrow \infty} var(\varepsilon_{it}^p) = \frac{\sigma_u^2}{1 - \rho^2} \Rightarrow \sigma_u^2 = (1 - \rho^2) \lim_{t \rightarrow \infty} var(\varepsilon_{it}^p) \quad (18)$$

Since we know the $var(\varepsilon_{it}^p)$ for arbitrary t , we can exploit equation (10) in the text to get

$$\rho^2 = \frac{var(\varepsilon_{it+1}^p) - \lim_{t \rightarrow \infty} var(\varepsilon_{it}^p)}{var(\varepsilon_{it}^p) - \lim_{t \rightarrow \infty} var(\varepsilon_{it}^p)} \quad (19)$$

These last two equations therefore deliver the parameters ρ and σ_u^2 . We have thus established the identification of both the pure EL and DPH models.

Figure 1: Simulated Covariances bw Pay and Performance

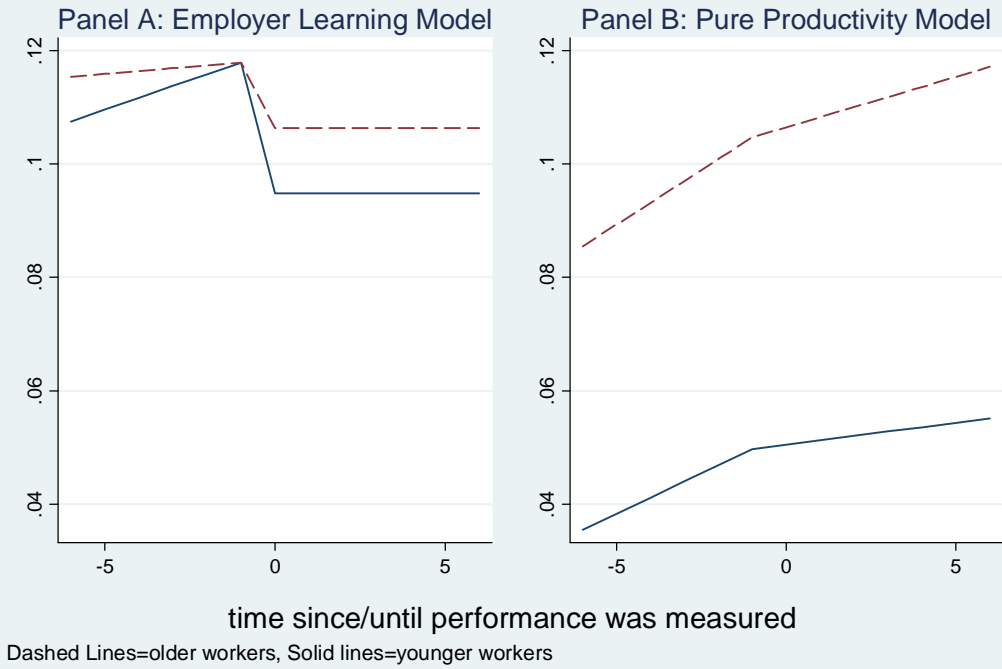


Figure 2: Log Wages and Performance, by Age
Controlling for education, race, gender, and year effects

