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

Preliminary and Incomplete

Lisa B. Kahn and Fabian Lange†
Yale University

January 8, 2009

Abstract

Two ubiquitous empirical regularities in pay distributions are that the variance of wages increases with experience and innovations in wage residuals have a large, unpredictable component. The leading explanations for these patterns are that over time, either firms learn about worker productivity but productivity remains fixed or workers' productivities themselves evolve heterogeneously. In this paper, we seek to disentangle these two models and place magnitudes on their relative importance. We derive a dynamic model of learning and productivity that nests both models and allows them to coexist. We estimate our model on a 20-year panel of pay and performance measures from a single, large firm (the Baker-Gibbs-Holmstrom data). Incorporating performance measures yields two key innovations. First, the panel structure implies that we have repeat measures of correlates of productivity, as opposed to empirical evidence on employer learning which uses one fixed measure. Second, we can separate productivity from pay, whereas the previous literature on productivity evolution could not.

We find that both models are important in explaining the data. However, the predominant effect is that worker productivity evolves idiosyncratically over time, implying firms must continuously learn about a moving target. Therefore wages differ significantly from individual productivity at all experience levels due to imperfect information, but the majority of pay dispersion is driven by variation in individual productivity. We believe this represents a significant reinterpretation of the empirical literature on employer learning.

*We are grateful for helpful comments from seminar participants at Columbia University, University of Southern California, University of Wisconsin, University of Rochester, Yale University, and Steffen Habermalz. We thank Mike Gibbs and George Baker for providing the data. Doug Norton provided able research assistance.

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

1 Introduction

Understanding how wages evolve as workers age and the reasons for wage dispersion in the population are among the central questions of labor economics. The predominant answers to these questions are based on the idea that wages perfectly reflect the worker's productivity. In recent decades, the literature on employer learning has offered a competing interpretation of how wages evolve as workers accumulate experience. This literature assumes that employers are imperfectly informed about worker productivity but learn as workers age. Changes in wage residuals with experience therefore reflect learning about worker productivity on the part of firms. An important and influential set of papers has used the finding that, as workers gain experience, wages increasingly correlate with variables that are hard to observe by the firm, as evidence of employer learning. In addition, this literature has proven successful in explaining two empirical regularities regarding wage residuals: the variance of wage residuals increases with experience and innovations in wage residuals have a large, unpredictable component.[1] However, these empirical regularities are also consistent with the hypothesis that wages equal productivity at all times but productivity itself is evolving. Indeed, without placing restrictions on the productivity process, one cannot use patterns in residual wage variance to reject the full information model.[2]

The main obstacle in distinguishing employer learning models from full information models is that most data sets do not contain direct measures of individual productivity that allow separating productivity and pay. In this paper, we provide new evidence on whether employer learning or changes in the productivity of workers drive changes in wage residuals over the life cycle. This evidence is based on firm-level data containing wages and performance evaluations.

Our data, a 20-year unbalanced panel of all managerial employees in one firm, were previously analyzed in Baker, Gibbs and Holmstrom (1994a and 1994b). These landmark studies provided early empirical evidence on the internal organization and pay dynamics

[1] These findings are intuitive. In learning models, wages equal expected productivity conditional on the information available at any age. The variance of conditional expectations increases as the conditioning set increases, implying that the variance of wage residuals increases as more information becomes available. Furthermore, because past wages are included in the firm's information set, wage growth will be uncorrelated over time. Finally, productivity measures that are observed in the data but not by the firm will increasingly be reflected in wages as firms learn about productivity itself. For more detail, see Farber and Gibbons 1996, Altonji and Pierret 2001 and Lange 2007.

[2] A separate literature (e.g., Hause 1980, MaCurdy 1982 and Baker 1997) analyzes the correlation in pay and pay changes over time to test for different patterns in the evolution of productivity, positing that pay equals productivity. By analyzing the structure of residuals in pay regressions which control for person-specific time trends, they can learn about the idiosyncratic component of productivity growth. For example, Baker uncovers parameters from an ARMA process. Evidence here is mixed, with correlations in wage growth varying widely.

of the firm.[3] These data have the crucial advantage that they contain both annual pay of workers as well as performance ratings. The panel structure allows us to observe past, current, and future performance ratings, giving us information about worker productivity that the firm was not able to exploit when setting wages. Our main innovation is in analyzing moments not before exploited in the data, correlations between pay and performance lags and leads, and using these moments to test the learning and pure productivity models.

To fully exploit these data, we write down a dynamic model of learning and productivity. In the model, firms set pay equal to expected productivity which they predict using noisy signals of productivity. In addition, worker productivity itself varies stochastically over time. It follows that the variation in wages is partially driven by changes in underlying productivity and partially by noise in the signals obtained by firms.

This model nests both of the competing explanations for how wages vary over the life cycle. This allows us to test pure versions of both models against each other. It also allows us to examine which features of the data are not reproduced by the pure learning or the pure productivity model. Finally, it allows us to estimate the models jointly and examine how the learning and productivity processes interact in setting wages.

In isolation, neither model can fully reproduce the moments of the data. The pure productivity model predicts that there are no major asymmetries of wage correlations with past and future performance measures. Observing that wages are more highly correlated with past rather than future performance ratings therefore leads us to reject the pure productivity model. The pure learning model does generate this asymmetry in the correlations of pay with past and future performance ratings. However, it also predicts that this difference declines with experience. We observe the opposite.

Estimating the full model, we find, quite intuitively, that firms do learn about worker ability and that productivity evolves over time. Somewhat surprisingly, we find that the initial variance in worker ability is quite small and that firms are well informed about the skills of workers at the outset of their careers. Over time, productivity evolves and firms do worse at predicting ability. We find that most of the changes in productivity cannot be predicted by past idiosyncratic productivity growth. Instead, productivity has a large random walk component. The firm must learn about an unpredictably moving target and consequently updates expectations over worker ability, even at high experience levels. This

[3] This work was extremely influential in the field of organizational economics. Their findings have inspired the well known contributions by Gibbons and Waldman (1999 and 2006) who reconcile most of the BGH findings by combining simple models of job (and later task) assignment, human-capital acquisition and learning. In addition, Gibbs (1995) describes the empirical relationship between pay, promotions and performance and DeVaro and Waldman (2007) use the data to test whether promotions signal worker ability to outside firms.

explains why we observe patterns consistent with learning models even at high experience levels. Overall, we find that wages differ significantly from individual productivity at all experience levels. Nevertheless, the majority of the observed dispersion in wage residuals is due to variation in individual productivity.

We believe that this reinterpretation of the role of learning represents a significant contribution to the empirical literature on employer learning. This literature based on the groundbreaking contributions by Farber and Gibbons (1996) and Altonji and Pierret (2001) interprets the employer learning process as uncovering a fixed, idiosyncratic productivity using repeated measures of productivity over time. We propose instead that employers need to continuously learn about a moving target: the productive ability of their workers as it changes over the life cycle.

The remainder of this paper is structured as follows. Section 2 describes the data. Section 3 presents the pure productivity model, the pure learning model, and the nested model. This Section also provides a reduced form evaluation of the pure productivity and pure learning models. In the Appendix, we show that these models are members of a larger class of models of learning. We derive a mapping between this larger class of models and the Second moment matrices of productivity measures and wages that allows estimating such models. In Section 4 we discuss the estimation and identification of the models developed in Section 3. We interpret our results and conclude in Section 5.

2 Data

In this paper, we analyze data first used in the canonical studies of Baker, Gibbs, and Holmstrom (1994a,b) on the internal organization of the firm (hereafter, BGH). The data consist of personnel records for 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 demographics including age, race, gender and education. The original sample contains 16,133 employees. Of these, we restrict attention to the 11,067 employees with a non-missing education variable who can be observed between the ages of 25 and 54. This age window allows us to focus on early years of experience (when employer learning and human capital accumulation should be most important) while still yielding a decent sample size.[4] An employee observation is useful to us if he or she can contribute to at least one comparison of

[4] Age 25 might be considered slightly old to begin the processes of employer learning and post-school skill accumulation for most education groups. However, our sample consists of workers who have already been promoted to the level of manager. As we have no way of learning about their labor market experiences before they enter this sample, we start as early as we can while still having a decent sample size. This is also why we extend the anaylsis to age 54.

the following kinds: an auto-correlation in either pay or performance ratings or a correlation between pay and a performance rating across a time gap of up to 6 years. We do not consider correlations across more than 6 years, because these are often estimated using very few individuals. 9,373 employees and 52,697 employee-years contribute to the moments analyzed in this paper. A given employee-year contributes an average of 8.5 correlations.

Summary statistics are reported in table 1. The sample are primarily white males with at least a college degree. Annual salary is cpi adjusted to 1988 dollars and measures base pay.[5] Workers earn on average $53,400. BGH (1994,b) present a detailed analysis of pay at this firm. They find that pay was higher in the firm, relative to industry average, likely due to this sample being managers. Pay inside the firm did fluctuate with market conditions, but by a smaller magnitude than the industry average. In their analysis, they find evidence of cohort effects, high variation in pay within a job level, serial correlation in pay growth, and a strong relationship between promotions and pay growth. They also find that nominal wage declines were almost nonexistent but real wage declines were common. Their paper does not, however analyze performance measures.[6]

Figure 1 illustrates both how the mean and variance log earnings residuals vary with age. The solid line graphs the log of annual salary by age, controlling for education, race, gender and year fixed effects. As can be seen the earnings profile is rising and concave, reflecting typical life-cycle patterns. The dashed line plots the squared residuals from a log wage regression which controls for the variables listed above as well as age fixed effects. The variance in pay around the age profile is substantial and increases almost linearly with age. It is only after age 45 that we observe a slow down. Understanding this variation and its increase over the life-cycle is the primary task of this paper.[7]

We recode the performance rating such that it ranges from 1 to 4 with higher ratings reflecting better performance.[8] From table 1, we see the average rating is a little over a 3. Less than 1% of workers receive a 1, the worst rating, 16% receive a 2, while half receive a 3 and a third receive a 4. This distribution of performance ratings is similar to those found in Medoff and Abraham (1980,1981) and Murphy (1991) in their studies of performance ratings

[5] 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. 22% of workers receive a bonus in the years 1981-1988. Conditional on receiving a bonus, the amount is on average 12% of base salary.

[6] In a subsequent paper, Gibbs (1995), does make use of these performance measures, in an effort to characterized within-job versus promotion-based incentives inside the firm. He shows that performance measures are correlated with current bonus and probability of promotion.

[7] It is worth noting that these variances are quite a bit lower than one would see in a cross-section (for example, the variance in log earnings residuals is 0.04 at age 25). This is because we are already restricting attention to workers in the same firm and occupation (broadly defined).

[8] In the data, the scale is inverted and ranges from 1 to 5. We combine the worst two ratings, since almost nobody receives the worst rating.

across various industries and firms. Further, Gibbs (1995) shows that these performance measures do contain meaningful information. For example, high performance ratings are correlated with higher raises and bonuses, and increased probability of promotions.[9]

Figure 2 shows the experience-performance profile both with and without worker fixed effects, controlling for education, race, gender in the first specification and year fixed effects in both. Focusing first on the solid line without fixed effects, we see that somewhat surprisingly, performance gets worse with experience. This is unexpected if we think part of the explanation for rising returns to experience is that workers are accumulating more skills However, this is a common finding in the literature. Medoff and Abraham (1980) interpret these performance measures as relative ranks within a comparison group. If ratings are relative then we could see any experience profile. For example, if workers are graded more harshly as they accumulate experience, we would see this negative slope. In our analysis, we follow the common practice in the literature and treat performance ratings as relative within experience groups.

As described above, the performance rating is a categorical ordered variable. We interpret these variables as arising from a latent signal on individual productivity. Equation 1 shows the mapping of the latent productivity signal p_{it}, for an individual, i, with experience t, onto the observed performance rating, $\widetilde{p}_{it}$.

$$\widetilde{p}_{it} = \sum_{j=1}^{j=K-1} \mathbf{1}\left(p_{it} \geq c_j\left(t\right)\right) \tag{1}$$

A worker is assigned the ranking $\widetilde{p}_{it} = k$, if his or her latent productivity signal falls between the two thresholds c_{k-1} and c_k. These thresholds differ across reference groups defined by level of experience, t. We could easily include demographics, such as race, gender and education, in forming these groups, though we have not done so here.[10]

Below, we make a number of assumptions that ensure that the latent signal p_{it} is normally distributed. These assumption allow us to estimate correlations of p_{it} with other normally distributed variables (such as log wage residuals) and with performance measures from other years using maximum likelihood methods. Of course, since the performance ratings are categorical variables without obvious unit, we cannot identify the variance of p_{it}. However, appendix figure 1 shows the distribution of the raw performance measures across age. It

[9] Gibbs finds higher magnitudes for these effects than do Medoff and Abraham (1980,1981) and Murphy (1991). The causes of these discrepencies are unclear, but Gibbs hypothesizes they might be due to the industries studied. Subsequent work by Gibbs and Hendricks (2004) is more consistent with the Gibbs finding.

