Experimentation and Job Choice

Kate Antonovics and Limor Golan* †

August 2, 2010

Abstract

This paper examines optimal job choices when jobs differ in the rate at which they reveal information about workers' skills. We then analyze how the optimal level of experimentation changes over a worker's career and characterize job transitions and wage growth over the life-cycle. Using the Dictionary of Occupational Titles (DOT) merged with the National Longitudinal Survey of Youth 1979 (NLSY79), we then construct an index of how much information different occupations reveal about workers' skills and document patterns of occupational choice and wage growth that are consistent with a tradeoff between information and wages.

1 Introduction

Workers enter the labor market with uncertainty about their skills, but learn through repeated observation of their on-the-job performance. Since information is valuable in making future job choices, it is natural to think that both workers and firms consider the

*Respectively, University of California, San Diego and Carnegie Mellon University.

†We thank Jan Eeckhout, Ronald Goettler, John Kennan, Dirk Krueger, Edward Lazear, Robert Miller, W. Bentley MacLeod, Victor Rios-Rull and Fallaw Sowell for their helpful comments, and Luis Zuluaga and Juan Vera for excellent research assistance. We also benefited from discussions with the participants of the SOLE 2004 meeting, the North American Econometrics Society 2004 meeting, Canadian Economic Theory conference 2005 and the seminar participants at Boston University, the St. Louis Federal Reserve, Stanford University, UCLA, UCSB and the University of Wisconsin . This research is supported by the Berkman Development Fund (Carnegie Mellon).

extent to which different jobs will reveal information about workers' skills when making job assignments. As a result, workers and firms may be willing to trade off current-period output for additional information. We refer to this tradeoff as experimentation, and in this paper we explore the theoretical and empirical implications of experimentation for occupational choice and wage growth over the life cycle.

The concept of experimentation has received little attention in labor economics, and one of this paper's primary contributions is to fill this gap. From a theoretical standpoint, we show that when you loosen a standard assumption underlying most learning models, experimentation naturally arises. We then analyze how the optimal level of experimentation changes over a worker's career and characterize how experimentation affects job transitions and wage growth over the life-cycle. Then, using data from the Dictionary of Occupational Titles (DOT) merged with the National Longitudinal Survey of Youth 1979 (NLSY79), we construct an index of how much information different occupations reveal about workers' skills and document patterns of occupational choice and wage growth which are consistent with the presence of experimentation.

The theoretical and empirical literature on uncertainty in the labor market primarily focuses on models of matching (see, for example, Jovanovic (1979) and Miller (1984)) and models of learning (see, for example, Gibbons and Waldman (1999), Neal (1999), Gibbons et al. (2005)). Our model differs from standard matching models because, in our model, workers' productivity is not match specific. In addition, our model differs from standard learning models in that we allow different jobs to convey different amounts of information about workers' skills. Only a handful of other papers also consider experimentation in the labor market. Ortega (2001) builds a two period model of job rotation and shows that expected productivity is higher when firms learn about workers through job rotation rather than through fixed job assignments. Ortega does not, however, fully characterize the optimal job rotation policy. Felli and Harris (1996, 2006) and Pastorino (2009) characterize wages and firm turnover in models of experimentation in strategic framework in which learning is inefficient due to competition over scarce talent. In contrast, our

paper characterizes the optimal level of experimentation in a non-strategic framework in order to focus on the trade-off between current-period productivity and information.[1] While few papers consider experimentation in the labor market, there is a large theoretical literature on optimal experimentation in other contexts. These papers, however, use different payoff functions and information acquisition processes, and thus their results are not directly applicable to our setting.

As mentioned above, we use data from the DOT to construct a measure of the extent to which different occupations reveal information about workers' skills. Using data from the NLSY79, we then relate this measure to job turnover and wage growth. Several recent papers also use data from the DOT to characterize occupations (see, for example, Autor, Levy and Murnane (2003), Ingram and Neumann (2006), Poletaev and Robinson (2008), Bacolod and Blum (2010) and Yamaguchi (2010)). These papers, however, do not consider whether employers will be able to easily observe whether workers possess the skills needed in a given occupation. In contrast, our primary goal is to capture the extent to which different occupations are likely to reveal information about workers' skills. Two recent papers Pastorino (2010) and Papageorgiou (2009) estimate models of experimentation. These papers do not use explicit measure of the extent to which different occupations reveal information about workers' skills. Thus, we are able to more directly establish the link between occupational choice, learning and wage dynamics.

In our model, workers choose a job in every period to maximize the expected present discounted value of lifetime income.[2] In each job, the more output depends upon the unobserved skill, the more information the job reveals about that skill. Thus, for example, a worker will learn the most about whether he or she is a good manager in jobs where

[1] Our information structure is similar to Felli and Harris (1996, 2006) in that jobs which depend more on skill are more informative. Pastorino (2009) explains the patterns of promotions of managers found in Baker, Gibbs and Holmstrom (1994), considering the case in which high level jobs are less informative about ability.

[2] Although the model is formalized as one in which workers choose jobs, we show that under perfect competition (with symmetric information), optimal job assignments and earnings are the same if instead firms choose job assignments.

performance depends heavily on managerial ability. To make the model tractable, we assume that there are two skills and that there is only uncertainty regarding one skill. For example, this might correspond the case in which a worker's ability as a programmer is known, but his or her managerial ability is not.

Workers value information because it increases the probability that they are assigned to the job at which they have a comparative advantage. Thus, not surprisingly, we find that workers experiment, foregoing expected current period output in order to learn about their skill. We find that at the beginning of a worker's career, when there is considerable uncertainty about skill, the optimal level of experimentation is relatively small, increases as worker's gain experience, and then declines again as workers become increasingly certain about their skills. The decline in experimentation at the end of a worker's career is intuitive; as workers gain experience, the degree of uncertainty about their skill level drops and, thus, the value of experimentation falls. The increase in experimentation in the early stages of a worker's career is driven by the fact that when there is a lot of dispersion in a worker's prior beliefs about his or her skill, marginal increases in information do little to increase the probability that workers are correctly assigned to jobs in the future.[3]

Superficially, the prediction that experimentation will be low in the early stages of a worker's career may seem to contrast the prediction of the matching model analyzed Miller (1984) in which the mean and variance of the prior distribution of match quality differ across occupations. Since workers learn about match quality more quickly in high-variance occupations than in low-variance occupations, workers may be willing to enter into occupations in which expected match quality is low as long as the prior variance of match quality is high enough. This form of experimentation primarily takes place early in a worker's career. Our model differs because instead of facing a binary decision about whether or not to learn (they either enter an occupation or they do not), workers can

[3]The fact that patterns of experimentation are often non-monotonic and complex is a manifestation of the of the famous Radner and Stiglitz (1984) result that the value of information is non-concave. This result is generalized by Chade and Schlee (2002).

choose both whether to learn and how much to learn. Thus, in our model workers are also most likely to experiment at the beginning of their career, but initial optimal level of experimentation is low.

In addition to characterizing patterns of experimentation over the life cycle, we also examine our model's implications for wage dynamics. We show that, unlike most standard models of wage growth in which wages increase because of either human capital accumulation or improvements in match quality, in our model, wage growth is also partially driven by the eventual decline in experimentation.[4] In addition, we examine the factors that influence a worker's lifetime earnings and show that job diversity increases both experimentation and lifetime earnings, suggesting that specialization may be suboptimal when jobs provide information about workers' skills. Further, we show that random productivity shocks can have long-lasting effects on both wages and wage growth. In particular, workers who receive negative productivity shocks may be reassigned to jobs that reveal little about their skill and where wage growth is slow. As a result, luck may lead to different career trajectories even in the long run. Further, since new information has the largest effect on prior beliefs when workers are young, productivity shocks have the largest impact early in worker's career.

We show that our model's predictions are consistent with observed wage and job mobility patterns in the NLSY79. We find that workers do not start in jobs that are likely to reveal a great deal about their skill. Over time, workers move to jobs which depend more on skills, but approximately one third of the sample, moves into jobs where output depends less on skill and where workers are likely to learn less about their skill. We also find that wage patterns conditional on moving to lower-skill or higher-skill occupations are consistent with a tradeoff between information and wages. Finally, we find that the extent to which a worker's initial job reveals information about the worker's skill has

[4] One exception is Harris and Holmstrom (1982) who examine an environment in which firms provide risk-averse workers with partial insurance against negative productivity shocks. In their model, wages rises over workers' careers because as uncertainty about worker ability falls, so does the cost of insuring the worker against future wage cuts.

long-lasting effects on wage dynamics.

The paper is organized as follows. Section 2 presents the model, Section 3 characterizes the solution to a two-period model and Section 4 describes the solution to the infinite horizon problem and the implied job transitions and wage dynamics over the workers' life-cycle. Section 5 describes the data, Section 6 presents the empirical regularities which are consistent with the model's predictions and Section 7 concludes.

2 Model

This section first describes the labor market, technology and the structure of the information. It then characterize the equilibrium wages. The optimal experimentation and jobs choice of workers is first characterized in a two-period model in order to highlight the tradeoff between information and current output on the worker's optimal job assignment. In order to characterize the dynamic aspects that are not captures in the two-period version of the model, we then solve the infinite horizon of the model.

2.1 Environment, Technology and Information

Consider an economy with infinitely-lived, risk-neutral workers and firms with a common discount factor δ. Workers differ in the set of skills that they possess. In principle, this skill set may be multidimensional and include skills such as creativity, diligence, adaptability, etc.. To focus on essentials, however, we examine a simple scenario in which each worker has only two skills: a known skill, k, and an unknown skill, θ, both of which are time-invariant.

Firms in the model offer a variety of jobs, and each job differs in the extent to which output depends upon k and θ. In addition, firms are identical so that a worker's expected output at a given job depends only upon the skill set of the worker and the characteristics of the job. In particular, there are N jobs, each completely characterized by a given value of α, where α denotes the degree to which output depends on θ, relative to k. Thus, choosing a job in period t is equivalent to choosing a value of α. Given this choice, we

assume output in period t is given by

$$y_t = \alpha_t \theta + (1 - \alpha_t)k + \epsilon_t, \tag{1}$$

where ϵ_t is an i.i.d. productivity shock, α_t denotes the value of α chosen by the worker at time t, and $\alpha_t \in \{\alpha^1, ..., \alpha^N\}$, where $\alpha^1 = 0$, $\alpha^N = 1$ and $\alpha^r > \alpha^s$ for all $r > s$.[5] Thus, there is one job in which output is only sensitive to θ ($\alpha^N = 1$) and one job in which output is only sensitive to k ($\alpha^1 = 0$). For the rest of the $N - 2$ jobs, the higher is α^j, the more output depends on θ.[6]

Information in the model is symmetric; firms and workers have common priors on θ, k is known to everyone and output is commonly observed. Workers and firms acquire additional information about a worker's unknown skill through successive observations of output. Notice, however, that even though both the known skill, k, and the unknown skill, θ, contribute to output, the contribution of θ cannot be observed separately from the contribution of the i.i.d. shock. Thus, having observed output, workers and firms calculate

$$x_t = \frac{y_t - (1 - \alpha_t)k}{\alpha_t} = \theta + \frac{\epsilon_t}{\alpha_t}, \tag{2}$$

where x_t serves as a signal of the worker's unobserved skill, θ. The noise of x_t is *not* independent of a worker's job choice. In particular, the higher is α_t, the higher is the signal-to-noise ratio and the more information about θ the market is able to extract from x_t. Further, this information is valuable since knowledge of θ enables workers to sort into the jobs to which they are best suited.

[5] Here we assume that skills are perfect substitutes in production. However, the solution to this problem is equivalent to the solution of one in which there is a Cobb-Douglas production function, where $y_t = \ln \tilde{y}$, $k = \ln \tilde{k}$, $\theta = \ln \tilde{\theta}$, $\epsilon = \ln \tilde{\epsilon}$ and

$$\ln \tilde{y}_t = \ln[\tilde{\epsilon}\tilde{\theta}^{\alpha_t}\tilde{k}^{1-\alpha_t}] = \alpha_t \ln \tilde{\theta} + (1 - \alpha_t) \ln \tilde{k} + \ln \tilde{\epsilon}_t.$$

[6] We assume that set of jobs is exogenously given. Our results below, however, suggest that if firms were costlessly able to choose as many jobs as they want, they would choose a continuum of jobs so that $\alpha \in [0, 1]$.

Under the assumption that the prior distribution of θ at time t is normal with mean μ_t and variance σ_t^2 and the distribution of ϵ_t is normal with mean zero and variance σ_ϵ^2, the posterior distribution of θ is known to be normal with mean μ_{t+1} and variance σ_{t+1}^2 where

$$\mu_{t+1} = \frac{\mu_t \sigma_{\tilde{\epsilon},t}^2 + x_t \sigma_t^2}{\sigma_{\tilde{\epsilon},t}^2 + \sigma_t^2} \tag{3}$$

and

$$\sigma_{t+1}^2 = \frac{\sigma_{\tilde{\epsilon},t}^2 \sigma_t^2}{\sigma_t^2 + \sigma_{\tilde{\epsilon},t}^2} \tag{4}$$

and where $\sigma_{\tilde{\epsilon},t}^2 = \frac{\sigma_\epsilon^2}{\alpha_t^2}$. Notice that, while both the mean and the variance of the posterior distribution of θ are influenced by α_t, only the posterior mean is influenced by the signal x_t.

In addition, we know that μ_{t+1} is itself normally distributed with mean, m_{t+1} and variance s_{t+1}^2 given by:

$$m_{t+1} = \mu_t \tag{5}$$

$$s_{t+1}^2 = \frac{\sigma_t^4}{\sigma_t^2 + \sigma_{\tilde{\epsilon},t}^2}. \tag{6}$$

Thus, the posterior mean of θ follows a martingale, and the more information x_t reveals about θ (the higher is α), the higher is the variance of the posterior mean.

Timing in the model is as follows: at the beginning of each period, workers announce a job choice, firms make take-it-or-leave it wage offers and each worker accepts an offer. We assume competitive markets and free entry into the labor market.

Thus, given their prior beliefs about θ and given a worker's job choice, α_t, firms pick a wage policy $w_t(\alpha_t, \mu_t, \sigma_t^2)$ to maximize the present discounted value of future profits, and workers choose α_t in each period to maximize the expected present discounted value of lifetime earnings.

2.2 Equilibrium Wages and Optimal Job Choice

First, we characterize equilibrium wages and then solve for optimal job assignments. We assume spot contracts. Workers' current period utility is given by $U_t = w_t$. Thus, workers

maximize the expected discounted life-time wages.