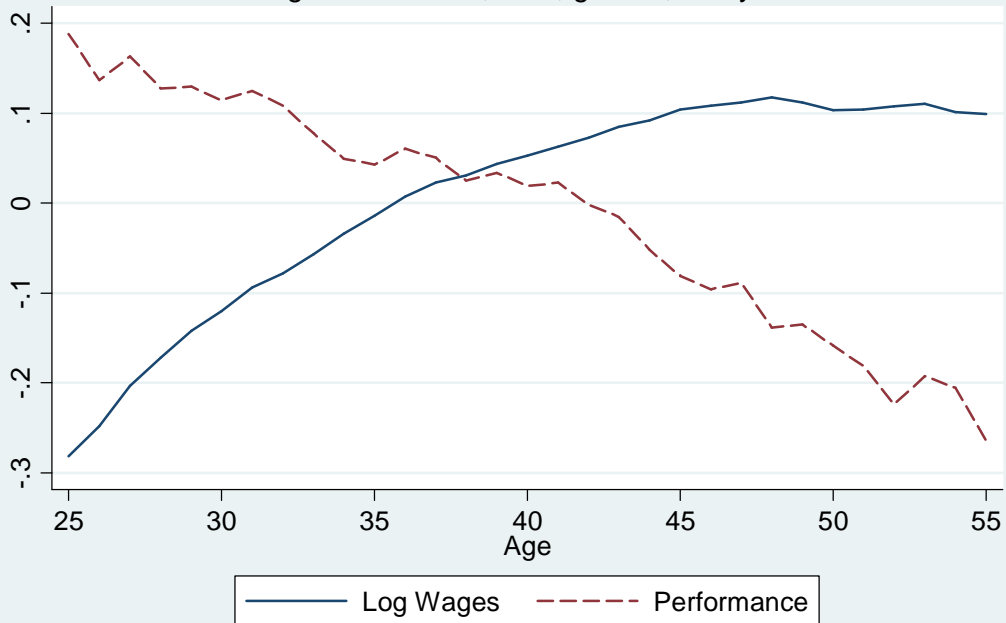
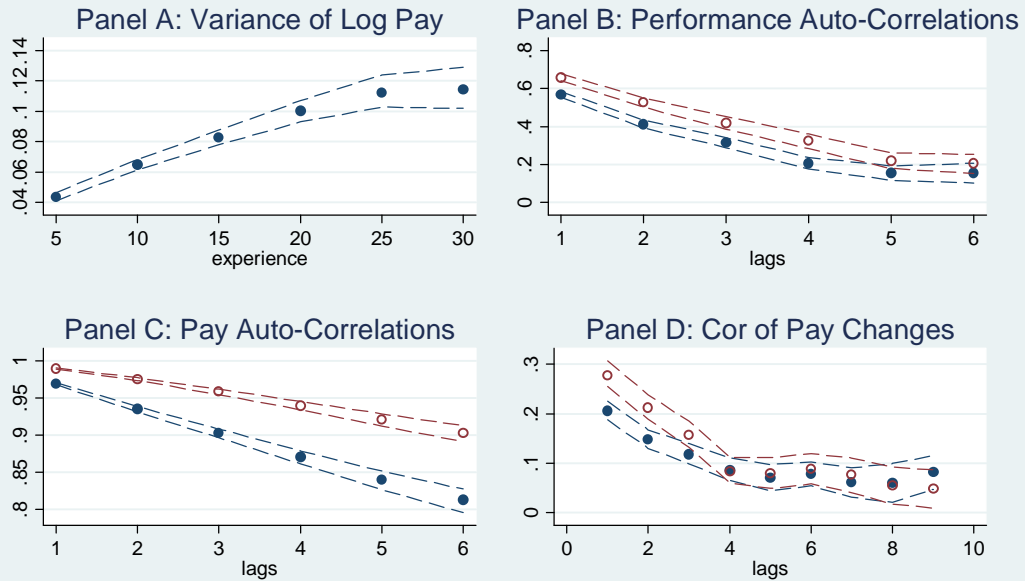
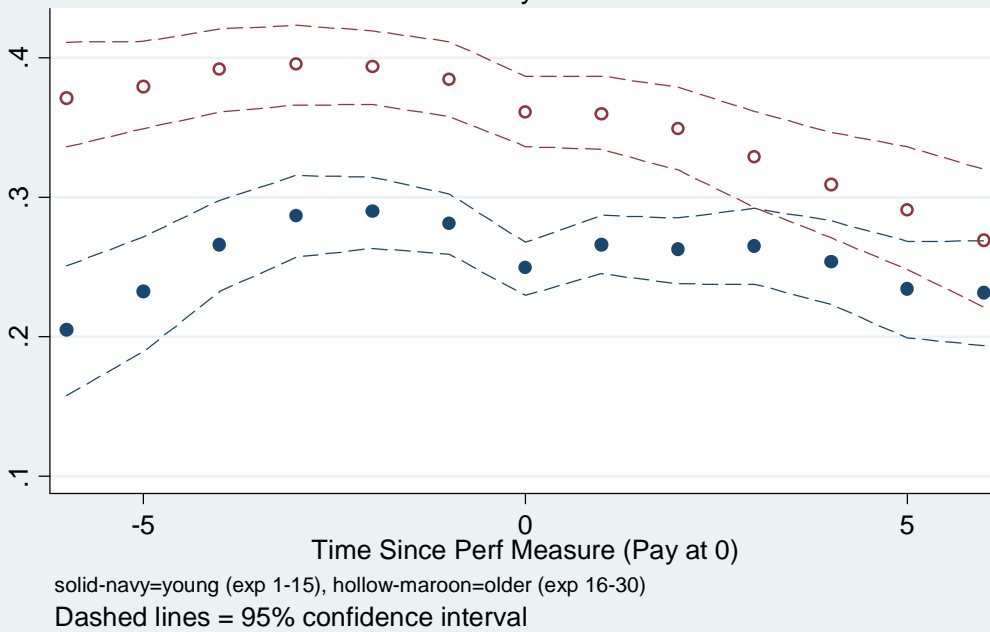