[10] These may not capture the exact reference group for a worker. A natural group might be job level. However, we did not want to residualize on a variable that is highly correlated with pay and may be the outcome of employer learning.

looks as though they are fairly evenly distributed, with older workers more likely to receive 2's and less likely to receive 4's, relative to younger workers.

Because we have data on only one firm, we may suffer from several selection problems. The primary source of selection, which we can do nothing about, is that all workers enter our sample (i.e., this firm) at some point. We are more concerned with the timing of exit from the sample, since nonrandom turnover could bias our results. To illustrate the problem, we estimate linear probability models of worker exit as a function of pay and performance in the current year.

Appendix table 1 reports these results. As can be seen, workers with better performance are far less likely to leave, even after controlling for salary. Salary itself has a small effect (higher pay implies less exit), which does not remain significant after controlling for performance. This firm appears to be a very good place to work, with relatively low turnover, where poor performance is the only predictor of exit. We discuss the implications of this nonrandom turnover in the conclusion.[11]

3 3 Models of Wage and Productivity Dynamics

We now consider three models about how productivity and information about productivity evolves over time. We begin with a pure productivity model in which firms have full information about workers productivity. In this model, wage changes simply reflect changes in productivity over time. We then turn to a pure employer-learning model, in which productivity is fixed over time. Initially, employers do not know worker productivity but they continuously update their expectations based on noisy signals of productivity. As employers update their expectations, wages evolve. We then present a third model that nests both the pure productivity and pure learning models. As we develop each model, we also derive some implications for the correlations in performance measures and wages. And, we will present some reduced form evidence based on these implications.[12]

A number of features are common to all three models that we discuss in this Section. These properties also apply to the more general class of learning and productivity models

[11] A second form of selection which may bias our results is non-random age of entry. This is because our exercise exploits differences in correlations between pay and performance across age levels. To test for selection on age fo entry, we analyze starting pay and starting performance as a function of entry age. We find no evidence that starting pay and performance are related to entry age, controlling for an age profile. We are therefore not worried about this form of selection.

[12] In the next section, we will present a more general class of models that includes all three models described here and we will show how to estimate the parameters of the more general class of models. This will form the basis of the estimation results presented later in the paper. For now however, we will limit ourselves to presenting the pure learning, the pure productivity model, and the model that nests both of these.

that we discuss in the appendix. Most importantly, we assume that labor markets are spot markets and that information is symmetric across all employers.[13] This implies that wages equal expected productivity in each period. Furthermore, we assume that firms know the structure of the economy and they update their expectations in a Bayesian manner. We make a number of normality assumptions that ensure that we can represent learning by employers using the tools of Kalman filtering. For now, we will also maintain the assumption that we observe in the data a signal of productivity p_{it} that is itself normally distributed around true productivity. As we discussed in the data section, we will allow the observed ordinal performance ratings to map into the signal p_{it}, though this process is described later.

We will generally assume that we can summarize a worker's productivity using a single variable $\widetilde{Q}_{it}$. Worker productivity varies with observed characteristics (x_i) and experience t. Thus, we let $\widetilde{Q}_{it} = Q(x,t) * Q_{i,t}$, where $Q(x,t) = E\left[\widetilde{Q}_{it}|x,t\right]$ and $Q_{i,t} = \widetilde{Q}_{it}/E\left[\widetilde{Q}_{it}|x,t\right]$ is the idiosyncratic component of individual productivity. Regardless what model we consider, the function $Q(x,t)$ is common knowledge, but in some models, the component $Q_{i,t}$ is only partially observed by firms.

Let $q_{it} = \log(Q_{it})$.[14] It is this idiosyncratic component of productivity that firms attempt to estimate. Our interest in this paper is primarily in how this idiosyncractic component is related to wages over the life-cycle.

3.1 Pure Employer Learning

The pure employer-learning model assumes that worker productivity $q_i = \log(Q_i)$ does not evolve over time (even though $Q(x,t)$ might) and that individual wage dynamics arise only because employers learn about worker productivity over time. If labor markets are spot markets and information is symmetric across all employers, then workers are paid their expected productivity in each period.

The flow of information to employers is modeled using three different signals. First, we

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

[14]We will generally follow the notational convention that upper case letters refer to variables measured in levels and lower case letter refer to variables measured in logs.

By construction q_{it} is mean zero and uncorrelated with the controls x. From now on, we will suppress the dependence on x.

allow for the possibility that firms have some knowledge about worker productivity at the beginning of the worker's career. This information is embodied in an initial signal z_{i0} and is not observed in the data. Furthermore, we assume that the firm observes two signals in each time-period: $\{p_{it}, z_{it}\}$. The only signal that is contained in our data is p_{it}.

As is standard in the learning literature, we impose a number of normality assumptions that allow us to exploit the convenient features of normal distributions.[15] In particular, we assume that log productivity q_i is distributed normally in the population with mean 0, standard deviation σ_q^2. We assume that all three signals are normally distributed around q_i and therefore have $z_{i0} = q_i + \varepsilon_{i0}$, $p_{it} = q_i + \varepsilon_{it}^p$, $z_{it} = q_i + \varepsilon_{it}^z$ where $\varepsilon_{i0} \tilde{} N\left(0, \sigma_0^2\right)$, $\varepsilon_{it}^p \sim N\left(0, \sigma_p^2\right)$, and $\varepsilon_{it}^z \tilde{} N(0, \sigma_z^2)$. Without loss of generality, we have imposed that $cov\left(\varepsilon_{it}^z, \varepsilon_{it}^p\right) = 0$.[16] Finally, all signals are assumed to reflect new information, i.e., the signal errors are uncorrelated across time. In summary, we face a standard normal signal extraction problem with three types of signals: an initial signal z_{i0} and a dynamic signal z_{it}, both of which are observed by firms but not in the data and a signal p_{it} that is observed both by employers and in the data.

Equation (2) shows the equilibrium log wage implied by this signal extraction problem, where I^t denotes information the firm has received up to time t.

$$w_{it}^* = E\left[q_i | I^t\right] = \chi_t + (1 - K_{t-1}) * E\left[q_i | z_{i0}\right] \tag{2}$$

$$+ K_{t-1} \frac{1}{t-1} \sum_{j=1}^{t-1} \left((1-\phi) p_{ij} + \phi z_{ij}\right) \tag{3}$$

$$K_t = \frac{t\sigma_q^2}{t\sigma_q^2 + \sigma_\phi^2}$$

The time effects χ_t capture both the common variation in log productivity over time and also how the variance of the prediction error varies with experience. A convenient feature of the normal learning model is that the variance of the prediction error does not depend on the observed signals and is instead common across all individuals with the same level of experience. The weight ϕ depends on the variance of the signal noise in both z_{it} and p_{it}. This weight combines the two signals z_{it} and p_{it} into a single scalar signal $(1-\phi) p_{it} + \phi z_{it}$ that represents a sufficient statistic for the information obtained in period t. The variance

[15] We impose these normality assumptions throughout the paper. One of the implications is that log wages include a term that reflects the variance of the expectation error arround worker productivity conditional on observable characteristics. By assumption this term is constant across individuals within experience levels and will be subsumed in Q(x,t). We describe this in more detail below.

[16] The information in correlated normal signals is identical to the information contained in orthogonalized signals. The correlations between p_{it} and wages are therefore identical, regardless of whether the firm observes a correlated signal or an uncorrelated signal.

of this scalar is denoted by σ_ϕ^2. The exact expressions of ϕ and σ_ϕ^2 are known, but are not of particular interest at this point.

Equation (4) shows what the pure learning model implies for the covariances between pay and performance measures across time.

$$cov(w_{it}^*, p_{i\tau}) = \left\{ \begin{array}{cc} K_{t-1}(\sigma_q^2 + \frac{1-\phi}{t-1}\sigma_p^2) & \tau < t \\ K_{t-1}\sigma_q^2 & \tau \geq t \end{array} \right\} \tag{4}$$

Three of these implications are particularly noteworthy.

First, for $\tau > t$, the $cov(w_{it}^*, p_{i\tau})$ is increasing with t, because K_{t-1}, the weight placed on the stream of performance measures, is increasing in t.[17] Intuitively, both wages and the performance ratings reflect measures of true productivity plus noise. As the firm learns, the wage becomes increasingly more correlated with underlying productivity. Since the noise in performance ratings does not change with experience, the two measures will become increasingly correlated.

Second, $cov(w_{it}^*, p_{i\tau})$ is larger for performance measures that occurred before the wage was set $(\tau < t)$, than for performance measures that were not yet observed when the wage was set $(\tau \geq t)$. This is because current pay incorporates the realizations of ε^p from previously observed performance measures, but not from future performance measures. Under the learning model, the relationship between $cov(w_{it}^*, p_{i\tau})$ and τ will be a step function. The size of the step can be obtained by differencing the two expressions in equation (4) and is equal to $K_{t-1}\frac{1-\phi}{t-1}\sigma_p^2$.

Third, the size of the step (reflecting the difference in covariances between wages and past, compared to future, performance measures) decreases in t. Mathematically, this is because $\frac{K_{t-1}(1-\phi)}{t-1}\sigma_p^2$, decreases in t. Intuitively, firms' expectations are based on substantially more productivity ratings when t is large and they therefore put less weight on any given signal p_{it} when setting wages.

To test these implications in the data, we need to learn about the covariance of pay and performance as a function of the timing of the performance measure. We first residualize pay and performance by age and year, both interacted with education, race and gender.[18] We will use these residuals throughout the paper. We then estimate separate regressions for current wage residual on each of 6 leads and lags of the performance measures, separately for two age groups, 25-39 and 40-54.[19]

[17] This is not necessarily true for $\tau < t$, because the weight placed on the measurement error component in $\tau < t$ declines with t.

[18] Specifically, we regress each variable on age, year, race and gender fixed effects where we also interact race and gender with a timetrend and a quadratic in age. We do this separately for each education group.

[19] These regressions are estimated separately for each performance rating so we do not have to restrict the

Figure 3 plots these coefficients as well as their 95% confidence intervals. The x-axis shows timing of performance measures where negatives indicate those that occurred before the current wage was set while 0 to 6 occurred after, separated by the black vertical line. The purple line shows the older age group while the blue line shows the younger. First, the purple line is above the blue line, meaning the relationship between pay and performance is stronger among the more experienced workers. This provides evidence in favor or the first implication of the learning model outlined above. Second, we observe that performance measures in the past are more highly correlated with pay than performance measures in the future. However, contrary to the third prediction above, it looks as though the difference between the impact of past, compared to future performance, on the wage is larger for the older age group. This larger step size seen in the older-worker sample suggests that the firm updates more on new signals for this group, compared to the younger workers. Thus we find some evidence in favor of the pure employer learning model, but this finding is inconsistent.[20]

3.2 Pure Productivity

In the pure productivity model, firms are perfectly informed about worker productivity and wage dynamics arise only because worker productivity itself evolves over time.

A simple yet flexible way of representing the evolution of individual productivity is given by equation (5). We assume $\kappa_i \sim N\left(0, \sigma_\kappa^2\right)$ and $\varepsilon_{it}^r \sim N\left(0, \sigma_r^2\right)$ and that the ε_{it}^r are uncorrelated over time and with κ_i. We initialize this difference equation in period 0 by assuming that q_{i0} is drawn from a normal distribution $N(0, \sigma_q^2)$ and is independent of κ_i.[21]

$$q_{it} = q_{it-1} + \kappa_i + \varepsilon_{it}^r \tag{5}$$

According to equation (5), the log of individual productivity q_{it} evolves following an experience profile with three sources of heterogeneity. The heterogeneity in the drift parameter κ_i captures that individuals may differ in the intensity with which they accumulate human capital over the life-cycle. Persistent differences in intensity would arise, for example, if

sample to individuals with non-missing values for all 13 comparisons.

[20] Note that figure 3 also informs us about the firm's pay-for-performance practices. If firms relied on the performance evaluations to set direct incentives, we would observe that pay and performance ratings correlate heavily for the current period. However, all other past performance evaluations, as well as all those observed in the future, should have no impact on pay. That is, we should see a large spike in figure 3 at -1. We find absolutely no evidence for direct incentives so conclude that is not a confounding factor. This is related to the fact that we do not use bonus data. When we restrict our sample to just the years where bonus data is available and incorporated the bonuses, we do find a small spike at -1.

[21] We adopt the convention that period 0 is a period prior to the first production.

individuals differ in either their preferences or ability to invest (Becker (1964), Ben-Porath (1967)). The heterogeneity in q_{i0} captures differences in the initial ability. Finally, the innovations ε_{it}^r represent time-variation in individual productivity that are not predictable. The i.i.d. assumption on the ε_{it}^r implies that the variation in these innovations does not decline with experience and that individual productivity diverges even for relatively experienced workers. There are various possibilities why worker productivity might evolve randomly over time. It is for instance plausible that at least a subset of workers is subject to health shocks that affect performance. A more intriguing possibility is that experience affects the tasks individuals are required to perform. If productivity on past tasks does not perfectly predict productivity on future tasks, then worker productivity would indeed be subject to unpredictable variation as individuals gain experience (Gibbons and Waldman 2006).

