Labor Market Imperfect Information, On-the-job Training and the Employer Size-Wage Effect

Shuaizhang Feng* Bingyong Zheng†

March 30, 2008

Abstract

Using NLSY79 data, we find that large firms pay more to workers and train a higher proportion of their workforce. In addition, wage premiums associated with large employers are lower for trained workers than for untrained ones. Existing theories can not explain these empirical findings simultaneously. We then develop a two-period model of imperfect information that can reconcile all three stylized facts.

Keywords: size-wage premium, imperfect information, On-the-job Training.

JEL classification codes: D83; J31.

1 Introduction

Large firms pay more on average to workers of similar characteristics than small firms. Studies have shown that the wage gain associated with working in a large firm or establishment remains statistically significant and practically large even after controlling for unobserved worker heterogeneity,[1] and is remarkably stable over time and across countries with different labor market institutions. To put the magnitude of the size-wage premium into perspective, we quote Brown and Medoff (1989), "if a typical worker went from an establishment with employment one standard deviation below average to an establishment with employment one standard deviation above average, the employee would enjoy a wage increase of 8-12 percent, about as large as the union-nonunion differential in these data."

*School of Economics, Shanghai University of Finance and Economics, 777 Guoding Rd., Shanghai, China 200433 Tel: 86-21-6590-3487. E-mail: shuaizhang.feng@gmail.com.

†School of Economics, Shanghai University of Finance and Economics, 777 Guoding Rd., Shanghai, China 200433 Tel: 86-21-6590-3124. E-mail: bingyongzheng@yahoo.com.

[1]Moore (1911) first documented this phenomenon for workers at Italian textile mills. For studies in U.S., see Lester (1967), Personick and Barsky (1982), Mellow (1982), Brown and Medoff (1989), Brown et al. (1990),Idson and Feaster (1990), Troske (1999). For studies in U.K. and other European countries, see Main and Reilly (1993), Green et al. (1996), Abowd et al. (1999), Albæ k et al. (1998), Winter-Ebmer and Zweimüller (1999). Similar findings are also reported in developing countries, see e.g. Velenchik (1997), Schaffner (1998), and Söderbom et al. (2005).

The existence of the size-wage premium is puzzling and hard to rationalize within the perfect competition paradigm (see e.g. Brown and Medoff, 1989). While some studies incorporate labor market imperfections to explain the size-wage effect, as Burdett and Mortensen (1998) do with job search frictions, others challenge this empirical finding directly, noting that the observed wage differential is due to inadequate control of workers' productivities. For example, Troske (1999) and Zabojnik and Bernhardt (2001) suggest workers in large firms receive more training, thus are paid more as they are more productive.

In this paper, we first re-examine the size-wage premium phenomenon with the 1979 Youth Cohort of National Longitudinal Studies (NLSY79), specifically controlling for workers' training status. Consistent with earlier studies, we find that large firms pay more than small firms, even for untrained workers, and they also train a higher proportion of their workforce. Thus, we are able to rule out the training hypotheses as explanations for the size-wage effect. In addition, we also find that wage differential for trained workers is smaller than for untrained workers. Brown and Medoff (1989, page 1089) have also noted that "A striking regularity among the professional, technical, and managerial workers is the tendency for the wage differential to decline with increasing skill levels." No existing theories, including the random search model of Burdett and Mortensen, can explain this piece of empirical evidence.

We then develop a two-period model of imperfect information that reconciles all three stylized facts. In this model, firms are identical ex ante, while workers are either of high type, with high productivity, or of low type, with low productivity. Firms post wages to attract workers they want to hire, and workers accept offers with the highest wage. At the beginning of the first period, neither firms nor workers know the type of any worker. However, when a worker applies for jobs, firms each observe a private signal that imperfectly indicates her type. In equilibrium, firms pay workers their expected productivities given the privately observed signals. Firms that post higher wages will hire more workers that are also more productive on average. Thus, workers of the same type are paid differently in different-sized firms, simply because they are pooled with workers of different average productivities. In the second period, firms receive another signal that imperfectly indicate workers' performance in the first period. Given the new information, incumbent firms make training decisions and post different wages for trained and untrained workers. In equilibrium, firms provide training only to those from whom they receive a good signal, but pay them according to pre-training expected productivities.

In identifying information imperfection as the source of size-wage premium, this paper falls into a large literature on labor market information that dates back to Stigler (1962) (see also Spence, 1973 and Gibbons and Katz, 1992). This is different from the existing theories of wage dispersion that either assume workers have no knowledge about particular jobs when applying, as in random search models, or restrict workers to apply to only a finite number of jobs at a time, as in the directed search / matching models.[2] According to our model, unobservable worker heterogeneity can explain a sizable portion of the observed size-wage premium, as reported in most empirical studies.[3] The remaining part of wage differentials that cannot be explained by unobservables is due to the pooling effect.