Figure 3a: Moments with 95% CI



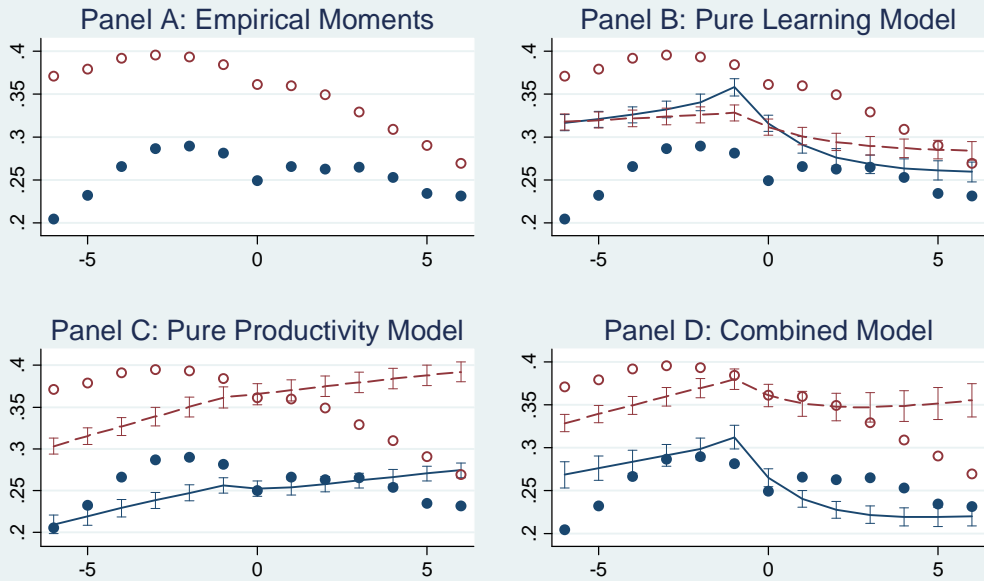
In panels B-D: solid-navy=young (exp 1-15), hollow-maroon=older (exp 16-30)
Dashed lines = 95% confidence interval