The pure productivity model imposes that wages are exactly equal to productivity. Therefore equation (5) represents a process by which wage growth follows a random walk with drift. Under these assumptions, it is easy to see that individual variation in κ_i will introduce persistent correlation in pay changes. As reported by BGHb, the data used in this project display positive correlation in pay changes. Using regression analysis, we confirm this finding in table 2 and find, for example, that last year's pay change predicts this year's change. The regression coefficient is approximately 0.21 and is statistically significant at the 1% level.

Thus the data exhibit patterns consistent with individuals differing in their rates of human capital accumulation, represented by κ_i. However, we can learn more about the relative importance of κ_i and the random walk component, ε_{it}^r. Equations (6) and (7) show what this formulation of the pure productivity model implies for the variance and covariances of pay changes.[22]

$$Var(w_{it}^* - w_{it-1}^*) = \sigma_\kappa^2 + \sigma_r^2 \tag{6}$$

$$Cov(w_{it}^* - w_{it-1}^*, w_{it+k}^* - w_{it+k-1}^*) = \sigma_\kappa^2 \tag{7}$$

Table 3 shows the empirical counterparts to equations (6) and (7) using the most recent three changes in log pay residuals. Here the variance in pay changes is 0.003 while the covariance is almost an order of magnitude smaller, equalling approximately 0.0007. A literal interpretation of the pure productivity model implies the variance in the random walk term is 3 times the variance in the linear growth term.

This finding is roughly consistent with both the previous literature on productivity cited

[22] The star on w_{it}^* is meant to represent the wage as measured without measurement error. We introduce measurement error in wages below.

above and the literature on employer learning (Farber and Gibbons 1996). Log wage changes seem to have a small persistent component but a sizeable random walk component. However, we do not know whether this large random walk component is driven by variation in productivity or by changes in the information available to employers.

3.3 A Nested Model of Learning about Changing Productivity

Above we described both a pure learning and a pure productivity model. We now present a nested model, combining the two. Nesting allows us to test and quantify the relative importance of both models for explaining wage and productivity dynamics.

We use the same dynamic specification for q_{it} that we also used in the pure productivity model. We therefore assume that q_{it} evolves according to equation (5) at the beginning of each period, including period 1 and maintain the same distributional assumptions described above. Again, equation (5) allows us to represent three sources of individual heterogeneity in productivity evolution: q_{i0} captures differences in initial productivity; κ_i captures heterogeneous growth rates; ε_{it}^r captures idiosyncratic shocks.

From the pure learning model, we adopt the idea that firms do not observe individual productivity directly. Rather, they observe correlates of worker productivity and use these to learn about worker productivity. As described in the learning model, we assume that firms receive signals of three types. Two of these $\left\{z_{i0}, \{z_{it}\}_{t=1}^{T}\right\}$ are not observed in the data and one $\{p_{it}\}_{t=1}^{T}$ is observed both by employers and in our data. The distributional assumptions for $\left\{z_{i0}, \{z_{it}\}_{t=1}^{T}\right\}$ are maintained from the pure employer learning model.

However, as we estimated the learning model, we realized that there is a relatively high degree of correlation in manager ratings that is difficult to explain with any learning or productivity model. We interpret this as a manager "chumminess effect": workers might be temporarily matched with managers that generally give higher ratings or that are particularly compatible with the worker. Such "chumminess" would generate temporarily high ratings that will not persist as individuals are reassigned in their careers. We model this effect by assuming that the ε_{it}^p evolve according to equation (8) :

$$\varepsilon_{it+1}^p = \rho \varepsilon_{it}^p + u_{it+1} \tag{8}$$

where the initial noise is $\varepsilon_{i1}^p = 0$ and $u_{it} \tilde{} N(0, \sigma_u^2)$. The parameter ρ governs the degree of persistence in manager ratings and will be estimated.[23]

[23] A different modeling assumption would be to put the auto-regressive component, ρ, directly into the productivity evolution equation. This would yield some auto-correlation in performance measures. However, because p_{it} contains noise terms, ε_{it}^p, the AR-1 process in performance would exhibit less persistence than the AR-1 process in productivity. In order to generate the relatively large auto-correlations between p_{it} and

Thanks to the nesting, we can get back to the pure productivity model by restricting the signal noise in z_{i0} and in $\{z_{it}\}_{t=1}^{T}$ to be zero, and to the pure learning model by restricting the variance in productivity innovations ε_{it}^{r} and growth heterogeneity κ_i to be zero. Estimating a model that leaves these parameters free allows for both productivity and learning dynamics.

As we move towards estimating this nested model, we need to address two measurement problems. We need to associate the ordinal performance ratings in our data with the signals p_{it} and we need to allow for measurement error.

The performance ratings in our data and the productivity signals p_{it} are clearly not identical. The ratings are reported on an ordinal scale with a finite set of (k) support points and they therefore do not follow a normal distribution. We therefore assume that the normal random variable p_{it} represents the probit index for an ordered probit variable, such that the signals p_{it} map into our observed manager ratings (denote P_{it}) as follows:

$$\tilde{p}_{it} = \sum_{i=1}^{k} 1\left(p_{it} \geq c_{kt}\right) \tag{9}$$

It is important to note that the intercepts c_{kt} differ by experience t. This implies that manager rankings are assumed to be relative to workers within their experience level. It is possible to refine the comparison group further, by demographic group or job level, for example.

To account for measurement error in wages, we write

$$W_{i,t} = W_{i,t}^{*}\Omega_{i,t} \tag{10}$$

where W_{it} is the observed wage, W_{it}^{*} is the wage measured without error and Ω_{it} represents the measurement error. Taking logs we get

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

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

This completes the description of the model that we are analyzing in this paper. This model is governed by 8 parameters $\left(\sigma_q^2, \sigma_r^2, \sigma_0^2, \sigma_u^2, \sigma_\omega^2, \sigma_\kappa^2, \rho, \sigma_z^2\right)$ and by imposing the appropriate restrictions, we can restrict this model to the pure productivity or the pure learning model and we can thus test these models against each other and against the unrestricted version.

p_{it-1}(we show below these are on the order of 0.6), we would need the signal noise in ε_{it}^{p} to be very small. But, if the ε_{it}^{p} were very precise, then we would necessarily require wages and performance signals to be very highly correlated. We show below that this is not something observe in the data.

The dynamic learning model that we developed in this section belongs to a much larger class of dynamic learning models. In the appendix, we describe this larger class of models and show how it can be estimated using correlations between productivity signals and wages. In Section 5 we discuss the estimation and identification of the 3 models developed in the current section.

4 Estimating and Identifying the Nested Model.

The results developed above allow us to estimate the parameters of the model of learning and productivity. A necessary condition for estimation is that the model is identified. While we have formally shown identification only for the pure learning model[24], we discuss identification simultaneously with the estimation results and the fit of the model.

To estimate the model we exploit the results from the general linear state space model that show how to derive the second moment matrices for the observable quantities. We transform these matrices into correlations for all moments that involve performance ratings. In a first step, we estimate the correlations between the latent performance rating p_{it} and with wages $w_{i\tau}$, where τ varies from t-6 to t+6. We also estimate the correlations between p_{it} and $p_{i\tau}$ for τ between t+1 and t+6. This is possible, because our model implies that the ordinal performance ratings are derived from the underlying normally distributed p_{it} and because wages are themselves normal. We can therefore estimate the correlations of p_{it} with $(p_{i\tau}, w_{it})$ using maximum likelihood.

The estimated correlations of the latent productivity measures p_{it} with wages and other productivity measures as well as the variance-covariance matrix of wages provide the moments that we will use to estimate our models.

We use data from 30 experience levels and we could therefore, in principle, match correlations in wages and performance ratings across 30 experience levels. To simplify the estimation, we chose 56 moments in the data that we think are particularly informative for distinguishing the learning and the productivity models. We match the variance in pay by 5-year experience groupings from 0 to 30 years experience, the auto-correlations of pay and 6 lags separately for two 15-year experience groupings from 0 to 30 years, the auto-correlations of performance and 6 lags separately for the same two experience groups, and the correlations of wages and current performance as well as 6 lags and leads of performance also separated by the same experience groups.[25]

[24] Results available upon request.

[25] The average auto-correlations at a given lag (or lead) and for a given experience group are obtained by averaging the auto-correlations at that lag (or lead) across years of experience levels weighted by the number of individuals for which we observe this auto-correlation.

These moments are plotted in figure 4 with 95% confidence intervals (obtained from bootstrapping with 500 repetitions). Where experience groups are separated, the red dots refer to the older group. Some key features of these moments are as follows. First, the variance in pay is increasing almost linearly in experience. Second the correlations between pay and performance increase with experience. Third, the correlations between pay and performance are higher for lagged performance, as also exhibited in figure 3. Fourth, the differences between the auto-correlations between current pay and future performance minus the auto-correlation of current pay and past performance increases in experience. These patterns in the data will drive much of our estimation results described in this section.

In appendix figure 2, we compare our performance moments, which estimate correlations using the latent continuous performance measure, to standard pearson correlations using the original performance residuals. The latter are represented with hollow dots alongside the original moments. As can be seen, the correlations look quite similar, so we are not worried that this method introduces any bias in our estimates.

Table 4 displays our parameter estimates for the three models which we obtain via method of moments with equal weights on all moments. Standard errors, obtained by bootstrapping with 500 repetitions, are shown in parentheses.[26] We also plot the implied fitted moments in figures 5-7. Each figure plots the sample moments as dots with lines represented the fitted moments from one of the three models. We now discuss how well each model fits the data. We will at the same time discuss how each model is identified and what features of the observed moments determines the observed parameter values for each model.

4.1 Pure Learning

In the pure learning model, the idiosyncratic component of productivity is constant and wages vary only because firms obtain new information about individual productivity. By imposing two restrictions on the nested model we obtain the learning model. We restrict the variance of the random walk and of the heterogenous growth component to 0: $\sigma_r^2 = \sigma_\kappa^2 = 0$. There are therefore 6 free parameters. These are the variance of initial productivity $\left(\sigma_q^2\right)$, the variance in the measurement error of wages $\left(\sigma_\omega^2\right)$, the variance in the noise of initial information $\left(\sigma_0^2\right)$, the variance in the signal observed by firms, but not in the data $\left(\sigma_z^2\right)$, and

[26] The exact bootstrapping procedure is as follows. We draw the sample randomly, with replacement and generate the bootstrapped moments. We then estimate the parameters to match these moments, taking as starting values the true parameters values shown in table 4. Because convergence is time consuming, we restrict the number of iterations to be 100 or less (using the Simplex Method in Judd XX). Approximately half of the bootstrapped samples actually converge. We do not search across starting values to find the global minimum for each of the 500 samples. However, we hope that since our starting values are the true estimates, we remain in the correct neighborhood.

the two parameters (ρ, σ_u^2) governing the variation in the signal observed both in the data and by firms.

Consider now how we can identify the parameters of the model using the moments presented in figure 4. In figure 5, we show these moments together with the fitted moments using the estimated parameters (table 4) of the pure learning model.[27]

We begin by noting that in pure learning models wages increasingly reflect true productivity and measurement error as individuals acquire experience. This means that we can identify the variance of productivity and of the measurement error by observing the variance and covariance of measured wages at high experience levels. In particular, we have that $lim_{t\to\infty}(v(w_t)) = \sigma_q^2 + \sigma_\omega^2$ and $lim_{t\to\infty}(cov(w_t, w_{t+1})) = \sigma_q^2$. Thus, the variance and covariance of wages at high experience levels identify the variances of both the measurement error and of idiosyncratic productivity.

The two panels on the left of figure 5 display the variance and auto-correlations of log wages. We find that the variance in log wages at 20-30 years of experience is close to 0.12 and that the auto-correlation in log wages at these experience levels is about 0.95. For the pure learning model, this implies estimates of the variance of productivity close to 0.12 and estimates of the measurement error in wages of about 0.005. These are indeed the estimates we obtain for $\left(\sigma_q^2, \sigma_\omega^2\right)$ and report in table 4 for the pure learning model.

