Move it and Lose it?
The Portability of On-the-Job Training
Please click here to download the most recent version.

Zachary Mahone

November 24, 2015

Abstract

This paper estimates the returns to on-the-job training and, in particular, asks how portable these acquired skills are from one job to the next. I build a model of a labor market with undirected, on-the-job search and counteroffers where firms may sign long-term contracts with workers that jointly determine wages and costly training (human capital acquisition). For any amount of training, some fraction of the human capital acquired is job-specific and is lost upon job termination. Estimating returns to training in this environment is complicated because wages do not reflect productivity; long-term contracts limit the degree to which training passes through to wages. Further, job transitions are endogenous; workers choose to leave only when a new job compensates them for lost skills. In this environment, even job transitions that induce human capital loss commonly feature wage gains.

Initial estimates suggests that skills acquired in training are fairly job-specific and that this plays a significant role in reducing job turnover. These findings have ramifications for the productivity costs of job loss, the sources of wage dispersion and the structure of policies that incentivize workforce training.
1 Introduction

Human capital is an important determinant of wages and productivity. While it can be accumulated in many ways, a primary source of skill acquisition is on the job (Mincer, 1962). But how transferable are these skills from one employer to the next? How are they rewarded both on the job where they are acquired and over the life cycle of a worker? And how strongly does human capital accumulation on the job influence separation decisions? The answers to these questions have significant implications for our understanding of the productivity costs of job loss, the sources of wage dispersion and the structure of policies to incentivize workforce training. They also have implications for understanding the effects of the observed decline in employer training since the 1990’s.

In this paper I take a structural approach to answering these questions and I focus on a particular form of observable on-the-job skill acquisition: employer training. Skills acquired in training can be either general to all jobs or specific to the job on which training is received. How portable, or transferable, training is reflects the composition of general and specific skills acquired through training. I consider a labor market environment where firms choose both when to provide skills to workers through costly training and how to optimally reward them using long term contracts. Employers and employees face uncertainty over which training and job search opportunities they will confront. I show that in the optimal contract wages will not respond to training unless forced to by a worker’s outside offer and that separation decisions are effected by training only if there is a specific component to the skills acquired. I exploit this property when I estimate the model on panel data of wages and employer training. I find that skills are quite specific; for workers with less than a highschool education on average about 56% of skills are portable across jobs. These findings confirm that training is an important source of productivity growth. Interestingly, the findings also highlight a potential pitfall in using wages to learn about how transferable skills are: training is estimated to be fairly specific even though wage regressions find the rewards to training are largely independent of the job on which it was acquired.

This paper makes two contributions to the literature. First, it develops a model with
on the job search where firms jointly determine wages and training according to an optimal contract. Second, it estimates the model on panel data of wages and training from the National Longitudinal Survey of Youth. The structural estimation is important because of the model environment. The existence of on the job search implies that outside employers reward non-productive training: workers choose to separate from the incumbent employer only if their outside offers are higher than the value of the current match (including specific capital). Further, the existence of long-term contracts mutes how changes in productivity are reflected in wages, effecting both estimated returns to training and tenure. From the perspective of a model environment with spot market wages, this dampens returns to tenure and amplifies those of experience, even if the human capital acquired is largely non-transferable.

The model environment is a version of Lentz (2010) augmented to include a training choice. Training options arrive randomly to each match in the form of a discrete option in units of time. This can be thought of as the time required to train workers to perform a specific task or learn a certain program. At the end of each period, workers can search on the job in an undirected manner, where any offers received are private information. If a worker receives an offer they consider taking, the incumbent firm is allowed to make a counter-offer as in Postel-Vinay and Robin (2002). I assume neither workers nor firms can commit to their response to arriving outside offers, however I do assume that once bargaining is settled, a firm can commit to the contractual provision of the negotiated outcome.

In this environment I show that wages are constant until renegotiation is triggered. They respond to training only insofar as the outcome of negotiations change for any given path of outside offers. In particular, this means that training may not be reflected in wages until a worker leaves their current match. I further show that separation decisions are informative about how portable training is: if some portion of training is match specific (i.e. specific to the employer-employee pair), then training will reduce worker mobility. In fact, in my environment training reduces worker mobility if and only if some portion of acquired skills is match specific.

The estimation strategy exploits this link between turnover and the specificity of training.
By comparing the separation decisions of trained versus untrained workers I can identify the total cost of separation implied by training. These costs are increasing in the value and productivity of skills and decreasing in how portable they are. I can then use the fact that on the job wage growth is increasing in both the productivity and portability of training to separately identify the two effects.

The distinction between general and specific human capital was popularized by Becker (1962). Early interest in the implications of this distinction focused on how the efficiency of skill provision in decentralized markets would depend on the specificity of skills and who paid for their acquisition. This line of investigation has been continued by Acemoglu and Pischke (1999) and most recently by Roys and Lentz (2015). An early paper that linked specific capital to labor market dynamics was the important contribution of Jovanovic (1979). He explores an equilibrium model of labor markets with investment in specific capital and used it to interpret the negative structural dependence of turnover on tenure. My results on the link between training and turnover will partially echo his. A more recent group of papers explore the quantitative implications of different types of specific capital. Violante (2002) examines the contribution of vintage-capital-specific skills and faster technology growth to rising wage inequality in the US. Kambourov and Manovskii (2009) and Yamaguchi (2012) argue that skills are occupation-specific, and this idea has been quantitatively explored in the context of wage inequality (Kambourov and Manovskii, 2009), employment dynamics during the great recession (Wiczer, 2015) and, most closely related to this paper, the costs of skill mismatch using task data from O*Net (Lise and Postel-Vinay, 2015; Guvenen, Kuruscu, Tanaka, and Wiczer, 2015).

Empirical evidence of the importance of training and its specificity has been provided in a variety of papers. Brown (1989), Duncan and Hoffman (1979) and Mincer (1988) use

---

1 The environment considered in Roys and Lentz (2015) shares several characteristics of the one presented below, although the papers were developed independently. More importantly, the models are used to different ends. Roys and Lentz (2015) are concerned primarily with providing very nice results on the efficiency question along with a quantitative example. This paper will provide some theoretical results but is primarily focused on measurement.

2 It should be noted that this implication is also embedded in the model of training and job search developed by Jovanovic (1979); here I explore the quantitative implications of it.
the Panel Study on Income Dynamics (PSID) and find training plays an important role in wage growth, largely explaining returns to tenure. Barron et al. (1989) find similar results using a unique, short panel on firm search, training and compensation collected by the National Institute of Education. A novel characteristic of their data is a separate measure of worker productivity (as reported by the employer), which yields an estimate that at least 50% of on the job training is specific. Later studies using the National Longitudinal Survey of Youth (NLSY), such as Parent (1999), find the returns to training are important but observe mixed evidence on how portable those skills are. Training from both current and past employers appears equally rewarded on the job while training also appears to reduce job to job mobility. Frazis and Loewenstein (2005) use the NLSY to estimate a functional form for training returns to wages. See Frazis and Spletzer (2005) for a nice review of training studies using the NLSY.

The question of how transferable skills are also maps to a broader empirical literature examining the wage-returns to tenure and labor market experience. The link between returns to tenure and specific capital is most explicit in Jovanovic and Mincer (1980) and Topel (1991), but the relative importance of experience versus tenure is commonly interpreted as reflecting the transferability of skills accumulated on the job. The evidence from this work is mixed, with some finding little contribution of tenure (Altonji and Shakotko, 1987; Altonji and Williams, 1997; Abraham and Farber, 1987) and others finding a more substantial component (Topel, 1991; Dustmann and Meghir, 2005).

A large literature has studied a variety of agreements between firms and workers that decouple the link between wages and marginal product. Under such contracts, observing how wages respond to training and job transitions may not be informative about productivity changes. Early studies such as Azariadis (1975) show that under commitment firms will perfectly insure their workers. Later papers have studied different environments with long term contracts that yield wage growth on the job, typically relaxing the assumption of worker commitment. Lazear and Rosen (1981) study wage growth through tournaments in an environment where workers can take private actions that effect their productivity.
Thomas and Worrall (1988) and Rudanko (2009) study environments in which neither side has commitment, resulting in wages that both rise and fall over the life of the job. Burdett and Coles (2003) and Shi (2009) explore environments with different types of on-the-job search, where promises of higher future pay are used to induce unobservable worker actions. Lamadon (2014) includes this feature as well as idiosyncratic risk and private action. The environment most closely related to mine is Lentz (2010), where the value of randomly arriving outside offers that firms respond to is endogenized by a bargaining protocol between the incumbent and outside firm.

In Section 2, I discuss the construction of the panel used in this paper, describe the characteristics of observed on the job training and present descriptive regressions of the data for comparison with other studies. In Section 3, I introduce the model and briefly characterize some properties relating to wages, training and turnover. In Section 4, I discuss identification of the model, the estimation strategy and present results.

2 Data

The panel data I use is obtained from the 1979 cohort of the National Longitudinal Survey of Youth. I use the nationally representative subsample. The panel is constructed on an annual basis since training is often longer than one quarter but rarely surpasses a year. The modeling assumption that workers are ex-ante homogeneous motivates separate estimation by education groups and restricting my attention to males only. It is reasonable to think that training decisions will be particularly sensitive to unobserved changes in workforce attachment, and for women in particular this may have been changing during the sample period. As a result, all further summary statistics will be reported for males only and conditional on one of two education groups: those who have attained at most 12 years of education and those with more (which I will refer to as “highschool” and “more than highschool”). In the estimation, additional heterogeneity will be removed through the use of residuals.

Wage, job characteristic and labor force data is constructed using the weekly Work His-
tory panel. As in Pavan (2011) I assign the primary job in a given year to that in which the most hours worked are reported. I define a worker as being attached to the labor force if at least three consecutive years are observed with a minimum of 30 hours worked each year. I restrict the sample to workers who I observe become attached to the labor force and eliminate those who I first observe attach to the labor force after the age of 30 due to concerns about missing earlier work experience (this is less than 25% of individuals and an even smaller percentage of observations). I treat those who I observe unattached to the labor force for three consecutive years as attrition from the sample. I also delete military and self-employed.\(^3\) Summary statistics for the sample are reported in Figure 1.