Proposition 1 *In equilibrium, in each period, 1) firms offer wages equal to each worker's expected productivity at the job chosen by the worker in that period, 2) each worker chooses the job that maximizes the worker's expected discounted lifetime productivity, and 3) if firms (instead of workers) made job choices, then firms would also assign workers the job that maximizes the worker's expected discounted lifetime productivity.*

The proof is in Appendix A. The first two points are straightforward. Competition among firms together with spot contracts imply that workers will be paid their expected productivity in every period. In addition, since wages equal workers' expected productivity in every period and since workers want to maximize their expected discounted lifetime earnings, the optimal job choice in each period maximizes workers' expected discounted lifetime productivity. The third point is more subtle. Suppose that firms rather than workers made job choices. Given that all information in the model is symmetric, when firms assign workers to jobs that reveal information about workers' skills, firms provide information to all other firms (and workers) in the market. Under perfect competition, however, when human capital is general, workers earn all the rents from this information. Therefore, if firms offer a worker any job other than the job that maximizes that worker's expected discounted lifetime earnings, then the worker would only accept the job if the firm offered a wage above the worker's expected productivity. This current-period loss cannot be optimal for the firm, however, because competition insures that the firm's future profits will never be greater than zero. Thus, one implication of Proposition 1 is that the solution to our model is the same regardless of whether we allow workers or firms to choose jobs.

3 Two-Period Problem

Since the infinite-period version of our model has no analytic solution, we first present a two-period version to build intuition. Consider a worker who works for two periods and

then retires. Assume there is a continuum of jobs $\alpha \in [0,1]$. For simplicity and without loss of generality, assume that $\epsilon \sim N(0,1)$. The worker's problem can be therefore written:

$$V(\mu_1, \sigma_1^2) = \max_{\alpha_t \in [0,1]} \alpha_1 \mu_1 + (1-\alpha_1)k + \delta E_1[\alpha_2 \mu_2 + (1-\alpha_2)k], \tag{7}$$

which we can solve recursively beginning from the second period.

Proposition 2 *The second period optimal choice of job is given by:*

$$\alpha_2(\mu_2) = \begin{cases} 1, & \textit{if } \mu_2 \geq k, \\ 0, & \textit{otherwise.} \end{cases} \tag{8}$$

The second-period job assignment is the solution to a static problem in which the worker maximizes his or her expected wage. Since workers are paid their expected productivity, the optimal job assignment will entail specialization and assignment to the job in which the expected productivity is the highest.[7]

Next, we solve for the optimal assignment in period one. Note that expected productivity in period two depends on the second-period belief, μ_2, which in turn depends on α_1 though x_1 (see equations (2) and (3)). We therefore rewrite the first-period problem as:

$$V(\mu_1, \sigma_1^2) = \max_{\alpha_1 \in [0,1]} \alpha_1 \mu_1 + (1-\alpha_1)k + \delta[\Phi(r)k + \int_k^\infty \mu_2 f(\mu_2)], \tag{9}$$

where $r = \frac{k-\mu_1}{s_2}$, $\Phi(\cdot)$ is the standard normal cumulative density function, and f is the normal probability density function with mean $m_2 = \mu_1$ and variance $s_2^2 = \frac{\alpha_1^2 \sigma_1^4}{\alpha_1^2 \sigma_1^2 + 1}$. The above equation makes clear that when $\mu_1 < k$, there is a cost associated with selecting $\alpha_1 > 0$ since expected current-period output will be less than k. Thus, when $\mu_1 < k$, workers must weigh the benefit of increasing α_1 in terms of expected future output against the cost in terms of expected current-period output.

To see why expected second period output is increasing in α_1, note that the expected value of any left-truncated normal random variable is increasing in the variance of that random variable. Thus, since s_2^2 is increasing in α_1, we know that expected second-period

[7] In the two-period problem, the prior-variance, σ_2^2, does not affect the optimal job assignment in the second period. This will not be true in the infinite-period version of our model

output must also be increasing in α_1. Intuitively, information is valuable because workers can insure themselves against the arrival of negative information about θ in the future by selecting future jobs with $\alpha = 0$, but can take advantage of the arrival of positive information about θ by selecting future jobs with $\alpha = 1$. Figure 1 describes second period output as a function of α_1.

If a worker chooses to forego expected current-period output in order to gain information about θ, then we say that the worker experiments.

Definition 1 *A worker experiments if $\mu_1 < k$, but the worker chooses $\alpha_1 > 0$.*

In addition, Proposition 3 establishes that for any $\mu_1 < k$, experimentation is beneficial if there is sufficient uncertainty about a worker's skill.

Proposition 3 *For every $-\infty < \mu_1 < k$ and any $\alpha_1^{'} > 0$ there exists a large enough $\sigma^{'}$ such that the value of experimenting is greater than the value of not experimenting. That is, $V(\mu, \sigma^{'}|\alpha_1^{'}) > V(\mu, \sigma^{l}|\alpha_1 = 0)$.*

The intuition is that even when θ is believed to be very low, if there is sufficient uncertainty about θ, then the probability that $\theta > k$ is high enough that it is worth foregoing current period output to gain additional information about θ's true value.

Next, we characterize the optimal solution.

Proposition 4 *The first order necessary condition for an interior solution is:*

$$\frac{\partial \phi(r)}{\partial \alpha_1} s_2 + \frac{\partial s_2}{\partial \alpha_1} \phi(r) + (k - \mu_1) \frac{\partial \Phi(r)}{\partial \alpha_1} = \frac{k - \mu_1}{\delta}, \tag{10}$$

where $\phi(r)$ is the standard normal pdf.

The above equation states that a necessary condition for an interior solution is that the marginal benefit of increasing α in terms of second period output be equal to the marginal cost in terms of first period output. Note that when $\mu_1 > k$, the cost of increasing α is negative (the right-hand side of equation (10) is negative). Thus, when $\mu_1 > k$, both first and second period expected output are increasing in α_1, and there will be a corner solution at $\alpha_1 = 1$.

The optimal job assignment can be characterized as an implicit function of the state variables,[8] Figures 2 and 3 illustrate the optimal job choice as a function of the state variables: σ_1 and μ_1. Holding σ_1 constant, the higher is the prior mean of θ, the higher is the optimal choice of α_1. The optimal choice of α_1, however, is not always increasing in the prior variance of θ.

To understand why there is a non-monotonic relationship between α_1 and σ_1, recall that in selecting α_1 workers weigh the benefit of increasing α_1 in terms of expected future output against the cost in terms of expected current-period output. Since current period expected output does not depend on σ_1^2, this implies that the non-monotonic relationship between σ_1^2 and the optimal choice of α_1 must depend solely on how increases in α_1 affect the expected future output. In particular, the effect of increasing α_1 on expected future output must be low both when σ_1^2 is small and when σ_1^2 is large.

It is fairly easy to see why the benefit of increasing α_1 is small when there is little uncertainty about θ. First, when σ_1^2 is small, the option value of new information is low because new information on θ is unlikely to have a large impact on the posterior mean of θ. Second, the expected loss of output due to incorrect future job assignments is small because the likelihood that θ is much different than μ_1 is small.

To see why the benefit of increasing α_1 is also small when σ_1^2 is large,recall that an increase in α_1 increases expected future output through its effect on s_2^2, the spread of μ_2. As is clear from equation 6, however, s_2^2 is also increasing in σ_1^2, and the marginal effect of an increase in α_1 on s_2^2 is small when σ_1^2 is large, leading increases in α_1 to have a small positive effect on expected future output.

Figure 4 describes the marginal effect of increase α_1 on second period output as a function of σ_1^2. Clearly, expected second-period output is increasing in α_1 when σ_1^2 is small, but decreasing in α_1 when σ_1^2 is large. Proposition 5 establishes this formally.

Proposition 5 *The marginal value of increasing α_1 in terms of second-period output shrinks to zero both as σ_1 becomes arbitrarily small and as σ_1 becomes arbitrarily large.*

[8] See Appendix A equation 16; Figure 5 shows the second order conditions.

That is, $\lim_{\sigma_1 \to 0} \frac{\partial E[y_2]}{\partial \alpha_1} = 0$, *and* $\lim_{\sigma_1 \to \infty} \frac{\partial E[y_2]}{\partial \alpha_1} = 0$.

4 Infinite Horizon Model

We now extend the model above to incorporate an infinite time horizon so that we can more fully characterize the evolution of wages and job assignments over the life-cycle. We assume that there are a finite number of job "types" (that is, α_t is discrete). Thus, we can generate comparative statics on the effect of the supply of jobs in the economy on workers' earnings. Since the worker's decision problem in period t is the same as the problem in period 1, except that the worker updates his or her prior beliefs about θ based on the history of productivity signals, $\{x_{t-1}, ..., x_1\}$, according to equations (3) and (4), the worker's problem is stationary. Thus, we can write the value function, $V(\cdot, \cdot)$, as the solution to a Bellman equation in which the control variable is α_t and in which the state variables, μ_t and σ_t^2, describe the prior distribution of θ. That is,

$$V(\mu_t, \sigma_t^2) = \max_{\alpha_t \in \{\alpha^1, ..., \alpha^N\}} \alpha_t \mu_t + (1 - \alpha_t)k + \delta \int V(\mu_{t+1}, \sigma_{t+1}^2) f(\mu_{t+1}) d\mu_{t+1}, \qquad (11)$$

where f denotes the normal probability density function with mean m_{t+1} and variance s_{t+1}^2; the dependence of σ_{t+1}^2 and s_{t+1}^2 on α_t and the state variables is given in equations 4 and 6. The first two terms on the right-hand side of (11) represent expected current period output, and the second term represents the continuation value, which incorporates the value of information obtained from observing x_t.

Similar to the two-period model, higher values of α_t are beneficial to workers in the sense that the higher is α_t, the more information on-the-job performance will reveal about a worker's skill. In expectation, this information is valuable because workers can insure themselves against the arrival of negative information about θ by selecting future jobs with $\alpha = 0$, but can take advantage of the arrival of positive information about θ by selecting future jobs with $\alpha = 1$. In the context of equation 11, the benefit of increasing α_t is reflected in the continuation value and the fact that s_{t+1}^2, the variance of μ_{t+1}, is increasing in α_t. In contrast to the two-period model, however, in the infinite-horizon

model, information affects not only future job assignments but also future experimentation patterns. As a result, the continuation value also depends on α_t through it's effect on σ^2_{t+1}, and the dynamic programming problem becomes more involved.

4.1 Solution method

A key feature of our model is that the productivity signal, x_t, provides information about a worker's productivity at many jobs. Thus, we cannot appeal to solution techniques developed in the literature on independent multi-armed bandit problems (because the "arms" in our problem are dependent).[9] Instead, we solve our problem numerically.

Let B denote the space of bounded, continuous, real-valued functions taking their domain on the space, $S = (-\infty, \infty) \times [0, \infty)$. For all $g \in B$, define the mapping $T : B(S) \to B(S)$ as

$$Tg(\mu, \sigma^2) = \max_{\alpha \in \{\alpha^1, \ldots, \alpha^N\}} \alpha\mu + (1-\alpha)k + \delta \int g(\mu^{'}, \sigma^{2'}) f(\mu^{'}) d\mu^{'}, \tag{12}$$

where $\sigma^{2'} = \frac{\sigma^2_{\tilde{\epsilon}}\sigma^2}{\sigma^2 + \sigma^2_{\tilde{\epsilon}}}$ and where f is the normal probability density function with mean $m = \mu$ and variance $s^2 = \frac{\sigma^4}{\sigma^2 + \sigma^2_{\tilde{\epsilon}}}$. As we show in Appendix B, Blackwell's sufficient conditions for T to be a contraction are satisfied. Thus, by the contraction mapping theorem, the value function in (11) is unique and can be obtained by iteration of T.

In our computations, we discretize the prior mean and use linear interpolation and extrapolation to evaluate the value function at arbitrary values of the prior mean. The prior variance is a deterministic function of the job assignment history and is weakly declining over time. Thus, we create a grid by calculating many possible variance points for all possible job assignment histories for large number of periods. As the number of periods becomes arbitrarily large, the variance drops to zero and thus zero is also on the grid. Since the value of the value function is known when the prior variance is zero, we use linear interpolation for variances greater than zero but too small to be on our grid.

[9] For further discussion, see Gittins and Jones (1974).

4.2 Optimal Job Assignments

In this section, we characterize optimal job assignments and verify that the basic qualitative properties of the two-period model hold. We refer to $\alpha = 0$ as the low-level job and $\alpha = 1$ as the high-level job. In addition, we refer to all jobs with $\alpha \in (0, 1)$ as intermediate-level jobs. Figure 6 illustrates optimal job choices in the N-job model when $\delta = 0.9$, $k = 7$, $\sigma_\epsilon^2 = 1$ and $\alpha \in \{0, 0.5, 1\}$. First, as long as $\sigma_t^2 > 0$, there are workers with $\mu_t < k$ who choose either $\alpha_t = 0.5$ or $\alpha_t = 1$. Thus, as in the two-period model there is a tradeoff between current wages and learning. In addition, similar to the two-period model, holding σ_t^2 constant, the higher is the prior mean of θ, the higher is the optimal choice of α_t. The property that the optimal choice of α_t is not always increasing in the prior variance of θ remains as well. In Figure 6, the frontier along which workers are indifferent between choosing $\alpha_t = 0.5$ and $\alpha_t = 1$ is negatively sloped when the prior variance of θ is relatively low but positively sloped when the prior variance of θ is relatively high.[10] Thus, as in the two-period model, experimentation is not strictly increasing in the prior variance of θ. To verify that our results hold as the number of jobs increases, Figure 7 demonstrates job assignments when ($\alpha = \{0, 0.2, 0.5, 0.7, 1\}$).

Discussion

Superficially, our model's predictions regarding experimentation may seem to contradict the prediction of the standard matching model in Miller (1984), that showed that young (high-variance) workers will gravitate towards risky jobs that reveal a great deal of information about workers' skills while older (low-variance) workers will shift towards safe jobs that reveal less about skill. Notice that, consistent with Miller, the negatively sloped frontier between $\alpha = 0$ and $\alpha = 0.5$ in figure 6 implies that the likelihood of experimentation falls as σ_t^2 falls. Thus, Miller's basic prediction is embedded in our results.

[10] Notice that the frontier between $\alpha_t = 0$ and $\alpha_t = 0.5$ is an "indifference" curve in which the value function is the same along the curve. However, along the frontier, between $\alpha_t = 0.5$ and $\alpha_t = 1$, the value function differs at different points on the frontier. In particular, the higher is the prior mean and prior variance, the higher is the value function.