We now show how 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. From equation (8), we get:

$$\lim_{t\to\infty} var(p_{it}) = \sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2} \tag{11}$$

We could thus identify $\frac{\sigma_u^2}{1-\rho^2}$ if we knew the variance in the performance signal in p_{it}, but unfortunately this variance is unobservable, because p_{it} is a categorical variable. However, we also have:

$$\lim_{t\to\infty} cor(p_{it}, p_{it-1}) = \frac{\sigma_q^2 + \rho\frac{\sigma_u^2}{1-\rho^2}}{\sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2}} \tag{12}$$

For relatively small σ_q^2, the correlations in p_{it} and p_{it-1} at high experience levels will identify the parameter ρ. With $\sigma_q^2 > 0$, the pattern of auto-correlations together will suffice to determine both (σ_u^2, ρ). The auto-correlation in p_{it} depends primarily on the parameter ρ. The tight link between ρ and the observed decline in the auto-correlations in $cor(p_{it}, p_{it-k})$

[27] Again, for all but the variances of wages, we show results for two experience groups: workers with less and with more than 15 years of experience. The color red indicates experience levels 15-30 and the color blue indicates experience levels 0-15.

at high k therefore determines ρ. Figure 4, shows that the first order auto-correlation in p_{it} at higher experience levels is about 0.66 and about 0.52 at 2 lags. Consequently, we report in table 4 an estimate of ρ of about 0.64 for the pure learning model.

To understand the identification of σ_u^2 using the auto-correlations in p_t, consider the limit as $t \to \infty$ for the auto-correlations at higher lags:

$$\lim_{t\to\infty} cor(p_{it}, p_{it-k}) = \frac{(1-\rho^2)\,\sigma_q^2 + \rho^k\sigma_u^2}{(1-\rho^2)\,\sigma_q^2 + \sigma_u^2} \tag{13}$$

Conditional on $\left(\rho, \sigma_q^2\right)$, these auto-correlations depend only on σ_u^2. Furthermore, the correlations in (13) are monotonically declining in σ_u^2 and we can therefore identify σ_u^2 at any lag given values for $\left(\rho, \sigma_q^2\right)$. The auto-correlations in (13) will be most responsive to changes σ_u^2 at longer lags and these longer lags are therefore particularly useful for identifying σ_u^2. Using the estimates of $\left(\rho = 0.64, \sigma_q^2 = 0.12\right)$ found above as well as an auto-correlation across 5 lags for p_{it} of about 0.22, we find from equation (13) an approximate value of 0.49 for σ_u^2 - reasonably close to our estimate of 0.65.[28]

This leaves only with two parameters that we need to identify: (σ_z^2, σ_0^2). These represent the noise in the dynamic and initial signals that are observed by firms, but not by us. The parameter σ_0^2 determines how much information the firm has about workers as they begin their careers. The parameter σ_z^2, together with (ρ, σ_u^2), determines how fast employers learn about worker productivity as they spend time in the labor market. To identify these parameters, we exploit the close link between the variance of wages and the amount of information that firms have at any moment in time. With fixed individual productivity, the variation in the variance of wages over the life-cycle informs us about how much information the firm has at any experience level. We can therefore use the life-cycle variation in wages to identify how informative z_0 and z_t are.

For young workers (0-4 years of experience), the variance in wage residuals is only about 0.04. This implies that at least initially the firm has little information about workers wages. To fit this fact, we will need the variance in the initial signal noise to be quite high and this is indeed what we find. Our estimate of $\sigma_0^2 = 0.58$ is almost 5 times as large as the variance in productivity, which results in very low variances in the wage for young workers. The increase in the variance of wages is then governed by the new information employers acquire through p_t and z_t. We do find that the variance in the signal noise in z_t is 0.49. Together, these parameter values reproduce the increase in the variance of the wage from about 0.04 to about 0.12 over the first 30 years of these individuals careers.

The learning model therefore does succeed in a number of ways. It matches the auto-

[28] The estimated value of 0.65 implies a correlation of 0.195 at 5 lags.

correlations in wages and the variance of wages at high experience levels, it matches the growth in the variance of wages with experience and it matches the auto-correlations in the performance measures using a small set of parameters.

However, the pure learning model fails to reproduce a number of patterns in the data. As evident from the top right panel in figure 5, the pure learning model does not match how observed performance measures correlate with wages and how these performance-pay correlations vary with experience. In our view, the pure learning model fails to match these correlations not because of any particular distributional assumptions. Instead, this failure of the learning model reflects a more general feature of pure learning models. In learning models, wages depend on past productivity signals. The weight given to any individual signal is larger when individuals are young and employers know relatively little about true productivity. By contrast, for higher experience levels, any given performance rating does not affect wages as much because firms have more precise expectations. For this reason, the learning model predicts that past productivity signals correlate more strongly with wages among younger rather than older workers.

The situation is reversed when we consider the correlations between current wages and future productivity signals. Because wages are only based on past productivity signals, the noise in future signals does not enter wage setting for young or for old workers. However, wages of older workers correlate more highly with true productivity and this leads to higher correlations of wages with future productivity signals for old rather than young workers.

The learning model therefore predicts a cross-over pattern when we compare correlations of wages with productivity signals at different leads and lags across different experience levels. There is no evidence for such a pattern in our data. Instead, the top right panel shows that wages and productivity signals are always more highly correlated for older workers than for younger. The therefore data suggests that firms rely more heavily on recent performance measures to set wages for their experienced employees than to set wages of young employees. This is inconsistent with the pure employer learning model.

4.2 The Pure Productivity Model

We next discuss how to identify the parameters of the pure productivity model which imposes that employers know individual productivity and that wages vary over the life-cycle because individual productivity varies. We impose that firms have perfect information by restricting the variance of the noise in the signals observed by employers (but not in our data) to 0: $\sigma_0^2 = 0$ and $\sigma_z^2 = 0$. We do not restrict the variance of the noise of the performance ratings (ε_t^p) in our data to equal zero because this restriction would imply that the performance ratings would be, absent measurement error in wages, perfectly correlated with wages.

Figure 6 displays the fit for the pure productivity model, where the solid line gives the implied moments from our parameter estimates from the second column in table 4.

We identify $(\rho, \sigma_u^2, \sigma_\omega^2)$ in much the same way as under the perfect learning model and limit our discussion to the parameters of the productivity process: $\left(\sigma_q^2, \sigma_r^2, \sigma_\kappa^2\right)$. To simplify the exposition, we assume that wages are measured without error. We then have

$$\begin{aligned} var\left(w_{i0}\right) &= \sigma_q^2 \\ cov\left(w_{it+1} - w_{it}, w_{it+k} - w_{t+k-1}\right) &= \sigma_\kappa^2 \\ var(w_{it+1} - w_{it}) &= \sigma_\kappa^2 + \sigma_r^2 \end{aligned}$$

Clearly, the variance of wages at the beginning of a career identifies the initial variation in productivity across individuals. Then, the covariance in wage growth across periods identifies the growth rate heterogeneity σ_κ^2. Finally, we can use the variance of wage growth together with σ_κ^2 to identify σ_r^2, the variance of the random walk component of productivity.[29] Indeed, we find that the variance in initial productivity is quite small (0.024), σ_κ^2 is approximately 0 and growth in productivity and wages is driven by a small random walk component. In 30 years variance in log pay (and therefore productivity in this model) rises by 0.08, which is just a bit smaller than would be implied by the random walk (whose variance is 0.004).

We note that we have only used the variances and covariance of wages to identify the parameters of the pure productivity model. The pure productivity model does however have additional implications for these correlations that allow testing the model. Most importantly, in the pure productivity model the variation in the productivity of individuals increases over time. In consequence, the pure productivity model predicts that (i) the correlation of the performance ratings with wages, (ii) the auto-correlations of performance ratings, and (iii) the auto-correlations of wages are all increasing with experience. This is exactly what we observe in the data. We find that the contemporaneous correlation of log wage residuals with performance ratings are about 0.25 for workers with 0-15 years of experience. For workers with 16-30 years of experience this correlation is about 0.35. The first order auto-correlation of performance ratings rises over the same time period from 0.57 to 0.67. The first auto-correlation of wages also increases from about 0.96 to 0.99. All of these aspects of the data are well matched by the performance model, both qualitatively and quantitatively.

However, other features of the data are difficult to match using a pure productivity model.

[29] As has been observed in MaCurdy (1982), Baker (1997) and many other papers that investigate the 2nd moment properties of log wages, the autocorrelation in wage growth identifies permanent heterogeneity in the wage growth. Farber and Gibbons (1996) propose testing the pure learning model using exactly this absence of autocorrelation in wage growth.

Most importantly, from our perspective, is that the pure productivity model predicts that the correlations of the current wage with future performance ratings exceeds the correlation of the current wage with past performance ratings. This feature is driven by the random innovations to productivity as individuals age. To see this, consider the correlations between wages and productivity signals when we set $\sigma_\kappa^2 = 0$, close to the estimated value. Then, compare the covariance $cov(w_t, p_{t-k})$ with $\text{cov}(w_t, p_{t+k})$:[30]

$$\begin{aligned} cov(w_t, p_{t-k}) &= cov(q_t + \varepsilon_\omega, q_{t-k} + \varepsilon_p) \\ &= cov\left(\sum_{j=1}^{t} r_j + \varepsilon_\omega, \sum_{j=1}^{t-k} r_j + \varepsilon_p\right) = (t-k)\,\sigma_r^2 \end{aligned}$$

The covariance between wages and future performance rating is

$$cov(w_t, p_{t+k}) = t * \sigma_r^2$$

For past productivity measures, the covariances between wages and productivity decline with the lag-size, whereas for future productivity measures the correlations are larger and identical across various leads. These features translate somewhat into the correlations that we can observe in the data and show in figure 5, but far from perfectly

An important difference between the productivity and the learning model arises when we consider correlations of wages and performance ratings that are almost contemporaneous. The learning model imposes an asymmetry in time, because past performance measures are used for setting current wages, while future performance measures can, by definition not be used in setting wages. The pure productivity model does not admit such an asymmetry. For small k, the performance ratings at t-k and t+k should be almost identically correlated in much the same way with log wages. However, in the data, we clearly observe, especially at higher experience levels, that future productivity levels are less correlated with log wages than are past productivity measures. As reported in figure 4, we observe among workers with 0-15 years of experience, that the correlation of the wage at t with performance ratings collected at t-3 exceeds the correlation with performance ratings at t+3 by about 0.03 points. For workers with 16-30 years of experience, the same difference is 0.07. These differences in the correlations of wages with past and future performance ratings and their increases are not predicted by the pure productivity model.

We have thus described how the parameters of the pure learning and the pure productivity models are linked to observable moments. We have also shown the features of the data that

[30] The problem is complicated by the fact that the standard deviations of the productivity signal also vary with k. However, empirically the covariances dominate the observed patterns in the correlations.

each of these models cannot match. Estimates from the joint model will in fact be able to match these features and will allow us to quantify what role the learning and productivity models play in setting wages.

4.3 The Combined Model

Finally, we consider how the nested model fits the data. Figure 7 displays the results for the nested model and parameter estimates are displayed in column 3 of table 4.

Comparing our estimates of the 3 productivity parameters $\left(\sigma_q^2, \sigma_\kappa^2, \sigma_r^2\right)$ across the pure productivity and the nested models, we find that the implied productivity processes of both models are almost indistinguishable. The estimates from both models admit very little heterogeneity in κ. Furthermore, both models allow for only modest heterogeneity in initial productivity. The standard deviation of initial productivity is 0.15 in the pure productivity model and 0.17 in the nested model. By contrast, we find substantial variation in productivity for older individuals. At 30 years of experience, the implied standard deviation in productivity is 0.38 and 0.41 in the pure and nested models, respectively. This rise in the dispersion of productivity is generated by the accumulation of random walk terms that have a standard deviation in each model of about 0.065, annually. In both models, therefore, individual productivity increases significantly and unpredictably over the life-cycle.

When we compare the learning parameters $(\sigma_0^2, \sigma_z^2, \sigma_u^2, \rho)$ across the pure learning and the nested model, we observe large differences in the estimated parameters and in the implied learning process. Only the parameters (σ_u^2, ρ) that govern the observable signal are similar across models. For both models, we find that ρ is around 0.64, reflecting the fact the the auto-correlations of the performance measures decline rapidly with higher lags.

However, we see large differences in the estimated variances of the signal noise in the unobserved firm signals z_{it} and z_{i0}. For the pure learning model, we estimate $\sigma_0^2 = 0.374$, implying that the initial signal z_{i0} is quite imprecise. This high degree of imprecision is required to match the low observed variance in initial wages. The nested model is able to fit this low variance in initial wages by imposing that the variance of idiosyncratic productivity itself is initially low. In the nested model, this allows for an estimate of σ_0^2 that is very close to 0 – implying that firms are almost perfectly informed about worker productivity at the beginning of the workers career.

For the variance in the dynamic signal, z_{it}, we find that the pure learning model yields a lot more noise than does the nested model. The learning model needs the dynamic signals to be relatively imprecise in order to allow the firm to continue learning about worker productivity even at higher ages. This in turn is required because we observe that the variance of wages continues to increase at all experience levels. The nested model by contrast explains the continuing increases in the wage variance as reflecting continued changes in individual productivity over the life-cycle.