The rest of the paper is structured as follows. Section 2 presents our empirical findings using the 1979 cohort of National Longitudinal Studies of Youth (NLSY79) data. We also briefly review some existing theoretical models but find none of them could explain our findings simultaneously. In Sections 3 and 4, we present a two-period labor market model of imperfect information and discuss its implications. Section 5 concludes.

2 Empirical Regularities from NLSY79

We base our empirical work on NLSY79 for three reasons. First, it is a longitudinal data set that contains detailed individual characteristics, including variables that are usually not available in comparable data sources, e.g., the AFQT (Armed Forces Qualification Test) score that indicates pre-market human capital (see Neal and Johnson, 1996). Second, as sample individuals were young when the survey started, we can examine size-wage relationships for labor market entrants. This turns out to be important for the purpose of differentiating some alternative hypotheses. Last, NLSY79 contains very detailed on-the-job training information, which allow us to compare size-wage premiums for trained and untrained workers. In fact, no previous study has examined returns to training, i.e., wage differences between trained workers and untrained ones, for large and small firms separately.[4]

2.1 NLSY79 Data

NLSY79 contains detailed information about training and has been used extensively in the empirical literature. It contains 12,686 individuals aged 14-21 in 1978. The respondents were interviewed annually from 1979 to 1994, and every two years since then. It consists of a nationally representative cross-sectional sample, a supplemental sample, and a military sample. In this paper, we restrict all analyzes to the cross-sectional sample, and focus on the period 1986-2000 since no establishment size (or firm size) information is available in the 1981-1985 surveys. We do not use post-2002 data because some survey questions have been restructured.

[2] See Moen (1997), Shi (2002), Shimer (2005) and Albrecht et al. (2006).

[3] See Brown and Medoff (1989) and other studies citied before.

[4] Most recent studies on return to training are based on NLSY79, see Lynch (1992), Veum (1995), and Frazis and Lowenstein (2005).

Despite its richness, training information has not been collected consistently in NLSY79 over time. The 1979-1986 surveys records only up to three formal training spells enrolled since last interview (thereafter called "current training spells") and up to two training spells that was still ongoing at last interview (thereafter called "previous training spells"). This was followed by a year of absence of training information in 1987. In the 1988-2002 surveys, up to four current and three previous training spells are recorded. Supplemental questions, such as who paid for the training and the usefulness of training programs, were only asked in some of these surveys.

For the present analysis, we try to consistently use the following information for each training spell: type of training, starting date, ending date, a dummy on whether the spell is censored and total training hours. The type of training information is used to restrict the analysis to on-the-job training only. A training spell is classified as on-the-job training if it is "company training (type=8)" during the 1979-1986 surveys, "formal company training run by employer (type=8)" or "training programs at work not run by employer (type=9)" during the 1984-2000 surveys.

For a current training spell still going on at the date of interview, we update the information from the next survey. For example, to update the information of a training spell still going on at the date of interview in the 1988 survey, we find the "previous training spell" information from the 1989 survey. If the types of the two spells match, we then update the ending date, and other information accordingly. As NLSY79 does not provide identifiers that could link training spells across surveys unambiguously, we only update the information once and do not go further even if the training spell was still going on after updating. In that case, the interview date is recorded as the pseudo-ending date and a dummy is used to record the spell as censored.

We use a training dummy (training incidence) instead of actual training hours in this paper. As noted in Barron et al. (1997), training incidence contains much less measurement errors than training hours. Moreover, similar to Veum (1995), our preliminary regressions also suggest that in most cases, after controlling for training incidence, number of training hours is no longer statistically significant.

Establishment size information are taken from the question "number of employees at location of current job" or "number of employees at location at respondent's job number 1". This information were recorded in all surveys except 1981-1985. The survey also asked for number of employees at other locations for the same years. However, this is only recorded as whether over 1000 employees or less than 1000 employees. Given that the firm size information is more likely to be problematic, only establishment size information is used in the analyzes. Brown and Medoff (1989) also found that the partial effect of establishment size on wage is larger than the firm effect.

The analysis uses current job (or CPS job, job number 1) information for the period 1986-2000. We first transform nominal hourly wages into real wages using CPI-U (1982-1984=100), then exclude values that are greater than 100 dollars or less than 1 dollar. For each current job hourly wage, the following information are recorded: year of tenure, total labor market experience, union status on the job, industry, part-time status, establishment size, a dummy whether completed any on-the-job training spell on the current job, total accumulated on-the-job training hours. Variables such as gender, race, education, and AFQT score are also included.