Figure 1: Sample Summary Statistics

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Full Sample</th>
<th>≤ High School</th>
<th>&gt; High School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>46,091</td>
<td>21,466</td>
<td>24,625</td>
</tr>
<tr>
<td>Individuals</td>
<td>2,039</td>
<td>965</td>
<td>1,074</td>
</tr>
<tr>
<td>Mean age at interview</td>
<td>34.7</td>
<td>34.2</td>
<td>35.1</td>
</tr>
<tr>
<td>Mean years schooling</td>
<td>13.8</td>
<td>11.5</td>
<td>15.8</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>7.0</td>
<td>8.4</td>
<td>6.2</td>
</tr>
<tr>
<td>% African-American</td>
<td>10.1</td>
<td>12.5</td>
<td>7.9</td>
</tr>
<tr>
<td>Mean tenure (weeks)</td>
<td>325.7</td>
<td>322.7</td>
<td>328.4</td>
</tr>
<tr>
<td>Median tenure (weeks)</td>
<td>206</td>
<td>202</td>
<td>211</td>
</tr>
<tr>
<td>Mean hours worked (annual)</td>
<td>2126.3</td>
<td>2089</td>
<td>2158.8</td>
</tr>
</tbody>
</table>

2.1 Training in the NLSY

The NLSY gathers information on various types of skill acquisition, including formal education, government sponsored training and “other” training, which is what this paper focuses on. “Other” training captures both employer and non-employer sponsored (self-paid, for example) training spells. These questions are asked at each interview and up to four spells may be asked about (two that were in progress last interview and two new ones). Informa-

\(^3\)Details on the construction of the sample and training heterogeneity are available in the Data Appendix A
tion on where the training was obtained, who sponsored it, how long it lasted and (if more than a week), how many weeks and how many hours per week were spent allows for the construction, for each training spell, of the total amount of hours spent in training.

There are two issues to be dealt with in constructing the panel of training spells. First, the NLSY does not directly match employers and training spells. Second, the questionnaire changed in 1988 in three ways: prior to this date, characteristics of training lasting less than one month were not gathered, questions about who paid for training were not asked and finally, weeks attended were not asked (i.e. training could span two months but only consist of one week per month). In both matching employers to training spells and dealing with the changes in 1988 I follow the approach of Frazis and Loewenstein (2005), briefly summarized here. Missing information prior to 1988 is imputed using the characteristics of similar spells post-1988. In particular, since on average a training session lasting \( n \) months consisted of \( 4n \) weeks of training ater 1988, these weeks attended were imputed accordingly. The same approach is used for hours when spells lasted less than one month, although these are computed conditional on age. Employer training was identified as all spells that were described as “company training” or lasted less than one month (most spells lasting less than one month post-1988 were employer sponsored). Training spells are matched to employers using the Employment History roster, information on when the training spells took place and where the worker was employed at that time. In the event that more than one main job is reported during the entire training spell, training is allocated to the main job. If more than one main job is reported during the time period, the training is re-allocated to non-employer paid. Approximately 75% of training spells that I observe are identified as employer paid, which likely underestimates the true fraction given the reassignment of sponsor status when no match can be found.

Below in Figure 2 I report summary statistics for those who receive employer paid training. The second row reports the number of individuals who we observed train at some point. For comparison, the third row reports the size of the full sample. The median hours trained ranges from one to two weeks, with means more than three times this, so the distribution
of training appears to be fairly skewed. While there are clear differences in the probability of being trained and in the length of training spells, the average age and tenure at which training is received does not differ noticeably between the two education groups. Note that in these summary statistics I report hours trained. In the model, these represent the fraction of average total hours worked across jobs, normalized to one. To map the data into the model I construct normalized training in terms of average annual hours worked for each education group to make the addition of training quantities across jobs consistent.

Figure 2: Training Summary Statistics

<table>
<thead>
<tr>
<th>Statistics (at training)</th>
<th>Full Sample</th>
<th>≤ High School</th>
<th>&gt; High School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Spells</td>
<td>3,232</td>
<td>843</td>
<td>2,389</td>
</tr>
<tr>
<td>Individuals</td>
<td>1,112</td>
<td>381</td>
<td>731</td>
</tr>
<tr>
<td>Full Sample</td>
<td>2,039</td>
<td>965</td>
<td>1,074</td>
</tr>
<tr>
<td>Mean age</td>
<td>33.9</td>
<td>33.3</td>
<td>34.1</td>
</tr>
<tr>
<td>Mean years schooling</td>
<td>14.9</td>
<td>11.8</td>
<td>16.1</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>6.3</td>
<td>7.1</td>
<td>5.9</td>
</tr>
<tr>
<td>% African-American</td>
<td>7.7</td>
<td>8.1</td>
<td>7.5</td>
</tr>
<tr>
<td>Mean tenure (weeks)</td>
<td>312.1</td>
<td>311.8</td>
<td>312.2</td>
</tr>
<tr>
<td>Median tenure (weeks)</td>
<td>219</td>
<td>216</td>
<td>221</td>
</tr>
<tr>
<td>Mean hours trained</td>
<td>171.8</td>
<td>137.5</td>
<td>183.9</td>
</tr>
<tr>
<td>Median hours trained</td>
<td>72</td>
<td>52.1</td>
<td>80</td>
</tr>
<tr>
<td>Mean training spells by job</td>
<td>1.9</td>
<td>1.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Median training spells by job</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

To briefly describe the panel characteristics of training and wages I present the results of a few simple wage regressions in Table 1. The dependent variable in all regressions is the reported log hourly wage. Other independent variables included but not reported are quadratic and cubic terms for tenure and experience along with race, unionization and occupation dummies. Occupational categories are collapsed into those constructed by Autor and Dorn (2013). Column (1) reports the regression on the entire sample and column (2) restricts the sample by eliminating training observations in the top centile (for both current and past training). Column (3) reports estimates on the entire sample using the cubic root
specification of training returns that Frazis and Spletzer (2005) identify as their preferred model. The linear specification of hours in the first two columns provides a simple interpretation of training effects on wages through percent change. With the cubic specification, returns depend on existing training. As Frazis and Spletzer (2005) find, this specification yields generally higher returns to training.

Table 1: Returns to Experience, Tenure and Training

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Full Panel</th>
<th>Restricted Panel</th>
<th>FL2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenure</td>
<td>0.0440***</td>
<td>0.0429***</td>
<td>0.0409***</td>
</tr>
<tr>
<td>Experience</td>
<td>0.0633***</td>
<td>0.0625***</td>
<td>0.0592***</td>
</tr>
<tr>
<td>Current Training</td>
<td>0.0001***</td>
<td>0.0003***</td>
<td>0.0267***</td>
</tr>
<tr>
<td>Other Training</td>
<td>0.0001***</td>
<td>0.0007***</td>
<td>0.0400***</td>
</tr>
<tr>
<td>N</td>
<td>36201</td>
<td>35475</td>
<td>36201</td>
</tr>
</tbody>
</table>

We are also interested in the mobility effects of training. To investigate this, I estimate a proportional hazard model using Cox’s approach. The results are presented for each education group in Table 2.

Table 2: Cox Proportional Hazard Estimates

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>≥ HS</th>
<th>&lt; HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Training</td>
<td>0.9547***</td>
<td>0.9228***</td>
</tr>
<tr>
<td>Past Training</td>
<td>1.0148***</td>
<td>1.0417***</td>
</tr>
<tr>
<td>Experience</td>
<td>0.9797*</td>
<td>0.9797*</td>
</tr>
</tbody>
</table>

| N | 4144 | 3914 |

Included independent variables that are not reported here are a quadratic experience term, education (within-group variation, which is not significant), race, unionization and occupation dummies as described above. For interpretability, I also normalized training by the median length of a training spell (80 hours) so that the effects of training are in units of the median spell rather than hourly. The estimates echo those of Parent (1999) although different sample selection and units of analysis yield differing coefficients. Training with the current employer reduces worker mobility while training with previous employers has the opposite effect. For both groups a median-length spell of current training has a larger effect.
on the hazard rate than an additional year of experience, although it is interesting to note that the effects of both current and past training appear larger for the less educated group.\footnote{I report in Figure A.2 the unconditional turnover rates by training and education groups. These tables demonstrate the substantial difference in mobility between trained and untrained workers but ignore all heterogeneity beyond tenure and an indicator for training.}

## 3 Model

In this section I describe a model of a labor market in which firms sign long term contracts with workers and workers may search on the job. Human capital is acquired through on the job training at the firm and some fraction of this is match-specific. Workers are assumed to lack commitment and may consider leaving a job for an outside opportunity. In this case, their current employer is allowed to counter the outside offer as in Postel-Vinay and Robin (2002).

### 3.1 Agents and Matches

#### 3.1.1 Workers

Workers enter the labor force with age 0 and exit with probability $\delta_x$ every period. Each worker’s state is captured by two parameters, specific human capital $h$ and general human capital $g$. I assume that all human capital is accumulated through on the job training, so new entrants to the labor force are indexed by $(h, g) = (0, 0)$. While employed, workers meet outside firms at a rate $\lambda_e$. While unemployed, this meeting rate is $\lambda_u$.

Workers are risk averse, discount the future at rate $\beta$, and maximize the present discounted value of an income stream, $\sum \beta^t u(w_t)$. Note that implicitly access to savings are ruled out, a standard assumption to motivate contracting.

#### 3.1.2 Firms

Firms are single-worker entities and are indexed by a fixed productivity parameter $p$. I will assume that upon reaching a bargaining outcome, firms can commit to a contract that
provides a certain level of expected utility. This is similar to environments explored by Lentz and Bagger (2015) and Lentz (2010).

Firms have access to a technology that combines general and specific capital, \( k(h, g) \) along with a production technology \( f(p, k) \). Both are assumed differentiable and increasing in their arguments. For initial estimates this paper will use a linear combination \( k(h, g) = h + g \) and production technology \( f(p, k) = p + k \xi_h \). The linearity simplifies the requirements of the estimation while decreasing returns will be important in informing returns to training.

In addition to their production technology, firms also have access to long term contracts \( C \) that determine wages, training decisions and future promised utility. These will be described shortly in the contracting problem.

### 3.1.3 Matches

A match is a worker-firm pairing. The state of a match at the beginning of each period is thus summarized by promised utility, firm type and the worker’s state \( (V, p, h, g) \). A productive match also has access to a training option \( \tau \) that arrives randomly from the set \( H(\tau) \). These values are denoted in time. If training occurs, the resulting human capital gain \( s(\tau, p) \) may depend both on time spent training and the productivity of the firm.