Because the nested model does not tie down the learning parameters to fit the variance of wages over time, it can instead use the them to obtain a better fit of the correlations between pay and performance measures. In particular, because productivity continues to evolve and because learning about any innovations in productivity is relatively rapid ($\sigma_z^2 = 0.08$), the nested model can fit the time-pattern in the productivity and performance auto-correlations quite well. It does underpredict the observed high auto-correlations of wages with past productivity measures at high experience levels, but it qualitatively does fit most of the observed patterns in how productivity and wages correlate.[31]

4.4 Interpretation

Our empirical analysis leads us to think of the life-cycle variation in the dispersion of wages as reflecting two contributing factors. First, productivity evolves along idiosyncratic paths and drives much of the increase in wage variance. However, the extent to which productivity is reflected in wages is limited by the information available to firms. Generally, the variance of wages is lower than the variance of underlying productivity because wages are conditional expectations based on information available to the firm. The difference between productivity and the wage is expectation error on the part of the firm. Given our estimates of the learning model, we can now examine how employer expectation error and the variance in productivity contribute to the overall variance of wages.

Figure 8 shows what the nested model implies for the variance of wages, the variance of productivity and the variance of the expectation error. First, we see that overall the variance of wages closely mirrors the variance of productivity. We found that the initial signals have almost no noise about initial productivity and therefore initially the variance of wages and the variance of productivity are nearly identical. However, as experience grows, productivity evolves stochastically and employers aim to learn about this evolution. Because the dynamic signals are noisy, employers start making errors and the variance of the expectation error starts increasing and then stabilizes at about 0.025. The overall variance in wages follows the variance in productivity over the life-cycle, but is somewhat smaller throughout.

[31] We have estimated our model restricting the sample to years in which bonus data is incorporate bonuses into our measure of wages. We find qualitatively similar results. However, this underprediction of the observed high correlations of wages with past productivity is even stronger. We attribute this to the notion that there is some direct pay for performance occuring via the bonuses that our model cannot fit. Since our primary analysis (using base pay) does not exhibit patterns consistent with direct incentives, we do not build this concept into our model. Incorporating incentive pay would be an interesting area for future work.

The nested model thus implies that the growth of the variance in wage residuals over the life-cycle can be primarily attributed to random variation in wages over the life-cycle and not to a slow discovery of some underlying fixed productivity characteristics. However, the model does not imply that employer learning is unimportant as an economic phenomenon. Instead, the standard deviation in the expectation error is about 0.15 for most of the life-cycle, which implies that firms make on average a mistake of about 12% of wages.

5 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, we do not find evidence that the wage dynamics overall are driven primarily by the learning process. Instead, our model suggests that random variation in productivity drives most of the observed increase in the variance of wages over the life-cycle. We believe these findings represent a significant reinterpretation of the employer learning literature.

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

A related caveat is that, because we only have data on one firm, we may have nonrandom selection out of the sample. We showed above that workers who leave are negatively selected on performance. We believe this may bias us against finding evidence of employer learning. When the firm observes low productivity, the learning model predicts that subsequent wages will fall. However, if a subset of these wages are not observed because low-performing workers left the sample, we would find a smaller correlation between pay and past performance. In principle, we can incorporate turnover into our model, and we hope to do so in future work.

Seemingly contradictory to most models of human capital accumulation (Becker 1964, Ben-Porath 1967), we find that productivity evolves unpredictably throughout the life cycle with almost no persistent-growth component. One explanation for this finding is that workers are assigned to different tasks throughout the life cycle and performance on past tasks does not predict performance on future tasks. This interpretation suggests that firms shift workers into job levels and tasks with little ability to predict worker success there.

References

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

[2] 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.

[3] 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.

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

[5] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[18] Kahn, L. (2009): "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy," *Labour Economics*, forthcoming.

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

[20] 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.

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

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

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

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

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

A A More General Class of Models

In Section 3, we have presented a model with a particular productivity process and a particular learning structure. In this section, we will 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.

A.1 The Productivity Process

In period 0 (before production starts), individuals are endowed with a $(n_q x1)-$vector of productivity parameters θ_{i0} with $E[\theta_{i0}] = 0$ and $E\left[\theta_{i0}\theta_{i0}'\right] = P_0$. In subsequent periods, productivity evolves according to a stochastic process represented by the stochastic difference equation:

$$\begin{aligned} \theta_{it+1} &= \Phi\theta_{it} + \varepsilon_{it+1}^{\theta} \\ & \varepsilon_{it+1}^{\theta} \tilde{} N(0, R_{\theta}) \end{aligned} \tag{14}$$

This implies that the productivity states in period 1, the first period of actual production are $\theta_{i1} = \Phi\theta_{i0} + \varepsilon_{i1}^{\theta}$.

A.2 Prediction in the Initial Period

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

$$\begin{aligned} z_{i,0} &= H_0'\theta_{i0} + \varepsilon_{i,0}^{z} \\ & \varepsilon_{i,0}^{z} \tilde{} N(0, R_{z,0}) \end{aligned} \tag{15}$$

The dimensions of $\left(H_0, \varepsilon_{i,0}^{z}, R_{z,0}, P_0\right)$ are implicitly defined to conform to $z_{i,0}$ and θ_{i0}. Based on the signal vector z_{i0} firms predict the state θ_{i0} :

$$\begin{aligned} \widehat{\theta}_{i,0|0} &= P_0 H_0 \left(H_0{}' P_0 H_0 + R_{z,0}\right)^{-1} z_{i,0} \\ &= K_z z_i \end{aligned} \tag{16}$$

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 (14). Firms best guess about productivity in period 1 is:

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

and the posterior variance of the expectation error is:

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

A.3 The Recursion

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

$$\begin{aligned}x_{i,t} &= H_x'\theta_{it} + \varepsilon_{it}^x \\ &\varepsilon_{it}^x \,\tilde{}\, N(0, R_x)\end{aligned} \tag{17}$$

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

$$\begin{aligned}\widehat{\theta}_{it|t} &= \widehat{\theta}_{it|t-1} + P_{t|t-1}H_x\left(H_x{}'P_{t|t-1}H_x + R_x\right)^{-1}\left(x_{it} - H_x'\widehat{\theta}_{it|t-1}\right) \\ &= \widehat{\theta}_{it|t-1} + K_t\left(x_{it} - H_x'\widehat{\theta}_{it|t-1}\right) \\ &= \left(1 - K_t H_x'\right)\widehat{\theta}_{it|t-1} + K_t x_{it}\end{aligned} \tag{18}$$

Again, firms account for the evolution in productivity described in equation (14) and therefore, firms best guess about productivity in period t+1 is:

$$\begin{aligned}\widehat{\theta}_{it+1|t} &= \Phi\widehat{\theta}_{it|t} \\ &= \Phi\left(1 - K_t H_x'\right)\widehat{\theta}_{it|t-1} + \Phi K_t x_{it}\end{aligned} \tag{19}$$

The variance of the expectation error then evolves according to

$$P_{t+1|t} = \Phi\left(P_{t|t-1} - K_t H_x' P_{t|t-1}\right)\Phi' + R_\theta \tag{20}$$

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

A.4 Wages

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

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

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

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

$$\begin{aligned} w_{it} &= \left(q\left(x,t\right) + \frac{1}{2}v\left(t\right)\right) + \widehat{q}_{it} + \omega_{it} \\ &= h\left(x,t\right) + \widehat{q}_{it} + \omega_{it} \end{aligned} \tag{21}$$

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

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

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

The linear state space model consists of three parts. First, we need to specify how the state evolves. This is done in equation (22). Second, we need to specify how the states map into observed variables. This measurement equation is given by (23). Finally, we need to specify the distribution of the initial state ξ_{i1}, the forcing variables v_{it}, and the unobservable noise in the measurement equation e_{it}.

$$\xi_{it+1} = F_t \xi_{it} + v_{it+1} \tag{22}$$

$$y_{it} = M \xi_{it} + e_{it} \tag{23}$$

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

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

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

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

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

The (K_z, K_t) −matrices were implicitly defined in equations (16) and (18) above.

A.6 The 2nd Moment Matrix of Observables

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

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

$$E\left[y_{it} y_{i\tau \geq t}^{'}\right] = ME\left[\xi_{it}\xi_{i\tau}'\right] M' + E\left[e_{it}e_{i\tau}'\right] \tag{24}$$

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

Tedious, but straightforward algebra yields

$$E\left[\xi_{it}\xi_{i\tau}'\right] = \sum_{j=2}^{j=t}\left\{\left(\prod_{l=j}^{l=t-1}F_l\right)E\left[v_{i,j}v_{i,j}^{'}\right]\left(\prod_{l=j}^{l=\tau-1}F_l\right)'\right\}+\left(\prod_{l=1}^{l=t-1}F_l\right)E\left[\xi_{i1}\xi_{i1}'\right]\left(\prod_{l=1}^{l=\tau-1}F_l\right)' \quad (25)$$

where

$$E\left[\xi_{i1}\xi_{i1}'\right] = \begin{pmatrix} \Phi K_z\left(H_0'P_0H_0+R_z\right)K_z'\Phi' & \Phi K_z H_0'P_0\Phi' \\ \Phi P_0 H_0 K_z'\Phi' & \Phi P_0\Phi' + R_\theta \end{pmatrix} \quad (26)$$

and

$$E\left[v_{i,j}v_{i,j}^{'}\right] = E\begin{pmatrix} \Phi K_{j-1}R_x K_{j-1}'\Phi' & 0 \\ 0 & R_\theta \end{pmatrix} \quad (27)$$

We have thus shown how to generate $E\left[y_t y_\tau\right]$ as functions of the parameters $(P_0, R_{z,0}, R_x, R_\theta, H_x, H_0, \Phi)$ and the measurement matrix for any dynamic specification of productivity that follows equation (14) and any normal learning model that follows equations (15) and (17).

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

In this Section, 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 3 is a special case of such a linear learning model. We now show in the remainder of the Section 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 (24) together with equations (25), (26), and (27) to generate the covariance matrices of the wage residuals and performance ratings.

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

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

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

The individual state evolves as

$$\begin{aligned}\theta_{it+1} &= \begin{pmatrix} q_{it+1} \\ \kappa_i \\ \varepsilon^p_{it+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} q_{it} \\ \kappa_i \\ \varepsilon^p_{it} \end{pmatrix} + \begin{pmatrix} \varepsilon^r_{it+1} \\ 0 \\ u_{it+1} \end{pmatrix} \\ &= \Phi\theta_{it} + \varepsilon^\theta_{it}\end{aligned}$$

The vector v_{it+1} is therefore given by $v_{it+1} = \begin{pmatrix} \Phi K_t \varepsilon^x_{it} \\ \varepsilon^\theta_{it} \end{pmatrix}$.

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^2_ω and thus $E\left[e_{it}e'_{it}\right] = \begin{pmatrix} \sigma^2_\omega & 0 \\ 0 & 0 \end{pmatrix}$.