Unlike Miller, however, in our model workers choose both whether to experiment ($\alpha = 0$ or $\alpha > 0$) and how much to experiment ($\alpha = .5$ or $\alpha = 1$). Once workers are given this option, we see that although the likelihood of experimentation falls as workers gain experience, the optimal *level* of experimentation is initially low and increases in the early stages of workers' careers (the frontier between $\alpha = .5$ and α=1 is positively sloped when σ_t^2 is large) and then eventually falls as workers gain experience (the frontier between $\alpha = .5$ and α=1 is negatively sloped when σ_t^2 is small).

The key feature of our model that generates the distinction between whether to experiment and how much to experiment is the fact that we allow productivity to be correlated across jobs. To see this, it is useful to examine a variant of our model in which in which $\alpha_t \in \{0, 1\}$, so that there are only two types of jobs: one in which output only depends on k ($\alpha = 0$) and another in which output only depends on θ ($\alpha = 1$). We refer to this model as the two-job model. Note that since a worker's output in the high-level job is uncorrelated with his or her output in the low-level job, observed output in one job reveals nothing about expected output in the other job. Thus, this version of our model can be viewed as an independent one-arm bandit problem and is a special case of Miller (1984) in which there are only two occupations: one in which expected productivity is known with certainty and another in which expected productivity is unknown.

Figure 8 illustrates optimal job choices as a function of prior mean and prior variance when $k = 0$, $\sigma_\epsilon^2 = 1$, $\delta = 0.90$, and $\alpha \in \{0, 1\}$. As the figure shows, the boundary line between the region in which the worker chooses the low-level job ($\alpha = 0$) and the region in which the worker chooses the high-level job ($\alpha = 1$) is negatively sloped so that as σ_t rises, workers are willing to take larger and larger current period wage cuts in order learn about themselves. That is, the greater is the uncertainty about the unobservable skill, θ, the more willing are workers to experiment.

4.3 Job Assignments Over the Life Cycle

Having discussed optimal job assignments at a given point in time, we now describe how optimal job assignments change over the life cycle. To do so, we simulate job choices over time for workers with $\mu_0 = 3$ and $\sigma_0 = 6$ in a model where $\delta = 0.9$, $k = 7$, $\sigma_\epsilon^2 = 1$, $\alpha \in \{0, 0.5, 1\}$. As Figure 9 reveals, given this initial mean and variance, all workers start out with $\alpha = 0.2$. Note, however, that this ia a transitory job. As workers gain experience and become more certain about whether θ is greater than or less than k, they increasingly sort into specialized jobs in which $\alpha = 1$ or $\alpha = 0$. In addition, notice that some workers who select $\alpha = 1$ do so because they wish to experiment, while others do so because $\alpha = 1$ maximizes their expected current period output. Workers who are assigned to $\alpha = 1$ but experiment ($\mu < k$), are marked by $\alpha = 1$ (E).

Proposition 6 *At the beginning of the life-cycle, workers may work in jobs that include a mixture of tasks ($0 \leq \alpha \leq 1$). As they accumulate experience, they sort into more specialized jobs, and in the limit, workers either choose $\alpha = 1$ or $\alpha = 0$.*

The proof is in Appendix A. The intuition for this proposition is clear. The first part follows directly from the optimal solution (see, for example, Figure 6), and the second part of the proposition follows from the fact that as t becomes arbitrarily large, σ_t^2 becomes close to zero for workers who are assigned to $\alpha_t > 0$ (once assigned to $\alpha_t = 0$, workers do not move). Thus, for these workers, the solution approaches the full information solution in which $\alpha_t = 1$ if $\theta > k$, and $\alpha_t = 0$ otherwise.

Since no information is revealed about skill when $\alpha = 0$, however, there will always exist a subset of workers for whom θ is never fully learned, even in the limit.[11] As a result, not all job assignments will be efficient. In particular, every period some workers choose $\alpha = 0$ even though $\theta > k$ because of negative prior beliefs about θ. Further, these mistakes are permanent since once a worker chooses $\alpha = 0$, no new information will be revealed about θ and the worker will never learn about his or her true ability. As a

[11] Thus, in the language of Aghion et al. (1991), learning is not adequate.

result, even good workers can get stuck in bad jobs. As Figure 13 reveals, the cumulative probability of finding a good worker in a bad job increases over time since there is always some chance that workers will experience a series of negative productivity shocks that lead them to revise downwards their beliefs about θ.

4.4 Experience-Wage Profiles

The non-monotonic between experience and experimentation means that wage growth in our model is driven by two factors. First, wages grow as workers gain information about the true value of θ and sort into the job at which their expected productivity is the highest. Second, since experimentation involves a loss in expected current period wages, wage growth later in a worker's career is also driven by the eventual decline in experimentation. To show our model's predictions for wage growth, we simulate the wage distribution across time assuming $\delta = 0.9$, $k = 7$, $\alpha \in \{0, 0.5, 1\}$, $\mu_0 = 3$, $\sigma_0 = 6$. Figure 10 shows the wage at different percentiles of the wage distribution for a given cohort of workers 10 periods into the future. Similar to most models of learning, Figure 10 shows that cohort wage dispersion increases over time. The bottom tail of the earnings distribution (any wage less than k) disappears as workers stop experimenting.

4.5 The Importance of Jobs

Our model suggests that economy-wide output is higher the larger is the number of jobs. To see this, Figure 11 shows the difference in the expected discounted value of lifetime earnings between a three-job model in which $\alpha \in \{0, 0.5, 1\}$ and a two-job model in which $\alpha \in \{0, 1\}$. As the figure reveals, this difference is always positive. Thus, workers can expect to earn more (and produce more) when there are more jobs to choose from.

As the number of jobs increases, expected earnings increases for two related reasons. First, workers are more likely to experiment. Second, among those who experiment, the cost of experimentation is lower. To see this, Figure 12 compares job choices when $\alpha_t \in \{0, 1\}$ to those when $\alpha_t \in \{0, 0.5, 1\}$. The black area shows the region in which

workers choose $\alpha_t = 0$ in the two-job model but $\alpha_t = 0.5$ in the three-job model. Thus, in an economy with more jobs, workers are more likely to experiment, choosing jobs which reveal information about θ. In addition, the gray area shows the region in which workers choose $\alpha_t = 1$ in the two-job model but $\alpha_t = 0.5$ in the three-job model. Thus, relative to the two-job model, in the three-job model, workers sacrifice less in terms of foregone current earnings in order to learn. The fact that expected lifetime earnings increase as the number of jobs in the economy increases implies that productivity will be lower in highly specialized economies than in economies where output reveals information about multiple dimensions of workers' skill sets.

4.6 The Importance of Early Labor Market Outcomes

Finally, our model suggests that i.i.d. productivity shocks, especially those early in a worker's career, have a persistent effect on earnings. It is a common feature of all learning models that past output realizations affect current beliefs about workers' skills, and on average, workers who receive positive productivity shocks ($\epsilon_t > 0$) will have higher wages than those who receive negative productivity shocks ($\epsilon_t < 0$), at least for any finite time horizon. In contrast with the previous literature, however, in our model, negative productivity shocks have longer-lasting effects on wages than do positive productivity shocks because workers who receive negative productivity shocks are more likely to choose jobs in which α_t is relatively low and in which little new information is revealed about the workers' skills. Since information helps workers optimally sort into jobs, workers who receive negative productivity shocks will not only have lower wages but also slower wage growth than will those who receive positive productivity shocks. Thus, experimentation serves as a propagation mechanism. Further, this effect is especially pronounced early in a worker's career since new information has the largest effect on beliefs when there is considerable uncertainty about θ.

The second column of Figure 13, which shows the probability of being incorrectly assigned to $\alpha = 0$ when $\theta > k$ in each period, starkly illustrates this point. As the table

reveals, the probability of these incorrect assignments is the greatest early in a worker's career. Further, these mistakes are permanent because once a worker is assigned to the low-level job, no additional information about that worker's skill is revealed and wages cease to grow. Thus, a worker's labor market experiences, especially early on, can have permanent impact on his or her career trajectory. This suggests that the period of time shortly following an individual's entry into the labor market may be critical to his or her future success.

Similarly, to illustrate the dispersion in wages caused by luck, we simulate our model for workers with $\theta = 8$ when $\mu_0 = 6$, $\sigma_0 = 4$, $k = 7$ and $\alpha = \{0, 0.5, 1\}$. Given that $\theta > k$, if θ were known, all workers would be assigned to $\alpha = 1$ and would earn a wage of \$8. Uncertainty about θ, however, leads to departures from this full-information outcome. Figure 14, for example, shows the percentiles of the predicted wage distribution for the first 10 periods of a worker's career. The probabilities on the line $k = 7$ reflect the cumulative probability of permanent mistakes in each period (that is, the probability of $\alpha_t = 0$). As the figure reveals, the initial assignment is $\alpha = 0.5$, and the initial wage is \$6.5.

The figure demonstrates several points. First, experimentation initially leads some workers to earn less than do workers assigned to $\alpha = 0$. Thus, for example, the wage at the 5th percentile is lower than $k = 7$ for the first 4 periods. Second, the exceptionally high wages captured by the 85th and 95th percentiles result from "good luck" (high realizations of ϵ) and the fact that productivity signals are highly influential early in a worker's career when the uncertainty is large. Continued learning, however, eventually leads wages for these individuals to converge to $\theta = 8$, and in the limit, all workers not assigned to $\alpha = 0$ will earn \$8. Third, convergence to a wage of \$8 is slower at the bottom than at the top of the wage distribution because learning is slower for those who initially experience bad luck and are assigned to relatively low-α jobs. Finally, since by period 7 the median worker earns a wage close to the true productivity, the long-run per period wage loss of workers incorrectly assigned to $\alpha = 0$ is roughly $\theta - k = 1$.

We repeat the above simulation changing the number jobs available from 3 to 5 with $\alpha \in \{0, 0.2, 0.5, 0.7, 1\}$. Figure 15 presents the results. Initially, all workers are now assigned to $\alpha = 0.7$ and earn a lower starting wage than in the previous simulation. Note, however, that wages converge more quickly to \$8, and there is a lower probability of permanent mistakes. This happens both because the initial level of experimentation is higher ($\alpha_0 = 0.7$ instead of $\alpha_0 = 0.5$) so learning is faster and because there is a low-cost experimentation option ($\alpha = 0.2$) so more workers experiment.

5 Data

To construct the data used in our empirical analysis, we use the Dictionary of Occupational Titles (DOT) in conjunction with the 1970 Current Population Survey (CPS) to construct a measure of α for each three-digit 1970 Census occupation code. We then merge this measure of α with occupational work histories created from the National Longitudinal Survey of Youth 1979 (NLSY79). Below we describe how we construct our measure of α from the DOT and the occupational work histories from the NLSY79.

5.1 The Dictionary of Occupational Titles

The DOT provides information on the skills, knowledge and work activities associated with different occupations. The data in the DOT are generated by job analysts who conduct both on-site job analyzes and who gather supplemental information from professional and trade associations. Initially, the DOT was intended to assist in career counseling, but it has been increasingly used by economists as a way of obtaining detailed information on occupational characteristics (see, for example, Autor, Levy and Murnane (2003), Bacalod and Blum (2010), Ingram and Neumann (2002) and Yamaguchi(2010)). This paper uses data from 1991 revised forth edition of the DOT, which describes over 12,000 occupations along 44 dimensions. In keeping with the character of our model, we derive our measure of α from job attributes we believe employers cannot readily observe. In particular, we focus on DOT variables describing the level of complexity required in dealing with infor-

mation, the level of complexity required in dealing with people, the level of educational development required in terms of reasoning, math and language, the necessary aptitudes in intelligence, verbal ability, numerical ability, spatial ability and form perception, and the occupations requirements in terms of an individuals adaptability to responsibility, to influencing others, to juggling a variety of tasks, to expressing feelings, ideas and facts, to working in situations requiring precise measurements or standards, to dealing with people beyond giving and receiving instructions and to making generalizations based on subjective criteria.[12] We have also experimented with deriving our measure of α using all of the available job characteristics in the DOT, and this has almost no effect on our results.

The occupational characteristics given in the DOT are linked to 1970 Census three-digit occupation codes in an augmented version of the April 1971 Current Population Survey (CPS). This augmented data file was compiled by the Committee on Occupational Classification and Analysis at the National Academy of Sciences and contains occupation codes from the forth edition of the DOT, which we update with the revised forth edition of the DOT.[13] Since there are more DOT occupation codes than 1970 Census occupation codes, we average the DOT job characteristics over the 1970 Census occupation codes using the sampling weights given in the CPS so that our averages are representative of the population employed in a given three-digit Census occupation. In order to make our empirical analysis match our theoretical model as closely as possible, we construct a single summary measure of an occupations skill requirement using principal component analysis. Since the DOT variables do not have a natural scale, we follow Autor, Levy and Murnane (2003) and transform each DOT measure into a percentile value corresponding to its ranking in the 1970 distribution of that job attribute in the population.[14] We

[12] In the approximate lexicon of the 1991 revised forth edition of the DOT, the 17 job characteristics we use are Data, People, GED-Reasoning, GED-Math, GED-Language, APT-Gengeral, APT-Verbal, APT-Numerical, APT-Spatial, APT-Perception, TEMP-Direction, TEMP-Influencing, TEMP-Variety, TEMP-Expression, TEMP-Tolerances, TEMP-People and TEMP-Judgement.

[13] We thank Shintaro Yamaguchi for helping us with this conversion process.

[14] This normalization is done prior to averaging over the 1970 Census occupation codes.

then calculate the first principal component and predict the first component score for each occupation. To further ease comparison with our theoretical mode, we normalize this predicted first component score by calculating its percentile ranking and dividing by 100. This normalized predicted score naturally takes on a value between zero and one, and higher values indicate a higher level of required skill. We use this normalized predicted score as our measure of α.[15] At every stage in constructing our measure of α we are careful to use the sampling weights given in the CPS. Thus, our measure of α best captures the variation in the occupational characteristics from the DOT for a nationally representative sample of men in the United States.[16]

5.2 The National Longitudinal Survey of Youth 1979

The model we develop focuses on the evolution of the skill requirements of jobs over the life-cycle. Thus, essential to our empirical analysis is a relatively complete occupational work history. In order to construct this occupational work history, we use the National Longitudinal Survey of Youth 1979 (NLSY79), which follows the labor market and other life experiences of individuals born between 1957 and 1964. We focus our empirical analysis on males in the cross-sectional sample. Although, the NLSY79 contains information on individuals' labor force activities for each week from 1978 through the most recent year in which a respondent was interviewed, we only rely upon labor market data from 1978 through 2000. We are unable to use the later waves of the NLSY79 because of a switch in occupational coding that occurred after 2000.[17] The weekly labor force information in the NLSY79 is gathered from an annual survey. If a respondent is not interviewed in a given year (or years), then at the next interview date, the respondent is asked to go back and retrospectively report their labor force activities so that relatively complete work histories can be constructed. For individuals who are working, the work history

[15] Our qualitative findings are not sensitive to this normalization.