Matches end with a random probability \( \delta_u \). In this case, the worker enters unemployment and the match dissolves.

It is worth noting briefly here that there is a sense in which matches are a combination of experience and inspection goods in this model. Firm and worker states are perfectly observable but the random arrival of training options means that the inside value of the match can evolve over time.

### 3.2 Human Capital Accumulation

I keep the law of motion for human capital as simple as possible (it can be thought of as a combination of Jovanovic (1979) and Violante (2002)). As indicated earlier, a fixed fraction \( \gamma \) of human capital accumulation through training will be general. Upon leaving, a worker
brings all their general capital with them and their stock of specific capital is destroyed. That is

\[ h' = \begin{cases} 
  h + (1 - \gamma)s(\tau, p) & \text{if no separation} \\
  0 & \text{if separation}
\end{cases} \quad (1) \]

\[ g' = g + \gamma s(\tau, p) \quad (2) \]

This formulation has several implications that will be useful in capturing the data. The first is that if \( \gamma < 1 \), then the more workers are trained, the larger the cost of leaving. Thus one force in the model will be that training (if specific to some degree) reduces turnover, consistent with observations in the data. The second implication is that the only effect of past training on a current match is to increase the general component of a worker’s human capital. In other words, past training will have no negative effects on separation probability (indeed, in the data we observe separation probabilities increasing in this value). It should be noted that this implies a minimum fraction of total human capital \( \gamma \) can be general, so proportional losses are bounded below.

### 3.3 Timing

Timing in this world is described in Figure 3. Training options arrive randomly at the beginning of the period (and are i.i.d). The contract specifies in all states whether this training occurs, how wages are set and what promised utilities are (which determine the probability the contract is renegotiated). If the match survives the unemployment and labor-force exit shocks, meetings randomly arrive to workers which may trigger bargaining and possibly separation, to be described in more detail below.

Figure 3: Intra-period Timing

<table>
<thead>
<tr>
<th>Training (( \tau ))</th>
<th>Unemployment, exit (( \delta_u, \delta_x ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>Produce, pay ( C )</td>
</tr>
<tr>
<td>( t )</td>
<td>Search, bargain, separate</td>
</tr>
<tr>
<td>( t+1 )</td>
<td></td>
</tr>
</tbody>
</table>

13
4 Choices and Value Functions

I will construct the sequence of intra-period choices and value functions working backwards from the end of the period. Consider a worker with state \((V', p, \hat{h}, \hat{g})\) who has already had a training opportunity in the current period, produced and consumed, has promised utility \(V'\) in hand and now may make contact with an outside firm.

4.1 Matching and Bargaining

Bargaining is over promised utility and follows the protocol in Postel-Vinay and Robin (2002) where it is assumed that firms obtain all surplus. Contact with firms arrive randomly to workers, drawn from the sampling distribution \(G\). These contacts are assumed to be private information on the part of the worker. After receiving an offer from the firm it has made contact with, a worker can choose whether or not to reveal this offer to its current employer. If they choose not to reveal the offer then the contract continues according to the state \((V', p, \hat{h}, \hat{g})\). If however the worker reveals the offer to the incumbent employer, the two competing employers enter into an alternating-offer bargaining process. This results in the sampling distribution being divided into three states: those contacts that are not worth revealing, those contacts with offers worth revealing but that will not induce separation and those contacts that will result in separation.

Define \(V^*(p, h, g)\) be the maximum promised utility a firm can offer a worker given \((p, h, g)\) while maintaining non-negative profits. We can then define three subsets of the sampling distribution:

\[
\Omega^0(V', \hat{g}) = \{p|V^*(p', 0, g) \leq V'\} \\
\Omega^1(V', p, \hat{h}, \hat{g}) = \{p'|V' < V^*(p', 0, \hat{g}) \leq V^*(p, \hat{h}, \hat{g})\} \\
\Omega^2(V', p, \hat{h}, \hat{g}) = \{p'|V^*(p', 0, \hat{g}) > V^*(p, \hat{h}, \hat{g})\}
\]

We define \(\hat{V}(p', V', p, \hat{h}, \hat{g})\) to be the promised utility accruing to a worker resulting from
an incumbent state \((V', p, \hat{h}, \hat{g})\) and meeting a firm \(p'\). The bargaining protocol then sets:

\[
\hat{V}(p', V', p, \hat{h}, \hat{g}) = \begin{cases} 
V & \text{if } p' \in \Omega^0(V', \hat{g}) \\
V^*(p', 0, \hat{g}) & \text{if } p' \in \Omega^1(V', p, \hat{h}, \hat{g}) \\
V^*(p, \hat{h}, \hat{g}) & \text{if } p' \in \Omega^2(V', p, \hat{h}, \hat{g}) 
\end{cases}
\] (6)

Continuation values for both worker and firm are not identical to promised utilities because of the option value of renegotiation, in which case the existing contract promises are thrown out. Given the outcome of bargaining described above, I define below the value of search.

### 4.1.1 Worker Continuation Value

Suppose a contract has specified expected continuation utility \(W\) and training such that at the time of search (post training and production) \((W, p, \hat{h}, \hat{g})\). We define the worker value under renegotiation as

\[
\hat{W}(W, p, \hat{h}, \hat{g}) = (1 - \delta_u) \lambda_e \int_{p' \Omega^0(W, \hat{g})} \hat{V}(s, p, \hat{h}, \hat{g}) dG(s) 
\] (7)

### 4.1.2 Firm Continuation Value

We first define

\[
\bar{J}(W, p, h, g) = \max_{V^*} E[J(V^*, p, h, g, \tau)] \quad \text{s.t. } E[V^*] \geq W 
\] (8)

From this we can then define the continuation value of the firm under renegotiation as

\[
\tilde{J}(W, p, \hat{h}, \hat{g}) = (1 - \delta_u) \lambda_e \int_{\Omega^1(V', p, h, \hat{g})} \bar{J}(V(s, V', p, \hat{h}, \hat{g}), p, \hat{h}, \hat{g}) dG(s) 
\] (9)

I will now use these to write the contracting problem faced by the firm.
4.2 Contracting Problem

Having defined search and bargaining consider the problem facing the firm upon entering the period with \( z = (V, p, h, g, \tau) \). A contract specifies \( \mathcal{C} = \{ w_i(z), W_i(z), V_i', \delta_i(z) \}_{i=t,nt} \), where \( \delta_i(z) \in (0,1) \) and \( \sum \delta_i(z) = 1 \). Contracts are measurable with respect to the match state, training option \( \tau \) and training decision \( \delta \). Also note that I am allowing for a two point lottery for each \( \tau \). This is a technical requirement to guarantee concavity of the problem.

\[
J(V, p, h, g, \tau) = \max_{\delta, w, W, V'} \sum \delta_i \left\{ f(p, k(h, g)) - w_i - c(\delta_i \tau) + \beta \left[ r(W_i, g_i) E_{\tau'} [J(V_i', p, h_i, g_i')] + J(W_i, p, h_i, g_i')] \right] \right\}
\]

s.t.
\[
\sum \delta_i \left\{ u(w_i) + \beta \left[ \delta_i U(g_i) + r(W_i, g_i) W_i + \tilde{W}(W_i, p, h_i, g_i) \right] \right\} \geq V \quad \text{(P1)}
\]
\[
\sum \delta_i \{ E_{\tau'} [V_i'] - W_i \} \geq 0
\]
\[
\sum \delta_i = 1 \quad \forall z, \tau
\]
\[
r(W_i, g_i) = [1 - \lambda_e + \lambda_e G(q_0(W_i, g_i))](1 - \delta_u)
\]
\[
h_{nt}(z, \tau) = h, \quad g_{nt}(z, \tau) = g
\]
\[
h_t(z, \tau) = (1 - \gamma)s(\tau, p) + h, \quad g_t(\tau) = \gamma s(\tau, p)
\]

The solution to the problem above can also be used to define implicitly the maximum willingness to pay \( V^*(p, h, g) \) as

\[
\hat{J}(V^*(p, h, g), p, h, g) = 0 \quad (10)
\]

4.3 Unemployed

An unemployed worker by definition has zero specific capital, so their state is \( h, g = (0, g) \). I will treat unemployment as being matched with the lowest productivity firm, \( p_{\text{min}} \) with zero
probability of training. This implies that there is no selection leaving unemployment; any meeting will result in a match. The value of unemployment then solves

\[
U(g) = f(p_{\text{min}}, k(0, g)) + \beta U(g) \Rightarrow U(g) = \frac{f(p_{\text{min}}, k(0, g))}{1 - \beta}
\]  

(11)

Note that while search occurs in unemployment, the assumption that firms obtain all surplus implies that from the workers perspective, getting hired will not change their level of promised utility.

Since I am not performing steady state analysis here I will not define nor prove the existence of an equilibrium. In my simulations I will not be requiring a steady state distribution, similar to Lise and Postel-Vinay (2015), so I will proceed immediately to a characterization of the optimal contract and its properties.

4.4 Properties of Contracts and Turnover

Lemma 4.1. The firm profit function \( J(V, p, h, g, \tau) \) is a continuously differentiable, concave and decreasing function in \( V \).

Proof. See appendix B.1

Concavity is ensured by the use of a lottery over the training choice. Although the continuation value of firms is complicated by the potential for renegotiation, promised utility only enters this function where renegotiation does not occur, which simplifies the characterization considerably.

Since I am interested in how wages, turnover and training are determined, the following statements characterize these movements in the theory.

Proposition 4.2. The contract defined in (P1) is unique and has the properties that:

1. Wages are constant unless renegotiation is triggered

2. Training is increasing in firm type \( p \) i.e.

   if \( \delta(V, p, h, g, \tau) = 1 \) then \( \delta(V, p', h, g, \tau) = 1 \ \forall p' > p \)
Proof. See appendix C.1

Because training is increasing in firm type, workers in matches where training is observed should on average be matched with more productive firms. To see clearly the implications of this, in Figure 4 I plot average employer productivity for a sample of simulated workers grouped according to whether or not they are on a training job (will be trained in that match).