Next,

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

Then

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

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

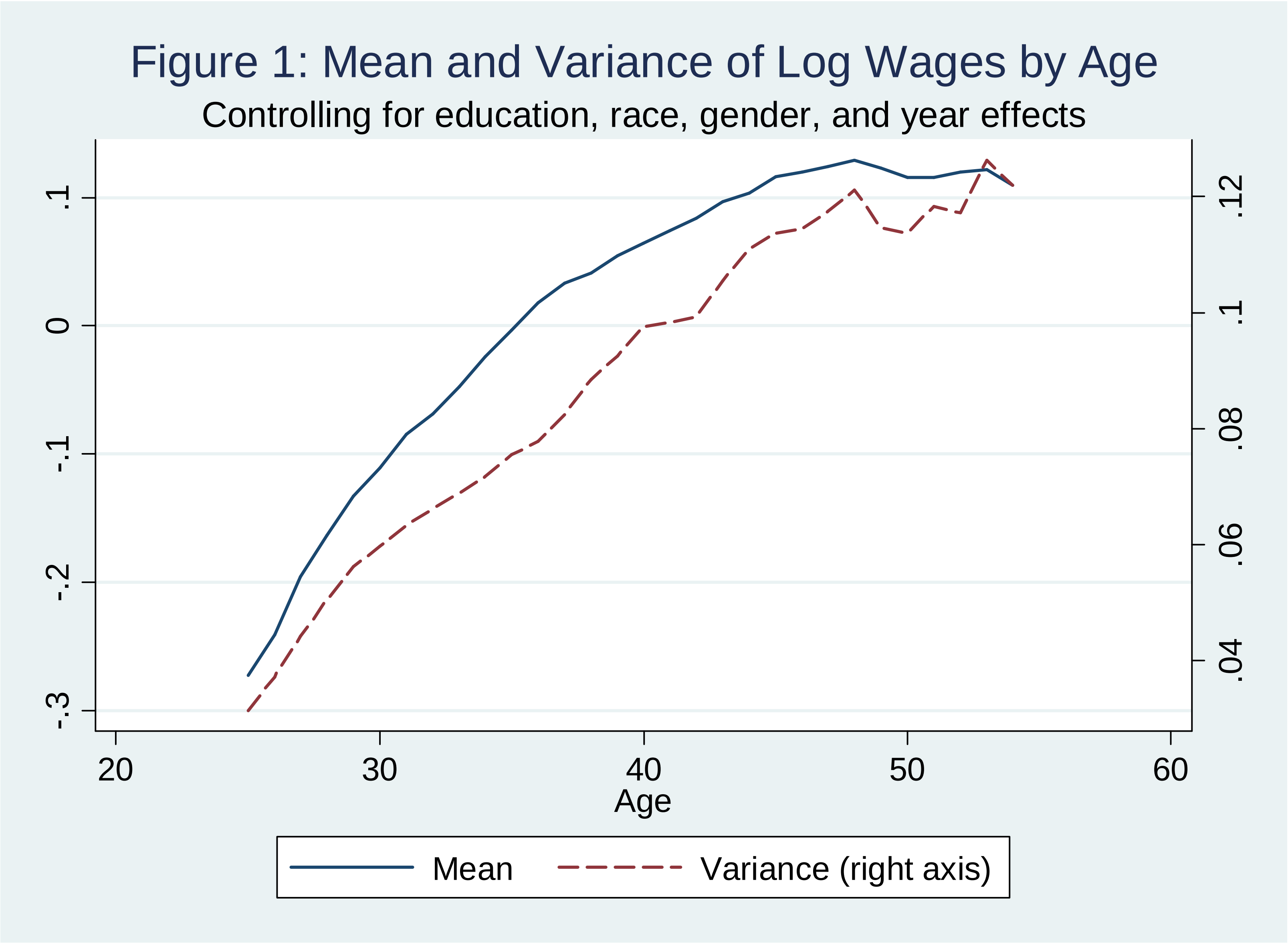
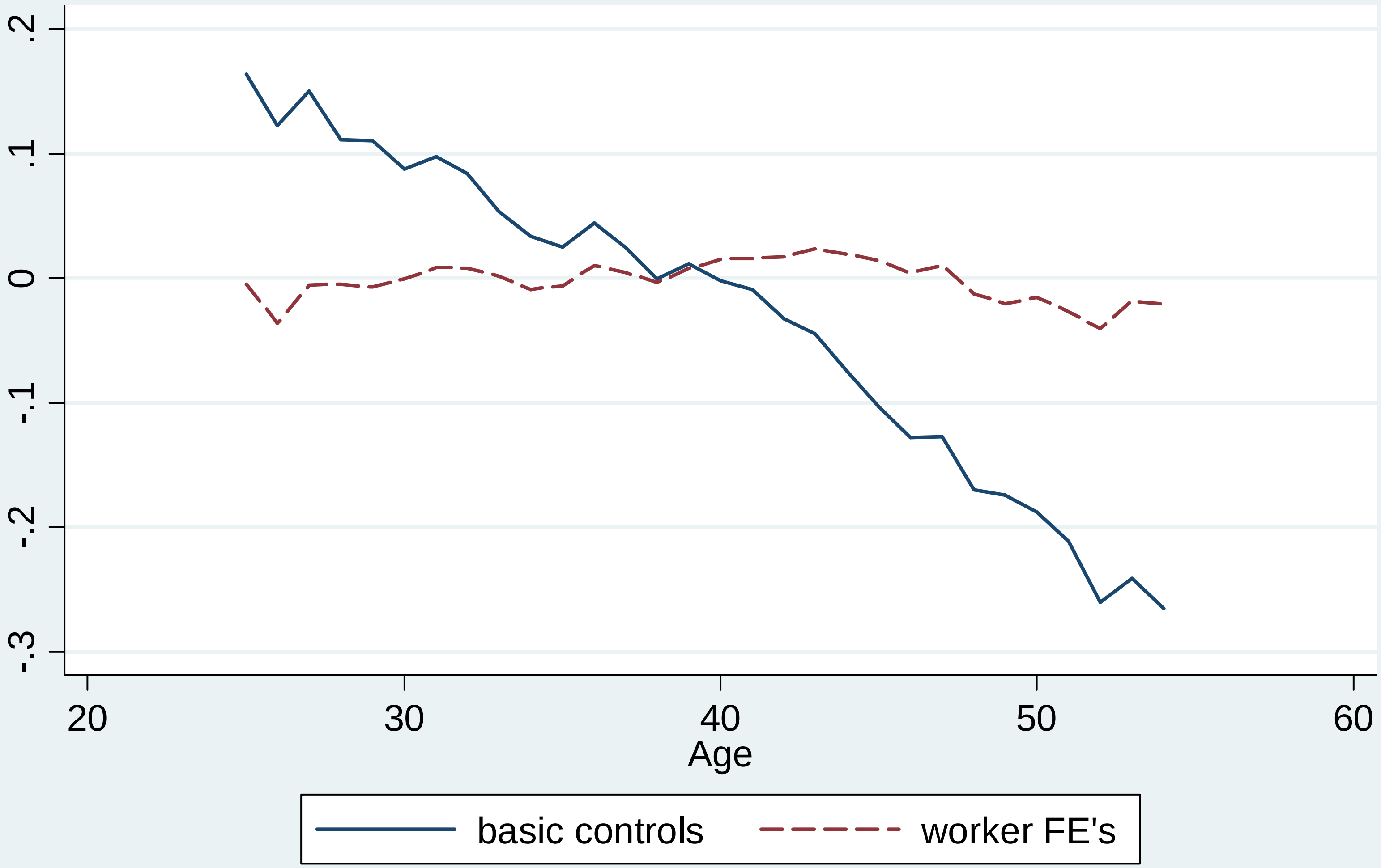


Figure 1: Mean and Variance of Log Wages by Age
Controlling for education, race, gender, and year effects
-.3
-.2
-.1
0
.1
.04
.06
.08
.1
.12
20
30
40
50
60
Age
Mean
Variance (right axis)

Figure 2: Average Performance Residual by Age

Controlling for education, race, gender, and year effects



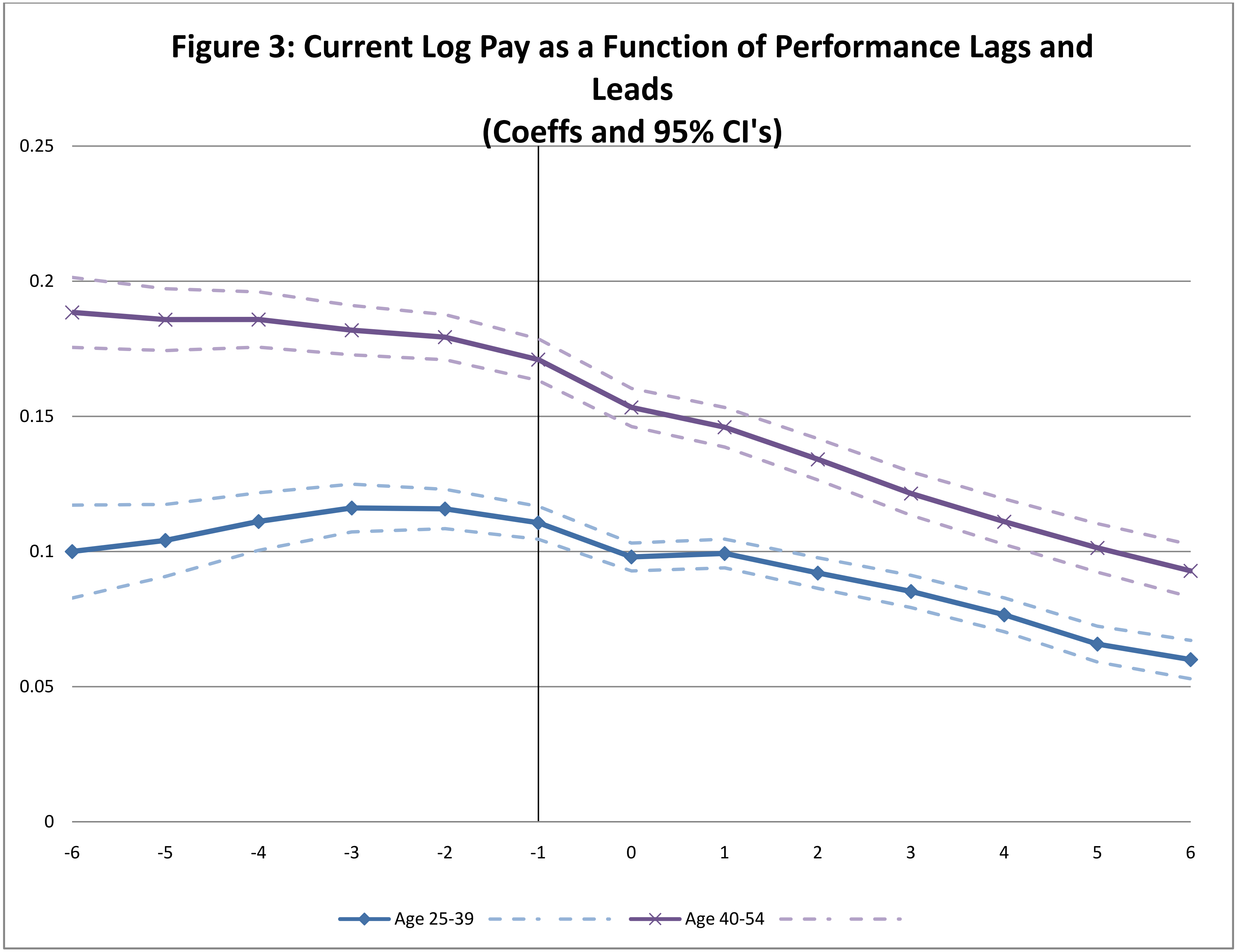
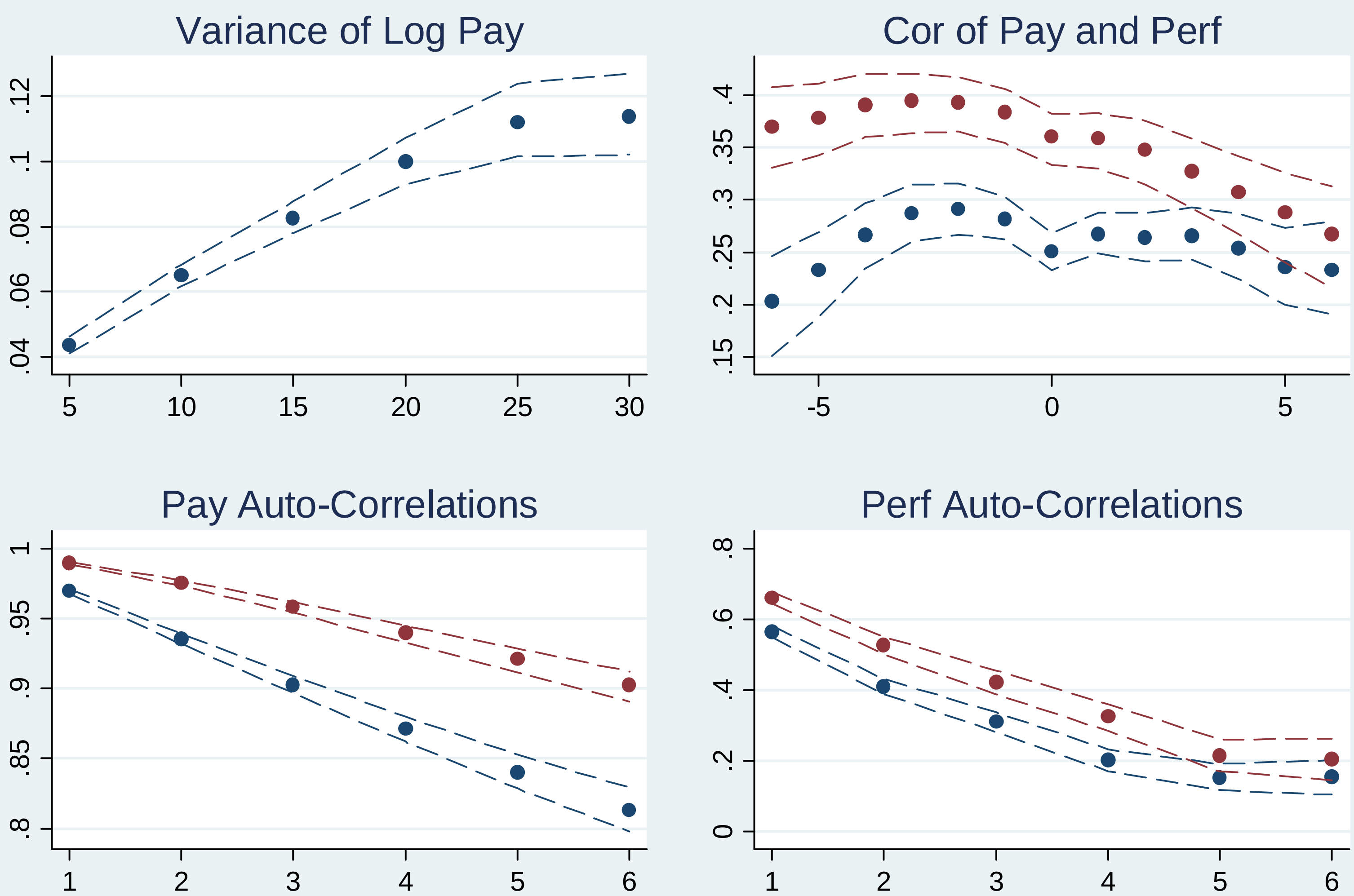