Figure 3b: Moments with 95% CI
Cor of Pay and Perf



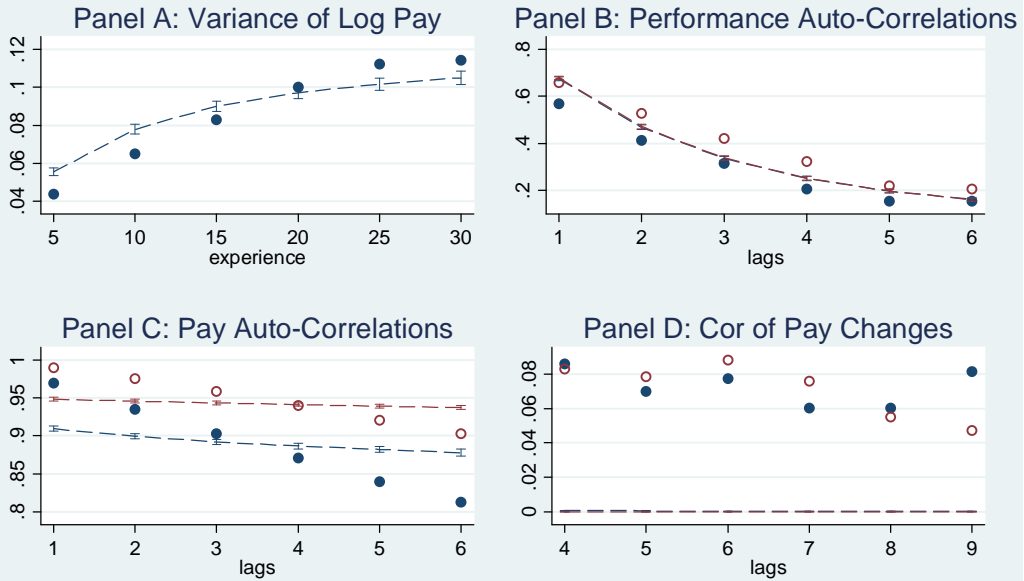
solid-navy=young (exp 1-15), hollow-maroon=older (exp 16-30)
Dashed lines = 95% confidence interval

Figure 4: Results - Correlations of Pay and Performance



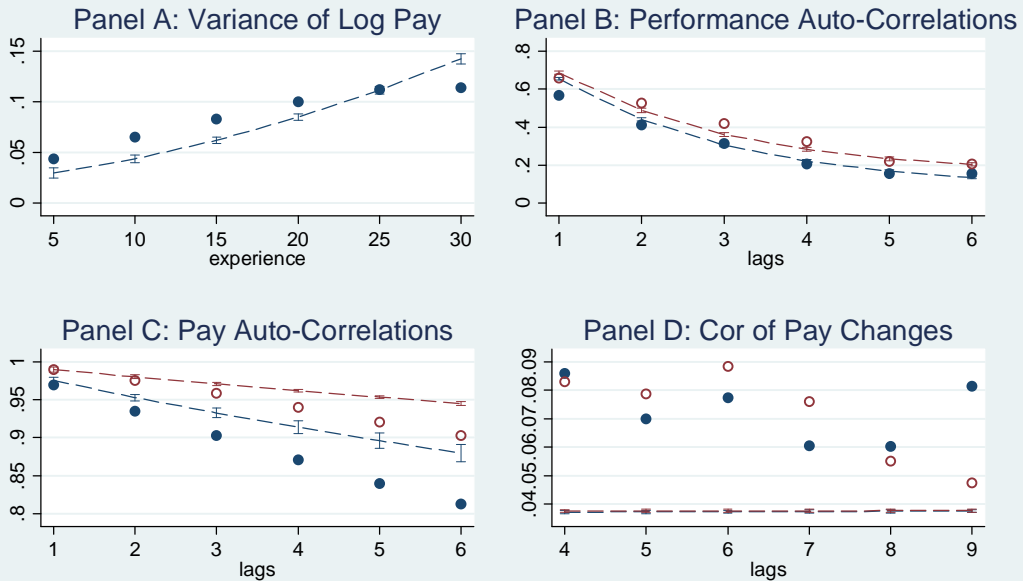
Dots=data, Lines=moments implied by model estimates, Standard Error Bars generated from 500 bootstraps. solid-navy=young (exp 1-15), hollow-maroon=older (exp 16-30)

