USC FBE APPLIED ECONOMICS WORKSHOP
presented by Lisa Kahn
FRIDAY, Oct. 23, 2009
1:30 pm - 3:00 pm, **Room: ACC-201**

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

PRELIMINARY AND INCOMPLETE: DO NOT CITE OR CIRCULATE

Lisa B. Kahn and Fabian Lange*
Yale University

October 19, 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.

*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,

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.

1 Introduction

Understanding how wages evolve over the life cycle and understanding the reasons for wage dispersion in the population are among the central questions of labor economics. The pre-dominant answers to these questions are based on the idea that wages reflect the worker's productivity. In recent decades, the literature on employer learning (EL) has offered a competing interpretation of how wage residuals evolve as workers accumulate experience. EL is based on the idea that employers are imperfectly informed about worker productivity and learn about worker productivity as workers age. 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]. Unfortunately, we cannot use these observations to reject the hypothesis that wages equal productivity at all times unless we restrict how workers productivity evolves over the life cycle.[2] Indeed, without any restrictions on the productivity process, any variation in wage residuals over the life cycle is consistent with the assumption that wages equal productivity.

The main obstacle in studying these questions is that most data sets do not contain direct measures of individual productivity that allow separating performance from pay. In this paper, we provide new evidence on whether employer learning or changes in the productivity of workers drive changes in

[1] 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.

wage residuals over the life cycle. This evidence is based on firm-level data containing wages and performance evaluations.

Our data, previously analyzed in Baker, Gibbs and Holmstrom (1994a and 1994b) consist of a 20-year unbalanced panel of all managerial employees in one firm.[3] These data have the crucial advantage that they contain both annual pay of workers as well as ratings reflecting the performance of workers on the job. The panel structure of the data is a great advantage for studying information models because we know what performance ratings were in the firm's information set in a given period. The panel data therefore provide us with information about worker productivity that the firm was not able to exploit when setting wages. This data structure allows us to test important implications of the learning and pure productivity models for how pay correlates with past and future performance ratings. For instance, the learning model predicts that wages correlate more with past rather than future performance measures. By contrast, the perfect information model does not imply such an asymmetry.

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. The variation in wages therefore 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 the 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 learning and productivity processes interact in setting wages.

We find that neither model can fully reproduce the moments of the data. The pure learning model for instance predicts that the difference between the correlations of wages with past and with future performance measures declines

[3] Our data have been used in the landmark studies of Baker, Gibbs and Holmstrom (hereafter BGH), to provide evidence on many features of internal labor markets (for example, ports of entry, cohort effects and fast tracks). This work was extremely influential in the field of organizational economics and 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. Our main innovation is in analyzing moments not before exploited in the data: auto-correlations in performance measures over time and the correlations between pay and performance measures (both past and future).

with experience. We however observe the opposite. 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.

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 explains why we observe that the correlations between performance rating and wages display properties of 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 wages 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 and section 3 provides some reduced-form evidence for both employer learning and producivity evolution. We then present our model, combining these two features, in section 4. We describe our estimation strategy and discuss the identification of the parameters of the model in section 5 [incomplete]. Results are presented in section 6 [missing], followed by our interpretation and conclusion in section 7 [missing].

2 Data

In this paper, we analyze data first used in the canonical studies of Baker, Gibbs, and Holmstrom (1994a,b) of the internal organization of the firm (hereafter, BGH). The data consists of personnel records for managerial employees of a large, 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 11,126 employees with a non-missing education variable who can be observed between the ages of 25 and 55. 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. An employee observation is useful to us if he or she can contribute at least one comparison of the following kinds: an auto-correlation in either pay or performance rating across a time gap of up to 6 years, or a correlation between pay with a performance rating also obtained 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. 9,426 employees and 53,466 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.[4] 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.

Figure 1 illustrates both how log earnings vary with age and how the variance around the earnings profile varies with age. The solid line graphs the log of

[4] We have information on bonus pay for some years (1981-1988) but do not include it in the analysis to remain consist. 20% of workers receive a bonus in these years. Conditional on receiving a bonus, the amount is on average 11% of base salary.

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 increasing almost linearly with age. It is only after age 45 that we observe the increase in the variance of wage residuals to slow down. The figure also illustrates that there is substantial variation in log wage residuals. Understanding this variation and its increase over the life-cycle is the primary task of this paper.

The performance ratings in the data range from 1 to 5, 5 being the worst rating. We combine 4 and 5 since the latter is almost never used and recode so that the highest rating is the best rating. 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 study of performance ratings 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.[5]

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. The dashed line in figure two shows the performance-experience profile is essentially flat when ratings have been residualized on person fixed effects. It could be that productivity grows with experience similarly for workers in the

[5] 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.

same comparison group, so relative productivity remains constant.

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, $\widetilde{p}_{it}$.

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

A worker is assigned the ranking $\widetilde{p}_{it} = k$, if her latent productivity signal falls between the two thresholds c_{k-1} and c_k. These thresholds differ across reference groups defined by observables, X, interacted with each level of experience, t. The vector X includes indicators for race, gender, and education, which implies that we assume that workers are being evaluated within these groups.[6]

Below, we make a number of assumptions that ensure that the latent signal p_{it} is normally distributed. This assumption allows 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 can not identify the variance of p_{it}.

3 Reduced-Form Analysis

We now consider two models in isolation. The first is a pure employer-learning model, in which productivity is fixed and unknown by firms at point of entry. Firms learn by receiving noisy signals of productivity and update in a Bayesian way. The second is a pure productivity model in which firms have full information but productivity evolves flexibly over time. We present pieces of evidence that are consistent with each model in isolation. In the following section, we nest the two models and estimate the parameters of the joint model.

3.1 Pure Employer Learning

Assume that worker productivity, q_i does not evolve over time but is distributed normally in the population with mean 0, standard deviation σ_q^2. Assume further

[6] 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.

that at point of entry, firms know nothing about an individual worker's ability. Instead, suppose firms observe signals, p_{it} each period, where $p_{it} = q_i + \varepsilon_{it}^p$ and $\varepsilon_{it}^p \sim N\left(0, \sigma_p^2\right)$. If the ε^p are independently distributed across time and independent of q_i then we have a standard signal extraction problem. Under symmetric information and perfect competition, firms will set pay equal to expected productivity conditional on their information set in a given period. The equilibrium wage is given in equation 2, where I^t denotes information the firm has received up to time t.

$$\begin{aligned} w_{it} &= E\left[q_i | I^t\right] = (1 - K_t) * 0 + K_t \left(\frac{1}{t} \sum_{j=0}^{t-1} p_{ij} \right) \\ K_t &= \frac{t\sigma_q^2}{t\sigma_q^2 + \sigma_p^2} \end{aligned} \tag{2}$$

This model has 3 important implications about the covariance between pay and performance measures summarized in equation 3:

$$cov(w_{it}, p_{i\tau}) = \left\{ \begin{array}{cc} K_t(\sigma_q^2 + \frac{1}{t}\sigma_p^2) & \tau < t \\ K_t\sigma_q^2 & \tau \geq t \end{array} \right\} \tag{3}$$

First, $cov(w_{it}, p_{i\tau})$ is increasing with t, because K_t, the weight placed on the stream of performance measures, is increasing in t. Intuitively, both wage and performance reflect measures of true productivity plus noise. As the firm learns and incorporates this learning into wages, the noise component of the wage falls. Since the noise in performance is not changing, 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 ($\tau < t$). This is because current pay incorporates the realizations of ε^p from previously observed performance measures. However, since ε^p are independently distributed, pay cannot incorporate future ε^p. This model predicts that the relationship between $cov(w_{it}, p_{i\tau})$ and τ will be a step function. For all $\tau < t$, covariances should be equal, and for all $\tau \geq t$ covariances should be equal, but the former will be larger in level.

Third, this step, i.e., difference in covariances between wages and past compared to future performance measures, should be decreasing in t. Mathematically, this is because the value of the step, $\frac{1}{t}\sigma_p^2$, is decreasing in t. Intuitively, as firms' expectations become more precise, through observing more signals, they will put less weight on any given signal, so the impact of a future signal will be small.