Figure 3: Current Log Pay as a Function of Performance Lags and Leads
(Coeffs and 95% CI's)
0.25
0.2
0.15
0.1
0.05
0
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
Age 25-39
Age 40-54

Figure 4: Moments and 95% CI



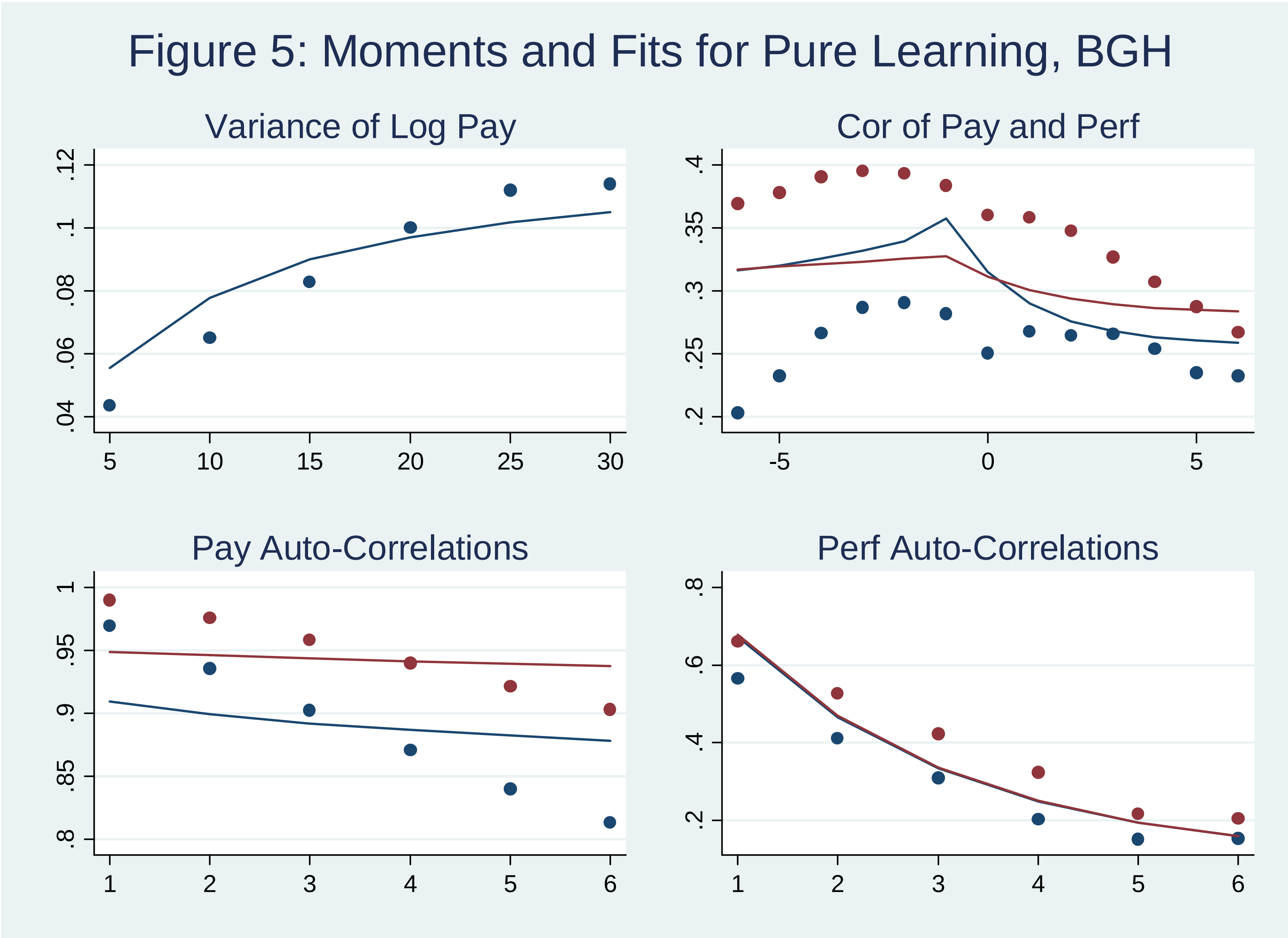


Figure 5: Moments and Fits for Pure Learning, BGH
Variance of Log Pay
.04
.06
.08
.1
.12
5
10
15
20
25
30
Cor of Pay and Perf
.2
.25
.3
.35
.4
-5
0
5
Pay Auto-Correlations
.8
.85
.9
.95
1
1
2
3
4
5
6
Perf Auto-Correlations
.2
.4
.6
.8
1
2
3
4
5
6

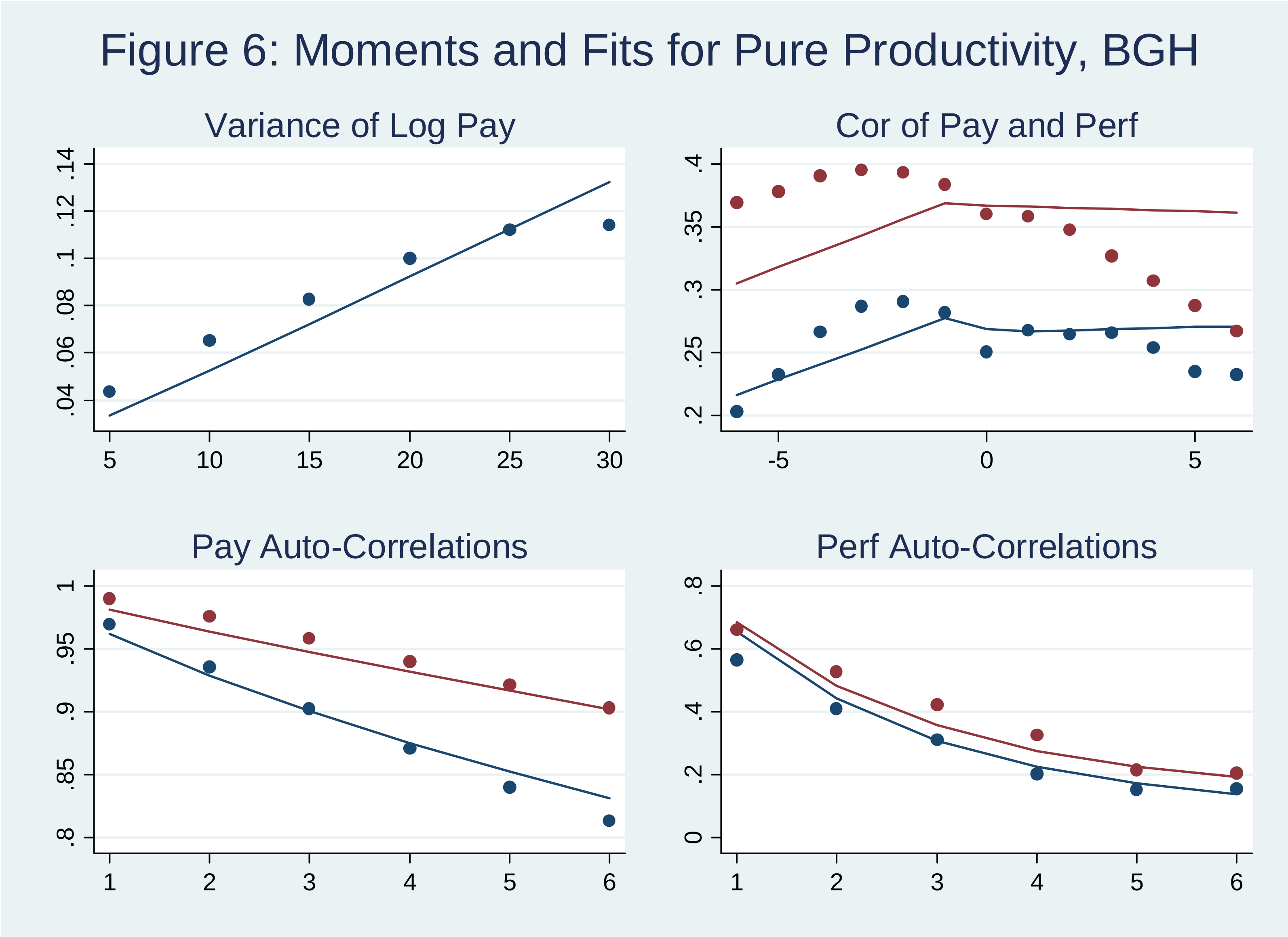


Figure 6: Moments and Fits for Pure Productivity, BGH
Variance of Log Pay
.04 .06 .08 .1 .12 .14
5 10 15 20 25 30
Cor of Pay and Perf
.2 .25 .3 .35 .4
-5 0 5
Pay Auto-Correlations
.8 .85 .9 .95 1
1 2 3 4 5 6
Perf Auto-Correlations
0 .2 .4 .6 .8
1 2 3 4 5 6

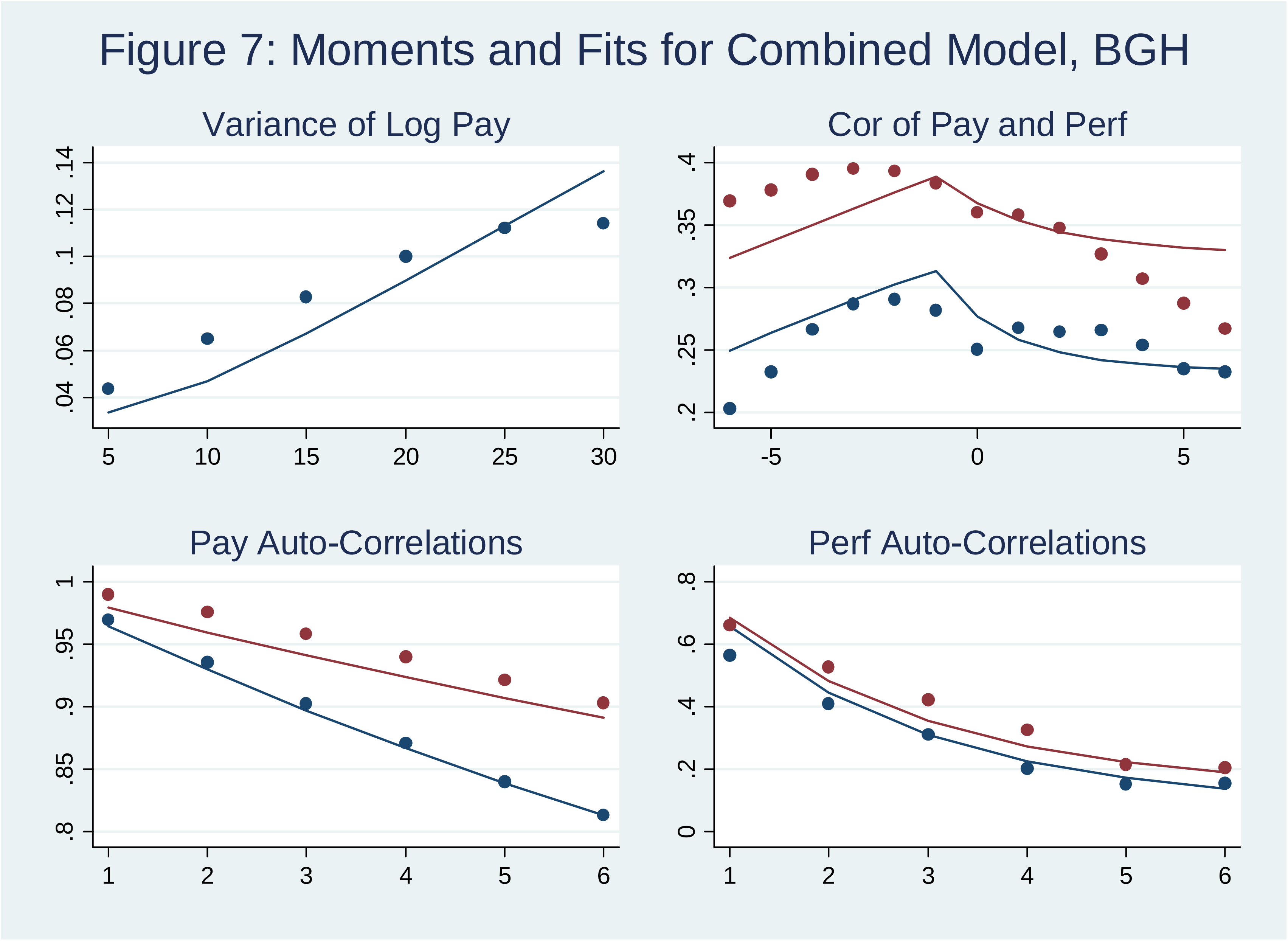


Figure 7: Moments and Fits for Combined Model, BGH
Variance of Log Pay
.04 .06 .08 .1 .12 .14
5 10 15 20 25 30
Cor of Pay and Perf
.2 .25 .3 .35 .4
-5 0 5
Pay Auto-Correlations
.8 .85 .9 .95 1
1 2 3 4 5 6
Perf Auto-Correlations
0 .2 .4 .6 .8
1 2 3 4 5 6


Productivity, Wage, and Error Variances
Full Model
0
.05
.1
.15
.2
0
10
20
30
Experience
Productivity
Wage
Firm Expectation Error

Table 1: BGH Summary Statistics

Years		1969-1988
Data Description		Managers of a medium-sized US firm in the service sector
# Employees[1]		9373
# Employee-years		52697
% Male		75.7%
% White		88.8%
Age		37.7 (7.67)
Education		
% HS		16.9%
% Some College		18.0%
% College		37.1%
% Advanced		27.9%
Salary[2]		$53,332 (24209) [n=50477]
Performance[3]		3.15 (0.706) [n=35856]
Performance Distribution		
	1	0.008
	2	0.162
	3	0.503
	4	0.328

Notes: Parentheses contain standard deviations.

