Experience and Worker Flows*

Aspen Gorry

University of California, Santa Cruz

aspen@ucsc.edu

January 5, 2011

Abstract

This paper extends the literature on learning in labor markets by parameterizing the amount of learning that transfers across jobs. Previous models have assumed that learning is either job specific as in Jovanovic (1979) or perfectly transferable across jobs as in Gibbons et al. (2005). By allowing some but not all learning to be transferred, this model generates novel predictions of a decline in job finding rates with age and a decline in the volatility of wages with experience that are consistent with observed worker outcomes.

Keywords: Learning, Job Finding Rates, Job Separation Rates, Wage Volatility.

JEL Codes: J64, J24, J31.

*Please send comments to aspen@ucsc.edu. I thank numerous seminar participants, Fernando Alvarez, Carlos Dobkin, Derek Neal, Ezra Oberfield, Ryan Oprea, Robert Shimer, Nancy Stokey, Carl Walsh, and Andy Zuppann for helpful comments. All mistakes are my own. Previously distributed as: "Labor Market Experience and Worker Flows." First draft: October 2008.

1 Introduction

Learning models account for many important features of labor market behavior. Jovanovic's (1979) early work explains broad features of worker turnover behavior: a hump shaped hazard of separation from a job by tenure and declining separation rates with age. Recent models use learning to understand wage dispersion, wage growth, and occupational mobility (for example Moscarini (2005), Farber and Gibbons (1996), Gibbons et al. (2005), and Papageorgiou (2007)[1]).

For analytical tractability, the literature on learning focuses on models that make stark assumptions about the form of learning. On one hand, matching models like Jovanovic (1979) assume that all learning is specific to a particular job. A worker's performance on a particular job provides information only about that job. On the other hand, sorting models take learning to be about a worker's ability. In these models, a worker's performance on one job generates an equivalent amount of knowledge about her performance on all other jobs. Workers then use their current belief about their ability to sort themselves into the most profitable job. These assumptions are stark as workers learn about their ability on a particular job and some but not all of this information is useful in determining their productivity in other potential endeavors.

This paper constructs a search model to bridge the gap between these extreme assumptions in the literature. The model extends the matching framework to allow agents to learn not only about their current match, but also allow past learning to be useful in discerning the quality of their prospective matches when unemployed. This initial screening is similar to Jovanovic (1984), however the amount of information contained in the signal depends on worker's past experience. Workers learn rapidly about their ability on a particular match and some but not all of this learning carries over into future matches. The model parameterizes how much learning from one job carries over to understand how productive the worker will be in other job opportunities.

[1]For a summary of the literature on learning and a summary of the stylized facts on the distribution of labor earnings see Neal and Rosen (2000).

The model explains how young workers transition from rapid turnover to stable employment over the life cycle. During the first ten years of labor market experience, workers transition from high job turnover into stable employment and have rapid wage growth. About two-thirds of lifetime job turnover and wage growth occurs during these early years (see Topel and Ward (1992), Flinn (1986)). Initial high turnover manifests itself in both high job finding and separation rates for young workers.

The model captures the well known decline in worker turnover with age (see Clark and Summers (1982)). Past models of turnover have focused on explaining the decline in job separation rates. However, less focus has been paid to the observed decline in job finding rates. The model in this paper replicates attractive features of previous learning models including the decline in unemployment and job separation rates with age and the rise in wages with labor market experience. Allowing experience to generate differential amounts of learning about current and future jobs generates a theory of the patterns of job finding rates and wage volatility by experience.

The calibrated model generates declining job finding rates with age as experience allows workers to distinguish between good and bad job offers. For inexperienced workers jobs are experience goods; they only learn about the quality of the match by trying it out. However, as workers gain experience jobs become inspection goods. Market experience influences decisions by unemployed workers about which jobs to accept. As their experience grows, they reject more bad jobs causing the job finding rate to decline. The past literature on learning does not generate any prediction on job finding rates. In matching models, learning is completely job specific so employment forms a renewal process as workers are in an identical situation each time they become unemployed. In sorting models, although information transfers between jobs, perfect transfer of information means that workers direct their search to the job that best fits their abilities. Embedding learning across jobs into a matching framework generates a mechanism for past experience to alter a workers search behavior and change their job finding rates.

The calibrated model is then used to generate novel predictions about the volatility of wages in new jobs. The model predicts wage volatility declines with experience. Intuitively, more experience from previous jobs generates more information about new matches. This implies that wages should vary less for workers starting a new job with more past experience. This new implication from the learning model is confirmed by examining wage data from the National Longitudinal Survey of Youth 1979 (NLSY79) data. Both the calibrated model and the data show about a 20% decline in one year wage volatility for a worker at a new job after 10 years of experience.

The model draws closely from Moscarini (2005) who assumes that jobs are drawn from a distribution of only two types. Moscarini (2005) and Moscarini (2003) use this trick to embed Jovanovic's (1979) model into a general equilibrium framework and explore implications for the wage distribution. Papageorgiou (2007) extends these models to explore occupational choices. The paper also relates to a literature that seeks to explain the decline in worker turnover as workers age. Neal (1999) presents a model where workers search for both a career and job specific match. The empirical implications of career and job matches for job turnover and wages are explored in Pavan (2007) and Pavan (2006) respectively. This paper generates observed declines in job finding and separation rates without adding the complexity of a second type of career match.

The paper proceeds as follows. Section 2 presents the model. Section 3 describes how the parameters of the model are chosen. Section 4 presents the results from the calibrated model about job finding and separation rates, unemployment and wage growth. Section 5 shows that the model predictions about wage volatility are consistent with data from NLSY79. Section 6 concludes.

2 Model

This section describes the economic environment of an individual making optimal decisions when faced with uncertain production opportunities (jobs). She searches for production opportunities and when confronted with one she learns about its quality.

2.1 Production

The infinitely lived worker has preferences given by:

$$U = \sum_{t=1}^{\infty} \beta^{t-1} c_t$$

There is no storage technology. The worker makes two decisions: when matched with an opportunity she decides between quitting to search for a new opportunity and continuing to produce and when unmatched she choose to accept or reject opportunities as she finds them.

Production occurs when a worker is matched with a productive opportunity. In each period, a match of type μ produces output:

$$x_t = \mu + z_t$$

where $z_t \sim N(0, \sigma^2)$ is independently and identically distributed noise on the output process. Therefore, $x_t \sim N(\mu, \sigma^2)$.

As in Moscarini (2005), the economy is composed of two types of opportunities: $\mu \in \{\mu_h, \mu_l\}$. Let $\mu_h > \mu_l$ so that μ_h denotes the productivity of a good opportunity and μ_l denotes the productivity of a bad one. All production opportunities are drawn independently from the same distribution where a fraction p_0 of them are of type μ_h.

2.2 Learning

The worker is uncertain about the quality of her current production. She learns about the quality of the match in two ways. First, while employed she observes her output in the current production opportunity and updates her beliefs about the quality of the match using Bayes' rule. Second, when an unmatched worker finds a new opportunity she receives a signal about its quality that depends on her past experience.

While matched, workers observe the output they produce in each period and update their beliefs. Given the normality of output noise, for any current belief, p, the expected density of output is given by:

$$\psi(x|p) = p\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_h}{\sigma}\right)^2} + (1-p)\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu_l}{\sigma}\right)^2}$$

With probability p output is drawn from a normal distribution with mean μ_h and variance σ, while with probability $1-p$ it is drawn from a normal with mean μ_l and the same variance.

Using this known distribution of output, the worker observers her production and uses it to update her belief about the probability that she has a good match using Bayes' rule. Given any current belief, p, and observed output for a given period, x, the updated belief, p', is is formed using Bayes' rule:

$$f(p,x) \equiv p' = Prob(\mu = \mu_h|p,x) = \frac{pe^{-\frac{1}{2}\left(\frac{x-\mu_h}{\sigma}\right)^2}}{pe^{-\frac{1}{2}\left(\frac{x-\mu_h}{\sigma}\right)^2} + (1-p)e^{-\frac{1}{2}\left(\frac{x-\mu_l}{\sigma}\right)^2}}$$

Here the numerator is proportional to the joint probability of observing output x and the match being good where the denominator is proportional to the total probability of observing output x.