Figure 4: Climbing the Job Ladder: Trained v. Non-trained

The selection of better firms emerges clearly at the beginning of the life cycle. Given a higher initial average productivity we would expect slower growth in firm productivities among trained workers. However, we also observe that approximately half-way through the life cycle the average productivity of non-training jobs surpasses that of the trained. This overtaking is the result of specific capital reducing the mobility of trained workers. There is a sense in which training with specific capital makes the job ladder “sticky”. It should be emphasized however that because match productivity is relatively high for trained workers, the gains to mobility are lower.
The result that wages within a contract are perfectly smooth until the arrival of a competitive outside offer has been obtained in many environments, (a well known example being Thomas and Worrall (1988)). It is of interest here primarily because of the implications it has for wage responses to training. Also note that these contracts imply that the path of wages on the job will feature a positive slope as in other contracting models with on the job search (Shi (2009), Burdett and Coles (2003)), although backloading arises in a slightly different manner. In the latter group, promised utility increases over time in anticipation of outside offers the firm cannot counter. Here, because firms can directly counter any offer, continuation values rise only when a worker has a competitive offer in hand.

To give a flavor of the implications these contracts have for wages and training, I plot in Figure 5 the time path of wages and productivity for a simulated worker. The black line represents wages as specified by the contract while the dashed blue line represents total human capital (i.e. \( h + g \)). The periods bookended by red bars denote job transitions, identified by the reduction in total human capital from the specific component of training. We can observe that in both periods of job transitions, losses of specific capital are associated
with significant wage gains. Further, on-the-job wage responses to human capital growth are muted. In this particular path plotted, wages rise once on the job at age five, where an outside offer and training coincide in the same period.

4.4.1 Returns to Training

In the returns to training are realized in wages only when competitive outside offers arrive. To clarify this point, consider the example in Figure 6.

In this example, the worker is trained two periods into the job. Note that wages do not initially react. In subsequent periods the worker receives no more training but enjoys two raises, the first on her initial job and the second from a job transition. Even without training, on-the-job search would imply rising wages at these junctures. The effect of training is to obtain a wage rise larger than what would have otherwise been (it contributes to backloading, in a sense). It is worth emphasizing here that the gains the worker enjoys from her job transition (the second gray shaded region) are by definition the remaining surplus from the match, which reflects the value of the worker to the incumbent, hence includes the value of specific capital. It should also be pointed out that the contract will anticipate expected gains to training when determining wages, so wages in jobs will reflect to varying degrees the value of the training option even in matches in which no training is observed, although I do not focus on that here.
4.5 Turnover

We are also interested in how the dynamics of turnover will relate to training theoretically. Below I emphasize two characteristics of training in the model that are of empirical interest.

**Proposition 4.3.** Define $\text{Sep}(V,p,h,g)$ as the separation probability implied by a match $(V,p,h,g)$ given some value function $J$ and sampling distribution $G$. Under the functional form assumptions of training and production, turnover has the following properties:

1. Separation probability falls in training i.f.f. $\gamma < 1$

2. $\partial \text{Sep}/\partial \tau$ is falling in $g$ (and past training).

**Proof.** See C.2

The first statement echoes the results discussed in Section III of Jovanovic (1979). If some portion of training is specific to the job, then increased training will reduce turnover. If on the other hand all training is general, then the only reason an incumbent and outside firm will value the potential match differently is based on their firm productivities $p,p'$, in which case training is irrelevant for determining turnover (it depends only on the relative ranking of $p \succ p'$). This will be the logic that I exploit in using separation rates to identify how transferable skills are. I include the second statement of the proposition because it is a nice corollary to findings on training and mobility reported in Table 2. Recall that training in a current job was estimated to reduce mobility while training in past jobs increased separation probabilities. The second statement is a narrow version of the effect of past training. In words, for highly trained workers the marginal productive value of training is low due to decreasing returns in the production function. This means that the retention effects of current training are reduced by past training. This mechanism, combined with a positive correlation between training probabilities on past and current jobs, can generate the empirical relationship observed in the data.
5 Computation and Estimation

In this section I turn to the computation and estimation procedure for the model.

5.1 Recursive Contract Formulation

Solving the model requires computing the entire set of promised utilities contingent on training options for each present state. This is burdensome. Instead, I re-formulate the problem in terms of marginal utilities as in Marcet and Marimon (1999). This approach exploits the fact that the optimal contract sets \( J(V_{i'\tau'}, p, h, g, \tau) = \lambda \ \forall \tau' \). Hence, as long as \( J \) is strictly concave, knowing \( \lambda \) is enough to compute the entire set of promised utilities.

\[
P(\mu, p, h, g, \tau) = \sup_V J(V, p, h, g, \tau) + \mu V \tag{12}
\]

With some simple substitutions (see Appendix D) we can rewrite the problem as

\[
P(\mu, p, h, g) = \inf_{\lambda} \sup_{\delta_i, w_i, W_i} \sum_i \delta_i \left\{ f(p, k(h'_i, g'_i)) - w_i - c(\delta_i \tau) + \beta \hat{J}(W_i, p, h'_i, g'_i) \right. \\
+ \mu [u(w_i) + \beta \left( r(W_i, g'_i)W_i + \hat{W}(W_i, p, h'_i, g'_i) \right)] \\
- \beta r(W_i, g'_i)\lambda_i W_i \\
+ r(W_i, g'_i)\beta E_{\tau'}[P(\lambda_i, p, h'_i, g'_i)] \}
\]

In this way we benefit from needing to keep track only of promised marginal utilities rather than the entire set of future promised utilities contingent on non-renegotiation.

5.2 Estimation

I use simulated method of moments. Below I report the functional forms I use:

In preliminary estimates I set \( \sigma = 2 \) and \( y = 0.1 \). I also normalize average starting wages
Table 3: Functional Forms

| Utility function                      | $\frac{c_1^{1-\sigma}}{1-\sigma}$ |
| Production technology                | $f(p, h) = p + y + h^{\xi_h}$      |
| Training cost                        | $c_2 \tau^{c_1}$                   |
| Skill accumulation                   | $s(\tau, p) = h + \tau^{\xi_p}$   |
| Sampling distribution (firm)         | Beta$(\alpha, \beta)$              |
| Sampling distribution (training)     | Uniform on $[0, 1]$                |

to one. Contact rates from unemployment are directly from the data. Attrition rates are set to $\delta_x = 0.02$ following Lise and Postel Vinay. The discount rate is set to 4% annually. The value of unemployment is set equal to being matched with the lowest productivity firm and receiving no training. Finally, I set $c_2 = 2$ for now. This leaves 8 parameters $\gamma, \lambda_r, \lambda_e, c_1, \xi_h, \xi_p, \alpha, \beta$ to be estimated.

No one parameter is identified by a single data counterpart, but I give a heuristic argument here about the sources of identification. The training cost parameter $c_1$ will govern the conditional mean of training received. The arrival rate of training opportunities ($\lambda_r$) will, all else equal, determine the average probability of training being received. The complementarity between training and firm type, $\xi_p$, will be determined by the conditional variance of training. The effect of training on reducing mobility will inform us about the costs of separation, determined jointly by $\xi_h$ and $\gamma$. Using wage growth on the job we can separate these two. The empirical value of average EE transitions will pin down $\lambda_e$. Finally, all other parameters held constant, the sampling distribution of $B(\alpha, \beta)$ is identified from the wages of workers entering employment. For a detailed example of this strategy in a two period world see Appendix E.1.

The targeted moments are the following: average on-the-job log wage growth, mean and variance of log wages, the conditional mean and variance of training, probability of receiving training and separation rates of trained and untrained. The last 5 moments are computed at two points in tenure, 2 and 5 years, for a total of 15 moments. For each set of parameters I simulate data for a sample of workers who start initially unemployed and are then observed for 35 periods. The set of 8 parameters are chosen to fit the 13 moments. I use a version of
differential evolution to perform global search.

My model ignores several dimensions of heterogeneity that might be of concern in the estimation. While important, I discuss these issues in further detail in Appendix E.2 to economize on space.

## 5.3 Estimation Results

The estimation is still in progress. An updated version of the draft will be available shortly. In Table 4 I report the values of the model and data for the target moments for each education group, currently only initial estimates for the less than highschool group are available and I emphasize they are preliminary.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Model(≥HS)</th>
<th>Data (≥HS)</th>
<th>Model(&lt;HS)</th>
<th>Data (&lt;HS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M[\ln(w'/w)])</td>
<td>-</td>
<td>.055</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>(M[\ln(w)])</td>
<td>-</td>
<td>.76</td>
<td>0.49</td>
<td>0.55</td>
</tr>
<tr>
<td>(V[\ln(w)])</td>
<td>-</td>
<td>.53</td>
<td>0.15</td>
<td>0.33</td>
</tr>
<tr>
<td>(\text{Prob}(\tau &gt; 0))</td>
<td>-</td>
<td>.08</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>(M[\tau</td>
<td>\tau &gt; 0])</td>
<td>-</td>
<td>.05</td>
<td>0.06</td>
</tr>
<tr>
<td>(V[\tau</td>
<td>\tau &gt; 0])</td>
<td>-</td>
<td>.01</td>
<td>0.01</td>
</tr>
<tr>
<td>(M[\text{Sep}</td>
<td>t])</td>
<td>-</td>
<td>.15</td>
<td>0.18</td>
</tr>
<tr>
<td>(M[\text{Sep}</td>
<td>nt])</td>
<td>-</td>
<td>.25</td>
<td>.26</td>
</tr>
<tr>
<td>(\text{Prob}(\tau &gt; 0))</td>
<td>-</td>
<td>.11</td>
<td>.07</td>
<td>.05</td>
</tr>
<tr>
<td>(M[\tau</td>
<td>\tau &gt; 0])</td>
<td>-</td>
<td>.09</td>
<td>0.09</td>
</tr>
<tr>
<td>(V[\tau</td>
<td>\tau &gt; 0])</td>
<td>-</td>
<td>.025</td>
<td>0.01</td>
</tr>
<tr>
<td>(M[\text{Sep}</td>
<td>t])</td>
<td>-</td>
<td>.14</td>
<td>.18</td>
</tr>
<tr>
<td>(M[\text{Sep}</td>
<td>nt])</td>
<td>-</td>
<td>.18</td>
<td>.09</td>
</tr>
</tbody>
</table>

The model does reasonably well on training and separation but has trouble capturing wages. In particular, the model has trouble matching wage growth and also wage variance.