1. Sample includes all employees who can be observed between the ages of 25 and 54, with a non-missing education variable and a non-missing value for at least one of the following comparisons: auto-correlation in current pay and up to 6 year lag in pay, auto-correlation in current performance and up to 6 year lag in performance, correlation between current pay and up to 6 year lags or leads in performance.

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: Serial Correlations of Pay Changes and Previous

	Log Pay Change[1]				
Last Year Change[2]	0.209**				
	[0.00770]				
2 Years Ago Change		0.154**			
		[0.00812]			
3 Years Ago Change			0.121**		
			[0.00852]		
4 Years Ago Change				0.0742**	
				[0.00938]	
5 Years Ago Change					0.0596**
					[0.0113]
Constant	0.0360**	0.00681**	0.0340**	0.0414**	0.0418**
	[0.0114]	[0.00151]	[0.0125]	[0.00693]	[0.0143]
Observations	33672	26999	21737	17488	14066
R-squared	0.078	0.059	0.051	0.043	0.037

Robust standard errors in brackets, clustered by worker.

** p<0.01, * p<0.05, + p<0.1

1. Equals log pay residual in year t minus log pay residual in year t-1. Pay are residualized by age interacted with education, race and gender and year interacted with these variables.

2. Equals log pay residual in year t-1 minus log pay residual in year t-2.

Note: Each column presents results from a separate regression. Sample selection criteria are based on non-missing log pay change and the specific lag change, as well as restrictions noted in table 1.

Table 3: Variance-Covariance Matrix of Pay Changes

(n=27,577)	Log Pay Change	Last Year Change	2 Years Ago Change
Log Pay Change[1]	0.0029		
Last Year Change[2]	0.00068	0.0029	
2 Years Ago Change	0.00051	0.00070	0.0029

1. Equals log pay residual in year t minus log pay residual in year t-1. Pay are residualized by age interacted with education, race and gender and year interacted with these variables.

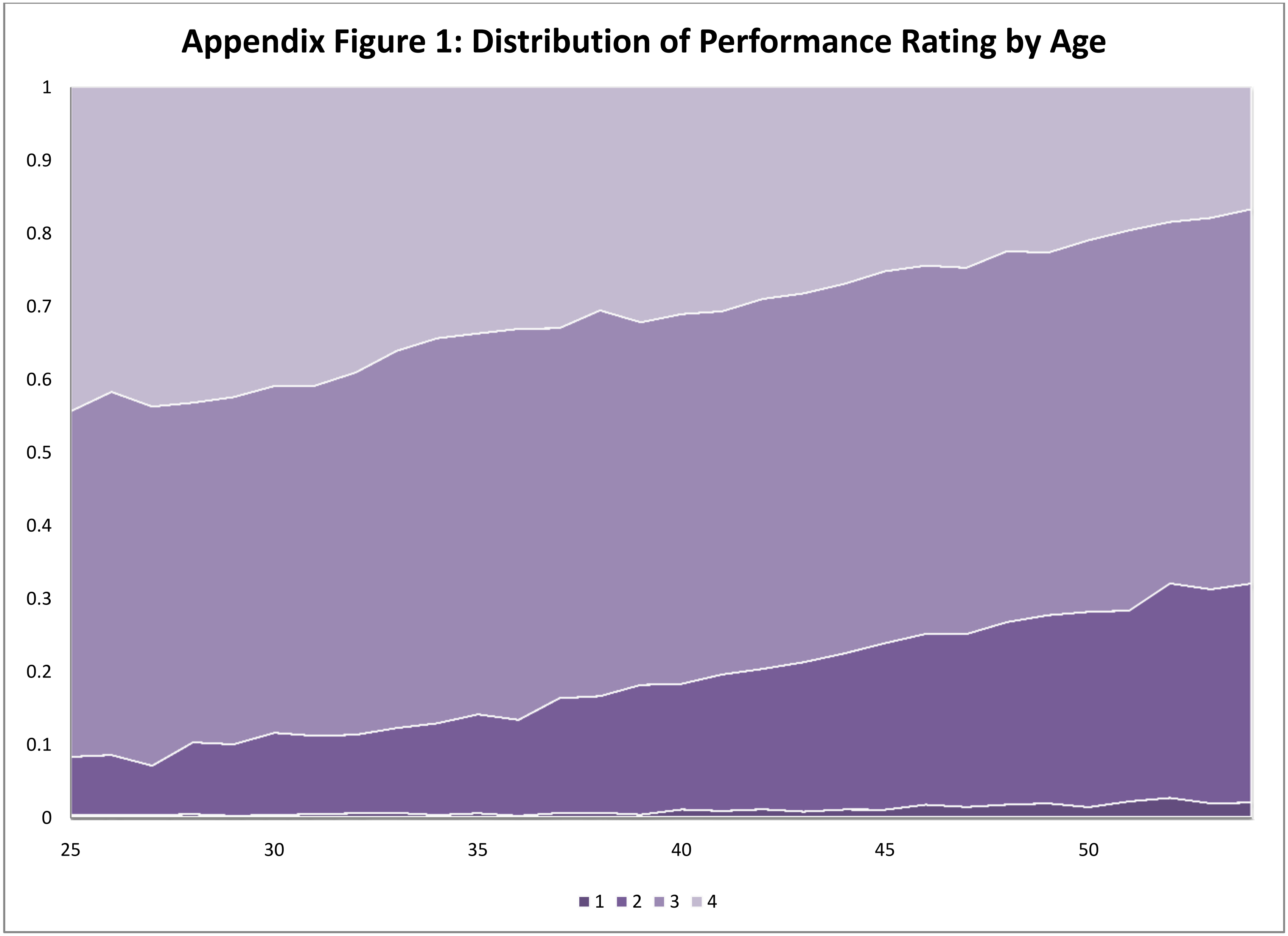
2. Equals log pay residual in year t-1 minus log pay residual in year t-2.

Note: Sample is restricted to those with non-missing values for all 3 pay changes, as well as restrictions noted in table 1.

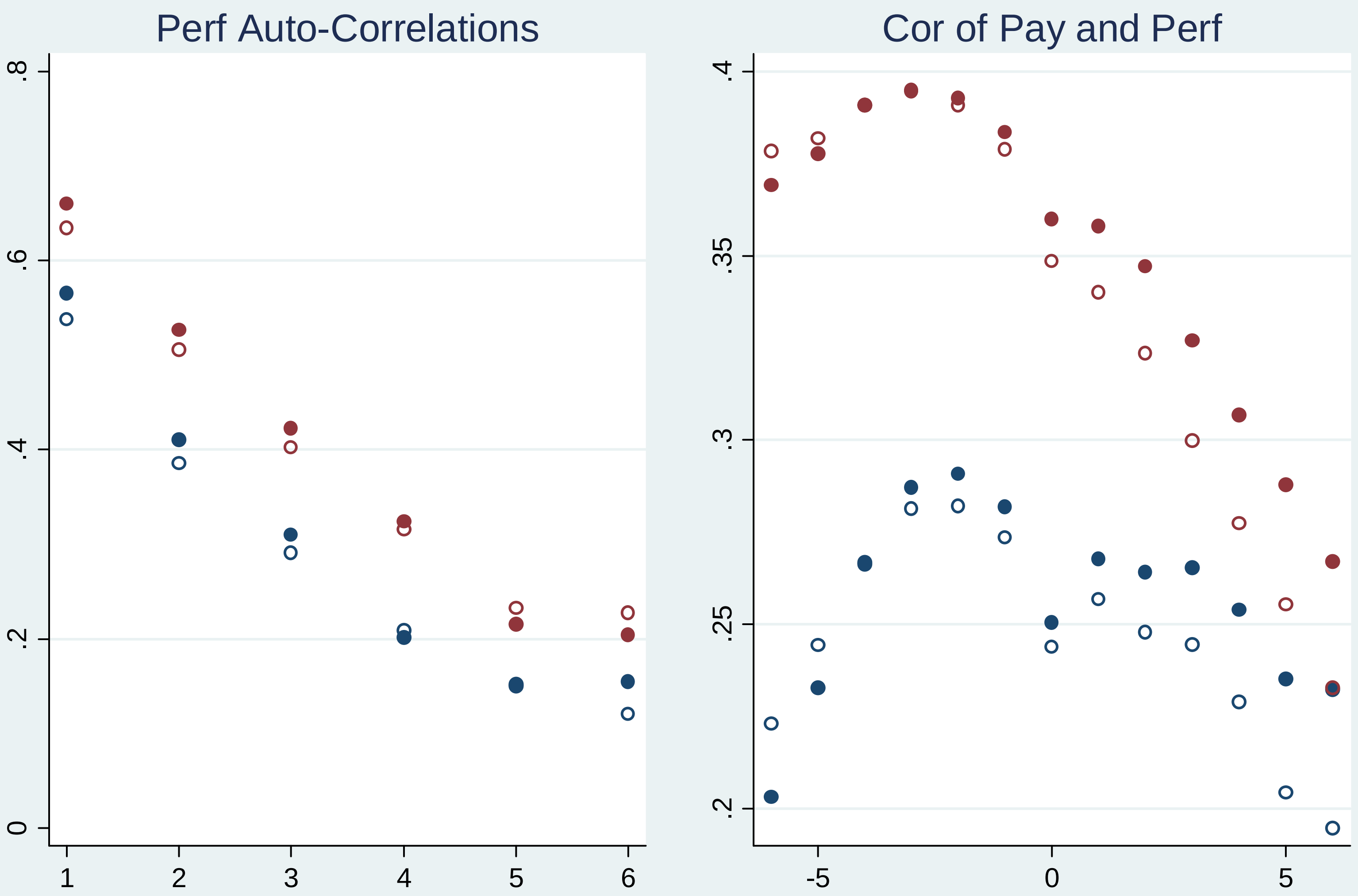
Table 4: Parameter Estimates for 3 Models

	Employer Learning	Productivity	Combined
σ_q^2	0.118 (0.0033)	0.024 (0.0028)	0.030 (0.0035)
σ_r^2	0	0.0040 (0.00026)	0.0045 (0.00027)
σ_0^2	0.374 (0.030)	0	0.000 (0.0031)
σ_u^2	0.653 (0.035)	0.409 (0.015)	0.502 (0.0024)
σ_ω^2	0.0049 (0.00035)	0.000 (0.00010)	0.000 (0.000)
σ_κ^2	0	0.000 (0.000)	0.000005 (0.00000031)
ρ	0.643 (0.0088)	0.634 (0.0072)	0.636 (0.0089)
σ_z^2	0.494 (0.074)	0	0.094 (0.032)

Reported are the parameter values for the pure employer learning model (Section 3.2), the pure productivity model (Section 3.1) and combined model (Section 3.3). 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.


Appendix Figure 1: Distribution of Performance Rating by Age
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
25
30
35
40
45
50
1 2 3 4

Appendix Figure 2: Polychoric vs. Pearson Correlations



Appendix Table 1: Probability of Exit as a Function of Pay and Performance

	I	II	III
Log(salary)	-0.017*		-0.0010
	[0.0082]		[0.013]
Perf=2		-0.088*	-0.116**
		[0.039]	[0.041]
Perf=3		-0.139**	-0.163**
		[0.038]	[0.041]
Perf=4		-0.162**	-0.189**
		[0.039]	[0.041]
Constant	0.167*	0.131**	0.170
	[0.085]	[0.050]	[0.133]
controls[1]	yes	yes	yes
Observations	21443	16444	15234
R-squared	0.185	0.165	0.169

** p<0.01, * p<0.05, + p<0.1

Standard errirs in brackets, clustered by individual.

1. Controls for gender, race, age and year fixed effects.