With this updating function, define the inverse function $f^{-1}(p'|p)$ to be the x required to have posterior p' given prior p. This function is given by:

$$f^{-1}(p'|p) = \frac{\sigma^2}{\mu_h - \mu_l}\ln\left(\frac{p'(1-p)}{(1-p')p}\right) + \frac{\mu_h + \mu_l}{2}$$

Define the distribution $G(p'|p)$ as the distribution of updated beliefs after observing one period of output given a current belief p. Then the p.d.f. of the G distribution, g, is given by:

$$\begin{aligned} g(p'|p) &= \psi(f^{-1}(p'|p)|p)\left|\frac{df^{-1}(p'|p)}{dp'}\right| \\ &= \psi(f^{-1}(p'|p)|p)\left(\frac{\sigma^2}{p'(1-p')(\mu_h - \mu_l)}\right) \end{aligned}$$

When meeting a new match the worker gets an initial signal about the quality of the match that depends on her past experience. She receives a signal that is equivalent to observing $\alpha\tau + k$ observations from the output process. Where τ is months of past work experience, $\alpha \in [0, 1]$ determines the fraction of experience that carries over from past jobs into information about new offers, and $k > 0$ sets the initial distribution of beliefs for new workers. The information that a worker with no experience gets is equivalent to observing k periods of output from the production process. The normality assumption makes non-integer observations well defined. Moreover, normality implies that to update beliefs after viewing t observations the worker only needs to know her prior belief p, the average value of the observation $\bar{x}$, and the number of observations observed t, not the entire list of observations $x_1, x_2, \ldots, x_t$. For a worker who observes t periods of output, the distribution of the average output per period, $\bar{x}$, is given by:

$$\tilde{\psi}(\bar{x}; p, t) = p \frac{1}{\frac{\sigma}{\sqrt{t}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_h}{\frac{\sigma}{\sqrt{t}}}\right)^2} + (1-p) \frac{1}{\frac{\sigma}{\sqrt{t}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_l}{\frac{\sigma}{\sqrt{t}}}\right)^2}$$

Using the same updating strategy, the posterior after observing the output from t periods is computed as:

$$\tilde{f}(p, \bar{x}, t) = \frac{pe^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_h}{\frac{\sigma}{\sqrt{t}}}\right)^2}}{pe^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_h}{\frac{\sigma}{\sqrt{t}}}\right)^2} + (1-p)e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu_l}{\frac{\sigma}{\sqrt{t}}}\right)^2}}$$

Again, inverting $\tilde{f}$ gives the value of $\bar{x}$ needed to generate posterior p': $\tilde{f}^{-1}(p', p, t) = \bar{x}$. Define $H(p'|\tau)$ as the distribution of initial beliefs from a new production opportunity. Hence the p.d.f. of the H distribution, h, is given by:

$$h(p'|\tau) = \tilde{\psi}(\tilde{f}^{-1}(p', p_0, \alpha\tau + k); p_0, \alpha\tau + k) \left(\frac{\sigma^2}{p'(1-p')(\mu_h - \mu_l)}\right)$$

where α and p_0 are parameters. p_0 is the prior probability that any new opportunity is good[2].

[2]The process of on the job learning can be generalized beyond the specified output process to be arbitrary distributions G and H. The distribution G must depend on the value of the current belief, p, so the

2.3 Wages

The worker's period payoffs from the production in the model are given by the expected value of output in each period. Given this output process, the wage is given by:

$$w(p) = p\mu_h + (1 - p)\mu_l$$

This wage process is an equilibrium in an environment where there are a continuum of production opportunities (firms) that have no cost of entering the market. The production opportunities must make zero expected profits. Under these assumptions any wage process that pays the average wage along with any matching rate, λ, between workers and production opportunities can be sustained as an equilibrium outcome[3].

2.4 Value Functions

This section defines the value functions for the worker's general problem. When employed the worker consumes her wage, $c_t = w(p)$, that depends on the probability that her job is good. The worker can separate from the job for two reasons. First, she could receive an unfavorable signal about the job quality and decide to quit. Second, in each period with

distribution of updated beliefs, p', is given by $G(p'|p)$. For a general learning process, two restrictions are made on G. First, G is non-degenerate so that the signal conveys some information about p. Second, G is restricted so that p is a martingale. This is a natural restriction since G is used to update an individual's current beliefs.

The distribution $H(p'|\tau)$ can be generalized beyond the specific normality assumptions described above. In general, for H to provide more information about jobs it must be weakly increasing in τ in terms of second order stochastic dominance. This means that for $\tau_1 > \tau_2$:

$$\int_0^x H(p'|\tau_1) - H(p'|\tau_2) dp' \geq 0 \quad \forall \;\; x \in [0, 1]$$

For higher values of τ workers get more initial information about the quality of a job. This increasing information for experienced unemployed workers is the novel feature of the model. A sufficient condition for second order stochastic dominance is that if $\tau_1 > \tau_2$ then $H(p'|\tau_1)$ is a mean preserving spread of $H(p'|\tau_2)$.

[3] This equilibrium concept assumes that production opportunities (firms) are unable to separate workers of different levels of experience when matching. The focus of this paper is to understand the implications of learning on worker's job decisions. Another interesting question would be to understand how firm's ability to select workers of different experience levels impacts equilibrium employment outcomes. Such a contribution is beyond the scope of this paper.

probability $\delta > 0$ an employed worker is exogenously separated from her job. δ captures reasons for job separations not captured by the endogenous quits that arise from learning. Possible examples include plant closures or geographic relocation by the worker.

Let $V(p, \tau)$ be the value function for an employed worker with belief p and experience τ. The value is written as:

$$V(p, \tau) = w(p) + \beta\delta U(\tau + 1) + \beta(1 - \delta) \int_0^1 \max\{U(\tau + 1), V(p', \tau + 1)\} G(dp'|p) \qquad (1)$$

A worker with belief p and experience τ gets her expected output $w(p)$. In the next period, she is separated from her job with probability δ, becoming unemployed with experience $\tau+1$. With probability $1-\delta$ she is not separated from her job and receives her updated belief from the distribution G. Depending on the realization of her updated belief she can choose to remain employed with belief p' and experience $\tau + 1$ or quit to become unemployed with experience $\tau + 1$.

Unemployed workers consume the unemployment value $c_t = b$. b is high enough that if a worker knows for certain that a job is bad it is optimal to quit and low enough so that if the worker knows that the job is good that she will work. These assumptions ensure that the worker's search problem is non-trivial.

When unemployed, the worker with experience τ gets an offer from the distribution of jobs $H(p'|\tau)$ with probability λ . She must choose between remaining unemployed and becoming employed with belief p'. If she does not receive a job offer she remains unemployed with the same experience.

Let $U(\tau)$ be the value function for an unemployed worker with experience τ. The value function is given by:

$$U(\tau) = b + \beta(1 - \lambda)U(\tau) + \beta\lambda \int_0^1 \max\{U(\tau), V(p', \tau)\} H(dp'|\tau) \qquad (2)$$

Finally, it is assumed that the maximum experienced that can be accumulated by a worker is T. While the infinite horizon model would in principle allow a worker to accumulated more experience, including a finite maximum experience attained simplifies the computation of the model and allows analytical results to be derived[4]. It can be justified on two separate grounds. First, T can be chosen to be large enough so that workers already have nearly perfect information about new production opportunities after T periods of past experience. Second, the finite nature of individual working lives means that workers only accumulate a finite amount of experience before retirement. The assumption implies that the marginal value of additional periods of experience is zero once a worker reaches T.

2.5 Model Characterization

The general learning framework described above embeds the the learning models of Jovanovic (1979) and Gibbons et al. (2005) into a matching framework so the implications for worker job finding and separation rates can be explored. When $\alpha = 0$ there is no learning across jobs and the model is identical to that of Jovanovic (1984) where workers search and get an initial signal about the quality of a match. In the case of $\alpha = 1$, a worker gets a signal about the quality of an initial opportunity of equal strength to their entire past job experience. All learning from a particular job carries over to future jobs. This case is similar to model of Gibbons et al. (2005) where all information transfers across jobs. The case of $\alpha = 1$ is however still different than Gibbons et al. (2005) as in their model workers direct their search to the most productive job, so experience does not interact with job finding rates. By integrating full information transfer across jobs into a matching framework, declines in job finding rates for part of the life-cycle can still be generated in the case where $\alpha = 1$. Different choices of $0 \leq \alpha \leq 1$ parameterize how much information transfers from one job to another.

The solution consists of a reservation level of productivity that depends on experience, $\bar{p}(\tau)$, such that workers will accept job offers or continue working as long as $p \geq \bar{p}(\tau)$

[4]An alternate assumption would be to use a finite horizon model. Using the finite horizon model expands the state space as the worker's age becomes an additional state variable. While the addition a state variable makes the model more cumbersome, the results are nearly identical beyond minor changes in the last couple years of working life. Since the novel implications of learning are concentrated at the beginning of the workers career the paper uses the more streamlined infinite horizon model.

and reject offers or quit otherwise. The general model is rich enough to allow for different relationships between experience, the reservation productivity level, the job finding rate, and wages that are explored quantitatively in the next sections of the paper. The rest of this section characterizes these relationships to build intuition about the workings of the model.

First, the reservation productivity level is solved for by setting the value function of a matched worker with the reservation productivity equal to the value of an unmatched worker for each level of experience. That is $\bar{p}(\tau)$ solves:

$$V(\bar{p}(\tau), \tau) = U(\tau) \tag{3}$$

The sign of $\bar{p}'(\tau)$ is indeterminate.

To see why $\bar{p}(\tau)$ might be decreasing in τ consider the following example. If a worker gets no extra information about the quality of jobs until she gains t units of experience then she gets a perfect signal after, there will be a space of experience just before t that the worker will be willing to accept worse and worse opportunities just to get the payoff from getting t units of experience. In this case, the option value of experience outweighs the current value to the worker and can generate decreasing reservation values.