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. We will use these residuals throughout the paper. We then estimate separate regressions for current wage residual on 6 leads and lags of the performance measures.[7] We also estimate these regressions separately for two age groups, 25-39 and 40-55, to test the first and third predictions.

Figure 3 plots these coefficients as well as their 95% confidence intervals. The x-axis shows timing of performance measures where negatives indicate performance measures that occured before the current wage was set while 0 to 6 occurred after. 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 for higher experience years. Second, impact of performance on pay is clearly larger for measures that occurred in the past. The black vertical line indicates the timing of the last performance measure observed before pay was set. For more experienced workers, we can clearly see a step to the right of this line.

Thus the first two predictions hold while the third fails. It does appear as though firms incorporate information from past performance measures which is consistent with a pure learning model. However, the firm is incorporating much more information about past performance for older workers, suggesting the firm is learning more about the more experienced rather than the younger workers. This finding is inconsistent with a pure employer learning model.

Before we present reduced form evidence on the pure productivity model, we 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 find that pay and performance ratings would correlate heavily for the current period. However, under direct incentives both past and future performance evaluations should have no impact on pay, only the most recent performance measure. In figure 3, we should see a large spike at -1. We find absolutely no evidence for direct incentives so conclude that is not a confounding factor.

[7] These regressions are estimated separatel for each performance rating so we do not have to restrict the sample to individuals with non-missing values for all 13 comparisons.

3.2 Pure Productivity

The key features of a pure productivity model are full information, that is the firm perfectly observes productivity at any moment in time, and productivity that evolves over time. Equation 4 shows a generalized form of evolving productivity used in the literature.

$$q_{it} = q_i * f(t) + \varepsilon_{it} \tag{4}$$

According to eq (4), individual productivity q_{it} evolves following a experience profile $f(t)$ but there are two sources of heterogeneity. The variable q_i allows for profile heterogeneity and ε_{it} allows performance to deviate from this profile following a random process. Typically, the literature restricts $f(t)$ to be linear and assumes that ε_{it} follow an ARMA process. Under these assumptions, it is easy to see that individual variation in q_i will introduce persistent correlation in pay changes within individuals.

Table 2 confirms the finding from BGHb that pay changes are positively correlated. For example the correlation between this year's pay change and last year's is 0.20, significant at the 1% level. This suggests that there is something predictable about productivity growth of individuals. This finding is consistent with human capital theory (Becker 1964, Ben-Porath 1967), augmented by an assumption that individuals differ in their preference or human capital production parameters and thus follow different human capital accumulation paths.

We next consider the variation in the idiosyncratic process ε_{it} in relation to the persistent component. We find it intuitive to consider this variation using pay changes rather than levels and thus consider the variance-covariance matrix of pay changes. Suppose productivity takes the particular functional form expressed in equation 5, where ε^r_{it} are uncorrelated over time and uncorrelated with κ_i. That is, productivity evolves as a random walk plus a linear heterogeneous growth term. Suppose $\kappa_i \sim N\left(0, \sigma^2_\kappa\right)$ and $\varepsilon_{it} \sim N\left(0, \sigma^2_r\right)$. Under these assumptions, the variance and covariances of pay changes are given in equations 6 and 7. We can therefore separate the human capital term from the idiosyncratic growth term.

$$q_{it} = q_{it-1} + \kappa_i + \varepsilon^r_{it} \tag{5}$$

$$Var(w_{it} - w_{it-1}) = \sigma^2_\kappa + \sigma^2_r \tag{6}$$

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

Table 3 gives the variance-covariance matrix for the most recent three changes in log pay residuals, where the sample is restricted to those with non-missing values for all three. Here the variance in pay changes is 0.003 while the covariance is almost an order of magnitude smaller, equalling approximately 0.0007. This 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 above and the literature on employer learning (Farber and Gibbons 1996). However, we cannot know whether this large random walk component is driven by an employer learning model or a productivity evolution model until we put structure on the problem and estimate each of the underlying parameters in the model. We do this next.

4 The Model

Our model describes how worker productivity evolves over time and how firms continuously learn about worker productivity. This model nests both a pure productivity model as well as a pure learning model.[8] Because we nest both models, we can test and quantify the relative importance of both models for explaining wage and productivity dynamics.

The model describes both how productivity and the information available to firms evolves stochastically over time. We assume that firms know the relevant parameters of the economy and update their expectations in a Bayesian manner. We also assume that labor markets are spot markets and employer

[8] A pure learning model refers to a model where firms learn about the idiosyncratic component of individual productivity, but this idiosyncratic component is fixed over time. The literature on employer learning (Farber and Gibbons (1996), Altonji and Pierret (2001), Lange (2007), and others) explores what such a model implies for the relationship between earnings and schooling as well as initial productivity signals over the life-cycle. The pure productivity model by contrast assumes that there is no uncertainty about individual productivity, but that individual productivity varies over the life-cycle. The human capital model combined with an assumption of heterogeneity in ability measures suggests that the idiosyncratic component of productivity varies systematically over the life-cycle. By contrast, some purely statistical models instead emphasize random walk shocks to individual productivity.

learning is common, implying that workers are paid their expected product.[9] This conceptual structure defines how productivity correlates available to the firm relate to individual wages.

To complete the description of the model, we need to also describe how the variables of the economic model map into the variables observed in our data. The main assumptions here are that we only observe a subset of the performance correlates that the firm observes. Furthermore, these observed performance measures are ordinal categorical variables, based on a subset of continuous performance correlates observed only by the firm. Finally, we allow for measurement error in wages.

We begin by describing how worker productivity evolves with experience. We assume that we can summarize a worker's productivity using a single variable $\widetilde{Q}_{it}$. Workers 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 $Q_{i,t}$ is the idiosyncratic component of individual productivity. Let $q_{it} = \log(Q_{it})$.[10] The function $Q(x,t)$ is common knowledge, but the component $Q_{i,t}$ is only partially observed by firms. [11]

We assume that q_{it} evolves stochastically according to the following difference equation.

$$q_{it+1} = q_{it} + \kappa_i + \varepsilon^r_{it+1} \tag{8}$$

where q_{i1} and κ_i capture the heterogeneity in initial productivity and subsequent productivity growth, respectively. They are drawn from independent normal distributions $N(0, \sigma^2_q)$ and $N(0, \sigma^2_\kappa)$.[12] The heterogeneity in the drift parameter κ_i captures differences in persistent ability or taste differences across individuals that lead to persistent differences in the intensity with which individuals accumulate human capital over their life-cycle. The heterogeneity in q_{i1}

[9] A large literature deviates from the assumptions of spot markets and symmetric information. We are sympathetic to this literature and believe it could be important in describing the labor market. However, it would be intractible to include features of these models in our estimation. What is important for us is despite evidence of the existence of these market imperfection, evidence also exists that firms are constrained by market forces. For example, BGH (1994b) find that this firm does not fully shelter pay from market fluctuations.

[10] 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.

[11] From now on, we will suppress the dependence on controls x.

[12] We make a number of normality assumptions as we develop the model. This allows us to derive the covariance structure of wages and observed performance measures in closed form as a function of the parameters. In turn, this means that we can estimate the moments of the model using relatively simple method of moment estimators.

captures differences in initial ability. Finally, the innovations ε_{it}^r capture time-variation in individual productivity that are not predictable. We assume that these innovations ε_{it}^r are iid Normal with mean 0 and variance σ_r^2. We therefore assume 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 will 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)

Individual, idiosyncratic productivity is not directly observed by firms. Instead, firms observe correlates of worker productivity. The learning process is divided into two parts. First, the firm draws a signal z_{i0} about the productivity of a worker at the beginning of his career. We might think of this initial signal as the information contained in the CV or obtained in the interview. We assume that

$$z_{i0} = q_{i1} + \varepsilon_{i0}^z \tag{9}$$

where the signal noise $\varepsilon_{i0}^z \tilde{} N\left(0, \sigma_0^2\right)$ is independent of all other variables.