In Table 5 I report the parameter values from the estimation.
Table 5: Parameters of the Model [Preliminary and Incomplete]

<table>
<thead>
<tr>
<th>Moment</th>
<th>≥HS</th>
<th>&lt;HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ</td>
<td>-</td>
<td>0.56</td>
</tr>
<tr>
<td>λ_r</td>
<td>-</td>
<td>0.13</td>
</tr>
<tr>
<td>λ_e</td>
<td>-</td>
<td>0.43</td>
</tr>
<tr>
<td>c_1</td>
<td>-</td>
<td>0.63</td>
</tr>
<tr>
<td>ξ_h</td>
<td>-</td>
<td>1.20</td>
</tr>
<tr>
<td>ξ_p</td>
<td>-</td>
<td>0.75</td>
</tr>
<tr>
<td>a</td>
<td>-</td>
<td>3.3</td>
</tr>
<tr>
<td>b</td>
<td>-</td>
<td>3.18</td>
</tr>
</tbody>
</table>

6 Empirical Implications

6.1 Returns to Training

As discussed in the contract characterization, the returns to training are captured in option values which may not be immediately reflected in wages. For comparison, consider a wage model of the form I estimate in the data section

\[
\log(w_{ijt}) = \beta_0 + \beta_1 \text{Tenure}_{ijt} + \beta_2 \text{Experience}_{it} + \beta_3 \text{Current}_{ijt} + \beta_4 \text{Past}_{ijt} + \bar{\beta}Z_{it} + \epsilon_{it} \tag{13}
\]

The coefficient \( \beta_3 \) is the average percent change in wages obtained from a marginal hour of training, which in this model is linear in hours trained. The model counterpart of this estimate is complicated by the fact that wage returns to training are not only non-linear but depend on the realization of outside offers. The immediate change in the option value of renegotiation is unobserved. One way to measure returns is in terms of life-time wages. Define a lifetime wage for promised utility \( V \) as the constant wage \( w^L(V) \) that solves

\[
\sum \beta^t u(w^L(V)) = V \tag{14}
\]

we can then define the lifetime wage gain for the average training spell.
Definition 6.1. The sample lifetime average return is defined as

\[ \hat{R}_L = \frac{w^L (V' (\bar{\Omega}_\tau))}{w^L (\bar{V})} \]  

(15)

Notice that this value already has rolled up into it the option values of renegotiation in the future. We do not need to adjust for promised utility growth as above because \( V' (\Omega) = V \).

In Table 6 I report both the simulated regression returns to training and the return to lifetime wages. The regression coefficient on both current and past training is approximately 0.13. The mean training spell on average is 0.04, hence we obtain the return to the mean training spell below.

Table 6: Returns to Training

<table>
<thead>
<tr>
<th>Statistic</th>
<th>≥HS</th>
<th>&lt;HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{R}_L )</td>
<td>-</td>
<td>0.005</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>-</td>
<td>0.005</td>
</tr>
</tbody>
</table>

The two measures are nearly the same, which may seem surprising at first but recall that current and past training are rewarded largely the same. In this case, from the perspective of wages at least, the marginal increase for training today \( \beta_3 \) is the same as a lifetime increase in wages, since once the worker transitions into another job, the training will now be with a past employer but rewarded equally. What is interesting to note about this is that the returns are similar even though the underlying portability of skills is relatively low (0.56).

6.2 Training and Turnover

This model features selection in which types of firms train. As described in the theory section, more productive firms are more likely to train. This means that the observed differences in turnover between trained and untrained workers are due in part to being higher up the ladder and in part to the cost of losing specific capital. We can decompose this for trained workers
in the following way

\[ Sep(T|t) = \left[ 1 - \int_p G(x)\mu(x|T, t) \right] + S_r(T, t) \] (16)

The bracketed term is the expected turnover rate for a given distribution of firm types \( p \) conditional on years of tenure \( T \) and receiving training at their current job. The remaining value \( S_r(T, t) \) is the sample average effect of training on turnover at tenure year \( T \).

In Figure 7 I plot this decomposition for a simulated sample of workers.

Figure 7: Effects of Match Quality v. Productivity

From this graph it is clear that the effects of training on turnover are primarily felt later in tenure. By six years into tenure, training reduces turnover by about half.

7 Conclusion

In this paper I structurally estimate the returns to on the job training as well as how portable the acquired skills are from one job to the next. The structural approach is important because in my environment wages will not directly reflect worker marginal productivities
and therefore cannot be used to read off how human capital evolves within and across jobs. Further, without a model we cannot disentangle the effects of firm type from training on both wage formation and separation decisions. While the focus on employer provided training takes a narrow view of how human capital is accumulated on the job, the lessons learned here are applicable to the wider debate on specific versus general human capital and the returns to tenure versus experience.

I estimate the model on panel data of both wages and firm training to evaluate both the returns to and portability of skills acquired through on the job training. I find that returns wage returns to training computed within the model largely coincide with those implied by wage regressions. I also find that about 56% of skills are transferable across jobs. This is particularly interesting because in reduced form estimates the returns to training are largely independent of the job on which the training occurred, suggesting that we should be cautious about how we interpret what the coefficients of wage regressions tell us about the specificity of human capital.

Important future directions for this work would be to allow for worker heterogeneity or worker-sponsored training. While most of the training spells observed in the NLSY are employer sponsored, useful information is contained in worker-sponsored training spells that are currently ignored. For example, if worker heterogeneity and worker-sponsored training are both allowed, it may be that self-training is used as a signaling device. It could be interesting to think about how search outcomes respond to this. Additionally, the model, or similar versions, could also be fruitfully applied [ongoing at the moment] to interpret the causes of the fall in employer provided training since the 1990s. This has repercussions both for wages and productivity and in particular may have meaningful implications for worker training programs if part of this decline is due to a change in how job-specific skills are.
A NLSY 1979

A.1 Additional Sample Selection

I observed workers employment history using the NLSY Employment History roster. For workers with spells between jobs, since my model does not include participation decisions I must either treat the worker as having exited the labor market altogether or as unemployed. While an ad hoc assumption, I use three years as a cutoff: those workers whom I do not observe re-enter the labor market within three years are interpreted as having left the market completely. I also require that I observe worker’s entrance into the labor force. I assume that no workers enter the labor force fully until 16 years old and for those greater than 16 require that I observe them initially unattached to the labor force. Jointly this leaves 8,236 individuals.

A.2 Training Spells and Employer Sponsorship

Training data is described by start dates, end dates and hours per week. Since my model interprets spells as occurring entirely in one period, if there is training spell that spans two years I assign the spell to the year in which the majority of training occurred. Alternatively, it is possible for a reported training spell to exceed 52 weeks (although never more than 104). In these scenarios, training is split between the two years.

The dataset of training spells is constructed as follows. All responses to training questions are collected. This includes variables describing the start and end dates of training (month and year), the weeks attended each month if training lasted more than a month, the hours per week for all training lasting a week or more, the purpose of training (what it was intended to do) and who paid for this training. Some spells are still in progress at the time of interview. Complete information is obtained on these spells by matching responses from following years using the reported start dates of training and the type/purpose of training to identify the corresponding end date. I drop in progress training spells that cannot be matched with an end date. I end with 27,424 identified spells of training.
I now consider specifically the set of training spells that are reported as being employer sponsored; these are 20,538 (75% of spells are reported/imputed as employer trained). The imputations for prior to 1988 follow Frazis and Lowenstein described in the main test. I merge descriptive data on employers with an annual panel of observations from the Employer History roster. I merge the interview year a training spell is reported with each observation of a job active during that interview. Training is allocated to an employer using the dates of employment and the dates of training. If more than one employer is active during the spell of training, precedence is given to the reported main job (again, following F&L). I successfully can match 18,161 spells. The remaining revert to non-employer paid spells. For the years prior to 1988, hours are not available for spells lasting less than one month. These are imputed following Frazis and Lowenstein to yield a complete dataset of training spells matched to individual employers and in units of hours trained.

A.3 Wages and other characteristics

The annual panel of jobs is constructed separately using the Work History weekly panel. I construct a week by week set of jobs for each worker. For each job, I match the job number with the reported characteristics of the match from the appropriate interview year using the crosswalk variable between interview year employer ID and unique ID provided in the Employment History roster. Using the reported average hours worked per week and the weeks reported worked in each job, I construct a measure of total hours worked for each job. I select as the main job the job for which most hours are worked in a given year. Wages are constructed using the reported pay rates for each job. The top and bottom percentile of earnings are set to missing to deal with outliers. Finally, the training data set (including a variable for each respondent of total training in a given job and total other training) is merged with this data set to yield a complete panel of wages and training.

Note that when merging the EH roster with WH roster we miss some observations. These are jobs numbered 6-10 in a given year (additional jobs) for which we do not have employer information (see NLSY notes on the EH roster). I eliminate all workers who we do not first
<table>
<thead>
<tr>
<th>Age</th>
<th>Hispanic</th>
<th>African-American</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 25</td>
<td>1.8</td>
<td>2.5</td>
<td>2.7</td>
</tr>
<tr>
<td>25-34</td>
<td>7.0</td>
<td>8.4</td>
<td>9.6</td>
</tr>
<tr>
<td>35-44</td>
<td>7.5</td>
<td>7.8</td>
<td>7.3</td>
</tr>
<tr>
<td>≥ 45</td>
<td>5.6</td>
<td>7.2</td>
<td>5.4</td>
</tr>
</tbody>
</table>

(a) Males, College

<table>
<thead>
<tr>
<th>Age</th>
<th>Hispanic</th>
<th>African-American</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 25</td>
<td>2.4</td>
<td>1.2</td>
<td>1.7</td>
</tr>
<tr>
<td>25-34</td>
<td>6.4</td>
<td>6.5</td>
<td>8.0</td>
</tr>
<tr>
<td>35-44</td>
<td>6.4</td>
<td>7.7</td>
<td>7.5</td>
</tr>
<tr>
<td>≥ 45</td>
<td>6.2</td>
<td>6.6</td>
<td>5.1</td>
</tr>
</tbody>
</table>

(b) Females, College

<table>
<thead>
<tr>
<th>Age</th>
<th>Hispanic</th>
<th>African-American</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 25</td>
<td>0.4</td>
<td>0.0</td>
<td>1.3</td>
</tr>
<tr>
<td>25-34</td>
<td>3.0</td>
<td>2.7</td>
<td>4.0</td>
</tr>
<tr>
<td>35-44</td>
<td>2.0</td>
<td>1.8</td>
<td>3.3</td>
</tr>
<tr>
<td>≥ 45</td>
<td>1.7</td>
<td>1.0</td>
<td>1.9</td>
</tr>
</tbody>
</table>