[16] Results of the principal components analysis are available upon request. The first component explains 52 percent of the variation in the 17 DOT variables we employ.

[17] Prior to 2000 the NLSY79 consistently reports the 1970 Census Occupational Code, but stops doing so after 2000.

data include information on each of up to five jobs a respondent may have held in a given week. We define an individual's occupation in a given week to be their occupation in the job at which they worked the most number of hours.

Although we do not model educational choices, we expect education to play an important role in an individuals' career path since beliefs about unobserved skills are likely to systematically vary by an individuals' level of education. As a result, we follow individuals' occupational histories starting with their first transition to full-time work after the completion of their highest degree. In particular, we identify the first week following the completion of their degree in which the individual is working at least 10 hours a week and in which they will continue to work at least 10 hours a week for at least 39 of the next 52 weeks. We then keep a running tab of a worker's actual labor market experience and the occupation they hold in each week in which they work.[18] In our empirical analysis, we focus on the first 350 weeks (about 6.7 years) of an individual's actual experience in the labor force. A drawback of the NLSY79 is that we lose a large number observations because of attrition from the sample and missing information.[19] These missing values are particularly problematic because they prevent us from being able to construct complete occupational histories. In particular, we lose 693 respondents because we cannot identify either their highest degree or the date at which they received their highest degree. The majority of these missing values occur because the respondent never completed high school or because the respondent had already left the sample by 1988, the first year in which the NLSY79 explicitly asks about highest degree. We additionally drop 350 respondents who completed their highest degree prior to the start date of the work history record and 254 observations who complete their highest degree relatively late in the life. In particular, we drop individuals who obtain their high school degree after age 21, their associates degree after age 24, their college degree after age 26 or their master's, doctoral or professional degrees after age 36. We do so because we worry

[18] In this tabulation, after making the sample selection rules discussed below, we treat individuals with missing occupation information as being out of the labor force.

[19] For example, sometimes relevant questions were skipped for invalid reasons.

that these workers already may have accumulated substantial labor market experience that could influence employers' beliefs unobserved skills. We then drop an additional 239 observations whose occupational history is relatively incomplete. In particular, we drop individuals who have more than 150 weeks in which the respondent either is not working or has missing occupation information during the first 500 weeks following their transition to full-time work. In other words, we give individuals 500 weeks in which to accumulate 350 weeks of valid occupation information, otherwise we drop them from the sample. We additionally drop 67 individuals who ever report an hourly wage of either over 100 or under 2.

After these restrictions have been made, we are left with 1,366 individuals. Relative to the initial sample, these individuals have a relatively strong attachment to the labor market, are relatively young (because we drop individuals who complete their highest degree prior to the start date of the work history information) and finish their schooling in the early stages of their careers. Finally, in order to verify that our measure of α accurately reflects the skill requirements of occupations, Table 1 shows the average value of α by highest degree and actual experience. As one would expect, the average value α increases with both education and experience, suggesting that people with higher education and greater experience have jobs with greater skill requirements. Similarly, Table 2 shows the average value of α by major Census occupation grouping and actual experience. We see that α lines up sensibly with these occupation groupings. For example, the average value α for Farm Managers and Foreman is .66, while for Farm Laborers and Foremen it is only 0.21. Within the major occupation categories, α remains relatively constant as actual experience increases, with slight changes being driven by shifts in the distribution of individuals across the specific occupations that make up each category.

6 Empirical Findings

In this section, we use our measure of the extent to which job performance depends on unobserved skill to document patterns of wage growth and job assignments over the

life-cycle. We discuss the extent to which some of our model's basic predictions are (and are not) supported by the data. We also relate our findings to existing models of wage dynamics and highlight a number of empirical patterns that are not easily explained by other theories.

6.1 Occupational Choice, Wage Dynamics and Experimentation

In our model, the optimal level of α is low at the early stages of workers' careers when there is considerable uncertainty about skill. Figure 16 describes the average value of α as a function of weeks of actual experience. On average, α increases with experience for both high school and college graduates. For high school graduates, the average α increases by 34 percent between week 1 and week 350 and by 14 percent for college school graduates. One interpretation of this finding is that the optimal level of experimentation is initially low but increases with experience.

As with any learning model, our model also predicts that job transitions will decline as uncertainty about workers' skills fades. We find support for this in the data. Figure 17 plots the absolute value of difference between the average value of α in any given week and the average value of α in week one. As the figure shows, this difference grows over time for both high school and college graduates, with most of the growth occurring in the first year.

Our model also predicts that patterns of sorting and wage growth will reflect the tradeoff between information and output. In particular, the model predicts that over time low-skill workers will move into low-α jobs in which output is less dependent on unobserved skill and that high-skill workers will move into high-α jobs in which output is more dependent on it. In addition, among workers who move to low-α jobs, some will experience real wage increases. That is, since experimentation means that not all workers will be assigned to jobs that maximize their expected current-period productivity (or wages), as the value of information falls, some will move into jobs in which α is lower but in which their current period wages are higher.

Indeed, we find that for roughly 30 percent of the workers in our sample α declines between week 1 and week 300.[20] Among these workers, the average decrease in α is 0.21, which is substantial given that the mean and standard deviation of α in week 1 are 0.43 and 0.29, respectively. Among workers who moved to jobs with a lower level of α, average wages increase by \$5.04, while for those who moved to a higher level of α, average wages increase by \$5.73. Thus, even though workers in our sample appear to move into occupations that have lower skill requirements, they still experience wage increases. These patterns of wage increases that occur with transitions to lower α occupations are also found in wages changes when we compare workers after shorter periods of employment as well as in wage changes that occur with occupational changes.[21]

Consistent with prior studies, we also find that both wages and wage variance increase over the life cycle, and these increases are larger for those with a college degree than for those with a high school degree. To see this, Figure 18 presents the average hourly wage by experience, and Figure 19 presents the standard deviation of hourly wages by experience.[22]

In our model, the more heavily workers' output depends on their unobservable skill, the more they will learn about themselves and the better matched they will be to jobs in the future. As a result, workers assigned to jobs where output depends heavily on unobserved skills will experience greater wage growth and greater wage dispersion than will those assigned to jobs that are less dependent on unobservable skill.[23]

To test whether wage growth is higher for those initially assigned to jobs where output depends heavily on unobservable skill, we regress a worker's hourly wage in week t

[20] For 54 percent of workers α increases, and for 16 percent α doesn't change.

[21] On average, wage changes of workers who switch to an occupation with a lower α is positive, but since the weekly wage data reported when the occupation changes may be noisy, we also looked at wage changes after various employment periods.

[22] In the first 350 weeks of employment, average hourly wages increase by 49 percent for high school graduates and 59 percent for college graduates.

[23] In the extreme case, for example, workers who are initially assigned to jobs with no information content ($\alpha_1 = 0$) will experience no wage growth. In contrast, workers assigned to jobs that depend upon skill ($\alpha_1 > 0$) will experience both wage growth and an increase in wage dispersion (see Figure 10).

on the worker's initial job assignment (α_0), their experience in week t and the interaction between α_0 and experience. Since workers' wage trajectories also depend on initial beliefs about their skill, we separately examine high school and college graduate and additionally control for workers' initial wage (w_0) and the interaction between initial wage and experience. We note, however, that we cannot fully control for μ_0 and σ_0, and so we view this analysis as suggestive.

Table 3 presents the results. Consistent with our model, we find that for both high school and college graduates, wage growth is significantly higher for workers who begin their careers in jobs in which α is high. For college graduates, for example, the coefficient estimate on the interaction between α_0 and experience implies that for each additional year of actual experience, the wages of those with α_0 close to one will be approximately 83 cents more than will the wages of those with α_0 close to zero. Given that the average starting wage of college graduates is \$13.80, this is a substantial percentage difference (approximately 5 percent). Interestingly, the coefficient on the interaction between initial wage and actual experience is negative and statistically significant, suggesting that those with high initial wages experience less wage growth than do those with low initial wages. This magnitude of this relationship, however, is very small. For college graduates, for example, each additional dollar in an worker's starting wage is associated with a 6 cent drop in their wage after a year of actual experience. These results hint that initial job assignments are extremely important in determining future wages and wage growth.[24]

To test whether the increase in the variance of wages is related to how sensitive output is to a worker's initial job assignment, we start by dividing the sample into education groups. Then, for high school and college graduates separately, we determine the quartile of the distribution of α in week one and the quartile of the distribution of wages in week 1 in which each worker falls. Thus, for each education category, each worker falls into one of 16 bins (4 α quartiles x 4 wage quartiles). Within each bin, we then calculate

[24]The negative coefficient on the interaction between wages and experience could be consistent with our model since wage growth is also partly determined by the variance in the initial beliefs about θ but which we cannot fully control for.

the difference between the wage at the 90th and the 10th percentile for every week of actual experience. Thus, for each week and for each bin, we have a measure of the spread of the wage distribution. We then regress this measure on the the quartile of the initial wage distribution, the quartile of the initial α distribution, experience, the interaction between experience and the quartile of the initial wage distribution and the initial α distribution. Table 4 reports the results of this regression. The coefficient on the interaction term between the quartile of the initial α distribution and experience is positive and statistically significant, suggesting that among workers with similar starting wages, the spread in the distribution of wages grows more quickly for those who initially in occupations with high α.

Our model predicts that for workers who continue to experiment as they gain experience (those for whom $\mu_t < k$), the optimal level of experimentation is initially low, increases over time, and then eventually declines. This non-monotonic relationship is difficult to test for empirically because the non-monotonicity only holds conditional on μ_t and for $\mu_t < k$, and because adequate measures of μ_t are not readily available in our data.[25] Standard models of wage determination in which workers are paid their expected marginal product often use wages to proxy for workers' expected skill. In our model, however, wages reflect not only expected skill, but also job assignments, and so there is no straightforward mapping between μ_t and wages.

Nonetheless, we look for a non-monotonic relationship between experience and experimentation by using workers' starting wage and the value of α associated with their first occupation to proxy for μ_0, and by focussing on individuals whose wages fall below a certain threshold to try to isolate individuals for whom $\mu_t < k$. In particular, for each week of experience, we only keep individuals whose wage, w_t is less than $\overline{w}_0^e$, where $\overline{w}_0^e$ is defined to be the average starting wage of workers in education group $e \in$ {high school graduate, college graduate}. Using this unbalanced panel, we then regress α_t on w_0, α_0,

[25]The simulation results in Figure 9 presents transition probabilities into jobs for workers who started with the same priors. The distribution of jobs of workers who experiment in each period is given by the probabilities of the jobs in which $0 < \alpha < 1$ and those in $\alpha = 1$ (E).

experience and experience squared.[26]

Table 5 presents the regression results. For college graduates the coefficient on experience is positive and the coefficient on experience squared is negative and statistically different from zero. Based on these estimates, figure 20 shows the predicted value of α by experience for individuals with a college degree. It shows a small decline in α after week 200, with α falling by approximately 4% between week 200 and 350, though this predicted decline is not statistically different from zero. For high school graduates, we find weaker evidence of a non-monotonic relationship between experience and experimentation (see figure 21, which shows the predicted value of α by experience for individuals with a high school degree). As discussed above, our model predicts that we should not find any evidence of a non-monotonic relationship between experience and experimentation for workers with $\mu_t > k$ expected. Indeed, when repeat the analysis in table 5 for high-wage workers, we find that α is strictly increasing with experience. Furthermore, when we further restrict the sample to workers who started in α_0 jobs below the mean the decline in experimentation becomes larger.[27] Given the difficulty involved in testing whether there exists a non-monotonic relationship between experience and α for workers with $\mu_t < k$, we interpret these findings as suggesting weak but inconclusive support for this pattern.

Finally, Proposition 6 states that workers will eventually sort into jobs with either $\alpha = 0$ or $\alpha = 1$. As discussed above, while we find that over time some workers sort into occupations both with significantly higher and significantly lower values of α, we do not find that individuals sort into occupations with values of α that are close to either zero or one (or to occupations with values of α that are in the tails of the distribution of α). For example, consistent with our model's predictions we find that over time workers

[26] We chose this specification to minimize the mean squared error. The inclusion of higher order terms of experience does not substantially change the mean square error, and the coefficient on higher order terms is statistically indistinguishable from zero. In addition, adding higher order terms does not change any of our qualitative findings.

[27] Results available from authors upon request.

are significantly more likely to be in occupations with values of α that are in the top 25 percent of the initial distribution of α[28], but we do not find that workers are significantly more likely to be in occupations with values of α that are in the bottom 25 percent of the initial distribution of α. There are a number of possible explanations for this finding which we discuss in the concluding section.

6.2 Relation to Existing Theories

Some of the regularities described above are consistent with existing models of wage dynamics. For example, our finding that workers eventually move into jobs where output depends more on skill (that α is increasing with experience) is consistent with theories of on-the-job training in which workers learn how to perform tasks and accordingly move up the job ladder (e.g. Jovanovic and Nyarko (1997)). These models, however, cannot explain our finding that over time a substantial fraction of workers move into jobs where output depends less on skill.[29]

Similarly, while theories of learning and sorting (e.g. Gibbons and Waldman (2006), Gibbons et al. (2005)) that incorporate on-the-job training can explain both why α increases over time on average (because experience augments the unobserved skill) and why α may decline for some individuals (some will receive negative information about their skill). In these models, however, workers have only one skill, and job assignments and wages are both determined by workers' expected skill. Thus, wages and α should move together. As discussed above, however, we find that average wages increase even for those who move into jobs in which output is less dependent on skill.

Thus, while we do not argue that our model explains every feature of the data, we do find evidence supporting many of our model's basic predictions, and some of these

[28]The initial distribution of α refers to the distribution of α in the first week following workers' transition to full-time employment.

[29]Of course, models of on-the-job training easily could be modified to allow skill to depreciate more quickly than it is acquired for some workers. You would expect, however, that the wages of workers in this group would also fall over time. As discussed above, however, we find that wages rise even among those who move into jobs with a lower α.

patterns are not easily explained by existing theories of wage dynamics.

7 Conclusion

This paper develops a life-cycle model of occupational choice and wage dynamics when jobs differ in the amount of information they provide about workers' skills. We find that at the beginning of a worker's career the optimal level of experimentation is relatively small, increases as worker's gain experience, and then declines again as workers become increasingly certain about their skills. We show that wage growth over the life-cycle is also partially driven by the eventual decline in experimentation. In addition, job assignment and experimentation can have a substantial effect on workers' lifetime earnings. Furthermore, random productivity shocks, especially when workers ar young, may lead to different career trajectories even in the long run, and can generate significant wage dispersion.