To make sure that we are analyzing a homogeneous group of workers, we restrict our sample to white males that worked full–time (defined as 35 hours work per week and above) at non-union jobs in non-agricultural sectors. Table 1 describes the variables used in this paper.

2.2 Employer size-wage effect

Many studies have shown that large firms pay more on average to their workers, even after controlling for a set of variables that reflect productivity. However, employer size-wage effect has not been uniform across all industries of the economy. For example, Lallemand et al. (2005) show that the effect is generally much stronger in manufacturing industries than in service industries for European countries. Although not extensively documented in the literature, similar patterns are also true in U.S.. Table 2 gives industry level firm size-wage relationship based on the 2002 data provided by U.S. Small Business Administration (SBA). The data show that for retail trade and a couple of service industries, larger firms do not necessarily offer higher wages, even though no controls have been included at all.

With NLSY79 data, the first empirical test is to see whether we can observe size-wage effect. Following SBA data, we exclude two industries from all analyzes: retail trade and professional and other services. Sample summary statistics are given in Table 3. Note that the unit of observation is a person-year. The sample mean of natural log of hourly wage is 6.77, which corresponds to a real wage rate of \$8.74. In terms of education, 56% of the sample has a high school diploma or less, while 23% of the sample has at least a Bachelor's degree. On average, workers have 4 years of tenure with current employers and 11 years of total labor market experience. 83% of the sample work in small establishments, while 17% work in large ones, where large establishments is defined as those with at least 500 employees. Table 4 gives sample means for small and large establishments, respectively. Average wage rate for large establishments is substantially higher. Those who work for large establishments earn \$11.6 per hour on average, more than 40% higher than those in small establishments. On the other hand, workers in large establishments also tend to be more productive, with more education, higher AFQT scores, longer tenure with current employers and more on-the-job and off-the-job training.

The following ordinary least squares (OLS) model controls for observable characteristics that affect productivities and wages.

$$Y_{it} = \gamma L\_EST_{it} + X'_{it}\beta + \epsilon_{it} \tag{1}$$

In equation (1), for worker i, Y_{it} is the log hourly wage in year t, L_EST_{it} is a dummy variable for large establishment, X_{it} is the vector of other explanatory variables, including AFQT, schooling dummies, marriage dummies, regional dummies, local unemployment rate, tenure with current job, total labor market experience, survey year dummies and industry dummies.

Equation (2) is a panel data model, with C_i being the unobserved person effect that does not change over time.

$$Y_{it} = \gamma L\_est_{it} + X'_{it}\beta + C_i + \epsilon_{it} \tag{2}$$

The model is estimated using both random effects and fixed effects approaches. Table 5 reports estimation results. We see that as expected, more schooling is associated with higher earnings. AFQT, tenure and labor market experience are all positively related to hourly wage rate, while workers in areas of high unemployment are paid less.

Based on the OLS result, workers in large establishments earn 14.6% more than those in small ones. When person effects are controlled for, the estimated coefficient is reduced to 6.6% in the random effects specification, and to 3.9% in the fixed effects specification. Nevertheless, they are all statistically significant at 1% level. Thus, it is reasonable to conclude that even though a substantial proportion of the size-wage effect is due to worker heterogeneity, the remaining part is statistically significant and practically large. This is consistent with the previous literature (e.g., Brown and Medoff, 1989).

2.3 Return to training by establishment size

To investigate how the size-wage premiums vary for trained and untrained workers, i.e., whether return to training is the same for workers from large and small establishments, we start with an OLS specification as in equation (3).

$$Y_{it} = \alpha TR_{it} + \gamma L\_EST_{it} + \theta TRL_{it} + X'_{it}\beta + \epsilon_{it} \tag{3}$$

In equation (3), $TR_{it} = 1$ if person i has finished at least one on-the-job training spell on the current job by year t, and $TR_{it} = 0$ otherwise. TRL_{it} is the interaction term of TR_{it} with L_EST_{it}. We also include interaction terms of L_EST_{it} with $Tenure$ and $TenureSQ/100$ in the vector X_{it} to control for possible different tenure effects in large and small establishments. The parameter of primary interest here is θ, which indicates the difference of returns to training between large and small establishments. The first column of Table 6 gives the OLS results. The coefficient on L_EST_{it} is 14.9%, which is close to that in equation (1), but now represents the wage difference between untrained workers in large and small establishments. Return to training is 6.8% in small establishments, while approximately equals 0% in large establishments. The coefficient of θ equals -6.6% and is statistically significant at the 5% level.