(c) Males, Non-College

<table>
<thead>
<tr>
<th>Age</th>
<th>Hispanic</th>
<th>African-American</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 25</td>
<td>0.7</td>
<td>3.3</td>
<td>2.2</td>
</tr>
<tr>
<td>25-34</td>
<td>4.4</td>
<td>4.2</td>
<td>5.0</td>
</tr>
<tr>
<td>35-44</td>
<td>2.7</td>
<td>1.5</td>
<td>4.6</td>
</tr>
<tr>
<td>≥ 45</td>
<td>2.9</td>
<td>2.1</td>
<td>3.9</td>
</tr>
</tbody>
</table>

(d) Females, Non-College

Figure A.1: Average % Reporting Training Spell by Age, Race, Education and Sex

see before the age of 25 and restrict observations to those who become attached to the labor force, defined as three consecutive years working 30 weeks per year.
Figure A.2: EE Transition Rates by Tenure and Training Jobs

<table>
<thead>
<tr>
<th>Tenure</th>
<th>(a) &lt;HS, Trained EE Rate</th>
<th>(b) &lt; HS, Untrained EE Rate</th>
<th>(c) HS, Trained EE Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>23.2</td>
<td>45.5</td>
<td>33.3</td>
</tr>
<tr>
<td>1</td>
<td>14.9</td>
<td>25.0</td>
<td>16.5</td>
</tr>
<tr>
<td>2</td>
<td>13.8</td>
<td>24.7</td>
<td>13.1</td>
</tr>
<tr>
<td>3</td>
<td>15.5</td>
<td>21.0</td>
<td>11.6</td>
</tr>
<tr>
<td>4</td>
<td>15.5</td>
<td>18.5</td>
<td>9.6</td>
</tr>
<tr>
<td>5</td>
<td>13.4</td>
<td>17.4</td>
<td>9.0</td>
</tr>
<tr>
<td>6</td>
<td>8.7</td>
<td>15.2</td>
<td>8.8</td>
</tr>
<tr>
<td>7</td>
<td>9.3</td>
<td>10.5</td>
<td>8.0</td>
</tr>
<tr>
<td>8</td>
<td>9.3</td>
<td>11.1</td>
<td>6.9</td>
</tr>
<tr>
<td>9</td>
<td>6.6</td>
<td>8.3</td>
<td>9.0</td>
</tr>
<tr>
<td>10</td>
<td>6.4</td>
<td>7.4</td>
<td>6.1</td>
</tr>
</tbody>
</table>

(d) HS, Untrained

<table>
<thead>
<tr>
<th>Tenure</th>
<th>EE Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>49.2</td>
</tr>
<tr>
<td>1</td>
<td>26.0</td>
</tr>
<tr>
<td>2</td>
<td>24.0</td>
</tr>
<tr>
<td>3</td>
<td>21.0</td>
</tr>
<tr>
<td>4</td>
<td>18.0</td>
</tr>
<tr>
<td>5</td>
<td>16.5</td>
</tr>
<tr>
<td>6</td>
<td>11.8</td>
</tr>
<tr>
<td>7</td>
<td>12.0</td>
</tr>
<tr>
<td>8</td>
<td>9.4</td>
</tr>
<tr>
<td>9</td>
<td>10.0</td>
</tr>
<tr>
<td>10</td>
<td>8.7</td>
</tr>
</tbody>
</table>
B Properties of the Profit Function

Proposition B.1. The firm profit function \( J(V, p, h, g, \tau) \) is a continuously differentiable, concave and decreasing function in \( V \).

Proof. Because promised utility effects the sets of competitive offers it is not immediately clear that the continuation value is concave in \( V \). I first establish this in a lemma and then show that the resulting problem is also concave. Define

\[
\tilde{J}(W, p, h, g) = \max_{V} E[J(V, p, h, g, \tau)] \quad \text{s.t.} \quad E[V] \geq W
\]

If \( J \) is concave in \( V \) then \( \tilde{J} \) is concave \( W \). This will also coincide with the contract promised utilities \( \{V\} \) so we can write the continuation value as

\[
K(W, p, h, g) = \left\{ G(\Omega^0(W, g))\tilde{J}(W, p, h, g) + \int_{\Omega^1(V, p, h, g)} \tilde{J}(\hat{V}(s, V, p, h, g), p, h, g)dG(s) \right\}
\]  

(B.1)

Lemma B.2. The function \( K(\cdot) \) is concave in \( V \) if \( J(\cdot) \) is concave in \( V \).

Proof. To show this, first note that \( \Omega^0(W, g) = [p_{\min}, q_0(W, g)] \) is increasing in \( W \). Dropping dependence on the other state variables, we want to show that \( K(V_\alpha) \geq \alpha K(W_1) + (1 - \alpha)K(W_2) \) for \( W_1 < W_2 \) and where \( W_\alpha = \alpha W_1 + (1 - \alpha)W_2 \). Write \( \Omega^1(W, p, h, g) = [q^0(W, g), q^1(p, h, g)] \) and notice that the upper bound of the support is independent of promised utility. We can then write (suppressing dependence of \( K, J \) on all except \( V \) )

\[
K(W) = G(q^0(W, g))\tilde{J}(W) + \int_{q^0(W, g)}^{q^1(p, h, g)} \tilde{J}(\hat{V}(x, p, h, g))dG(x)
\]  

(B.2)

Rewrite

\[
K(W_\alpha) = G(q^0(W_1, g))\tilde{J}(W_\alpha) + [G(q^0(W_\alpha, g)) - G(q^0(W_1, g))]\tilde{J}(W_\alpha) + \int_{q^0(W_\alpha, g)}^{q^1(p, h, g)} \tilde{J}(\hat{V}(x, p, h, g))dG(x)
\]  

(B.3)
\[ \alpha K(W_1) + (1 - \alpha)K(W_2) = G(q^0(W_1, g)) [\alpha \tilde{J}(W_1) + (1 - \alpha)\tilde{J}(W_2)] \]
\[ + [G(q^0(W_2, g)) - G(q^0(W_1, g))] (1 - \alpha) \tilde{J}(W_2) \]
\[ + \int_{q^0(W_2, g)} q^0(p, h, g) \tilde{J}(\hat{V}(x, p, h, g)) dG(x) + \alpha \int_{q^0(W_1, g)} q^0(W_2, g) \tilde{J}(\hat{V}(x, p, h, g)) dG(x) \]
(B.4)

Setting the inequality, we first use the assumption that \( J \) is concave in \( V \) to show
\[ G(q^0(W_1, g)) [\alpha \tilde{J}(W_1) + (1 - \alpha)\tilde{J}(W_2)] \leq G(q^0(W_1, g)) \tilde{J}(W_1). \]
We then subtract out the shared component of integration and are left with the claim
\[ [G(q^0(W_\alpha, g)) - G(q^0(W_1, g))] \tilde{J}(V_\alpha) + \int_{q^0(W_\alpha, g)} q^0(W_2, g) \tilde{J}(\hat{V}(x, p, h, g)) dG(x) \]
\[ \geq [G(q^0(W_2, g)) - G(q^0(W_1, g))] (1 - \alpha) \tilde{J}(W_2) + \alpha \int_{q^0(W_1, g)} q^0(W_2, g) \tilde{J}(\hat{V}(x, p, h, g)) dG(x) \]
(B.5)

Notice however that we can bound the RHS above by
\[ [G(q^0(W_\alpha, g)) - G(q^0(W_1, g))] [(1 - \alpha)\tilde{J}(W_2) + \alpha \tilde{J}(W_1)] + [G(q^0(W_2, g)) - G(q^0(W_1, g))] (1 - \alpha)\tilde{J}(W_2) \]
\[ + \alpha \int_{q^0(W_\alpha, g)} q^0(W_2, g) \tilde{J}(\hat{V}(x, p, h, g)) dG(x) \]
(B.6)

Applying the concavity of \( \tilde{J} \) to this we again have
\[ [G(q^0(W_\alpha, g)) - G(q^0(W_1, g))] \tilde{J}(W_\alpha) \geq [G(q^0(W_\alpha, g)) - G(q^0(W_1, g))] [(1 - \alpha)\tilde{J}(W_2) + \alpha \tilde{J}(W_1)]. \]
Subtracting out the integrals, this leaves the claim
\[ (1 - \alpha) \int_{q^0(W_\alpha, g)} q^0(W_2, g) \tilde{J}(\hat{V}(x, p, h, g)) dG(x) \geq [G(q^0(W_2, g)) - G(q^0(W_1, g))] (1 - \alpha)\tilde{J}(W_2) \]
(B.7)

This is clearly true since on the set \([q^0(W_\alpha, g), q^0(W_2, g)]\) all bargained values \( \hat{V}(\cdot) \) lie at or below \( V_2 \) and \( J \) is decreasing in promised utility. Thus, referring to Eqn (13) as BOUND,
we have shown \( \text{RHS}_{eqn12} \geq \text{BOUND} \geq \text{LHS}_{eqn12} \) and have thus established concavity of continuation value \( K \) in promised utility \( V \).

**Lemma B.3.** \( J(V, p, h, g, \tau) \) is concave and decreasing in \( V \).

*Proof.* Here the proof is straightforward. Consider separately the choice to train or not. First note that concavity of the objective function is obvious given the proof of \( K \). Hence second order conditions are sufficient. The envelope condition tells us that the derivative of \( J \) is equal to one over MU. This immediately yields \( J \) as decreasing. For contradiction, suppose the second derivative is not negative. Then \( V_1 < V_2 \) implies \( w_1 > w_2 \) (using envelope). Then \( V_1' < V_2' \) by PK. However the FOC for \( V' \) simplifies to \( G(q^0(W))[\hat{J}'(W) + \lambda_j] = 0 \). \( \lambda_1 > \lambda_2 \) by FOC of wages. Since \( G(\cdot) > 0 \) for both \( W_1, W_2 \) this implies \( \hat{K}'(W_1) < \hat{K}'(W_2) \). But given that \( \hat{K} \) is concave and decreasing, this requires \( W_1 > W_2 \), a contradiction. Hence \( J \) is concave in \( V \).

Including the training choice is just a two-point lottery, concave by a straightforward argument. See for example Appendix F in Menzio and Shi (2009). Clearly each function is decreasing in \( V \). [TBC]

**Lemma B.4.** \( J(V, p, h, g, \tau) \) is continuously differentiable.

*Proof.* Almost everywhere differentiable is satisfied by concavity. Everywhere differentiable is established using the standard arguments in Benveniste and Scheinkman. [TBC]

**Lemma B.5.** Firm profits \( J \) are increasing in \( p \) and \( h \), as is maximum willingness to pay \( V^*(p, h, g) \).