We then use data from the DOT to construct a measure of how much information different occupations reveal about skills and matches this measure to data from the NLSY79 to examine whether observed patterns of occupational choice and and wage dynamics are consistent with our model. We find that, on average, jobs depend more on the unobserved skill when workers are more experienced, which is consistent with the prediction that initial intensity of experimentation is comparably low. Over time the distance between the degree to which job depends on skills relative to the initial job increases. The decline in uncertainty suggests, however, that workers will not only move into jobs that depend more on skills but that some workers will receive negative signals and sort into jobs that depend less on skills, and we find strong support for this. The trade-off between information and current wage workers make implies that wages grow over time partly because of the decline in experimentation. Indeed we find that on average, wages of workers who sort into jobs which depend less on skill also grow substantially. In addition, we find support to the model's prediction that the information content of jobs initially affect the wage distribution.

Taken together, the above patterns broadly suggest experimentation is important in explaining labor market outcomes, and existing models do not account for all these regularities. Nonetheless, we note that without estimating the model's fundamental parameters, it is hard to fully test the model's predictions. For example, fully characterizing experimentation over the life-cycle requires, among other things, knowledge of the initial beliefs about workers' skills, which cannot be ascertained without estimating the model.

While we find evidence that workers sort over time into jobs which depend either substantially less or substantially more on the unobserved skill, they do not sort into occupations that specialized in one of the tasks. Extending the model to include more elaborate demand formulation can account for these job assignment patterns. For example, costs involving search frictions in which offer arrival of jobs from similar occupations is higher imply that workers will not sort into specialized jobs. In addition, the production function of jobs itself, and which tasks are performed in each job is not only driven by experimentation considerations and can affect the demand for the type of jobs.

An obvious extension to our model would be to allow workers to accumulate task-specific human capital on the job. While it is possible that this extended model would account for the regularities we find, doing so will alter the cost of experimentation since switching jobs will not only involve a change in the rate at which information accrues but also an opportunity costs of loss of task-specific human capital. In addition, it raises more subtle identification issues of human capital and information. While we think these issues are worthwhile exploring, we leave identification issues of the different extensions to future work.

References

[1] Aghion, Philippe; Bolton, Patrick; Harris, Christopher and Jullien, Bruno. "Optimal Learning by Experimentation," *Review of Economic Studies*, July 1991, *58*(4), pp. 621-654.

[2] David H. Autor, David; Levy, Frank and Murnane Richard, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, November 2003, *118*(4), pp. 1279-1334.

[3] Bacolod, Marigee and Blum, Bernardo, Two Sides of the Same Coin: U.S. Residual Inequality and the Gender Gap, *Journal of Human Resources*, Winter 2010, *45*(1), pp. 197-242.

[4] Baker, George; Gibbs, Michael and Holmstrom, Bengt. "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, November 1994, *109*(4), pp. 881-919.

[5] Baker, George; Gibbs, Michael and Holmstrom, Bengt. "The Wage Policy of a Firm," *Quarterly Journal of Economics*, November 1994, *109*(4), pp. 921-55

[6] Chade, Hector and Schlee, Edward. "Another Look at the Radner-Stiglitz Nonconcavity in the Value of Information," *Journal of Economic Theory*, December 2002, *107*(2), pp. 421-452.

[7] Felli, Leonardo and Harris, Christopher. "Learning, Wage Dynamics, and Firm-Specific Human Capital," *Journal of Political Economy*, August 1996, *104*(4), pp. 838-868.

[8] Felli, Leonardo and Harris, Christopher. "Firm-Specific Training," *mimeo*, London School of Economics, March 2006.

[9] Gibbons, Robert; Katz, Lawrence; Lemieux, Thomas and Parent, Daniel. "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 2005, *23*(4), pp. 681-723.

[10] Gibbons, Robert and Waldman, Michael. "A Theory of Wage and Promotion Dynamics Inside Firms," *Quarterly Journal of Economics*, November 1999, *114*(4), pp. 1321-58.

[11] Gibbons, Robert and Waldman, Michael. "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," *Journal of Labor Economics*, January 2006, *24*(1), pp. 59-107.

[12] Gittins, J. and Jones, D. "A Dynamic Allocation Index for Sequential Design of Experiments," in Joseph Gani, Karoly Sarkadi and Istvan Vincze, eds., *Progress in Statistics.* Amsterdam: North-Holland, 1974, pp. 161-173.

[13] Harris, Milton and Holmstrom, Bengt. "A Theory of Wage Dynamics," *Review of Economic Studies*, July 1982, *49*(3), pp. 315-33.

[14] Ingram, Beth and Neumann, George, "The Returns to Skill," *Labour Economics*, February 2006 *13*(1), pp. 35-59, .

[15] Jovanovic, Boyan. "Job Matching and the Theory of Turnover," *Journal of Political Economy*, October 1979, *87*(5), pp. 972-990.

[16] Jovanovic, Boyan and Nyarko, Yaw. "Stepping Stone Mobility," *Carnegie Rochester Conference Series on Public-Policy*, June 1997, *46*(0), pp. 289-325.

[17] Miller, Robert. "Job Matching and Occupational Choice," *Journal of Political Economy*, December 1984, *92*(6), pp. 1086-1120.

[18] Neal, Derek. "The Complexity of Job Mobility Among Young Men," *Journal of Labor Economics*, April 1999, *17*(2), pp. 237-261.

[19] Ortega, Jaime. "Job Rotation as a Learning Mechanism," *Management Science*, October 2001, *47*(10), pp. 1361-1370.

[20] Papageorgiou, Theodore. "Learning Your Comparative Advantages," *mimeo*, Penn State University, June 2009.

[21] Pastorino, Elena. "Learning in Labor Markets and Job Mobility," *mimeo*, University of Iowa, January 2009.

[22] Pastorino, Elena. "Learning and Career Dynamics: Estimating the Value of Firm Experimentation" *mimeo*, University of Iowa, May 2010

[23] Poletaev, Maxim and Robinson, Chris, "Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000," *Journal of Labor Economics*, July 2008, *26*(3), pp. 387-420.

[24] Radner, Roy and Stiglitz, Joseph. "A Nonconcavity in the Value of Information," in M. Boyer and R. Kihlstrom, eds., *Bayesian Models of Economic Theory*, Elsevier, 1984, pp. 33-52.

[25] Yamaguchi, Shintaro "Career Progression and Comparative Advantage," *Labour Economics*, August 2010, *17*(4), pp. 679-689.

Appendix A

Proof of Proposition 1

1. Competition and spot contracts imply that workers will never accept a wage w_t such that $w_t < Ey_t$ and that firms will never offer a wage w_t such that $w_t > Ey_t$. Thus, $w_t = Ey_t$ for all t.

2. This follows directly from the fact that workers choose α_t to maximize the present discounted value of their lifetime earnings and $w_t = Ey_t$ for all t.

3. Suppose that instead of workers announcing job choices and firms making wage offers, firms simultaneously offer (α_t, w_t) and workers either accept or reject. Let α^* denote the solution to the worker's problem (see equation 11)

$$\max_{\alpha_t \in \{0,\ldots,1\}} \alpha_t \mu + (1-\alpha_t)k + \delta E_t[V(\mu_{t+1}, \sigma_{t+1})]$$

Suppose a firm offers a worker (α', w') where $\alpha' \neq \alpha^*$. A worker will accept α' if and only if

$$w' + \delta E_t[V(\mu_{t+1}, \sigma_{t+1})|\alpha'] > \alpha^* \mu + (1-\alpha^*)k + \delta E_t[V(\mu_{t+1}, \sigma_{t+1})|\alpha^*].$$

Optimality implies that

$$\alpha' \mu + (1-\alpha')k + \delta E_t[V(\mu_{t+1}, \sigma_{t+1})|\alpha'] < \alpha^* \mu + (1-\alpha^*)k + \delta E_t[V(\mu_{t+1}, \sigma_{t+1}|\alpha^*)].$$

Thus, (α', w') will be accepted if and only if $w' > \alpha' \mu + (1-\alpha')k$. However, if firms make this wage offer, then their expected profits will be negative since workers will never accept any future wage offers less than their expected productivity.

Proof of Proposition 2

The second period payoff is given by

$$U_2(\mu_2, \sigma_2) = w_2(\alpha_2, \mu_2, \sigma_2) = \alpha_2 * E(\theta|\mu_2, \sigma_2) + (1-\alpha_2) * k = \alpha_2 * \mu_2 + (1-\alpha_2) * k$$

If $\mu_2 > k$ then $w_2 = \mu_2 > \alpha_2 * \mu + (1-\alpha_2) * k$. The opposite holds if $\mu_2 < k$. If $\mu = k$ all jobs pay the same. Thus, the decision rule is optimal.

Proof of Proposition 3

We need to show that for any $\alpha_1^{'} > 0$ and $-\infty < \mu_1 < k$, there exists a large enough σ_1 such that $V(\mu_1, \sigma_1|\alpha_1^{'}) > V(\mu_1, \sigma_1|\alpha_1 = 0) = k(1+\delta)$. Using the well-known properties of a truncated normal distribution, it can be shown that this condition holds iff

$$s_2\phi(r) + \mu_1 + (k-\mu_1)\Phi(r) > \frac{(k-\mu_1)\alpha_1^{'} + \delta k}{\delta}, \tag{13}$$

where $r = \frac{\mu_1 - k}{\sigma_1}$. Since $\mu_1 < k$, and $\Phi(r) > 0$, it it is suffices to show that

$$s_2\phi(r) + \mu_1 > \frac{(k-\mu_1)\alpha_1^{'} + \delta k}{\delta}. \tag{14}$$

The above equation holds iff

$$s_2 > \frac{(k-\mu_1)(\alpha_1^{'} + \delta)}{\delta\phi(r)}. \tag{15}$$

Since

$$\frac{\partial\phi(r)}{\partial\sigma_1} = exp\{\frac{(k-\mu_1)^2(\sigma_1^2\alpha_1^2+1)}{2\sigma_1^4\alpha_1^2}\}\frac{(k-\mu_1)^2(\sigma_1^2\alpha_1^2+2)}{\sqrt{2\pi}\sigma_1^5\alpha_1^2} > 0,$$

the right hand side of equation (15) is continuous and decreasing in σ_1. In addition,

$$lim_{\sigma_1\to\infty}\frac{(k-\mu_1)(\alpha_1^{'}+\delta)}{\delta\phi(r)} = \frac{\sqrt{2\pi}(k-\mu_1)(\alpha_1^{'}+\delta)}{\delta}.$$

Further, since

$$s_2 = \frac{\alpha_1\sigma_1^2}{\sqrt{\alpha_1^2\sigma_1^2+1}},$$

the left-hand side of equation (15) is continuous and increasing in σ_1 and $lim_{\sigma_1\to\infty}s_2 = \infty$, there exists $\sigma_1 < \infty$ such that equation 13 holds.

Proof of Proposition 4

Again, using the well-known properties of a truncated normal distribution, the value function, given the optimal assignment rule in Proposition 2, can be written as

$$(1-\alpha_1)k + \alpha_1\mu_1 + \delta\{\Phi(r)k + (1-\Phi(r))[\mu_1 + s_2\frac{\phi(r)}{1-\Phi(r)}]\},$$

which simplifies to,

$$k + \alpha_1(\mu_1 - k) + \delta\mu_1 + \delta[(k - \mu_1)\Phi(r) + s_2\phi(r)],$$

where $r = \frac{k-\mu_1}{s_2}$ and $\phi(t)$ and $\Phi(r)$ are standard normal density and standard cumulative density functions, respectively. Equating the derivative of the above equation with respect to α_1 to zero gives the condition in equation (10).

$$\frac{\partial\phi(r)}{\partial\alpha_1}s_2 + \frac{\partial s_2}{\partial\alpha_1}\phi(r) + (k - \mu_1)\frac{\partial\Phi(r)}{\partial\alpha_1} = \frac{k - \mu_1}{\delta}.$$

Proof of Proposition 5

It can be shown that $\frac{\partial\phi(r)}{\partial\alpha_1}s_2 = -(k - \mu_1)\frac{\partial\Phi(r)}{\partial\alpha_1}$, thus the first order condition in equation (10) reduces to

$$\frac{\sigma_1^2 exp\{-\frac{1}{2}\frac{(\sigma_1^2\alpha_1^2+1)(k-\mu_1)^2}{\sigma_1^4\alpha_1^2}\}}{\sqrt{2\pi}(\sigma_1^2\alpha_1^2 + 1)^{1.5}} = \frac{k - \mu_1}{\delta}. \tag{16}$$

1.

$$\lim_{\sigma_1\to 0}\frac{\sigma_1^2 exp\{-\frac{1}{2}\frac{(\sigma_1^2\alpha_1^2+1)(k-\mu_1)^2}{\sigma_1^4\alpha_1^2}\}}{\sqrt{2\pi}(\sigma_1^2\alpha_1^2 + 1)^{1.5}} = \lim_{\sigma_1\to 0}\frac{\sigma_1^2}{\sqrt{2\pi}(\sigma_1^2\alpha_1^2 + 1)^{1.5}} * \lim_{\sigma_1\to 0} exp\{-\frac{1}{2}\frac{(\sigma_1^2\alpha_1^2 + 1)(k - \mu_1)^2}{\sigma_1^4\alpha_1^2}\} = \frac{0}{\sqrt{2\pi}} * 0 = 0$$

2.

$$\lim_{\sigma_1\to\infty}\frac{\sigma_1^2 exp\{-\frac{1}{2}\frac{(\sigma_1^2\alpha_1^2+1)(k-\mu_1)^2}{\sigma_1^4\alpha_1^2}\}}{\sqrt{2\pi}(\sigma_1^2\alpha_1^2 + 1)^{1.5}} = \lim_{\sigma_1\to\infty}\frac{\sigma_1^2}{(\sigma_1^2\alpha_1^2 + 1)^{1.5}} * \lim_{\sigma_1\to\infty}\phi(r) = 0 * \frac{1}{\sqrt{2\pi}} = 0$$

Since for any $0 < \sigma_1 < \infty$ and $0 < \alpha \leq 1$, $\frac{\sigma_1^2 exp\{-\frac{1}{2}\frac{(\sigma_1^2\alpha_1^2+1)(k-\mu_1)^2}{\sigma_1^4\alpha_1^2}\}}{\sqrt{2\pi}(\sigma_1^2\alpha_1^2+1)^{1.5}} > 0$ and continuous, it is increasing and eventually decreasing in σ_1.