Figure 5: Results - Pure Learning Model



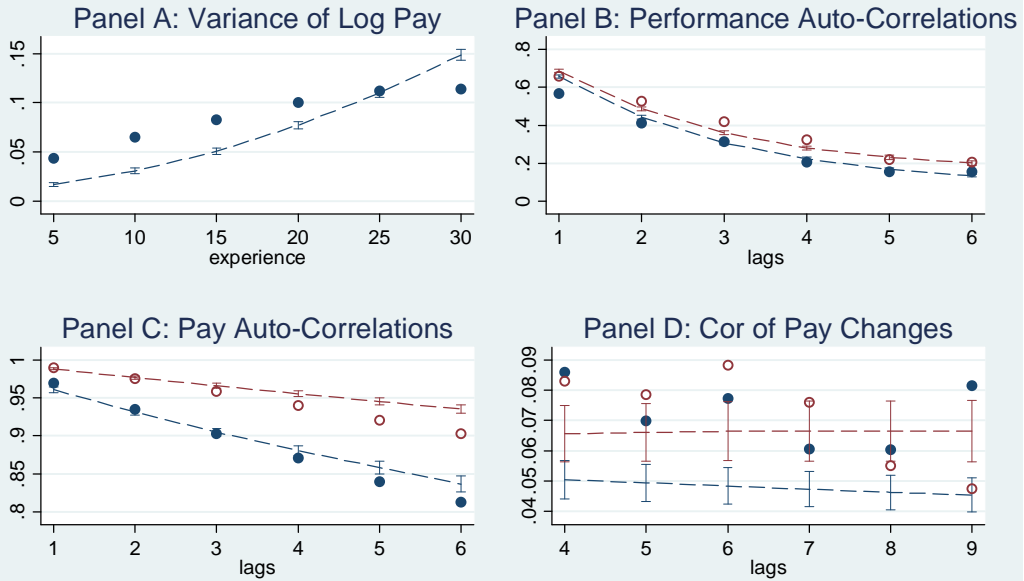
Dots=data, Lines=moments implied by model estimates, standard error bars generated from 500 bootstraps. solid-navy=young (exp 1-15), hollow-maroon=older (exp 16-30)

Figure 6: Results - Pure Productivity Model



Dots=data, Lines=moments implied by model estimates, standard error bars generated from 500 bootstraps. solid-navy=young (exp 1-15), hollow-maroon=older (exp 16-30)

Figure 7: Results - Combined Model



Dots=data, Lines=moments implied by model estimates, standard error bars generated from 500 bootstraps. solid-navy=young (exp 1-15), hollow-maroon=older (exp 16-30)