The reservation productivity level will be increasing if the marginal value of information while employed at the reservation belief is less than the marginal value of information when unemployed. The reservation value increases when $U'(\tau + 1) \leq U'(\tau)$ because extra experience can only impact a worker when unemployed seeking a new job. This condition can be interpreted as requiring that the marginal value of experience for unmatched workers is declining. The direct benefit from the additional unit of experience has to be greater than the option value of the unit of experience for getting more experience later in life.

This intuition is formalized in the following proposition:

Proposition 1 *If $U'(\tau + 1) \leq U'(\tau)$, then $\bar{p}'(\tau) > 0$ for all $\tau \in \{0, 1, \dots, T\}$.*

Proof. See Appendix. ■

Although, Proposition 1 does not reduce the sign of $\bar{p}'(\tau)$ to restrictions on model parameters, it provides clear intuition for when the reservation belief will be increasing in experience. The condition that guarantees $\bar{p}'(\tau) > 0$ is:

$$V_\tau(\bar{p}(\tau), \tau) \leq U'(\tau)$$

Using the results on the worker's reservation decision above, it is useful to consider the behavior of the job finding rate as a function of experience, $f(\tau)$. The job finding rate is determined by the exogenous rate of matches combined with the workers willingness to accept production opportunities:

$$f(\tau) = \lambda(1 - H(\bar{p}(\tau), |\tau))$$

Proposition 2 *If $U'(\tau + 1) \leq U'(\tau)$ and $\bar{p}(\tau) \leq p_0$, then $f'(\tau) > 0$.*

Proof. Taking the derivative of $f(\tau)$ with respect to τ gives:

$$f'(\tau) = -\lambda h(\bar{p}(\tau)|\tau)\bar{p}'(\tau) - \lambda H_\tau(\bar{p}(\tau)|\tau)$$

where $h(\bar{p}(\tau)|\tau)$ is the pdf of H. By Proposition 1 $U'(\tau + 1) \leq U'(\tau)$ implies that $\bar{p}'(\tau) > 0$. This condition guarantees that the first term is negative. The second term is also negative when $\bar{p}(\tau) \leq p_0$ because $H(p|\tau)$ is a mean preserving spread around p_0. ■

The conditions in Proposition 2 generate declining job finding rates early in workers lives where $\bar{p}(\tau) \leq p_0$ as workers accept most jobs to gain experience[5]. While generating a decline in job finding rates does not provide a test of the model, the strength of the model is that it generates a theory about when job finding rates will be increasing or decreasing. Job finding rates are likely to decline early in workers careers when additional information has diminishing returns.

[5] The focus of this paper is to model the initial decline in job finding rates for young workers. In certain parameterizations, job finding rates can be increasing later in life when $\bar{p}(\tau) > p_0$, however for most parameter values the increase in job finding rates occurs after longer horizons than the typical length of a worker's career.

The final implications of the model involve the process of the worker's current belief p while employed. Given the binary structure of productive opportunities in the model, the density of output $\psi(x|p)$ is a mixture density of two normal distributions with means μ_h and μ_l and weights of p and $(1-p)$. The properties of the mixture density are characterized in the following proposition:

Proposition 3 *The distribution characterized by the mixture density $\psi(x|p)$ has mean $\mu_{mix} = p\mu_n + (1-p)\mu_l$ and variance $\sigma^2_{mix} = p(1-p)(\mu_h - \mu_l)^2 + \sigma^2$ where σ^2 is the variance of each of the two distributions in the mixture.*

Proof. See Appendix. ■

For special case with two jobs types considered in this paper the standard deviation of expected output depends both on the standard deviation of the output process σ^2 and the uncertainty regarding which type of job the worker has. While the noise from the output process is constant the uncertainty depends on the current belief p. This implies that the standard deviation of expected output peaks at $p = 0.5$ and decreases as p converges to zero or one. While the model does not provide a closed form solution for the standard deviation of $G(p'|p)$, the above intuition shows that it is decreasing in p if $p > 0.5$.

This result is important to understand how workers learn. Unlike in Jovanovic (1979) the standard deviation of output does not decrease monotonically as the worker learns more. Instead, the standard deviation depends on the current value of p and will on average decrease for workers as their current belief converges to either zero or one.

Next, the expected initial belief based on initial experience τ is characterized in the following proposition:

Proposition 4 *If $\tilde{\tau} > \tau$, $\bar{p}(\tau) \leq \bar{p}(\tilde{\tau})$, and $\bar{p}(\tau) \leq p_0$, then the expected value of the initial beliefs for an accepted offer is higher for the worker with more experience. That is:*

$$E[H(p|\tau)|p \geq \bar{p}(\tau)] \leq E[H(p|\tilde{\tau})|p \geq \bar{p}(\tilde{\tau})]$$

Proof. See Appendix. ■

Given the wage process:

$$w(p) = p\mu_h + (1 - p)\mu_l$$

the behavior Proposition 3 and Proposition 4 can be used to make predictions about volatility of wages on new jobs. The novel feature of the model is that a worker with more experience who starts a new match will have more information about the quality of that match than a worker with less experience. In the case where $\bar{p}(\tau) > 0.5$ and is increasing, the model would predict that more experience translates to on average a higher value of p at the start of a new job and hence lower variation in the path of future wages. These implications are quantitively evaluated with simulations of the model.

While the above propositions derive the novel implications of the model, the model generates many other predictions that are present in other search and learning models that are consistent with empirical findings. Two of these implications that are particularly worth emphasizing are that the wage tenure gradient should be declining in experience and that expected duration of jobs is increasing in experience. The model generates a declining wage tenure profile with more past experience as workers have more information about their ability. This prediction is not tested as it is difficult to distinguish this channel from standard models of human capital with diminishing returns that can generate similar wage profiles. The implication that more past experience leads to longer job durations has been captured in McCall's (1990) finding that longer tenure in the first job implies lower hazard rates in future employment as experience allows workers to reject poor second matches.

3 Calibration

To parameterize the model, assume that there are a large number of workers facing identical decision problems. Each worker faces a different history of idiosyncratic shocks. Averaging outcomes across workers, aggregate data are constructed from the model. In computations, simulated data over a 40 year career is compared to actual worker outcomes. The period length is one month so that parameters are chosen to match monthly data on job finding and separation rates in the United States.

To compute the model there are ten parameters that must be chosen: the maximum amount of experience T, the initial signal k, the discount factor β, the job offer rate λ, the expected output from a good match μ_h, the expected output from a bad match μ_l, the probability that a match is good p_0, the variance of output noise σ, the proportion of experience used for new matches α, the exogenous separation rate δ, and the value of leisure b.

μ_h is normalized to one and μ_l is normalized to zero. Given these normalizations the evolution of p will be determined by the variance of output noise, σ. The evolution of p is fully determined by the signal to noise ratio: $\frac{\mu_h-\mu_l}{\sigma}$. Because the model period is one month, β is set to 0.9966 which corresponds to an annual interest rate of 4%. $T = 480$ to corresponding to a maximum level of experience of 40 years. This is a reasonable upper bound as it corresponds to the normal length of work for individuals in the U.S. Increasing the maximum level of experience has no effect on the results. Finally, $k = 1$ give a non-degenerate initial distribution of beliefs about a first job while still being concentrated around p_0.

The remaining parameters are chosen to match features of the decline in job finding and separation rates in the U.S. The left panel of Figure 1 shows the decline in the job separation rate with age in the U.S. for workers aged 18-57[6]. The separation has a sharp initial decline from age 18-25 followed by a gradual decline later in life.

The right panel of Figure 1 shows the decline in the job finding rate. Similarly, the job finding rates fall fastest for the first 8-10 years, but the initial decline is less dramatic than the separation rate and finding rates continue to decline at a greater rate for the remainder of the workers' careers. Notice that while job separation rates fall by about a factor of 10,

[6]This data was constructed by Robert Shimer using CPS monthly microdata from 1976 to 2005. The procedure used follows Shimer (2007) to create a time series of job separation and finding rates for individuals of each age. The time series is used to create average unemployment, job finding, and job separation rates for each age group. For additional details, please see Shimer (2007) and his webpage http://robert.shimer.googlepages.com/flows.