In subsequent periods, the firm continues to obtain information about individual productivity. Our data contains one correlate of individual productivity, a manager rating of individual employees. However, we do want to allow for firms to learn about workers in ways that are not observed by us. For this reason, we assume that the firm obtains two signals on individual productivity in each period, one of which is (partially) observed by us. These signals are (z_{it}, p_{it}), where p_{it} represents the signal unobserved to us. We assume, without loss of generality, that z_{it} is orthogonal to p_{it}.[13]

$$z_{it} = q_{it} + \varepsilon_{it}^z \tag{10}$$

$$p_{it} = q_{it} + \varepsilon_{it}^p \tag{11}$$

The ε_{it}^z are iid and distributed normally with mean zero and variance σ_z^2. As we estimated the learning model, we realized that there is a relatively high

[13] The information contained in two correlated normal signals (z_{it}, p_{it}) is the same as the informatoin contained in two uncorrelated signals (rz_{it}, p_{it}), where rz_{it} is the component of z_{it} orthogonal to p_{it}.

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 12

$$\varepsilon_{it+1}^{p} = \rho\varepsilon_{it}^{p} + u_{it+1} \tag{12}$$

where the initial noise is $\varepsilon_{i1}^{p} = 0$ and $u_{it} \tilde{} N\left(0, \sigma_u^2\right)$. The parameter ρ is one of the parameters that we will estimate. Our data contains manager performance ratings. We associate these ratings with the signals p_{it}. The observed ratings do not however follow a normal distribution, but are instead ordinal variables with a finite set of (k) support points. We 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:

$$P_{it} = \sum_{i=1}^{k} 1\left(p_{it} \geq c_{kt}\right) \tag{13}$$

Note at this point, that the intercepts c_{kt} differ by experience t. This implies that the manager rankings are assumed to compare workers within their experience level. It is possible to refine the comparison group further. For instance, we could impose that workers are being ranked against workers with similar experience and education or against workers with similar experience, education, and job level.

The firm will try to predict productivity q_{it} at each t. Because we assume that labor markets are frictionless spot markets and all information is common, we have that wages will equal expected productivity: $W_{it} = E\left[Q\left(x,t\right)Q_{it}\right] = E\left[Q\left(x,t\right)\exp\left(q_{it}\right)\right]$. However, in order to know how to form expectations over productivity, the firm will also need to predict the individual growth term κ_i and the current "chumminess" of the manager. The former is required to update expected productivity next period based on current productivity. The estimate of the latter affects the weights placed on different signals that the firm receives. In each period, the firm will therefore form expectations over three objects: $\left(q_{it}, \kappa_i, \varepsilon_{it}^{p}\right)$. Denote this vector as θ_{it} and let $\widehat{\theta}_{it} = E\left[\theta_{it}|I_{it}\right]$ represent the expected value of θ_{it} based on all information available to firms about individuals at time t.

As already mentioned, we make the necessary assumptions (frictionless labor markets with common assumption) that ensure that wages equal expected productivity. We allow for measurement error in our data and write:

$$W_{i,t} = W_{i,t}^{*}\Omega_{i,t} \tag{14}$$

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

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:

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

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

$$\begin{aligned} w_{it} &= \left(q(t) + \frac{1}{2}v(t)\right) + \widehat{q}_{it} + \omega_{it} \\ &= h(t) + \widehat{q}_{it} + \omega_{it} \end{aligned} \tag{15}$$

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

At this point, it will be useful to point out how this model nests the pure productivity and the pure learning model. There are 8 parameters $\left(\sigma_q^2, \sigma_r^2, \sigma_0^2, \sigma_u^2, \sigma_\omega^2, \sigma_{q,\kappa}, \sigma_\kappa^2, \rho, \sigma_z^2\right)$. The pure productivity model is a model with perfect information on the part of firms. We can represent this by restricting the signal variance of unobserved signals z_{it} to equal 0, so that firms have perfect signals about individual productivity. The pure productivity model therefore sets $\sigma_0^2 = 0$ and $\sigma_z^2 = 0$. The role of the signal variance σ_p^2 and of ρ in this model is simply to allow observed ratings to display a distribution conditional on wages. By contrast, in the pure learning model, the idiosyncratic component of productivity is fixed conditional on the initial productivity parameter. The only variation in this model over time comes from the learning process. We can represent this model by setting $\sigma_r^2 = \sigma_\kappa^2$. Therefore the above model nests both the pure productivity model and the pure learning model and can be used to test these against each other.

In Appendix A, we represent the updating problem faced by the firm as a Kalman filter. We describe how our model is a special case of a more general recursive formulation. In this appendix, we also solve for the covariances of w_{it} and performance indices p_{it} as functions of the parameters. This mapping from the parameters into the second moments of wages and performance indices allows us to apply methods of moments to estimate the parameters. However, while the autocovariances of wages are observed, the covariances of performance measures p_{it} with wages and the autocovariances of p_{it} are not directly observed. Instead, we can estimate the correlations of p_{it} with w_{it} and the autocorrelations of p_{it} using the observed performance measures P_{it}. Our model therefore delivers a mapping from parameters to observed moments that allows us to estimate the parameters of the model.

5 Estimation and Identification

To simplify our empirical problem, we match 56 moments in the data. These include the following: the variance in pay for 5-year experience groupings from 0 to 30 years expereience, the autocorrelations of pay and 6 lags separately for two 15-year experience groupings from 0 to 30 years, the autocorrelations 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 the same experience groups.

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 and concave in experience. Second both pay and performance auto-correlations are higher for the older experience group. Third, the correlation between pay and performance is on average higher for lagged performance, as exhibited in figure 3. We also again see that correlations are higher for the older experience group and the step size between past and future performances is larger for the older group. When we discuss identification below, we will explain the intuition for these patterns as well as how they dictate the estimate.

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 and are shown in parentheses. 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 fit and identification for each model in turn.

5.1 Pure Learning

In the pure learning model, the idiosyncratic component of productivity does not vary over the life-cycle and wages vary only over the life-cycle because firms learn about individual ability. To impose constant productivity on our model, we restrict the variance of the random walk and of the heterogenous growth component to 0: $\sigma_r^2 = \sigma_\kappa^2 = 0$. The free parameters are those that determine how firms learn about worker productivity. They are 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 the two parameters $\left(\rho, \sigma_u^2\right)$ governing the variation in the signal observed both in the data and by firms.

We begin by noting that in pure learning models the variance in measured wages asymptotes to the variance of measurement error plus the variance in idiosyncratic productivity: $lim_{t\to\infty}\left(v\left(w_t\right)\right) = \sigma_q^2 + \sigma_\omega^2$. Because the measurement error is classical, the covariances in log wages at high experience levels goes to the σ_q^2. Thus, using the covariance and the variance of wages at high experience levels, we can obtain the variances of idiosyncratic productivity and of the measurement error in wages.

In figure 4, we report moments of 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 0.13 and that the auto-correlation in log wages at these experience levels is about 0.9. 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 0.01. These are indeed the values for our estimates of $\left(\sigma_q^2, \sigma_\omega^2\right)$ we report in table 4 for the pure learning model.

We now turn to show how the auto-correlations of p_{it} with p_{it+k} at different leads and lags k inform us about the parameters $\left(\rho, \sigma_u^2\right)$ that govern the signal noise ε_{it}^p. The limiting distribution of p_{it} as t grows depends on the parameters

ρ and σ_u^2 in the autoregressive specification of p_{it}. From eq **??**, we get:

$$\lim_{t\to\infty} var\left(p_{it}\right) = \sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2} \tag{16}$$

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:

$$\lim_{t\to\infty} cor(p_{it}, p_{it+1}) = \frac{\sigma_q^2 + \rho\frac{\sigma_u^2}{1-\rho^2}}{\sigma_q^2 + \frac{\sigma_u^2}{1-\rho^2}} \tag{17}$$

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 $\left(\sigma_u^2,\rho\right)$. The autocorrelation in p_{it} depends primarily on the parameter ρ.The tight link between ρ and the observed decline in the autocorrelations in $cor\left(p_{it}, p_{it+k}\right)$ at high k therefore determines ρ. Figure 4, shows that the first order autocorrelation 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 0.65 for the pure learning model.

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

$$\lim_{t\to\infty} cor(p_{it}, p_{it+k}) = \frac{\left(1-\rho^2\right)\sigma_q^2 + \rho^k\sigma_u^2}{\left(1-\rho^2\right)\sigma_q^2 + \sigma_u^2} \tag{18}$$

For high k, with $\rho < 1$, this correlation will increasingly depend on σ_u^2. Using the estimates of $\left(\rho = 0.65, \sigma_q^2 = 0.12\right)$ found above as well as an autocorrelation across 5 lags for p_{it} of about 0.22[14], we find from eq. (18) an approximate value of 0.68 for σ_u^2. Our estimated value for σ_u^2 for the pure employer learning model is 0.70.