Proof of Proposition 6

1. Suppose that in time t it is optimal for a worker to choose α' where $0 < \alpha' < 1$ and stay in that job forever. We know that at time $t + n$

$$\mu_{t+n} = \frac{\mu_t\sigma_{\tilde{\epsilon},t}^2 + \sigma_t^2\sum_{i=0}^{n-1} x_{t+i}}{\sigma_{\tilde{\epsilon},t}^2 + n\sigma_t^2}.$$

Further, by the Law of Large Numbers, we know that

$$\lim_{n\to\infty} \mu_{t+n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} x_{t+i} = \lim_{t\to\infty} \theta + \frac{1}{n} \sum_{i=0}^{n-1} \frac{\epsilon_{t+i}}{\alpha'} = \theta.$$

Thus, the worker's problem approaches the benchmark case in which θ is known. However, when θ is known, workers will optimally choose $\alpha = 1$ if $\theta > k$ and $\alpha = 0$ otherwise. Thus, in the limit, it cannot be optimal for the worker to stay in a job in which $0 < \alpha' < 1$.

Appendix B

We prove that the Bellman equation in 11 is a fixed point of the corresponding contraction mapping. Let $S \in \{R \times R^+\}$ be the vector of state variables (μ, σ). Let $g(S) \in B$ be the space of all bounded functions. Note that B is complete (the set of bounded functions defined on the real line is complete). We can write the value function as

$$V(S) = \sup_{\alpha \in \{\alpha^1, \ldots, \alpha^N\}} \{\hat{Y}(S) + \delta \int V(S^{'}) f(S^{'}|S) dS^{'}\}. \tag{17}$$

and show that it is a fixed point of

$$Tc(S) = \sup_{\alpha \in \{\alpha^1, \ldots, \alpha^N\}} \{\hat{Y}(S) + \delta \int g(S^{'}) f(S^{'}|S) dS^{'}\}. \tag{18}$$

We first show that T is a contraction mapping by showing it satisfies Blackwell's sufficient conditions for contraction. First, we show that T maps every function which belongs to the space of bounded functions into a function in the same space: $T : B(S) - > B(S)$. Note that $\hat{Y}(S) = E[\theta\alpha + k(1-\alpha) + \epsilon] = \mu\alpha + k(1-\alpha)$ and therefore, it is bounded. Since g is bounded, the expectations over it: $\delta \int g(S^{'}) f(S^{'}|S) dS^{'}$ is a bounded function as well; thus, T maps any bounded function into a bounded function. Now we need to show that monotonicity and discounting conditions are satisfied.

Monotonicity: let $h, g \in B(S)$. If $g \geq h$ then $Tg(S) \geq Th(S)$.

$$Tg(S) - Th(S) = \sup_{\alpha} \delta \int [g(S^{'})|S] p(S^{'}|S) dS^{'}\} - \sup_{\alpha} \delta \int [h(S^{'})|S] p(S^{'}|S) dS^{'} \geq 0$$

This is straightforward: if the function g is weakly greater then h in every point then the sup of g is weakly greater too.

Discounting: $T(g(s) + c) - T(g(s)) \leq \delta c$. where $0 \leq \delta \leq 1$

$$T(g(s) + c) - T(g(s)) = \sup_{\alpha} \delta \int ([g(S^{'})|S] + c) p(S^{'}|S) dS^{'}\} - \sup_{\alpha} \delta \int [g(S^{'})|S] p(S^{'}|S) dS^{'}$$

This is less or equal than

$$\sup_{\alpha} [\delta \int ([g(S^{'})|S] + c - [g(S^{'})|S]) p(S^{'}|S)) dS^{'} = \delta c$$

Table 1: Average Value of α by Highest Degree and Actual Experience

	Actual Experience			
	1 Week	100 Weeks	200 Weeks	300 Weeks
High School Degree	0.304	0.352	0.379	0.389
	(0.007)	(0.008)	(0.008)	(0.009)
	[813]	[814]	[816]	[816]
Associate Degree	0.477	0.535	0.572	0.573
	(0.027)	(0.027)	(0.025)	(0.025)
	[83]	[81]	[83]	[83]
BA or BS (or higher)	0.659	0.709	0.733	0.750
	(0.013)	(0.011)	(0.011)	(0.010)
	[466]	[471]	[470]	[468]
All	0.436	0.486	0.512	0.524
	(0.008)	(0.008)	(0.008)	(0.008)
	[1,362]	[1,366]	[1,369]	[1,367]

Note: The standard error of the mean is in parentheses. The cell size in is brackets.

Table 2: Average Value of α by Major Census Category and Actual Experience

	Actual Experience			
	1 Week	100 Weeks	200 Weeks	300 Weeks
Professional and Technical	0.884	0.867	0.868	0.879
	(0.007)	(0.007)	(0.007)	(0.007)
	227	249	256	270
Managers and Administrators	0.779	0.791	0.789	0.790
	(0.006)	(0.006)	(0.005)	(0.004)
	104	152	173	200
Sales Workers	0.555	0.576	0.589	0.585
	(0.013)	(0.013)	(0.013)	(0.012)
	69	82	99	94
Clerical and Unskilled	0.410	0.430	0.433	0.445
	(0.009)	(0.011)	(0.012)	(0.013)
	143	125	132	104
Craftsman	0.481	0.494	0.504	0.505
	(0.009)	(0.008)	(0.008)	(0.008)
	188	231	254	249
Operatives Except Transport	0.223	0.201	0.189	0.194
	(0.010)	(0.010)	(0.009)	(0.009)
	141	140	150	126
Transport Operatives	0.192	0.206	0.192	0.159
	(0.017)	(0.014)	(0.015)	(0.013)
	61	76	65	79
Non-Farm Laborers	0.099	0.098	0.095	0.108
	(0.003	(0.004	(0.006	(0.006
	189	137	93	105
Farmers and Farm Managers	0.660	0.660	0.673	0.676
	(0.000)	(0.000)	(0.013)	(0.016)
	[2]	[5]	[6]	[5]
Farm Laborers and Foremen	0.211	0.200	0.231	0.246
	(0.013)	(0.000)	(0.031)	(0.039)
	[39]	[19]	[16]	[18]
Service Workers	0.270	0.323	0.334	0.323
	(0.011)	(0.013)	(0.015)	(0.015)
	[199]	[150]	[124]	[116]
Private Household Workers	.	.	0.180	0.180
	.	.	.	.
	[0]	[0]	[1]	[1]

Note: The standard error of the mean is in parentheses. The cell size in is brackets.

Table 3: Relationship Between Wage Growth and Initial Job Assignment

VARIABLES	(1) High School	(2) College
Initial wage	0.587***	0.690***
	(0.0368)	(0.0514)
Initial alpha	0.266	0.0366
	(0.512)	(1.063)
Actual experience (in years)	1.020***	1.515***
	(0.101)	(0.234)
Initial wage x actual experience	-0.0507***	-0.0551***
	(0.0104)	(0.0212)
Initial alpha x actual experience	0.433**	0.827**
	(0.174)	(0.322)
Constant	3.682***	4.606***
	(0.356)	(0.689)
Observations	267921	155675
R-squared	0.170	0.266

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Relationship Between Growth in Wage Dispersion and Initial Job Assignment

VARIABLES	(1) High School	(2) College
Initial wage quartile	0.0126***	-0.1093***
	(0.0043)	(0.0058)
Initial alpha quartile	0.0142***	-0.0961***
	(0.0043)	(0.0058)
Actual experience (in years)	0.0135***	0.0320***
	(0.0041)	(0.0056)
Initial wage quartile x actual experience	0.0044***	-0.0105***
	(0.0011)	(0.0015)
Initial alpha quartile x actual experience	0.0134***	0.0175***
	(0.0011)	(0.0015)
Constant	0.6762***	1.2507***
	(0.0160)	(0.0217)
Observations	5600	5600
R-squared	0.371	0.397

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Relationship Between Job Assignment and Experience, Low-Wage Workers

VARIABLES	(1) High School	(2) College
Initial wage	0.0014	-0.0007
	(0.0022)	(0.0021)
Initial alpha	0.3502***	0.3884***
	(0.0392)	(0.0413)
Actual experience (in years)	0.0167**	0.0296**
	(0.0076)	(0.0126)
Actual experience squared	-0.0015	-0.0038**
	(0.0012)	(0.0019)
Constant	0.1839***	0.3912***
	(0.0208)	(0.0331)
Observations	102458	48833
R-squared	0.109	0.193

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Second-Period Output as a Function of α_1

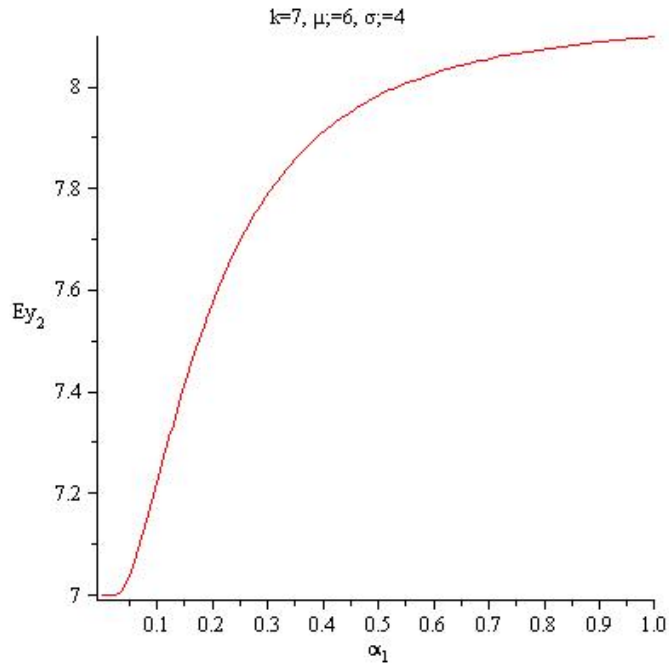


Figure 2: Optimal α_1 as a Function of μ_1, Two-Period Model

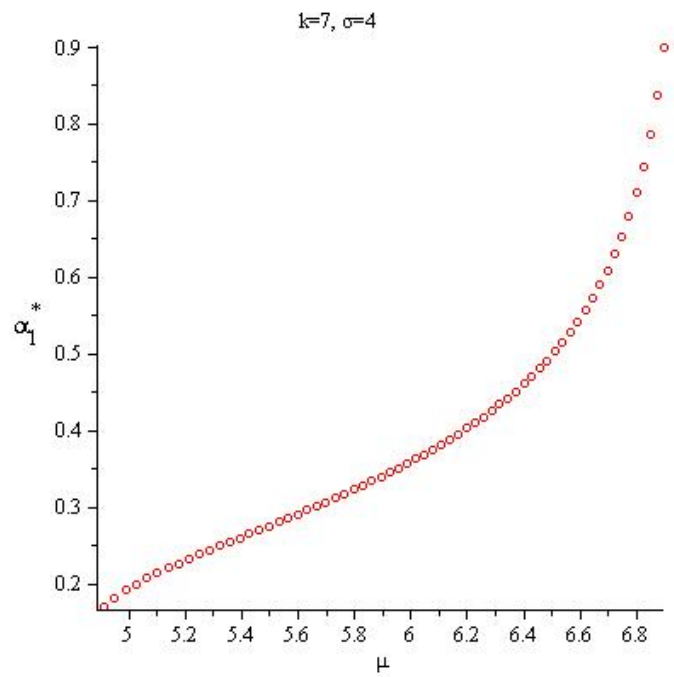


Figure 3: Optimal α_1 as a Function of σ_1, Two-Period Model

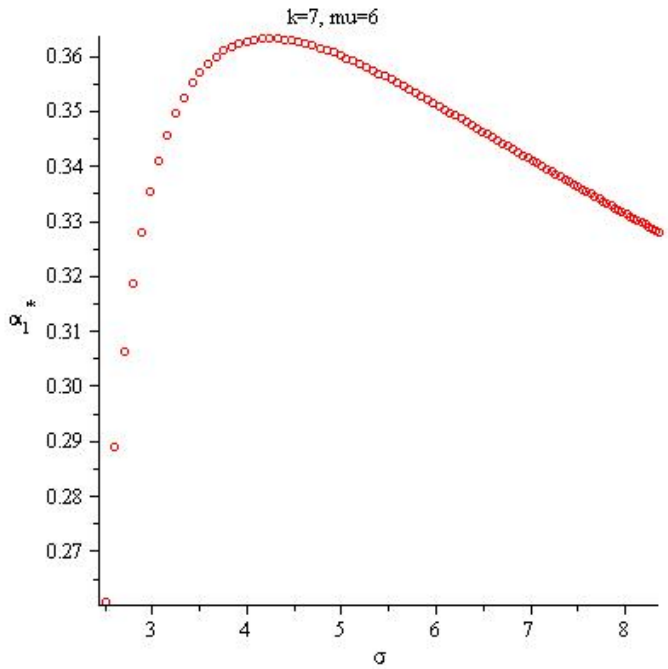


Figure 4: Marginal Change in Second-Period Output as a Function of σ_1

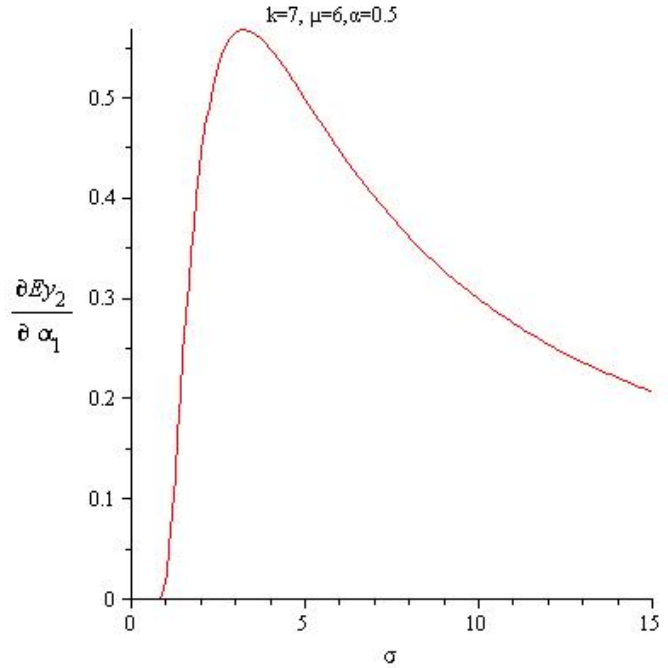


Figure 5: Second-Order Condition as a function of σ_1 and α_1. $\mu = 6$, $k = 7$