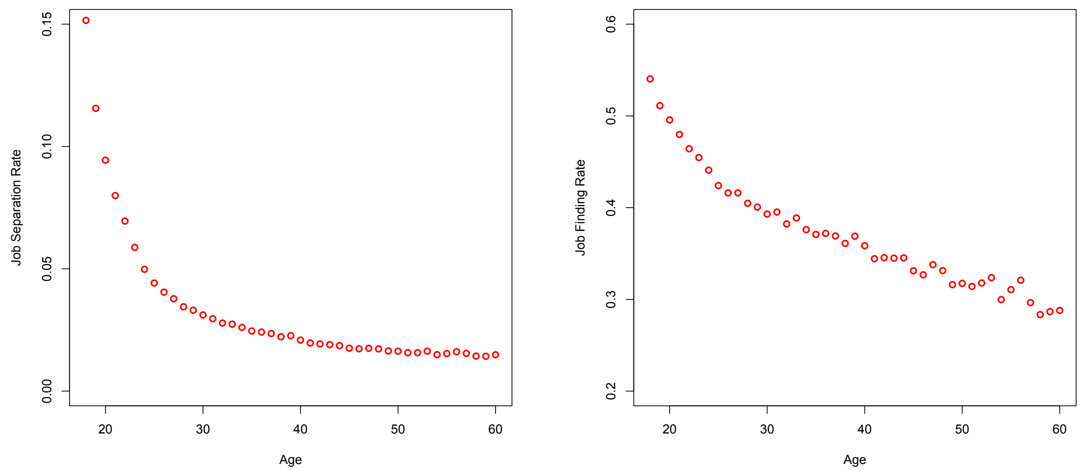

Figure 1: *Monthly job separation rate in left panel and monthly job finding rate in right panel by age for the U.S. economy.*

the job finding rates only decline by about a factor of 2 over the life cycle. Taken together, the steeper decline in the separation rate implies that the unemployment rate declines with age.

λ is chosen to match the worker's rate of job offers. λ provides an upper bound for the job finding rate in the model as workers with little experience will accept nearly any productive opportunity that they find. In the data, 17-year-old workers have a job finding rate of 0.57. To match this feature of the data, λ is set to 0.6.

p_0 determines the portion of good jobs in the economy. Since a worker with perfect information about the quality of jobs will only accept good ones, p_0 determines the amount of decline in the job finding rate over the worker's life. p_0 is chosen to match the decline in the job finding rate found in the data. It is set to 0.7 which allows the model to match the job finding rate of 0.30 for 57 year old workers in the data.

Next, σ is the amount of output noise. Higher values of σ imply that workers learn slowly about the quality of their matches. In the limit, $\sigma = 0$ implies that workers perfectly observe the quality of the match with one observation while as $\sigma \to \infty$ workers have no learning. $\sigma = 4$ is chosen to match the shape of the decline in job finding rates. Higher values of σ imply that workers learn more slowly. Slower learning implies that it takes longer to distinguish bad matches, and generates a flatter decline in job finding rates. Higher values of σ imply that the decline in job finding rates is quick.

α determines the amount of experience that carries over in learning about new job opportunities. It is natural to restrict α to be in $[0, 1]$. $\alpha = 0$ is analogous to the standard Jovanovic (1979) model where individuals learn nothing about future jobs and the employment is a pure renewal process. $\alpha = 1$ is the limit where all learning carries over to future jobs. Higher values of α imply that workers learn faster about future jobs and therefore have a steeper decline in both job finding and separation rates. Model results for various values of α are shown. With the model period set to be a month, $\alpha = \frac{1}{30}$. This corresponds to getting one month worth of information about a new job for every two and a half years of labor market experience. Higher values of α predict a steeper initial decline followed by less learning later. This parameter is sensitive to the choice of σ. The chosen value of σ implies that individuals learn quickly by observing output. Surprisingly, very low values of α generate large changes in the patterns of job finding rates.

δ is the rate of exogenous job separations. An upper bound on the value of δ is lowest observed monthly job finding probability in the data is 0.014 for 59-year-olds. A lower value of $\delta = 0.0075$ is chosen.

The final parameter is b. This parameter determines the relative desirability of being employed in a bad job compared to searching for a new job. Higher values of b make unemployment more attractive. $b = 0.3$ is chosen to match the level of unemployment over a workers lifetime.

Table 1 summarizes the chosen parameters and their values.

Name	Parameter	Value	Target
Max Age	T	480	Max 40 Years Experience
Initial Information	k	1	Normalization
Discount Factor	β	0.9966	4% Interest Rate
Job Offer Rate	λ	0.6	Peak of Finding Rate
Good Output	μ_h	1	Normalization
Bad Output	μ_l	0	Normalization
Probability of Good Job	p_0	0.7	Decline in Finding Rate
Output Noise	σ	4	Curvature in Finding Rate
Experience Rate	α	$\frac{1}{30}$	Various Values Shown
Exogenous Separation Rate	δ	0.0075	Minimum of Separation Rate
Value of Leisure	b	0.3	Level of Unemployment

Table 1: *Calibrated values of the model parameters.*

$\leq$18	19	20	21	22	23	24	$\geq$ 25
29.6	24.9	18.8	11.4	8.1	4.8	1.7	0.7

Table 2: *Percent of populations first employment spell by age. From Topel and Ward (1992).*

4 Simulated Results

This section documents the implications from the calibrated model. The novel feature of allowing learning to transfer between jobs is that past experience now has implications for workers while unemployed through their job search behavior. To document this, the value functions are computed to generate reservation probabilities for workers at each experience level. Using these decision rules, employment outcomes are simulated for individual workers. The outcomes for 10,000 simulated workers are computed from the date that workers enter the labor force. Monthly employment, job finding rates, job separation rates, wages, tenure, and total experience are recorded.

To compare with labor force data, outcomes by age are constructed by entering workers into the labor market at the age they get their first full time employment. Topel and Ward (1992) compute the percentage of workers who enter the labor force at a given age by assuming that workers enter when they attain their first employment that lasts at least 2 quarters. This measure leaves out workers who take summer jobs and then return to school. Table 2 replicates their table showing the percentage of workers who enter the labor force at each age. When constructing the data from the model, all workers in the ≤ 18 category enter at age 18 and all workers in the ≥ 25 category enter at age 25. The remainder of this section compares the simulated data from the calibrated model with employment data from the CPS.

4.1 Job Finding Rate

The first result is that the calibrated model is able to match the decline in job finding rates with age. Since experience allows individuals to learn about the quality of new matches, experienced workers are more selective about which jobs they choose to accept. This feature allows the job finding rate to decline over a worker's lifetime.

Figure 2 plots the decline in job finding rates from the simulated model against the data. Simulated job finding rates start out slightly higher and decline to match the rates observed in the data. The two series are almost identical after age 30. The calibrated model is able to capture the initial steep decline in job finding rates and continued gradual decline later in life. No previous models of learning generated any change in job search behavior so their predicted job finding rate is constant.

To compare the fits of the model with the data a goodness of fit is computed:

$$Fit = 1 - \frac{\sum_{a=18}^{57}(\varepsilon_a - \bar{\varepsilon})^2}{\sum_{a=18}^{57}(y_a - \bar{y})^2}$$

This is similar to an R^2 measure, where ε_a is the difference between the model and the data for age a, $\bar{\varepsilon}$ is the average difference, y_a is the level of the data for age a, and $\bar{y}$ is the average level of the data. The numerator give the sum of squared errors between the data and the model and the denominator gives the sum of squared deviations in the data. The calibrated model has a fit of 0.95.

To see how changes in α affect the results of the model, Figure 3 plots the job finding

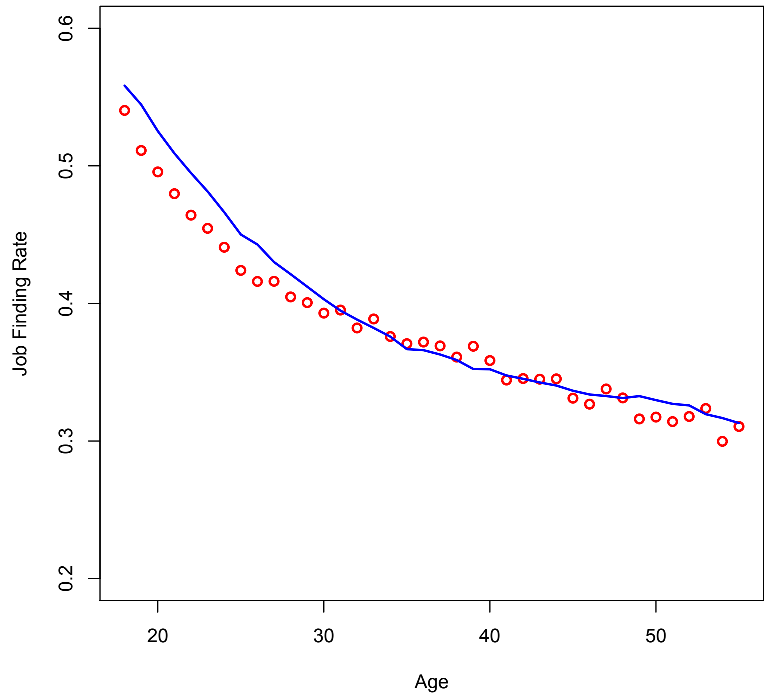

Figure 2: *Job finding rate by age. Data: dots; Calibrated Model: Line.*

rate from the calibrated model for different values of α. Low values imply a steeper initial decline in job finding rates as the worker is able to distinguish between good and bad jobs more quickly. In the cases close to the calibrated value of $\alpha = \frac{1}{30}$ the simulated job finding rates decline throughout the worker's simulated working life. However, for high values of α the simulated job finding rate is not monotonic. In particular, when $\alpha = 1$ there is a large portion of the worker's life for which the job finding rate is increasing. Recall that for a worker with perfect information, the job finding rate is determined by the arrival rate of productive opportunities λ multiplied by the fraction of those opportunities that are good p_0. This gives a job finding rate of 0.42 for the current calibration. The initial rapid decline occurs as some information about the quality of the job initially makes the worker much pickier about which jobs to accept. Over time better information pushes a greater portion