This leaves us with only the parameters for the signal of σ_z^2 and σ_0^2 to identify. The parameter σ_0^2 determines how much information the firm has about workers as these begin their careers. The parameter σ_z^2, together with $\left(\rho, \sigma_u^2\right)$, determines how fast employers learn about workers productive abilities. To identify these parameters, we exploit the close link that the learning model establish between the variance of wages and the amount of information that firms have

[14] See Table @

at any moment in time. Consider therefore the variance of wages at t=0 and the speed with which the variance of wages grows as experience accumulates.

At the beginning of workers careers (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 the idiosyncratic component. 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 0.57, about the same size as the variance in the signal noise of z_{i0}. Together, these parameter values reproduce relatively accurately the increase in the variance of the wage from about 0.04 to about 0.11 over the first 30 years of these individuals careers. These values also reproduce the slightly concave shape we observe in the experience profile of wage residuals.

The learning model therefore does succeed in a number of ways. It matches the autocorrelations 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. It also matches, at least very roughly, the level of the autocorrelations in performance and wages.

However, the pure learning model also fails to reproduce a number of patterns in the data. Most importantly, the pure learning model does not match how the correlations in observed performance measures with wages at various lags vary with experience. The contribution of this paper is to exploit the information from observed performance ratings to the analysis of the dynamics of wage residuals, and it is exactly this new information that the pure employer learning model fails to match. To be specific, the pure employer learning model has difficulties with two patterns in the data. We find that the auto-correlations between performance measures and wages are much higher at high experience levels than at low experience levels. And, we also find important differences when we compare across experience how wages correlated with past and with future performance ratings.

The learning model predicts that the correlation of wages with future performance measures increases with experience. This is because wages are progressively based on more and more information about underlying productivity. Future performance measures are noisy measures of this constant productivity.

Therefore the common component in wages and future performance measures increases with experience and correlations of wages and performance ratings in the future should increase as workers age.

The situtation differs when we consider the correlations of wages with past performance ratings, that are already observed when wages are set. When workers are young, firms will place a lot of wait on any given performance rating that has already been observed. This implies that wages and past performance ratings are more highly correlated among young workers than among old workers.

The patterns in the data do not correspond to these predictions of the learning model. Instead, we find that wages are much more closely related with past performance measures (at constant lags) when individuals are older than when they are young. As predicted by the pure learning model, wages are also more closely correlated with future performance measures. However, the difference of the correlation of current wages with past and with future performance measures is substantially larger for high experience levels than it is for low levels of experience. The data suggests that firms rely more heavily on recent performance measures to set wages of their experienced employees than to set wages of young employees. This is inconsistent with the pure employer learning model.

5.2 The Pure Productivity Model

We next discuss how to identify the parameters of the pure productivity model. The pure productivity model 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 impose the variance of the noise of the performance ratings (ε_t^p). This restriction would imply that the performance ratings would be, absent measurement error in wages, perfectly correlated with wages.[15] We can however use the performance ratings as additional evidence regarding the underlying productivity process.

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 $(\sigma_q^2, \sigma_r^2, \sigma_\kappa^2)$ that govern how productivity varies over the life-cycle. To simplify the exposition, we assume

[15] WIth perfect information, there is no reason for firms to elicit manager evaluations about their employees. One might take this as evidence against the pure productivity measure.

that wages are measured without error. We then have

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

In this setting, the variance of wages at the beginning of a career fix the initial variation in productivity across individuals. 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. Finally, the variance in wage growth identifies the, together with σ_κ^2, the variance in the random walk component.

We have not relied on the correlations between wages and performance ratings to identify the parameters of the pure productivity model. The pure productivity model does however have 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 (i) that the correlation of the performance ratings with wages, (ii) the autocorrelations of performance ratings, and (iii) the autocorrelations 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 autocorrelation of performance ratings rises over the same time period from 0.57 to 0.67. The first autocorrelation of wages is likewise increasing 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, there are other features of the data that the pure performance model has difficulties matching. Most importantly, from our perspective, is that the pure productivity model predicts that the correlations of the current wage with past and future performance ratings are (nearly)[16] symmetric in time. The learning model imposes an asymmetry in time, because past performance measures are used for setting current wages, while future performance measures

[16] This statement is not strictly correct, because the variance of productivity increases over time.

can, by definition not be used in setting wages. The pure productivity model does not admit such an asymmetry. Performance ratings at t-k and t+k should be correalted in much the same way with log wages, as long as k is being held constant. 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 difference in the correlations of wages with past and future performance ratings and their increases are not predicted by the pure productivity measure.

Another dimension along with the pure productivity measure does not match the data is with respect to the concave relationship between the variance of wages and experience. The pure productivity model with heterogeneity in κ_i predicts the variance of the wage to follow a strictly convex relationship with experience. If there is no heterogeneity in κ_i, then we would expect a linear relationship. Instead, the data suggests that at relatively high experience levels (above 20), the increase in the variance of wages slows down, so that the variance of wages follows a linear relationship in experience.

We have thus described how the parameters of the pure learning and the pure productivity models are linked to observable moments. We also shown the features of the data that each of these models can not match. The joint model will in fact be able to match these features and estimates from this joint model will allow us to quantify what role the learning and the productivity model play in setting wages. We will next present and interpret the parameter estimates obtained from all three models.

5.3 The Combined Model

6 References

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

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

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

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

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

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

[16] 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.

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

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

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

7 (INCOMPLETE:) Appendix A - The Firm Updating Problem

We have assumed normality throughout. This means we can rely on standard results from Kalman updating to represent the firms learning process. The firms updating problem is of the following type:

The firm forms expectations over the distribution of a dynamically evolving random vector θ_{it} :

$$\begin{aligned} \theta_{it+1} &= \Phi\theta_{it} + \varepsilon_{it}^{\theta} \\ &\varepsilon_{it}^{\theta}\tilde{\ }N(0, R_{\theta}) \end{aligned} \tag{19}$$

At t, the firm holds a prior expectation $\widehat{\theta}_{it|t-1}$, where the subscript t|t-1 denotes that this prior is based on information prior to the receipt of new signals at t. The target θ_{it} is distributed $\theta_{it}\tilde{\ }N\left(\widehat{\theta}_{it|t-1}, P_{t|t-1}\right)$, where $P_{t|t-1}$ denotes the prior variance over θ_{it}.

Information arrives in the form of a signal vector x_{it} :

$$\begin{aligned} x_{i,t} &= H_x'\theta_{it} + \varepsilon_{it}^{x} \\ &\varepsilon_{it}^{x}\tilde{\ }N(0, R_x) \end{aligned} \tag{20}$$

Based on this signal, the firm updates its expectation so that:

$$\begin{aligned}\widehat{\theta}_{it|t} &= \widehat{\theta}_{it|t-1} + P_{t|t-1}H_x\left(H_x{}'P_{t|t-1}H_x + R_x\right)^{-1}\left(x_{it} - H_x'\widehat{\theta}_{it|t-1}\right) \\ &= \widehat{\theta}_{it|t-1} + K_t\left(x_{it} - H_x'\widehat{\theta}_{it|t-1}\right) \\ &= \left(1 - K_tH_x'\right)\widehat{\theta}_{it|t-1} + K_tx_{it}\end{aligned} \tag{21}$$

This formulation shows that the firm will update by comparing the signal to the linear expectation of the signal based on the prior held by the firm. The weight place on this signal depends on the relative variance of the prior ($P_{t|t-1}$) and the signal variance (R_x). If the firm is certain of its prior ($P_{t|t-1} \to 0$) or if there is little information in the signal ($R_x \to \infty$), then little weight is placed on the signal. By contrast, if the prior is diffuse $\left(P_{t|t-1} \to \infty\right)$ relative to the signal ($R_x \to 0$), then the updated prior will depend almost entirely on the new signal.