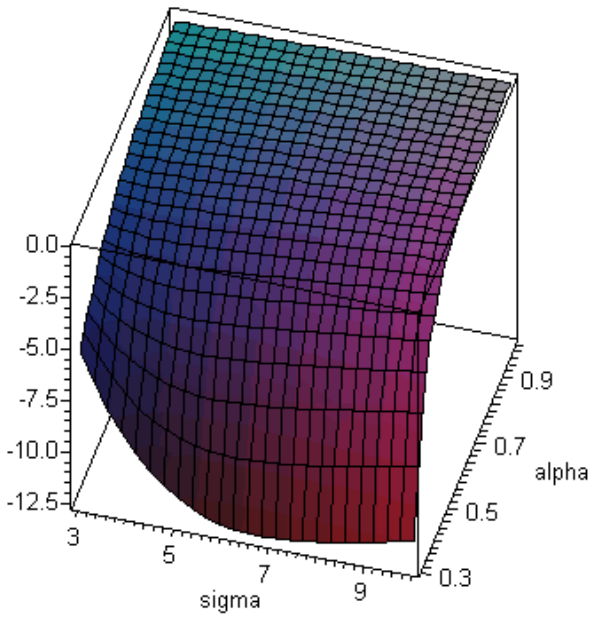


Figure 6: Optimal Job Choice: Three-Job Model

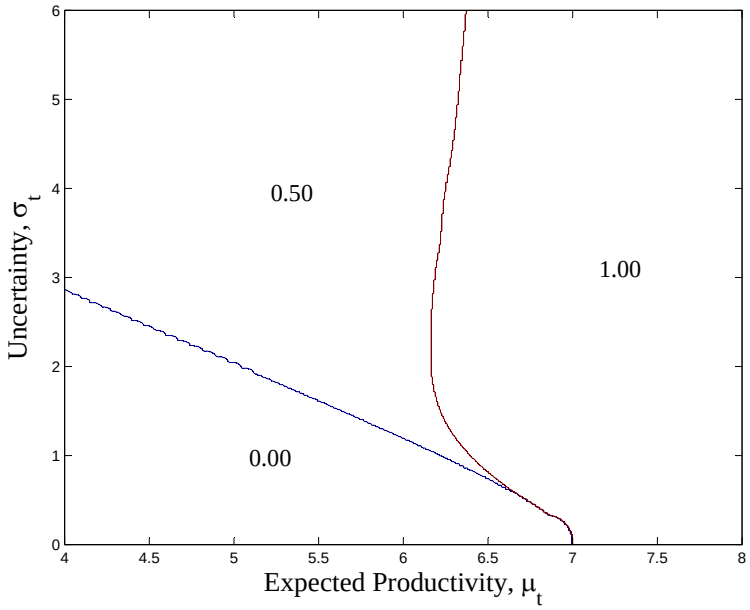


Figure 7: Optimal Job Choice: Five-Job Model

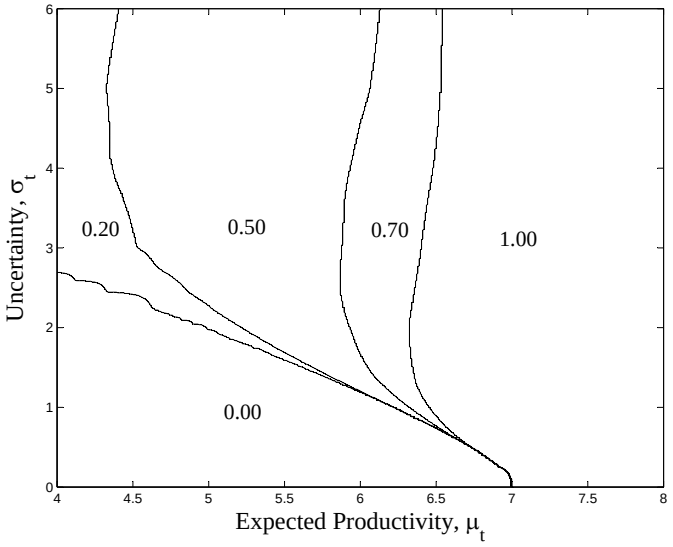


Figure 8: Optimal Job Choice: Two-Job Model

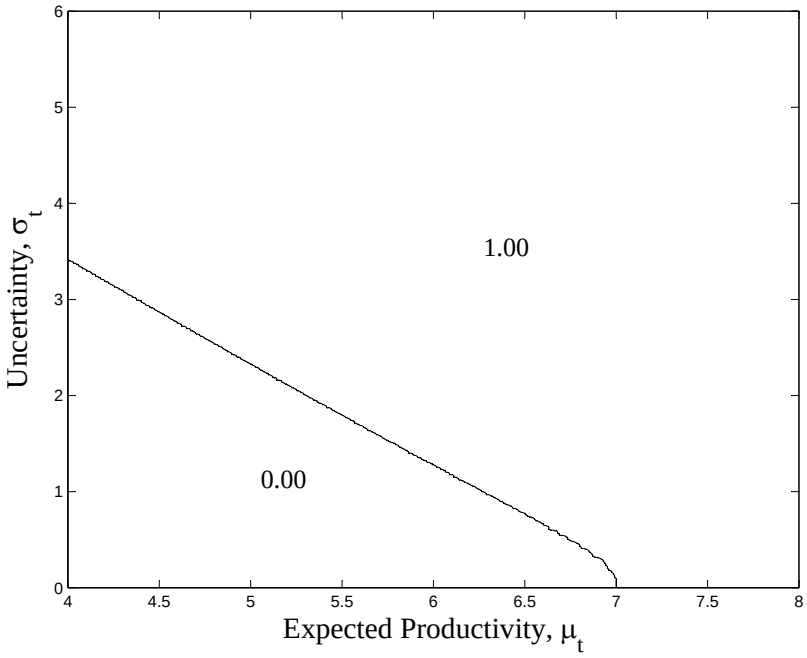
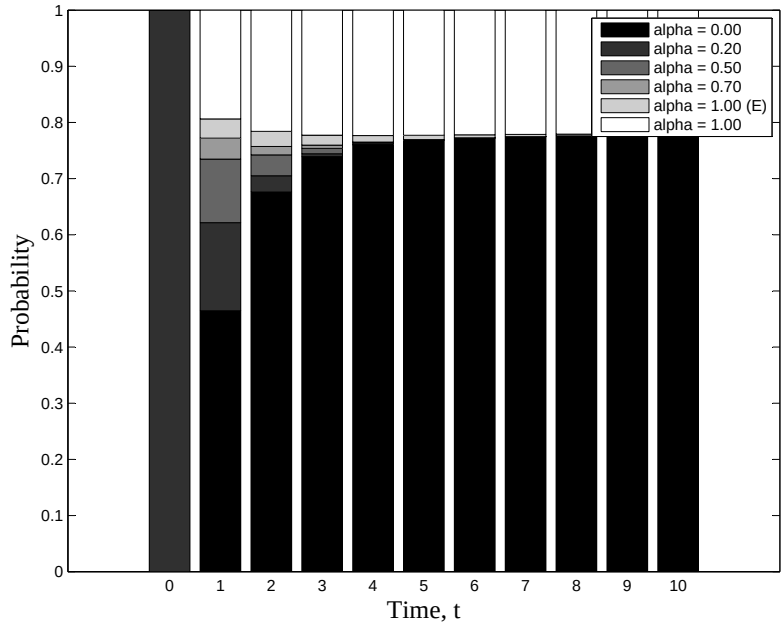


Figure 9: Job Assignments Over the Life-Cycle When $\mu_0 = 3$ and $\sigma_0 = 6$



Note: The light gray region shows the likelihood that $\alpha_t = 1$ but $\mu_t < k$ (experimentation) while the white region shows the likelihood that $\alpha_t = 1$ and $\mu_t \geq k$ (no experimentation) .

Figure 10: Percentiles of the Wage Distribution When $\mu_0 = 3$ and $\sigma_0 = 6$

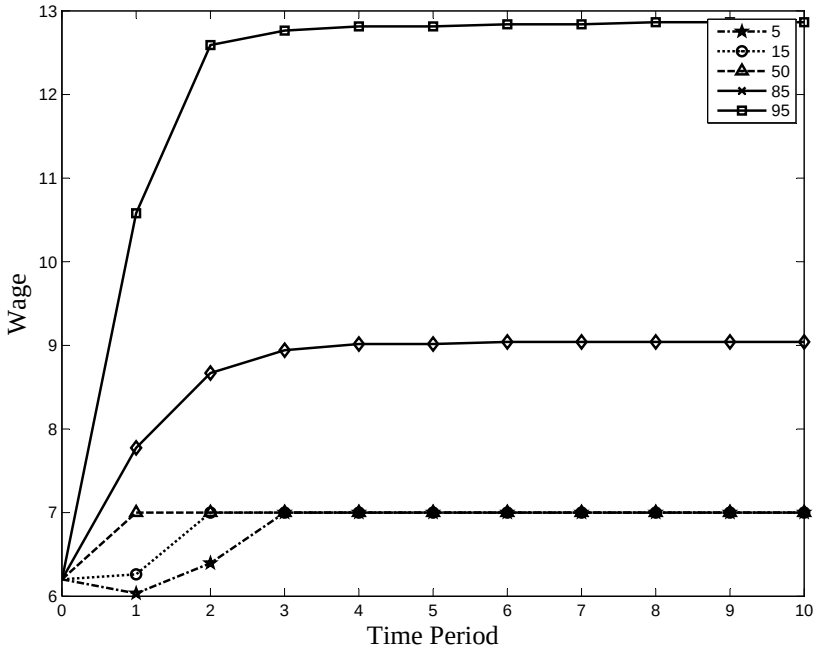


Figure 11: Difference in Value Function Between a Three-Job and Two-Job Model

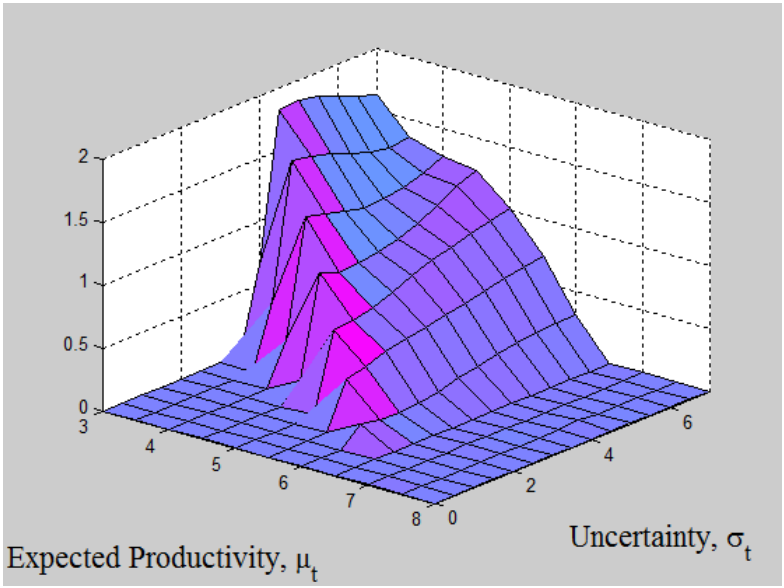
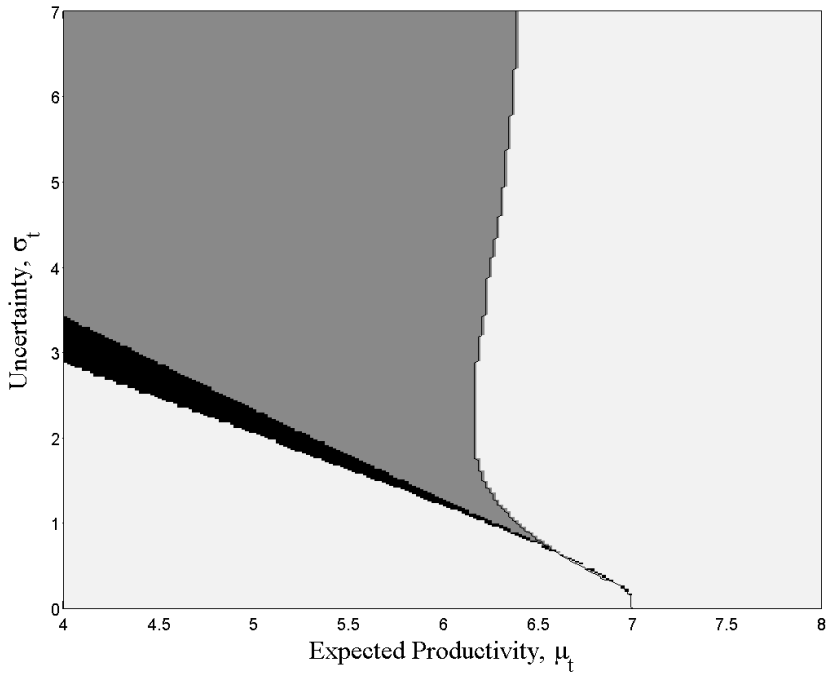


Figure 12: Comparison Between Job Choices in the Three-Job and Two-Job Model



Note: Black region: $\alpha = 0$ in two-job model, but $\alpha = 1$ in three-job model. Dark gray region: $\alpha = 1$ in two-job model, but $\alpha = 0.5$ in three-job model.