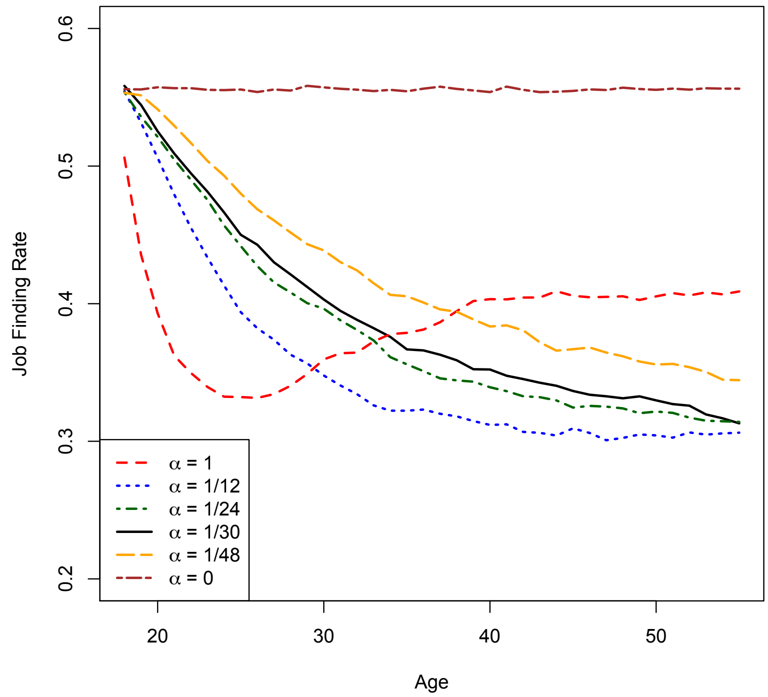

Figure 3: *Job finding rate by age for various values of α.*

of jobs above the threshold to increase the job finding rate. The case of $\alpha = \frac{1}{48}$ shows that even small amounts of learning across jobs can have dramatic effects on the predicted worker search behavior over the life cycle. Finally, the case of $\alpha = 0$ is shown to be flat. This corresponds to the Jovanovic (1979) model where no learning transfers across jobs and workers have a flat job finding rate for their entire life.

4.2 Job Separation Rate

Figure 4 shows the decline in separations for the calibrated model compared with the data. The model exhibits an initial decline in the separation rate that is steeper than the data, but is unable to generate the highest levels of separations for young workers. Some of the high rates are due to workers moving in and out of the labor force for schooling that is not

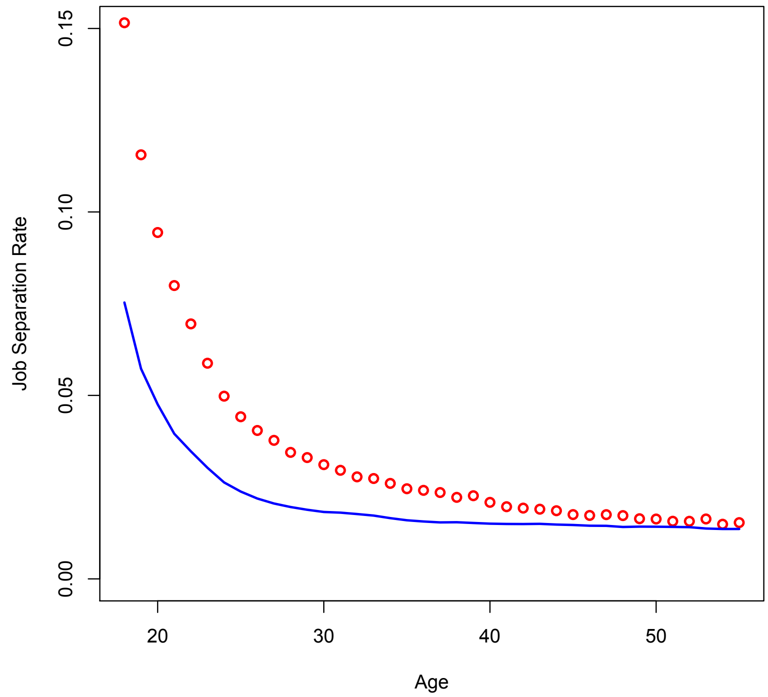

Figure 4: *Job separation rate by age. Data: dots; Calibrated Model: Line.*

captured by the model. The decline in job separations happens in the model for two reasons. First, as in Jovanovic (1979) the job separation rate declines as workers sort themselves to good jobs which last longer on average than bad jobs. Second, experience allows older workers to match with better jobs than younger workers reducing the chance of separation for new jobs acquired later in life.

The fit of the calibrated model is 0.68. Despite not capturing the high level of initial separations in the data, the model with learning is still able to capture most of the decline in separations. This is consistent with the predictions in other models of learning.

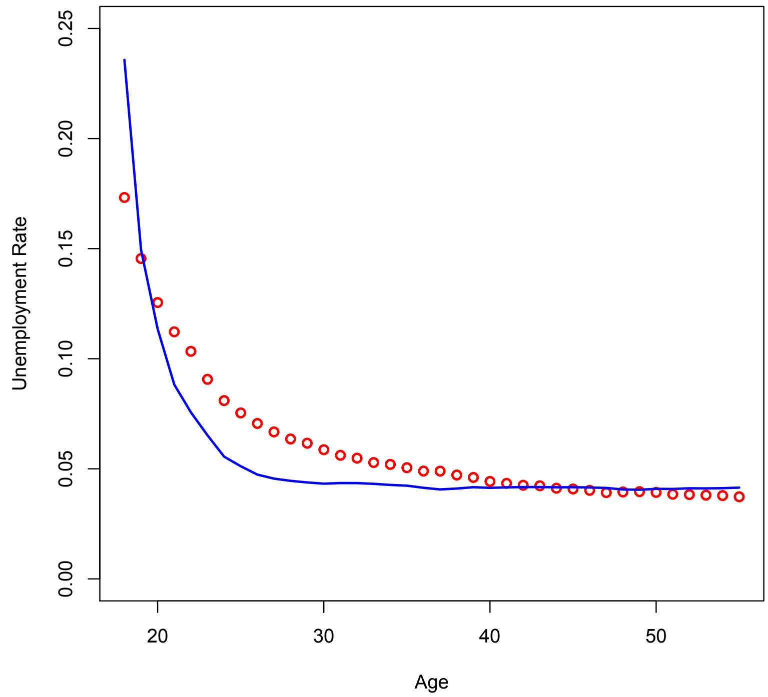

Figure 5: *Unemployment rate by age. Data: dots; Calibrated Model: Line.*

4.3 Unemployment

It is well known that young workers face higher unemployment rates than prime aged workers. The model is able to capture this decline in unemployment with age.

Figure 5 shows the average annual unemployment rate by age. The dots depict the decline in unemployment found in the data where the solid line depicts the results from the calibrated model. The data show a steady decline in unemployment with age. Unemployment declines from about 17% for 18-year-old workers to between 3.5 and 4% for prime aged workers. The calibrated model captures a similar decline over the life cycle, with 18-year-old workers experiencing unemployment of 23% and declining to 4.1%. The initial decline in unemployment is steeper in the calibrated model than in the data reflecting all 18 year old worker entering the market unemployed. The fit calibrated model on the unemployment data is 0.75.

The predicted decline in unemployment from the model can be understood by combining the results about declining job separation and job finding rates. The decline in job separation rates drives most of the decline in unemployment while the decline in job finding rates tends to slightly increase the unemployment rate. However, the decline in separations dominates as it goes from about 7.5% to 1.3% over the worker's life while the job finding rates only decline by about a factor of 2 from about 56% to 30%.

4.4 Wage Growth

Flinn (1986) argues that wage growth and turnover are related for young workers. This model presents a theory that accounts for both phenomena. Topel and Ward (1992) document a number of features of wage profiles during worker's first 10 years of experience. They document that the first 10 years of the career account for two-thirds of lifetime wage growth. Job changes explain about one-third of wage growth. Moreover, wages on the job approximate a random walk. The model qualitatively replicates the behavior of wages over the life cycle.

Figure 6 shows the average annual wages by age from the model. The pattern of wage growth from the model is endogenous. The model generates rapid wage growth during the first 10 years of experience and then levels off. While the model matches the general pattern of wage growth, it doesn't generate quite as much wage growth as found in the data where wages about double over the lifetime. This should be expected as the model generates only wage growth from workers moving to better matches with firms and does not include wage growth from learning by doing or other forms of human capital gained while working. A model would need to include these other forms of wage growth to fully account for wages over the life cycle.