No new information becomes available between t and t+1, and therefore the posterior in t determines the prior in t+1 based on the dynamic equation (19)

$$\widehat{\theta}_{it+1|t} = \Phi\widehat{\theta}_{it|t} \tag{22}$$

Finally, the prior variance in t+1 can be updated as follows:

$$P_{t+1|t} = \Phi\left(P_{t|t-1} - K_tH_x'P_{t|t-1}\right)\Phi' + R_\theta$$

where K_t is implicitly defined in eq. (21).

This recursion determines the time-series of firm expectations based on signals observed. We can operationalize this structure by mapping the particular assumptions on the learning model introduced in eqs. (8) – (11) above into the matrices defined in equations (19) − (22). This implies that[17]

$$\theta_{it} = \begin{pmatrix} q_{it} \\ \kappa_i \\ \varepsilon^p_{it} \end{pmatrix}$$

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

$$\varepsilon^\theta_{it} = \begin{pmatrix} \varepsilon^r_{it+1} \\ 0 \\ u_{it+1} \end{pmatrix}$$

$$R_\theta = \begin{pmatrix} \sigma^2_r & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2_u \end{pmatrix}$$

$$x_{it} = \begin{pmatrix} p_{it} \\ z_{it} \end{pmatrix}$$

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

$$\varepsilon^x_{it} = \begin{pmatrix} \varepsilon^z_{it} \\ 0 \end{pmatrix}$$

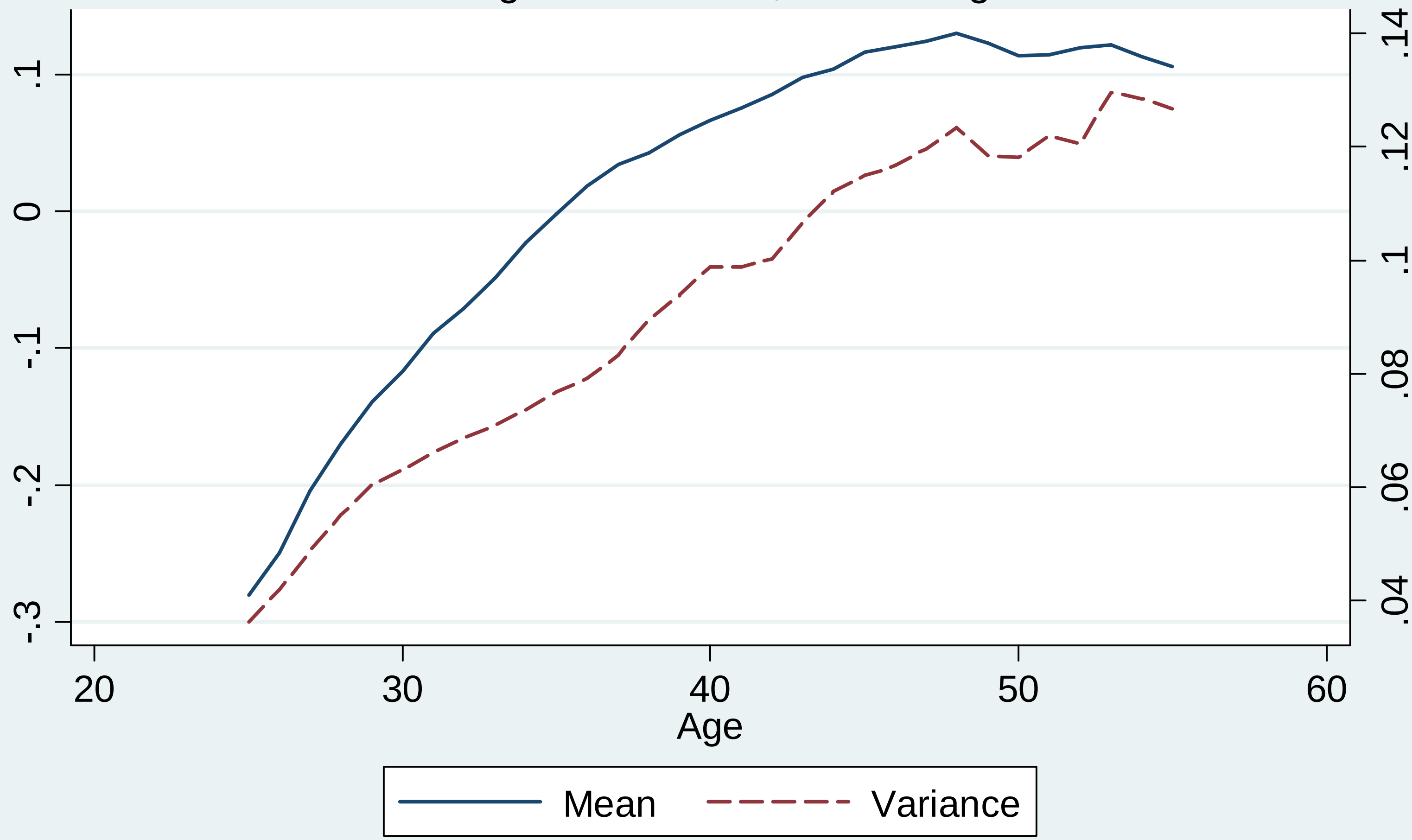
$$R_x = \begin{pmatrix} \sigma^2_z & 0 \\ 0 & 0 \end{pmatrix}$$

This recursive structure fully describes how firms learn about worker productivity on the basis of 2 signals (z_{it}, p_{it}). Combined with eq. (15) and the assumption that workers are paid their expected products, we have a description of the evolution of wages over time. Equation (13) describes how the signals map into the observed manager ratings.

[17] There are various equivalent ways of mapping the model into this learning structure. In examining the matrices below, note that the state contains itself the noise of one of the signals. This leads to the unusual from for H_x and R_x. The updating equations are however unaffected.

Figure 1: Log Wage Residual Means and Variances by Age

Controlling for education, race and gender



Variance are squared residuals, residualized on all of the above and age fixed effects.