Figure 13: Cumulative Probability of Permanent Mistakes. $\mu = 6$, $\sigma = 4$, $\alpha \in \{0, 0.5, 1\}$

Time Period	Probability $\alpha = 0$ and $\theta > k$	Marginal Change
t = 1	0.00%	
t = 2	1.70%	1.70%
t = 3	2.67%	0.97%
t = 4	3.17%	0.50%
t = 5	3.45%	0.28%
t = 6	3.63%	0.19%
t = 7	3.75%	0.12%
t = 8	3.84%	0.08%
t = 9	3.91%	0.07%
t = 10	3.97%	0.06%

Figure 14: Wage dispersion Due to Luck with 3 Jobs, $\alpha \in \{0, 0.5, 1\}$

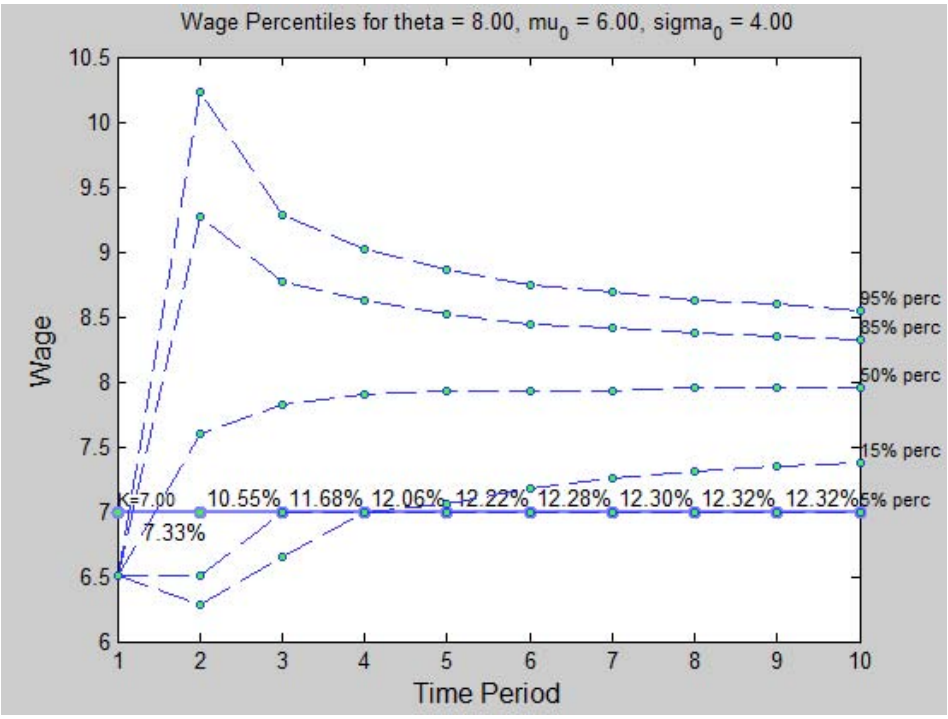


Figure 15: Wage Dispersion Due to Luck with 5 Jobs, $\alpha \in \{0, 0.2, 0.5, 0.7, 1\}$

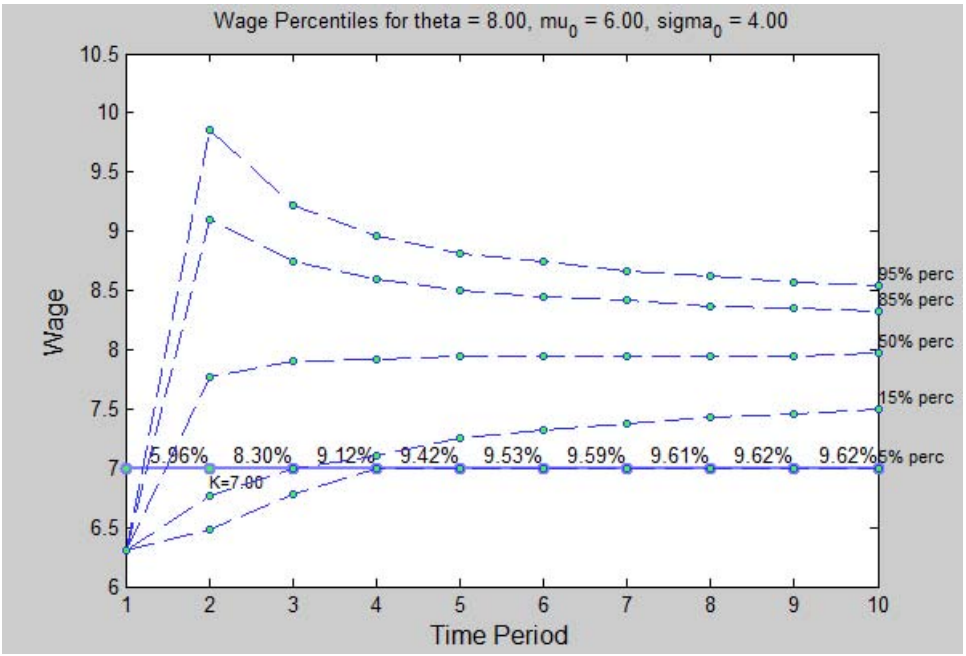


Figure 16: Average α by Actual Experience

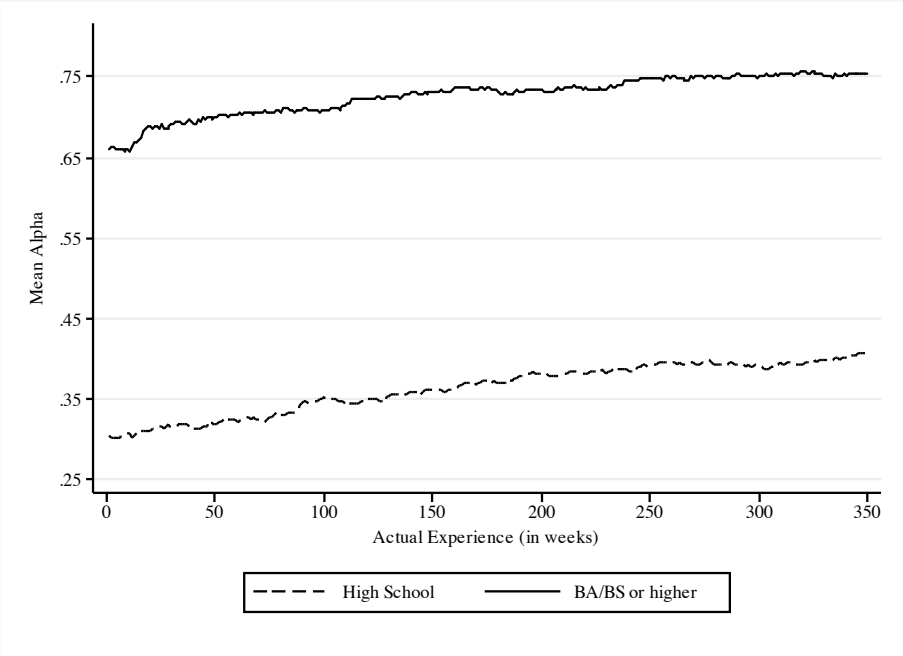


Figure 17: Distance from Initial α by Actual Experience

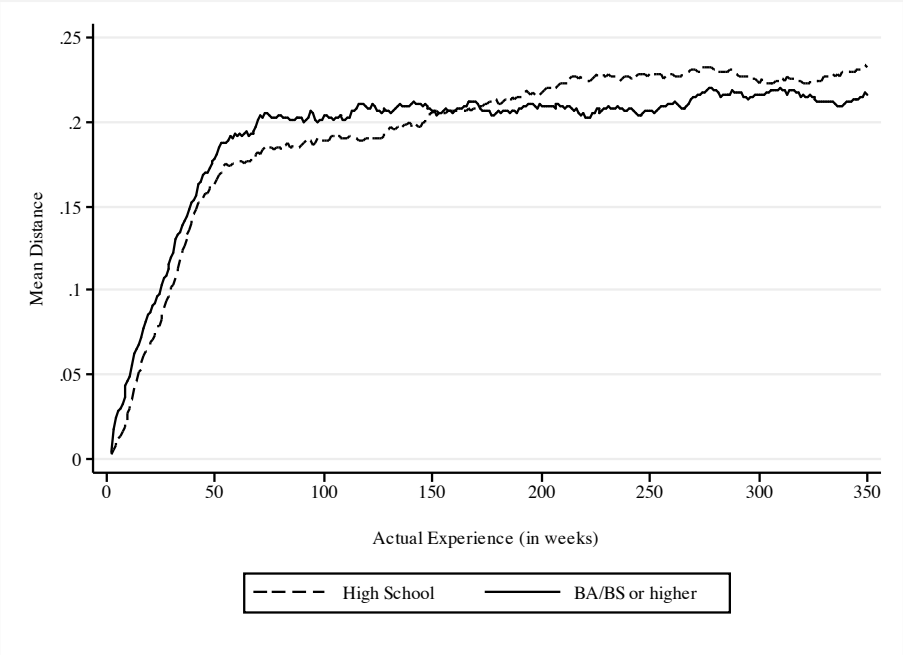


Figure 18: Mean Wage by Actual Experience

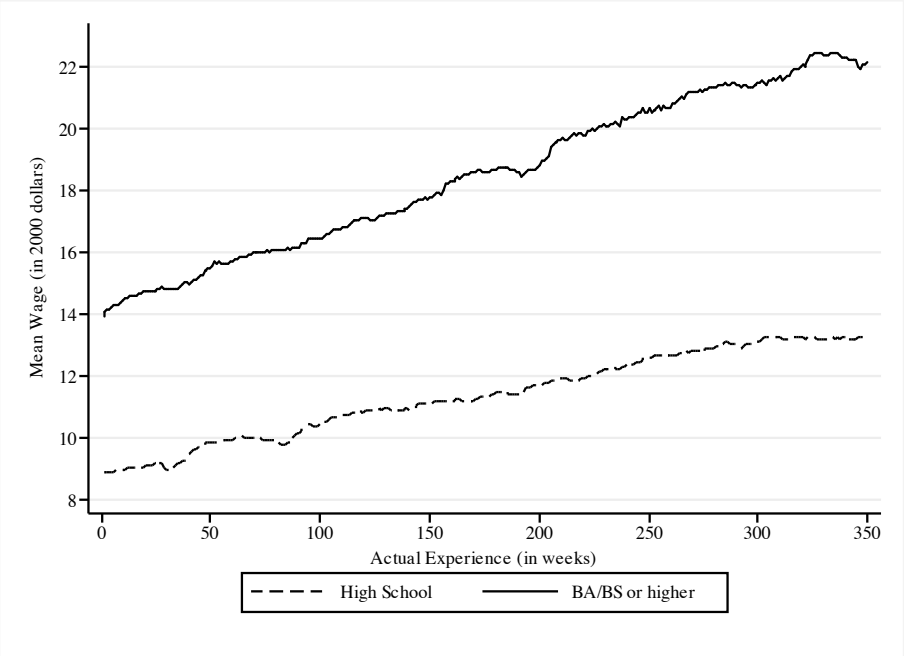


Figure 19: Standard Deviation of Wages by Actual Experience

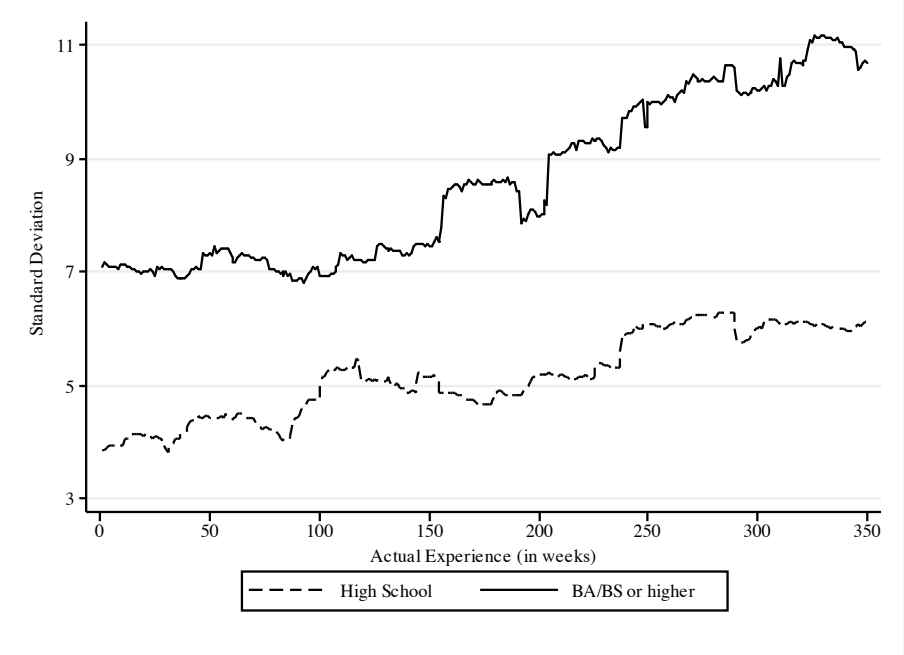


Figure 20: Predicted α by Experience, Low-Wage Workers, College Graduates

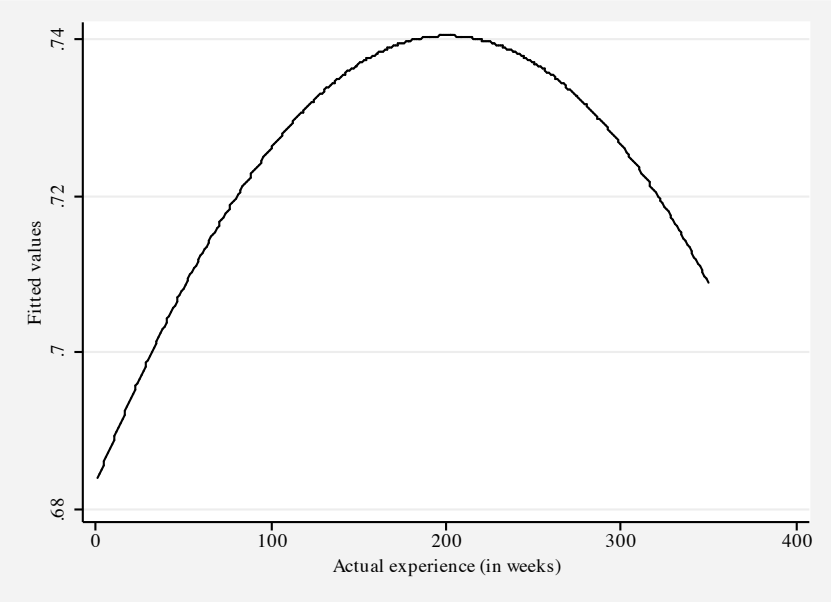
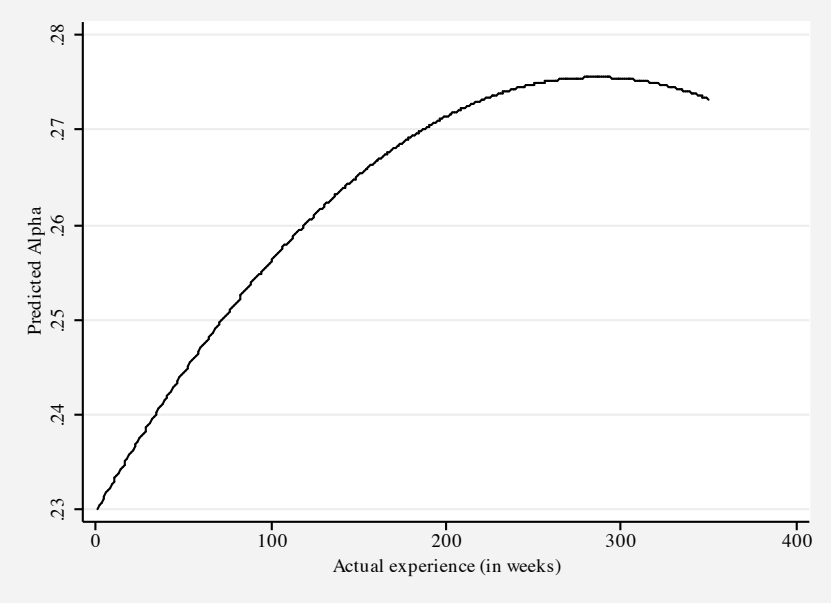



Figure 21: Predicted α by Experience, Low-Wage Workers, High School Graduates