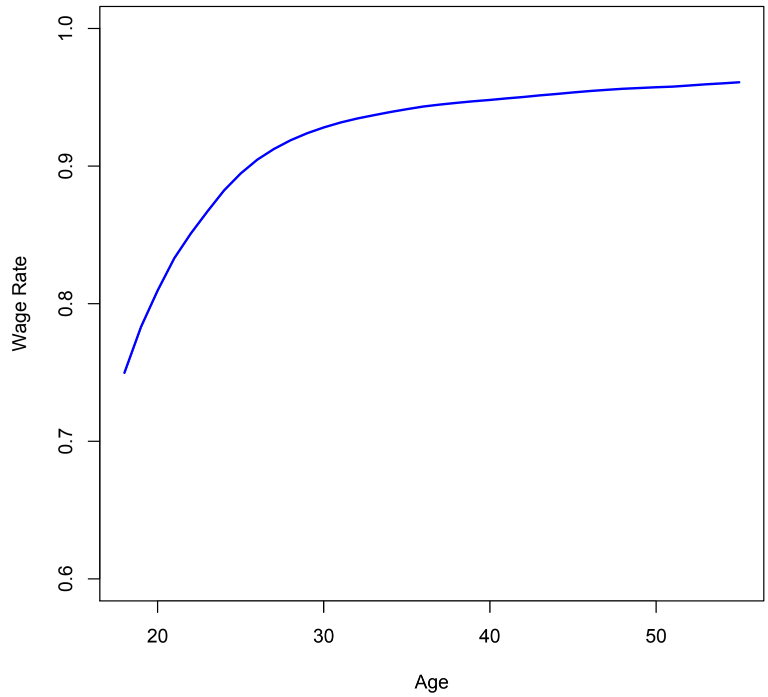

Figure 6: *Wages by age in calibrated model.*

5 Wage Volatility

This section quantitatively evaluates the model's predictions on wage variation using data on wages and job switching from NLSY79. The novel feature of this model is that workers who start jobs with more experience have better information about the quality of their new job. For standard parameterizations, this information means that experienced worker's on average start with a higher p and hence their wages should display less variation in subsequent periods. These predictions are first confirmed using observations simulated from the model then the same results are documented using data from the NLSY79.

The NLSY79 is a nationally representative longitudinal survey conducted by the Bureau of Labor Statistics that samples 12,686 individuals who were between the ages of 14 and 22 years old when first surveyed in 1979. The individuals continued to be surveyed every year until 1994 when the survey switched to every two years. The sample is restricted to be from 1970-1994 as yearly differences in wages are needed to construct measures of wage volatility. NLSY79 provides a rich set of panel data for tracking worker's career outcomes. To avoid miscalculation of past experience, the sample is limited to workers who are 17 years old or younger at the time of the first interview.

To construct job variables the NLSY79 provides a variable for the total number of past jobs that the respondent has held. In the NLSY79 a job is defined as a relationship between an individual employer and the worker. That is changes in position within a firm are not considered new jobs. If the total number of jobs in year t is greater than in year $t-1$, then there is a new job observation. For each job observation the wage in each year is given by the CPS wage[7]. Finally, experience can be constructed by taking a cumulative sum of the weeks worked in the past year variable. The number of past weeks worked is divided by 52 so that results can be presented in terms of years of experience. To compare observed outcomes from the NLSY79 with the model, 25 years of annual observations are simulated for the worker's employment status, past experience, accumulated job number, and wage from the model. To make the samples comparable, both the simulated and NLSY79 data are restricted to jobs where workers start with less than 15 years of prior experience.

Each worker's employment history is broken into jobs that are characterized by a wage for each year of tenure on the job and the initial experience level when starting the job. Wage volatility is measured as the absolute deviation from the worker's expected wage growth path. The simplest measure of wage volatility is to take the absolute value of the difference in log wages at each tenure level from the initial wage on the job. However, this measure does not control for the expected levels of wage growth that occur at different levels of tenure

[7]Since wages are only recorded each year it is possible that the wages could have already changed from the initial wage at the time of first observation. Despite this measurement issue, the same issue arises when annual data is taken from the simulated model. In the simulated model a worker's wages change every month based on their updated beliefs. Treating the simulated data the same as the NLSY79 observations should yield similar biases.

and experience. To control for this, the wage volatility measure used is:

$$v_t = |\log(w_t) - \log(w_0) - \bar{w}_{et}|$$

v_t is the volatility of wages at tenure t on a given job[8]. w_t is the wage observed at tenure t, w_0 is the initial wage observed on the job, and $\bar{w}_{et}$ is the median log wage deviation $(\log(w_t) - \log(w_0))$ observed for the two year initial experience group e and tenure level t[9]. By construction the volatility is zero for the initial wage observation (worker's tenure of zero). Note that the concept of wage volatility here is at the individual level rather than a cross section across individuals. Higher volatility implies that a given individual experiences larger changes in her wages on a given job. By subtracting $\bar{w}_{et}$ the measure of volatility used in this paper controls for median wage gains in each year of tenure at a particular job and with experience. While most workers get wage increases from year to year, subtracting of the expected wage growth at the tenure and experience level means that many workers are both above and below the expectation.

While the effect of experience on wages has been explored by a large theoretical literature (See Neal and Rosen (2000), Gibbons et al. (2005)), the previous literature has not explored the impact of experience on individual within job wage volatility. Understanding the features of the individual income process is important to explain a wide array of individual behavior (see Meghir and Pistaferri (2004)). This paper shows that past job experience has a predictable effect on individual wage volatility. Experience is shown to decrease individual level uncertainty about wages while cross sectional heterogeneity may increase within group

[8] v_t is the t year measure of volatility, so v_3 measures the volatility of wages over a three year increment from starting the job. Another measure of volatility, $v(t)$ can be defined as the one year volatility for each year t from the previous years observed wage:

$$v(t) = |\log(w_t) - \log(w_{t-1}) - \tilde{\bar{w}}_{et}|$$

Where $\tilde{\bar{w}}_{et}$is the median log wage deviation $(\log(w_t) - \log(w_{t-1}))$ for the two year initial experience group e and tenure level t. The rest of the analysis focuses on the first year volatility on a job, so these two measures are identical.

[9] The median is a more appropriate measure here as it is robust to truncation. This is especially important as there is a zero lower bound of observed wage volatility. The results are similar if the mean is used.

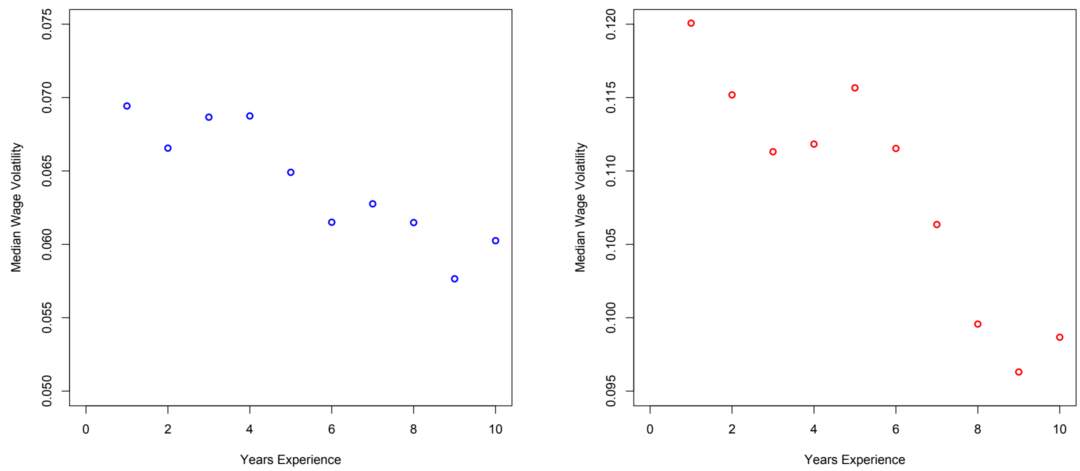

Figure 7: *Median wage volatility (v_1) by years of past experience. Simulated model in left panel NLSY79 data in right panel.*

wage variation. The model is able to account for the individual level declines in volatility.