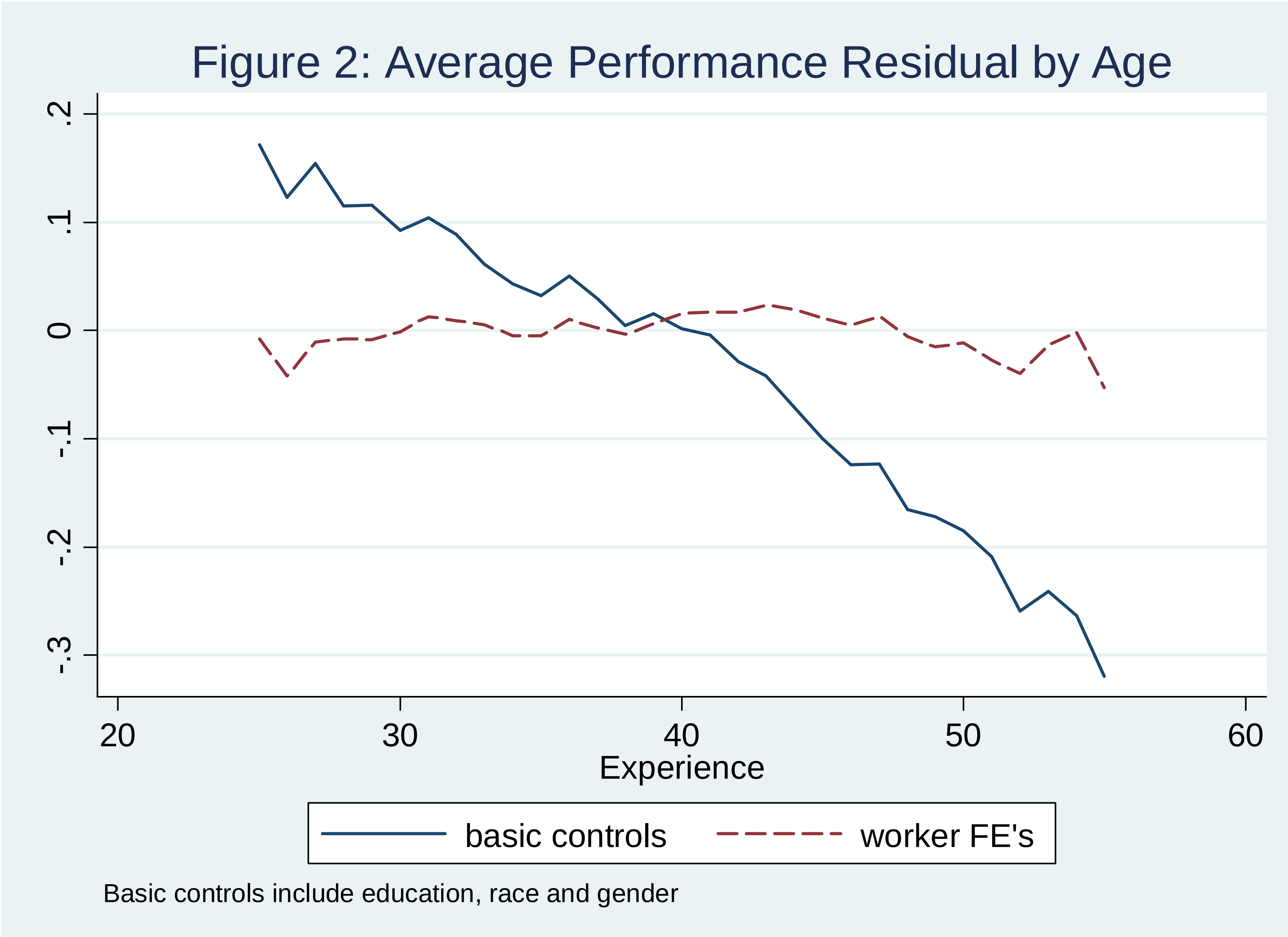


Figure 2: Average Performance Residual by Age
.2
.1
0
-.1
-.2
-.3
20
30
40
50
60
Experience
basic controls
worker FE's
Basic controls include education, race and gender

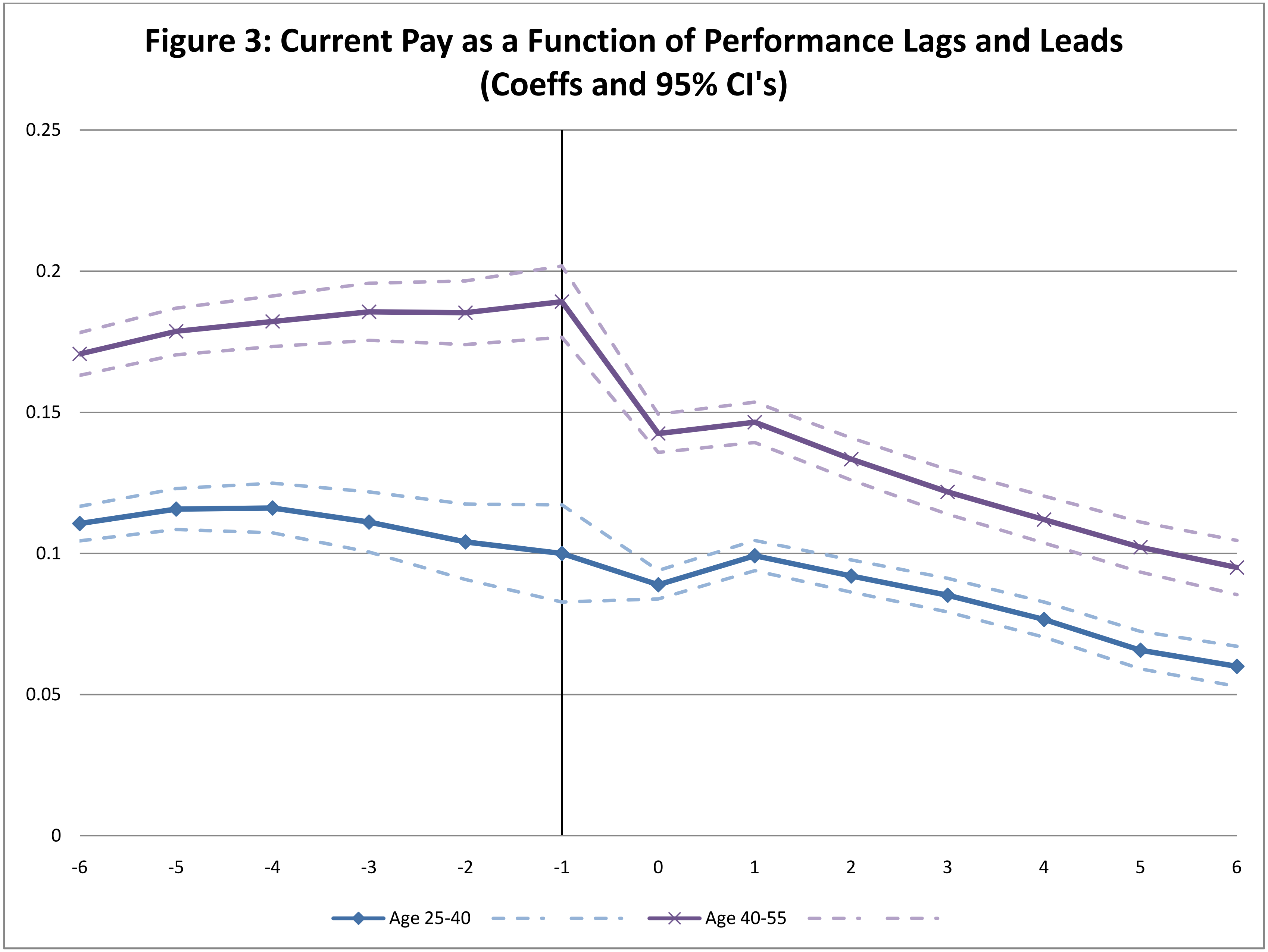


Figure 3: Current Pay as a Function of Performance Lags and Leads
(Coeffs and 95% CI's)
0.25
0.2
0.15
0.1
0.05
0
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
Age 25-40
Age 40-55