*Proof.* Consider a match \((p, h, g)\) with promised utility \( V \). We will rewrite contracts in the following sequential form. Define a history \( s^t = (s_t, s_{t-1}, \ldots, s_0) \) where \( h_t = (\tau_t, p'_t) \) as the history of exogenous training and meeting draws. For any \( V \), contracts can be defined as a sequence of functions \( \hat{C} = \{w_t, \delta_t, h_t, g_t, D_t\} \) which are all functions mapping the set of possible histories \( S^t \) to their respective spaces and describes for every possible history the
values and decisions taken. The function \( D_t \) denotes whether separation has occurred after some history \( s^t \).

We can then write the value to the firm of some contract promising \( V \) as the discounted sum of the expected value of period returns at every possible point in future histories.

\[
\sum_{t=0}^{\infty} \sum_{s^t} \pi(s^t) D(s^t) \beta^t \left[ f(p, k[h_t(s^t), g_t(s^t)]) - w_t(s^t) - c_t(s^t) \right]
\] (B.8)

Assume we have a contract of this form that optimally delivers \( V \) in a match \((p, h, g)\). Now consider a firm \( p' \) and suppose they offer the exact same contract. Profits are strictly higher in every period. In particular, expected continuation values at any point are strictly higher, hence the firm \( p' \) would always want \( D(s^t) = 1 \) if the contract for firm \( p \) stipulates this. The only cases in which \( \hat{C} \) may not be feasible is where the firm \( p' \) would prefer the match continues. Since this only occurs when profits are positive, the optimal contract for firm \( p' \) can do at least as well, hence profits are strictly increasing.

Recall that the outside offer a firm \( p' \) will make given a worker \((h, g)\) is the value at which a match yields zero expected profits, \( V^*(p', 0, g) \). This is independent of \( h \). Profits are thus higher for every continuation of the match.

A corollary of these two statements is that \( V^*(p, h, g) \) is increasing in each of \( p, h \). This is obvious through the definition of maximum willingness to pay. \( \square \)

## C Contract Characterization

**Proposition C.1.** The contract defined in \((P1)\) is unique and has the properties that:

1. Wages are constant unless renegotiation is triggered
2. Training is increasing in firm type \( p \)
3. Wages decrease in \( p \)

**Proof.** Letting \( \lambda \) be the multiplier on the first PK constraining and \( \mu \) on the second, the first order conditions characterizing optimality are:
• \( \lambda = 1/u'(w) \)

• \(-\beta r(W,g) J_v(V_{i\tau}, p, h, g) = \mu\)

• \(\beta \lambda r(W,g) = \mu\)

• \(J_v(V, p, h, g) = \lambda \) [Envelope]

I have already simplified a few expressions using definitions.

1. Flat wages: The result is standard in contracting literature when match continuation does not depend on agent choices. It can be seen directly from the first order condition for continuation values and the envelope condition, using the fact that wages in any period will be independent of training. The FOC for continuation values implies \( J_v(V_{i\tau}, p, h, g) = -\lambda \). Then using the envelope condition and FOC for wages, \( \frac{1}{u'(w)} = J_v(V_{i\tau}) = \frac{1}{u'(w')}. \) By the assumption of strictly concave utility, this implies that \( w = w' \). Notice that this occurs only if \( V_{i\tau} \) is the value of promised utility tomorrow, implying that no renegotiation occurred at the end of the period.

2. Training: First note that the FOC implies training happens whenever positive profits are earned (in the tradeoff between output and provision of promised utility). Suppose a contract for firm type \( p \) implies training under some history \( s^t \). We can make a similar argument as in B.5. For any firm \( p' \), if we implement the exact same contract, continuation values are strictly higher. The only histories for which this contract will not be implementable is when firm \( p' \) would want to continue the match while the contract requires the worker leave. But this is only because the match earns weakly positive profits for the firm. Since this augmented contract yields higher profits to training for firm \( p' \), the optimal contract can do no worse. But then it must be the case that the optimal contract for \( p' \) stipulates training, since positive profits are earned.

3. Wages fall in \( p \): Continuation values are increasing in both \( p \) and training (both \( (h, g) \)), so by promise keeping, for any level of promised utility \( V \), wages fall.
Proposition C.2.  1. Separation probability falls in training i.f.f. $\gamma < 1$

2. $\partial \text{Sep}/\partial \tau$ is falling in $g$ (and past training).

Proof.  1. Separation occurs only when $V^*(\bar{p}, 0, g) > V^*(p, 0, g)$. By the proposition above, $V^*$ is increasing in $p$ and $h$. If $\gamma = 1$ then $h \equiv 0$ always and separation is determined only by $\bar{p} \geq p$. If $\gamma < 1$ then training implies both $h, g$ increase. The $h$ raises the incumbent willingness to pay while leaving unchanged outside willingness to pay (which responds only to $g$).

2. Take the argument for $V^*$ increasing in $h$ above. Notice that the gain implied depends on the marginal increase in productivity to $h$. With decreasing returns and effective human capital linear in $h$ and $g$, this yields the change in marginal willingness to pay for an increase in $h$ lower as $g$ increases, which is equivalent to past training.
D Recursive Contracts

Consider the following problem. For simplicity I assume no exogenous separations and workers meet with a firm with probability one each period.

\[ J(V, p, h, g, \tau) = \max_{\delta_i, w_i, W_i, V_{i,\tau'}} \sum_i \delta_i \left\{ f(p, k(h_i', g_i')) - w_i - c(\delta_i \tau) + \beta \left[ r(W_i, g_i') E_{\tau'} [J(V_{i,\tau'}, p, h_i', g_i', \tau')] + \tilde{J}(W_i, p, h_i', g_i') \right] \right\} \]

s.t.
\[ \sum_i \delta_i \left\{ u(w_i) + \beta \left[ r(W_i, g_i') W_i + \tilde{W}(W_i, p, h_i', g_i') \right] \right\} \geq V \quad (P1) \]
\[ \sum_i \delta_i \{E_{\tau'}[V_{i,\tau'}] - W_i \} \geq 0 \]
\[ \sum_i \delta_i = 1 \quad \forall z, \tau \]
\[ h_{nt}(z, \tau) = h, \quad g_{nt}(z, \tau) = g \]
\[ h_t(z, \tau) = (1 - \gamma)s(\tau, p) + h, \quad g_t(\tau) = \gamma s(\tau, p) \]

We define \( V^*(p, h, g) \) implicitly such that \( \tilde{J}(V^*(p, h, g), p, h, g) = 0 \)

To simplify the computation I use the approach pioneered by Marcet and Marimon among others. We can write the Pareto problem as follows

\[ P(\mu, p, h, g, \tau) = \sup_V J(V, p, h, g, \tau) + \mu V \quad (D.1) \]
Substituting in the problem above and the promise keeping constraining for $V$ we obtain

\[
P(\mu, p, h, g, \tau) = \max_{\delta_i, w_i, W_i, V'_{i', \tau'}} \sum_i \delta_i \left\{ f (p, k(h'_i, g'_i)) - w_i - c(\delta_i \tau) + \beta \left[ r(W_i, g'_i) E_{\tau'}[J(V'_{i', p, h'_i, g'_i, \tau'})] + \hat{J} (W_i, p, h'_i, g'_i) \right] \right. \\
\left. + \mu \left[ \sum_i \delta_i \left\{ u(w_i) + \beta \left[ r(W_i, g'_i) W_i + \hat{W}(W_i, p, h'_i, g'_i) \right] \right\} \right. \right. \\
\left. \left. \text{s.t.} \right. \right. \\
E_{\tau'}[V'_{i', \tau'}] \geq W_i \\
\sum_i \delta_i = 1 \ \forall z, \tau \\
h_{nl}(z, \tau) = h, \ g_{nl}(z, \tau) = g \\
h_i(z, \tau) = (1 - \gamma)s(\tau, p) + h, \ g_i(\tau) = \gamma s(\tau, p) \\
\]

Subbing in the restrictions on $\delta_i$ and appending $\beta r(W_i, g'_i)$ to $\lambda_i$ we obtain

\[
P(\mu, p, h, g, \tau) = \inf_{\lambda} \sup_{\delta_i, w_i, W_i, V'_{i', \tau'}} \sum_i \delta_i \left\{ f (p, k(h'_i, g'_i)) - w_i - c(\delta_i \tau) + \beta \left[ r(W_i, g'_i) E_{\tau'}[J(V'_{i', p, h'_i, g'_i, \tau'})] + \hat{J} (W_i, p, h'_i, g'_i) \right] \right. \\
\left. + \mu \left[ \sum_i \delta_i \left\{ u(w_i) + \beta \left[ r(W_i, g'_i) W_i + \hat{W}(W_i, p, h'_i, g'_i) \right] \right\} \right. \right. \\
\left. \left. + \beta r(W_i, g'_i) \lambda_i \left[ E_{\tau'}[V'_{i', \tau'}] - W_i \right] \right\} \]

Under the assumption that randomization is not needed (numerically I have verified this
to be the case under my parameterizations), we can rearrange this as the following

\[ P(\mu, p, h, g, \tau) = \]

\[ \inf_{\lambda} \sup_{\delta, w_i, W_i, V'_{i,\tau}} \sum_i \delta_i \left\{ f(p, k(h_i', g_i')) - w_i - c(\delta_i \tau) + \beta \hat{J}(W_i, p, h_i', g_i') + \mu \left[ u(w_i) + \beta \left( r(W_i, g_i')W_i + \hat{W}(W_i, p, h_i', g_i') \right) \right] - \beta r(W_i, g_i') \lambda_i W_i + r(W_i, g_i') \beta E_{\tau'}[P(\lambda_i, p, h_i', g_i')] \right\} \]

Where we can re-express the last term and move the choices to obtain

\[ P(\mu, p, h, g, \tau) = \]

\[ \inf_{\lambda} \sup_{\delta, w_i, W_i} \sum_i \delta_i \left\{ f(p, k(h_i', g_i')) - w_i - c(\delta_i \tau) + \beta \hat{J}(W_i, p, h_i', g_i') + \mu \left[ u(w_i) + \beta \left( r(W_i, g_i')W_i + \hat{W}(W_i, p, h_i', g_i') \right) \right] - \beta r(W_i, g_i') \lambda_i W_i + r(W_i, g_i') \beta E_{\tau'}[P(\lambda_i, p, h_i', g_i')] \right\} \]

E Identification and Heterogeneity

In this section I consider a two period version of a simplified model to make a clearer argument for the identification strategy. I will also explicitly discuss how my model will interpret additional heterogeneity in the data.