Figure 8: Productivity, Wage, and Error Variances
Full Model

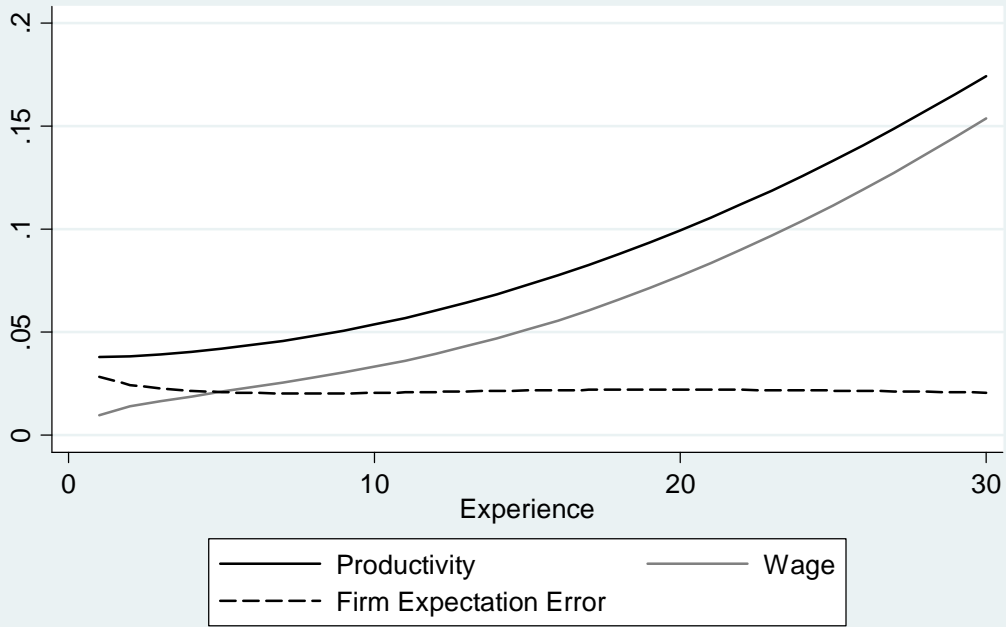
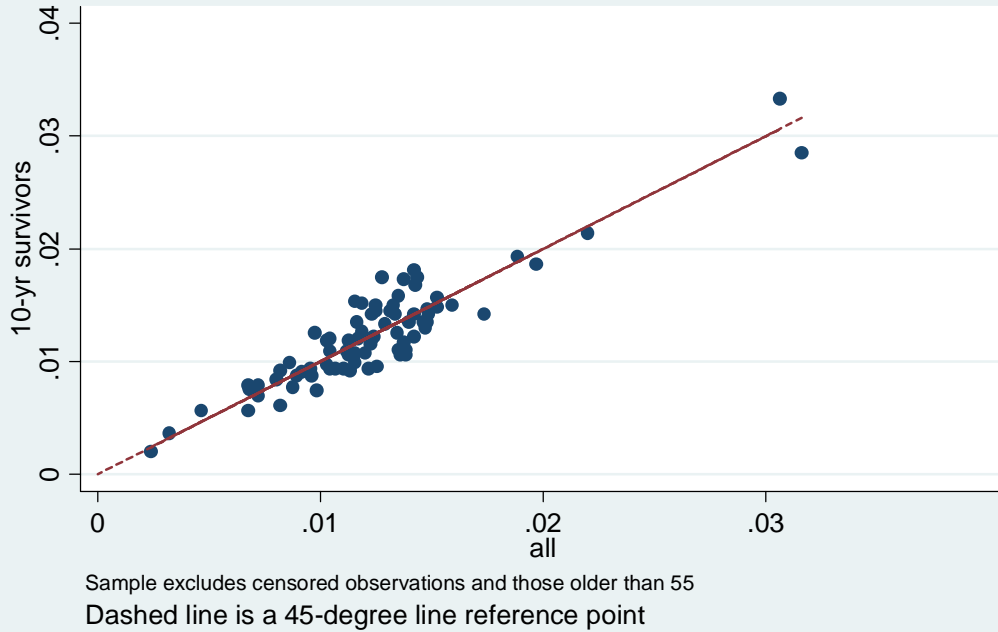
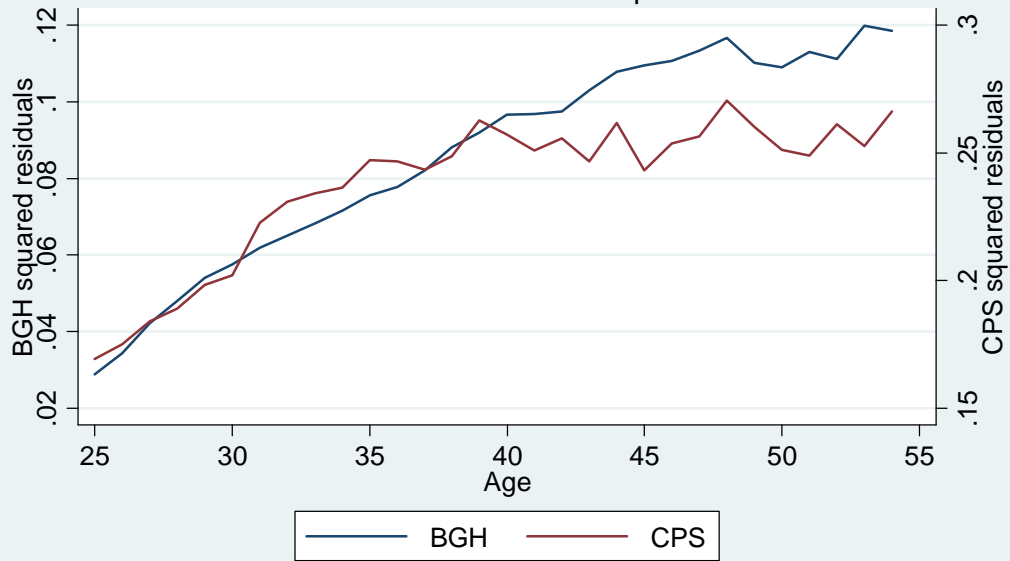