The novel prediction of the model is that volatility should decline with more past job experience. To explore this prediction, it is sufficient to just look at the first year wage volatility on each new job, v_1. The left panel of Figure 7 plots the median wage volatility for the first year in each job binned by years of past experience. The model generates a decline in wage volatility for workers with more past work experience. Note that the model predicts that for inexperienced workers the one year change in wages will be about 7%. Median volatility declines to under 6% for workers with 10 years of experience. The right panel of Figure 7 plots the median wage volatility for the first year in each job with experience binned into yearly groups for the NLSY79 data. Just as in the simulation the data shows a decline in wage volatility for workers with more past work experience. For inexperienced workers, the one year change in wages is about 12%. Volatility declines at a steeper rate to under 10% for workers with 10 years of experience. The patterns of volatility with experience are consistent with those found in the model. Note that there is a scale shift between the two panels in the figure. Since the magnitudes of wage volatility are higher in the data, the

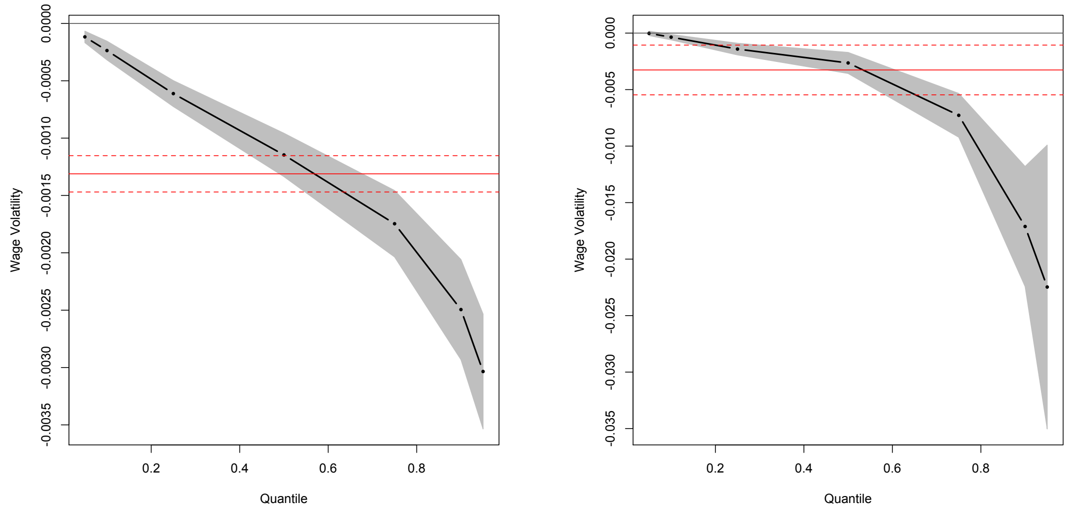

Figure 8: *Plot of quantile regression of first year wage volatility on past experience with error band. Thin lines gives OLS estimate with dashed error bands. Simulated model in left panel NLSY79 data in right panel.*

model does not capture all of the observed wage volatility. However, ignoring the differences in levels shows that in both the model and the data the median wage volatility drops by about 20% with 10 years of experience.

To more formally assess the model's predicted effect of experience on wage volatility, the year one wage volatility measure v_1 is regressed on the past experience associated with each new job event. Quantile regressions are valuable for two reasons. First, they are robust to the lower bound issues in observed volatility. Second, they provide a more details predictions about how experience influences volatility at different points in the distribution so that the model predictions can be more closely compared to the evidence found in the data. This gives additional feedback about how experience influences workers at different parts of the volatility distribution. Quantile regressions are run for the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 95^{th} percentiles[10].

The left panel of Figure 8 shows the results for the OLS regression plotted against the

[10] See Koenker and Hallock (2001) for other examples of the quantile regression procedure.

	Simulated Model		NLSY79 Data	
Quantiles	Intercept	Experience	Intercept	Experience
OLS	0.0888***	−0.00131***	0.256***	−0.00326**
	(0.000691)	(0.0000963)	(0.0120)	(0.00136)
0.05	0.00640***	−0.00012***	0.00932***	−0.00002
	(0.00024)	(0.00003)	(0.00071)	(0.00012)
0.10	0.0125***	−0.00024***	0.0208***	−0.00036
	(0.00039)	(0.00005)	(0.00094)	(0.00015)
0.25	0.0327***	−0.00061***	0.0577***	−0.00141***
	(0.00057)	(0.00007)	(0.00202)	(0.00032)
0.50	0.0701***	−0.00115***	0.122***	−0.00264***
	(0.00090)	(0.00012)	(0.00336)	(0.00057)
0.75	0.128***	−0.00175***	0.296***	−0.00728***
	(0.00137)	(0.00018)	(0.00757)	(0.00118)
0.90	0.192***	−0.00249***	0.617***	−0.0171***
	(0.00195)	(0.00027)	(0.0196)	(0.00323)
0.95	0.233***	−0.00303***	0.932***	−0.0225***
	(0.00246)	(0.00030)	(0.0395)	(0.00746)

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors clustered by individual in parentheses.

Table 3: *OLS and quantile regression results for the simulated model and the data.*

quantile regression results. The graph shows that both the OLS and median quantile regression confirm a negative and significant effect on wage volatility. The quantile regressions at other points in the distribution show that at higher percentiles of the distribution experience causes larger decreases in wage volatility. The right panel of Figure 8 plots the regression results of wage volatility on experience from the NLSY79 data. Again, the patterns in the regressions confirm the model predictions. Both the OLS and median quantile regression show a significant negative effect of experience on wage volatility. The magnitude of the decline is between 23 and 32 basis points per year of experience depending on using the median quantile regression or the standard OLS estimate.

The regression results for both the model and NLSY79 data are presented in Table 3. The estimated effect for the median and mean in the model is that an extra year of experience decreases the volatility of wages by about 13 basis points. The coefficient on experience is negative and significant in all cases except for the 0.05 quantile. This is expected as wage volatility is close to zero for the low quantiles and hence cannot decrease much further. The larger declines in the upper quantiles imply that there is a greater effect of experience for jobs

		Specification	
Variable	I	II	III
Intercept	0.256*** (0.0120)	0.264*** (0.00780)	0.256*** (0.0120)
Experience	−0.00326** (0.00136)	−0.00341** (0.00137)	−0.00330** (0.00136)
Female		−0.0160* (0.00862)	−0.0161* (0.00870)
College Degree			0.00291 (0.0106)
Graduate Degree			0.00342 (0.0151)
Race Dummies	No	No	Yes

*** p<0.01, ** p<0.05, * p<0.1.

Robust standard errors clustered by individual in parentheses.

Table 4: *OLS regression results for NLSY79 data with controls.*

with higher wage volatility. The results from the NLSY79 data confirm this general pattern. The OLS estimate indicates that a year of experience decreases volatility by about 33 basis points while the median decreases by about 24. The data are negative and significant for all points in the quantile regression except for the .05 and .1 quantiles with slightly higher magnitudes than generated from the model.

Finally, Table 4 presents additional regression specifications for the NLSY79 data. Specification I shows the baseline results from above. Specification II includes a female dummy variable. The estimate remains similar and the result shows that women have about 1.6% less volatility than men. Finally, specification III includes education and race dummies. None of the dummies are significant and the effect of experience remains unchanged. The education dummies are important as the impact of experience on individual wage volatility remains the same even though different education groups are known to have different wage experience profiles[11]. The table confirms a robust negative relationship between past work experience and observed wage volatility in the first year on a new job.

[11] See for instance Farber and Gibbons (1996) and Lange (2007) for a more recent discussion.

6 Conclusion

This paper presents a model of learning that can explain changes workers' job finding rates over their life cycle. Workers' learning about the quality of their match is important for both observed outcomes while employed like wages and employment durations and outcomes while unemployed. This insight motivates the model where experience gives workers both knowledge about the quality of their current job and the ability to distinguish between good and bad jobs when unemployed.

A model with learning about both the quality of the current match and future matches has rich implications for labor market outcomes. It is consistent with the age profiles of unemployment, job finding rates, job separation rates, hazard rates of separation with tenure, wage dispersion, and wage growth. Having a model that has consistent predictions about a broad range of labor outcomes makes it ideal to analyze the effects of policy on these outcomes. The model is used to generate new predictions about individual worker's wage volatility on jobs based on their past level of experience. The prediction of lower volatility with more past experience is found to hold in NLSY79 data.

While learning both within and across jobs accounts for many of the observed patterns found in the evidence on individual labor earnings, it generally does not capture the entire wage growth observed over the life cycle. Learning that transfers between jobs can be thought of as one specific type of human capital that agents acquire while working. To account for the entire wage patterns observed in the data it is important to distinguish between learning and other forms of specific human capital.

A Proof of Proposition 1

Claim 1 *If $U'(\tau+1) \leq U'(\tau)$, then $\bar{p}'(\tau) > 0$ for all $\tau \in \{0, 1, \ldots, T\}$.*

Proof. Differentiating equation (3) with respect to τ gives:

$$\bar{p}'(\tau)V_p(\bar{p}(\tau), \tau) + V_\tau(\bar{p}(\tau), \tau) = U'(\tau)$$

$$\bar{p}'(\tau) = \frac{U'(\tau) - V_\tau(\bar{p}(\tau), \tau)}{V_p(\bar{p}(\tau), \tau)}$$

Then $\bar{p}'(\tau) > 0$ if $V_\tau(\bar{p}(\tau), \tau) \leq U'(\tau)$. It suffices to show that $V_\tau(p, \tau) \leq U'(\tau+1)$ for all $\tau \in \{0, 1, \ldots, T\}$ and $p \in [0, 1]$. We will proceed by backward induction starting from $\tau = T$. For $\tau = T$:

$$V_\tau(p, T) = U'(T+1) = U'(T) = 0$$

For $T-1$:

$$\begin{aligned} V_\tau(p, T-1) &= [\beta\delta + \beta(1-\delta)G(\bar{p}(T)|p)]\, U'(T) + \beta(1-\delta)\int_{\bar{p}(T)}^{1} V_\tau(p', T)G(dp'|p) \\ &= 0 = U'(T) \end{aligned}$$