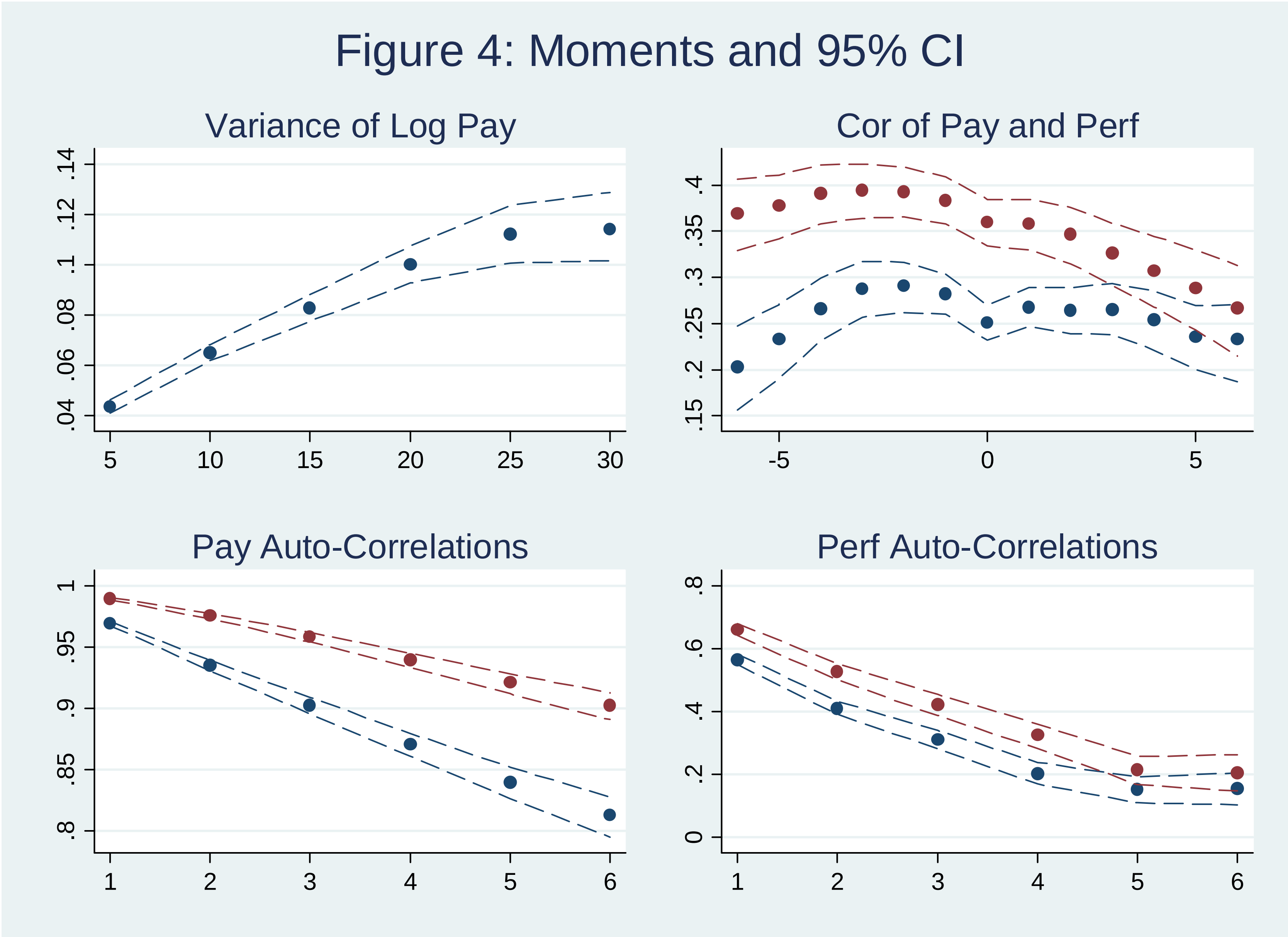


Figure 4: Moments and 95% CI
Variance of Log Pay
.04
.06
.08
.1
.12
.14
5
10
15
20
25
30
Cor of Pay and Perf
.15
.2
.25
.3
.35
.4
-5
0
5
Pay Auto-Correlations
.8
.85
.9
.95
1
1
2
3
4
5
6
Perf Auto-Correlations
0
.2
.4
.6
.8
1
2
3
4
5
6

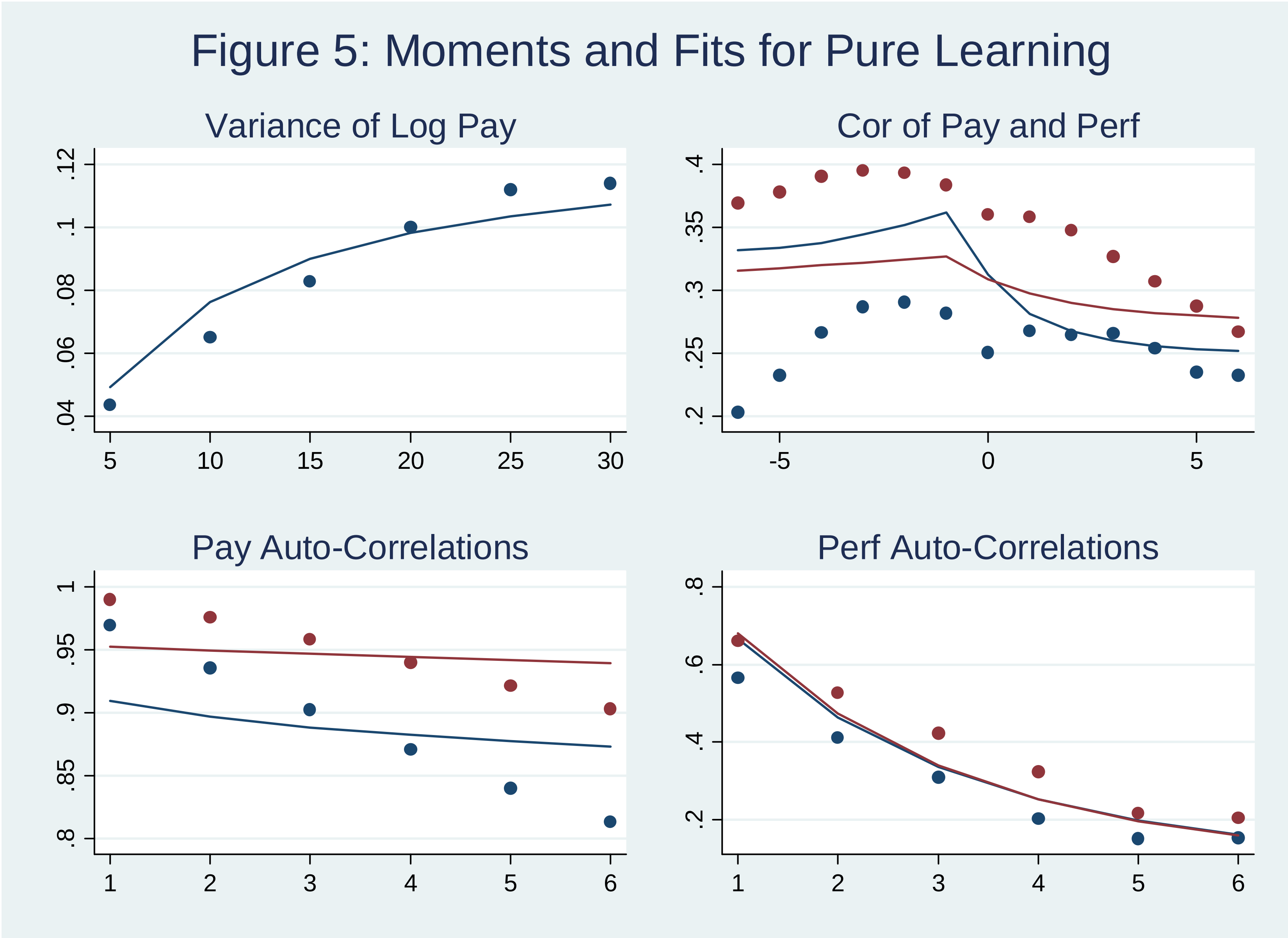
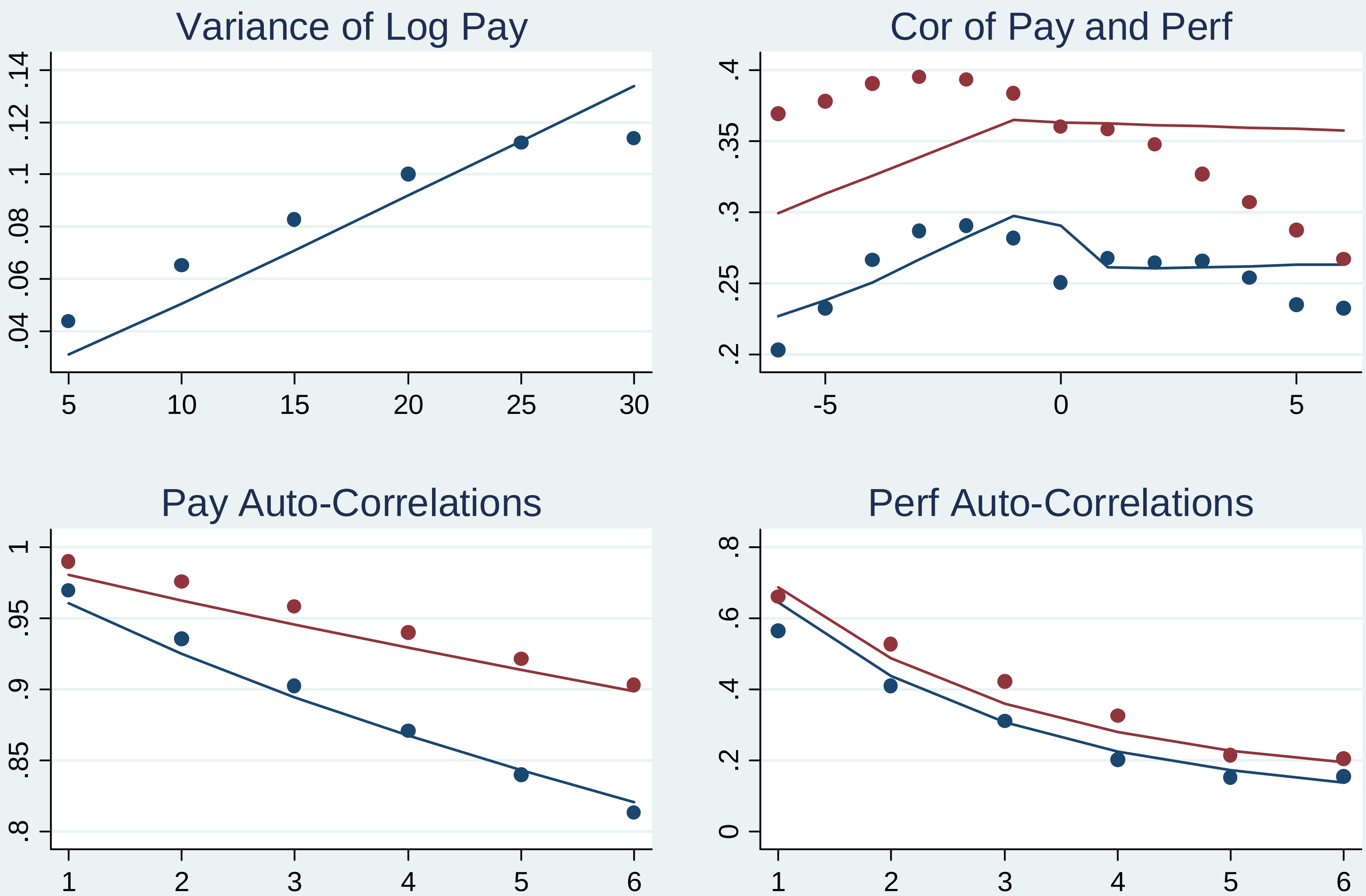