Figure 9: Performance-Wage Distribution in Base Year
All Workers and 10-Year Survivors



Appendix Figure A1: Residual Wage Variances
BGH and CPS March Comparison



CPS March Supplement 1970-1989: Full-time, full-year, private sector income of HS graduates
CPS weighted to BGH age-education-gender-race distribution

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	56231
% Male	76.4%
% White	89.6%
Age	39.57 (9.47)
Education	
% HS	17.9%
% Some College	19.8%
% College	36.1%
% Advanced	26.3%
Salary ² (\$1,000s)	\$54.003 (25.562) 3.13
Performance ³	(0.71) [n=36569]
Performance Distribution	
	1 0.008
	2 0.175
	3 0.500
	4 0.317

Notes: Parentheses contain standard deviations.

1. Sample includes all employees who have a pay or performance measure between the ages of 25 and 65 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)

Auto-Correlation 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)

Auto-Correlations 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)

Auto-Correlations in Pay Changes for lags 4-9						
Lags						
Experience	4	5	6	7	8	9
1-15	0.086	0.07	0.077	0.06	0.06	0.081
	(0.011)	(0.013)	(0.013)	(0.015)	(0.020)	(0.017)
16-30	0.083	0.079	0.088	0.076	0.055	0.047
	(0.013)	(0.016)	(0.015)	(0.018)	(0.019)	(0.020)

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

Correlation of Wages in t with lags and leads in performance

		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
Experience 1-15		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)	(0.013)	
		Leads						
		0	1	2	3	4	5	6
Experience 16-30		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)

The table displays the second moments of wages and performance measures that form the basis of the estimation described in the paper. The same moments are displayed in figure 3a and 3b. 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 (0.005)	0.024 (0.010)	0.024 (0.014)	0.001 (0.018)	-0.021 (0.023)	-0.029 (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 (0.004)	0.033 (0.009)	0.046 (0.014)	0.063 (0.017)	0.070 (0.021)	0.080 (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 comparison, 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
productivity parameters:			
σ_q^2 (initial productivity)	0.118 (0.0057)	0.025 (0.0051)	0.037 (0.0072)
σ_r^2 (random productivity innovations)	-	0.0040 (0.00032)	0.00049 (0.00040)
σ_k^2 (heterogeneous growth)	-	0.0000825 (0.0000023)	0.00015 (0.000016)
information parameters:			
σ_0^2 (noise on initial, private signal)	0.383 (0.061)	-	0.114 (0.071)
σ_u^2 (noise on performance measure)	0.650 (0.062)	0.405 (0.031)	0.488 (0.051)
σ_z^2 (noise on repeat, private signals)	0.506 (0.131)	-	0.206 (0.075)
measurement parameters:			
σ_w^2 (measurement error in wages)	0.0049 (0.00021)	0.00030 (0.00048)	2.83e-12 (4.95e-12)
ρ (auto-correlation in performance)	0.645 (0.0084)	0.634 (0.0084)	0.640 (0.009)

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

Experience	Discount Factor R			
	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.