For $T-2$:

$$\begin{aligned} V_\tau(p, T-2) &= [\beta\delta + \beta(1-\delta)G(\bar{p}(T-1)|p)]\, U'(T-1) \\ &\quad + \beta(1-\delta)\int_{\bar{p}(T-1)}^{1} V_\tau(p', T-1)G(dp'|p) \\ &= [\beta\delta + \beta(1-\delta)G(\bar{p}(T-1)|p)]\, U'(T-1) \\ &= \beta\,[\delta + (1-\delta)G(\bar{p}(T-1)|p)]\, U'(T-1) < U'(T-1) \end{aligned}$$

Finally, assuming $V_\tau(p, T-n) \leq U'(T-n+1)$, we can solve for $T-n-1$:

$$\begin{aligned} V_\tau(p, T-n-1) &= [\beta\delta + \beta(1-\delta)G(\bar{p}(T-n)|p)]\, U'(T-n) \\ &\quad +\beta(1-\delta)\int_{\bar{p}(T-n)}^{1} V_\tau(p', T-n)G(dp'|p) \\ &\leq [\beta\delta + \beta(1-\delta)G(\bar{p}(T)|p)]\, U'(T-n) \\ &\quad +\beta(1-\delta)(1-G(\bar{p}(T-n)|p))U'(T-n+1) \\ &\leq [\beta\delta + \beta(1-\delta)G(\bar{p}(T)|p)]\, U'(T-n) \\ &\quad +\beta(1-\delta)(1-G(\bar{p}(T-n)|p))U'(T-n) = U'(T-n) \end{aligned}$$

The first inequality comes from the induction and the second comes from the hypothesis that $U'(\tau+1) \leq U'(\tau)$ for all $\tau \in \{0, 1, \ldots, T\}$. ■

B Proof of Proposition 3

Claim 2 *The distribution characterized by the mixture density $\psi(x|p)$ has mean $\mu_{mix} = p\mu_n + (1-p)\mu_l$ and variance $\sigma^2_{mix} = p(1-p)(\mu_h - \mu_l)^2 + \sigma^2$ where σ^2 is the variance of each of the two distributions in the mixture.*

Proof. The density $\psi(x|p)$ is the mixture of two normal distributions with means μ_h and μ_l and the same variance σ^2. Letting w_i denote the weights in each distribution we have that the mean of the mixture distribution μ_{mix} is given by:

$$\mu_{mix} = \sum_{i \in \{h,l\}} w_i \mu_i = p\mu_h + (1-p)\mu_l$$

The variance σ^2_{mix} is given by:

$$\begin{aligned}
\sigma^2_{mix} &= \sum_{i\in\{h,l\}} [w_i(\mu_i^2+\sigma^2)] - \mu^2_{mix} \\
&= p(\mu_h^2+\sigma^2) + (1-p)(\mu_l^2+\sigma^2) - (p\mu_h + (1-p)\mu_l)^2 \\
&= p(1-p)\mu_h^2 + (1-p)(1-(1-p))\mu_l^2 - 2p(1-p)\mu_h\mu_l + \sigma^2 \\
&= p(1-p)(\mu_h-\mu_l)^2 + \sigma^2
\end{aligned}$$

■

C Proof of Proposition 4

Claim 3 *If $\tilde{\tau} > \tau$, $\bar{p}(\tau) \leq \bar{p}(\tilde{\tau})$, and $\bar{p}(\tau) \leq p_0$, then the expected value of the initial beliefs for an accepted offer is higher for the worker with more experience. That is:*

$$E[H(p|\tau)|p \geq \bar{p}(\tau)] \leq E[H(p|\tilde{\tau})|p \geq \bar{p}(\tilde{\tau})]$$

Proof. First, note that since $\tilde{\tau} > \tau$ $H(p|\tilde{\tau})$ is a mean preserving spread of $H(p|\tau)$. By definition of a mean preserving spread we have:

$$\int_0^x H(p|\tau)dp \leq \int_0^x H(p|\tilde{\tau})dp$$

for any value of $x \in (0,1]$. Note that the lower bound on the integrals is zero as the support of the distribution H is from zero to one. Integrating the above equation by parts on each side gives:

$$xH(x|\tau) - \int_0^x ph(p|\tau)dp \leq xH(x|\tilde{\tau}) - \int_0^x ph(p|\tilde{\tau})dp$$

Subtracting the bound x from each side and adding the mean of the distribution H p_0 to each side of the equation gives:

$$\begin{aligned} xH(x|\tau) - x + p_0 - \int_0^x ph(p|\tau)dp &\leq xH(x|\tilde{\tau}) - x + p_0 - \int_0^x ph(p|\tilde{\tau})dp \\ -x[1 - H(x|\tau)] + \int_x^1 ph(p|\tau)dp &\leq -x[1 - H(x|\tilde{\tau})] + \int_x^1 ph(p|\tilde{\tau})dp \end{aligned}$$

Factoring and simplifying each side gives:

$$\begin{aligned} [1 - H(x|\tau)]\left[\frac{\int_x^1 ph(p|\tau)dp}{1 - H(x|\tau)} - x\right] &\leq [1 - H(x|\tilde{\tau})]\left[\frac{\int_x^1 ph(p|\tilde{\tau})dp}{1 - H(x|\tilde{\tau})} - x\right] \\ \frac{\int_x^1 ph(p|\tau)dp}{1 - H(x|\tau)} &\leq \frac{1 - H(x|\tilde{\tau})}{1 - H(x|\tau)}\left[\frac{\int_x^1 ph(p|\tilde{\tau})dp}{1 - H(x|\tilde{\tau})} - x\right] + x \\ E[H(p|\tau)|p \geq x] &\leq \frac{1 - H(x|\tilde{\tau})}{1 - H(x|\tau)}[E[H(p|\tilde{\tau})|p \geq x] - x] + x \end{aligned}$$

The above equation holds for any $x \in (0, 1]$. Now if $x = \bar{p}(\tau)$ we have:

$$\begin{aligned} E[H(p|\tau)|p \geq \bar{p}(\tau)] &\leq \frac{1 - H(\bar{p}(\tau)|\tilde{\tau})}{1 - H(\bar{p}(\tau)|\tau)}[E[H(p|\tilde{\tau})|p \geq \bar{p}(\tau)] - \bar{p}(\tau)] + \bar{p}(\tau) \\ &\leq [E[H(p|\tilde{\tau})|p \geq \bar{p}(\tau)] - \bar{p}(\tau)] + \bar{p}(\tau) \\ &\leq E[H(p|\tilde{\tau})|p \geq \bar{p}(\tau)] \\ &\leq E[H(p|\tilde{\tau})|p \geq \bar{p}(\tilde{\tau})] \end{aligned}$$

The second inequality follows from $\bar{p}(\tau) \leq p_0$ and the definition of a mean preserving spread and the fourth inequality comes from $\bar{p}(\tau) \leq \bar{p}(\tilde{\tau})$.

■

References

Clark, Kim B and Lawrence H Summers, "Labour Force Participation: Timing and Persistence," *Review of Economic Studies*, 1982, *49* (5), 825–44.

Farber, Henry S. and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 1996, *111*, 1007–1048.

Flinn, Christopher J., "Wages and Job Mobility of Young Workers," *Journal of Political Economy*, June 1986, *94* (3), S88–S110.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent, "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, October 2005, *23* (4), 681–724.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy*, October 1979, *87* (5), 972–90.

_ , "Matching, Turnover, and Unemployment," *Journal of Political Economy*, 1984, *92* (1), 108–22.

Koenker, Roger and Kevin F. Hallock, "Quantile Regression," *Journal of Economic Perspectives*, Fall 2001, *15* (4), 143–156.

Lange, Fabian, "The Speed of Employer Learning," *Journal of Labor Economics*, 2007, *25*, 1–35.

McCall, Brian P, "Occupational Matching: A Test of Sorts," *Journal of Political Economy*, February 1990, *98* (1), 45–69.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 01 2004, *72* (1), 1–32.

Moscarini, Giuseppe, "Skill and Luck in the Theory of Turnover," February 2003. Unpublished, Yale University.

_ , "Job Matching and the Wage Distribution," *Econometrica*, 2005, *73* (2), 481–516.

Neal, Derek, "The Complexity of Job Mobility among Young Men," *Journal of Labor Economics*, April 1999, *17* (2), 237–61.

_ **and Sherwin Rosen**, "Theories of the distribution of earnings," in A.B. Atkinson and F. Bourguignon, eds., *Handbook of Income Distribution*, Vol. 1 of *Handbook of Income Distribution*, Elsevier, 2000, chapter 7, pp. 379–427.

Papageorgiou, Theodore, "Learning Your Comparative Advantage," December 2007. Unpublished, Yale University.

Pavan, Ronni, "Career Choice and Wage Growth," July 2006. Unpublished, University of Rochester.

_ , "The Role of Career Choice in Understanding Job Mobility," July 2007. Unpublished, University of Rochester.

Shimer, Robert, "Reassessing the Ins and Outs of Unemployment," October 2007. Unpublished, University of Chicago.

Topel, Robert H and Michael P Ward, "Job Mobility and the Careers of Young Men," *The Quarterly Journal of Economics*, 1992, *107* (2), 439–79.