E.1 Two Period Identification

For this example, I assume workers have linear preferences, \( \lambda_c = 1 \) (all workers receive offers) and wages are constant unless renegotiation occurs. I will also change the timing of training so that it rewards production only in the following period (so that we may assume no training occurs in period two). Let \( U \) be the outside option of workers in this environment. Starting
with period two wages, they will be

$$\min \{ p + (h + g)^{\xi_h}, p' + g^{\xi_h} \}$$

(E.1)

I now make an argument for joint identification in this simplified world. Firm training for non-trained workers equates marginal returns with costs (in one period this would be $\xi_h(\tau p^{\xi_p})^{\xi_h-1} p^{\xi_p} = c'(\tau)$). Given $\gamma, \xi_h, \xi_p, G(p), \lambda_r$, the cost of training is pinned down by the conditional mean of training received. Similarly, holding costs and all else fixed, $\lambda_r$ determines the probability of training being received.

Given $G(p), \gamma, \xi_h$ and the cost function $c(\cdot)$, the conditional variance of training received pins down $\xi_p$.

Given our knowledge of $G(p), \text{cost function } c(\cdot)$ and $\xi_p$, the last two parameters are then jointly identified in the following manner. Let $T = h + g$.

Turnover at the end of period 1 will occur i.f.f a worker in match $(p, h, g)$ receives an offer $p'$ such that

$$p' + (\gamma T)^{\xi_h} \geq p + T^{\xi_h}$$

(E.2)

For any $\xi_h, \gamma$ we can compute the joint distribution at period one of $\mu_1(p, T)$. Given this distribution, we want the implied average separation rate for trained workers to be

$$\text{Sep}(p, T; \gamma, \xi_h) = \int [1 - G(\tilde{p} + (1 - \gamma^{\xi_h})\tilde{T}^{\xi_h})]d\mu_1(\tilde{p}, \tilde{T}|t) = \text{Sep}(p, T)^{data}$$

(E.3)

Notice that both lower $\gamma$ and higher $\xi_h$ will reduce turnover rates of trained workers. To separate the two we need a second equation where the two forces work in the same direction. Consider returns in period two to training on the job from period one. Wages will rise for
an outside offer $p'$ according to

$$w_{ijt+1}(p', p, T) = \begin{cases} w_{ijt}(p) & \text{if } p' \in \Omega^0 \\ p' + (\gamma T)\xi_h & \text{if } p' \in \Omega^1 \\ p + T\xi_h & \text{if } p' \in \Omega^2 \end{cases} \quad (E.4)$$

since the incumbent firm will match the willingness to pay of the outside firm (and no training occurs in period 2). Initial wages must satisfy

$$u(w(p)) + \beta E[w_{ijt+1}(p', T)|p] = U$$

Since $w_{ijt+1}(p', T)$ is increasing in $\gamma$ and $\xi_h$ then for a fixed $U$, $w(p)$ must be decreasing in both. Jointly then, average returns are increasing in both parameters. Given data on average wage growth for trained workers, we then set

$$W_\Delta(t; \gamma, \xi_h) = \frac{\int w_{ijt+1}(p', \tilde{p}, \tilde{T})d\mu_1(p', \tilde{p}, \tilde{T})}{w(p)} = W^\text{data}_\Delta(t) \quad (E.5)$$

I do not have a formal proof of existence in this environment, but I can state the following:

**Lemma E.1.** For a given separation rate and wage growth for trained workers, if there exists a $\xi_h^*, \gamma^*$ such that both equations (E.3) and (E.1) hold, this pair is unique.

**Proof.** Sketch [TBC]. Returns on the job depend on $V^*(p, 0, g)$, which is increasing in $\xi_h$ and decreasing in $\gamma$. We will denote the function that describes the loci satisfying the value of returns as $\xi^R_h(\gamma)$, and we have $\partial \xi^R_h / \partial \gamma < 0$. Further, we can show that the parallel function for the separation rate is increasing $\partial \xi^S_h / \partial \gamma > 0$. Suppose $\gamma^*$ is the smallest value such that these two functions coincide in the $[\gamma, \xi_h]$ space. Note that because the derivatives of these two functions are of strictly opposing signs, we easily arrive at a contradiction if we assume they will coincide again. Thus this $\gamma^*$ is unique. \qed

43
E.2 Additional Heterogeneity

E.2.1 Productivity

In the model environment I develop, I ignore several dimensions of heterogeneity. To the extent that I can, I address the identification issues raised by these simplifications here. In particular, I will consider augmenting the model with three sources of productivity: \( \hat{p} = (p_j, z_i, \phi_{ij}) \); firm, worker and match productivities, respectively. If we assume that in both \( f(\hat{p}, h, g) \) and training \( s(\tau, \hat{p}) \) firm and match productivities are perfect substitutes, then for any joint distribution \( \Pi(p, \phi) \) we can construct an alternate distribution for \( \tilde{p} = p + \phi \). Under this assumption then, while I will not be able to separately identify match productivities, it will not effect estimates of either returns to training or turnover.

Worker heterogeneity \( z_i \) is only partially dealt with by estimating the model separately for different education groups. I use wage residuals after controlling for other observable heterogeneity to correct for this bias in the wage data. Nonetheless, worker type may still effect turnover through a firm’s willingness to pay. Below however I show that under the functional form assumptions this is not the case.

Lemma E.2. \( V^*(\hat{p}, h, g) \) is separable in \( z \)

Proof. Sketch [TBC]. Guess and verify. \( J^0(V, p, z, h, g) \) is assumed to be separable and \( f() \) is linear as in the functional form assumptions. Optimal wages, and PU are determined by derivatives of \( V \) and PK, neither of which depend on \( z \). Training decision depends on a max operator from which the fixed value for \( z \) can be subtracted. Thus no decisions depend on \( z \) and the value function can be separated into a component from production, \( z \), and the rest.

The implication of the above lemma is that worker type will not interact with turnover decisions, and so it will not bias estimates even if it is no accounted for. This relies on the assumption in particular that worker type does not effect returns to training beyond the average effect by education. Under these assumptions, however, I can restrict my attention to \( p \) alone.
E.2.2 Mobility

One particular form of unobserved heterogeneity that is relevant for this problem and that I do not account for in the model is individual or firm specific heterogeneity in exogenous match survival probabilities (Jarosch (2015) explores the implications of the latter in a search model of the labor market). For example, we might be concerned that if some individuals simply prefer to move less between jobs then these are the ones who will be trained. In this case we would be imputing to firm productivity the training that was in fact due to an individual specific trait.

To investigate heterogeneity on the worker side, I ask if there are observable differences among those workers who are eventually trained and those who are not prior to observing their training (I consider only observations prior to the first training spell since job mobility afterwards may be effected by training on prior jobs). First, I define an indicator variable equal to one if an individual is observed being trained at any point in the sample. I then run a logistic regression on the choice of separation conditional on wages, education, age, tenure and the indicator. It is instructive to compare results of this logistic regression across the inclusion or exclusion of the jobs in which workers are first trained (not all training happens immediately on a job). As can be seen in Table XX (note that coefficients on hourly pay, education, age and tenure are not reported for brevity), when the jobs in which training is first observed are included the effects of being a trained type imply a reduction in separation probability. When excluded, the trained type indicator implies instead an increase in the turnover rate. One way of interpreting this is the following. When training jobs are included, the indicator is simply picking up that many of individuals are in jobs in which they will eventually be trained; we know these have lower separation rates. When these observations are excluded, we find that among those who have not yet been trained, knowledge that they will be trained in future jobs implies increased turnover. This implies that it requires a different job for training to be obtained. In other words it is match or firm characteristics that are most important for training rather than some unobserved individual traits.

In considering firm heterogeneity in match survival, the appropriate parallalel analysis
would require panel data on firms or a matched employer-employee set, which I do not have. If these shocks are unforseeable however, we might suspect that they would play a role in transitions to unemployment (UE). A comparison of these transition rates by tenure across jobs that train and do not show that early in the match non-training jobs have higher rates of UE transition. These fall over time, however, suggesting that the annual aggregation may be misidentifying some EE transitions as UE. A more constructive comparison is to compare transition rates when they have become approximately constant (after about ten years on the job). A comparison of EU transition rates across groups at these levels of tenure shows they are practically the same. I have not reported the tables here for brevity.

E.2.3 Amenities

The last source of heterogeneity I will address is amenities; i.e. non-wage utility that accrues to the worker from a particular job. Taber (DATE) examines this in detail. We might be concerned that training is related to unobserved amenities that lead to lower separation rates (but do not effect productivity). To see the implications of amenities, consider the two period example augmented with firm or match amenity $\nu$. A worker now separates if:

$$p' + (\gamma T)^{\xi_h} + \nu' \geq p + T^{\xi_h} + \nu$$

(E.6)

where linear preferences make utility and productivity enter as equivalents. In particular, there may be pair $(p', \nu')$ such that separation occurs while $p' < p$ (and $\nu' > \nu$). That is, a worker may accept a wage loss in exchange for the amenities provided by the firm. If amenities make training more likely by reducing turnover, then we may be concerned that observed separation rates are misidentifying the effects of training.

If this were the case, we would expect training jobs to have higher amenities on average,
which would also imply that we would expect more wage losses to be accepted for EE transitions out of non-training jobs as compared to training jobs. Checking this implication empirically, I find that trained workers are on average more likely to accept a job loss than untrained workers (43% to 37%), which is a significant difference under the assumptions of a standard t-test. We should be careful however because losses are more likely to occur when separations happen at higher tenure. Since trained jobs have on average higher tenure, we may be conflating two effects. As a simple control for this, I consider only EE transitions at five years or more of tenure. The trained group is still more likely to accept a loss although by only a few percent, and this difference is no longer significant.

Examining wage changes entering into training jobs, I find that workers entering a training job are less likely to accept a wage loss than workers entering non-training jobs. Using a five year cutoff for the same reasons as above, this difference remains significant (about 8%), and is the opposite of what one would expect if amenities played an important role. Jointly I take this as evidence that amenities are not a significant factor in determining training and as such will not bias results.

References