Figure 5: Moments and Fits for Pure Learning
Variance of Log Pay
.04
.06
.08
.1
.12
5
10
15
20
25
30
Cor of Pay and Perf
.2
.25
.3
.35
.4
-5
0
5
Pay Auto-Correlations
.8
.85
.9
.95
1
1
2
3
4
5
6
Perf Auto-Correlations
.2
.4
.6
.8
1
2
3
4
5
6

Figure 6: Moments and Fits for Pure Productivity



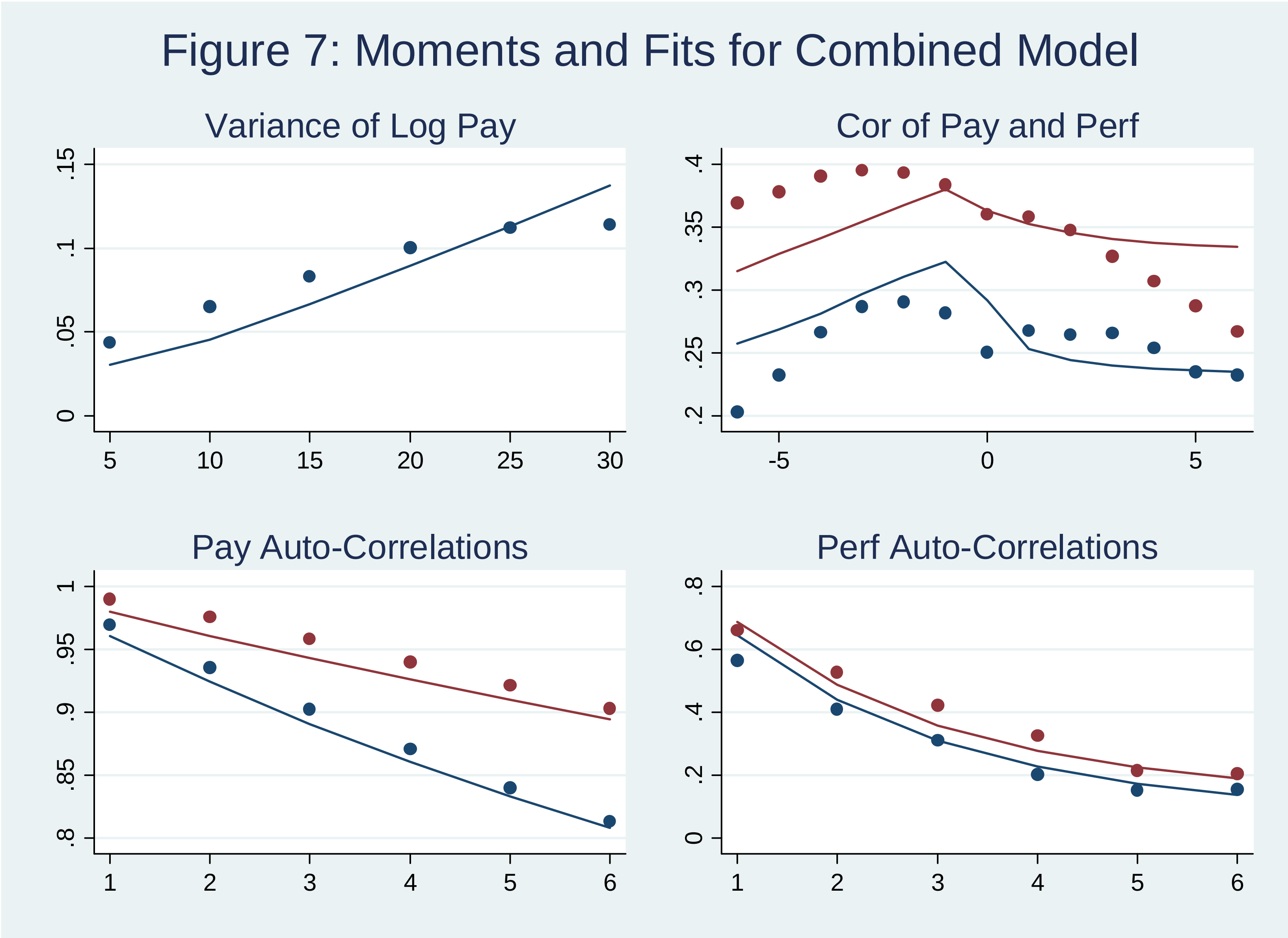


Figure 7: Moments and Fits for Combined Model
Variance of Log Pay
0
.05
.1
.15
5
10
15
20
25
30
Cor of Pay and Perf
.2
.25
.3
.35
.4
-5
0
5
Pay Auto-Correlations
.8
.85
.9
.95
1
1
2
3
4
5
6
Perf Auto-Correlations
0
.2
.4
.6
.8
1
2
3
4
5
6

Table 1: BGH Summary Statistics		
Years		1969-1988
Data Description		Managers of a medium-sized US firm in the service sector
# Employees[1]		
# Employee-years		
% Male		75.7%
% White		88.9%
Age		38.0 (7.88)
Education		
% HS		17.0%
% Some College		18.3%
% College		37.0%
% Advanced		27.7%
Salary[2]		$53,421 (24346) [n=51199]
Performance[3]		3.145 (0.708) [n=36383]
Performance Distribution		
	1	0.008
	2	0.164
	3	0.502
	4	0.325

Notes: Parentheses contain standard deviations.

1. Sample includes all employees who can be observed between the ages of 25 and 55, 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]	0.206**				
	[0.00772]				
2 Years Ago Change		0.153**			
		[0.00779]			
3 Years Ago Change			0.116**		
			[0.00840]		
4 Years Ago Change				0.0712**	
				[0.00914]	
5 Years Ago Change					0.0546**
					[0.0115]
Constant	0.0364**	0.0108**	0.0459**	0.0415**	0.0422**
	[0.0114]	[0.00149]	[0.00573]	[0.00691]	[0.0143]
Observations	34675	28015	22658	18270	14707
R-squared	0.077	0.058	0.049	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=30,558)	Log Pay Change	Last Year Change	2 Years Ago Change
Log Pay Change[1]	0.0030		
Last Year Change[2]	0.00074	0.0030	
2 Years Ago Change		0.00075	0.0030

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

	σ_q^2	σ_r^2	σ_0^2	σ_u^2	σ_ω^2	σ_κ^2	ρ	σ_z^2
Pure Employer Learning	0.124	0	0.598	0.698	0.0045	0	0.648	0.576
Pure Productivity		0.0042	0	0.415	0.0000	0.0000	0.646	0
Combined	0.032	0.0043	0.0080	0.491	0.0000	0.000012	0.6410	0.064
	(0.0043)	(0.00026)	(0.0103)	(0.033)	(0.0000)	(0.00036)	(0.0095)	(0.034)