One could argue that unobserved worker heterogeneity would bias estimates from the OLS specification. Suppose that the omitted person effect is positively related to whether a person is trained or not (TR_{it}), and whether a person is employed in a large establishment (L_EST_{it}), then the OLS estimates of α and γ would be upward biased. However, a priori whether the estimate of θ is biased, and if biased, the direction of the bias, are both hard to tell. Thus, we consider the following panel model formally in equation (4).

$$Y_{it} = \alpha TR_{it} + \gamma L\_EST_{it} + \theta TRL_{it} + X_{it}'\beta + C_i + \epsilon_{it} \tag{4}$$

Equation (4) is also estimated with both fixed and random effects approaches. The results are given in Columns two and three in Table 6. In both cases, the coefficients of L_EST_{it} are substantially smaller than that in OLS. Return to training in small establishments is 3.7% in the random effects specification and 3.0% in the fixed effects specification, suggesting that workers are selected to be trained based on unobservable characteristics. Nevertheless, the difference between returns to training for small and large establishments (θ) stays at about -6% and statistically significant. Thus return to training is substantially lower in large establishments even after unobserved person heterogeneities are taken into account.

The NLSY79 data also consist of information regarding the starting and ending date of each job spell, allowing us to explicitly consider job-match effects that do not change during a job spell. The model is as follows.

$$Y_{it} = \alpha TR_{it} + \gamma L\_EST_{it} + \theta TRL_{it} + X_{it}'\beta + C_i + \eta_{iJ_{it}} + \epsilon_{it} \tag{5}$$

In equation (5), $\eta_{iJ_{it}}$ is the unobserved job-match effect, with J_{it} indexing the employer for person i at year t. The idea is that when a worker and a firm is a good "match", the productivity level would be higher than otherwise. This is in addition to the person specific wage effect captured by C_i. We can estimate this model in three different ways. The first is the usual random effects approach as in equation (4), but at the person-job level. The second is the fixed effects approach which uses only within job spell variations. In both of the previous two cases, C_i can be simply dropped because the person effect is included in $\eta_{iJ_{it}}$. The last one is a two-level mixed model, accounting for both C_i and $\eta_{iJ_{it}}$ in the variance-covariance matrix .

Table 7 reports the results. The random effects results based on job match effects are not greatly different from those based on person effects as in equation (4). Again, return to training is smaller in large establishments by 4% and this difference is statistically significant with a p-value of about 6%.

The fixed effects specification for equation (5) does not give sensible results, but are listed nevertheless for completeness. For example, tenure effect is estimated to be negative and labor market experience effect is very large. This is because that within a job spell, tenure and total labor market experience are perfectly collinear. Also, size-wage effect (L_EST_{it}) is estimated to be zero. Again, this is because that within a given job spell, variations on establishment size either comes from changes in the size of current establishment, or more likely, measurement errors. Nevertheless, the coefficient on TRL_{it} is both negative and large (-3.1%), suggesting return to training to be smaller in large establishments.

Results from the two-level mixed model are similar to those of random effects approach. Return to training is 2.9% in small establishments. The coefficient on TRL_{it} is -4.4% and significant at the 5% level. Also, as shown by the estimated coefficients of L_Tenure and $L\_TenureSQ/100$, with both random effects and two-level mixed effects specifications, there does not seem to be a difference in terms of tenure effects in small and large establishments.

It should also be pointed out that there does not exist a single specification that is preferred to others. While fixed effects estimator is consistent even if person or job-match effects are correlated with the explanatory variables, it uses only within variations and the measurement error problem can be quite severe. Nevertheless, in all specifications there is strong evidence that return to training is lower in large establishments.

Our empirical results are not sensitive to small modifications in model specification, such as including occupation dummies, adding or dropping quadratic terms. We have also preformed two additional robustness checks, one that also includes off-the-job training information in the model, and another that restricts all analyzes to the period of 1987-2000 and uses only employer paid on-the-job training spells. In both cases the main finding of lower size-wage premium for trained workers is unaffected. The results are not listed in this paper, but available from the authors upon request.

Some caveats may apply to our empirical analyzes. The first one concerns possible problems caused by measurement errors. However, Barron et al. (1997) suggest that measurement errors in training variables are unrelated to establishment size. Thus, the difference in returns to training between small and large establishments would be even larger if there were no measurement errors, as first differencing attenuates the magnitude of all coefficients. The second one is related to a point raised by Gibbons and Katz (1992). If person effects or job-match effects are not constant for a worker or during a job spell, then even the fixed effects estimator is not consistent. However, a priori there is no reason to expect the effects to be different for different sized establishments. Thus, the possible impact on the estimated coefficient might be small. The third one is possible bias due to endogenous labor mobility, as discussed in Gibbons and Katz (1992). For example, model (4) estimated by fixed effects approach will only give effects for those who change jobs. Nevertheless, while the estimated parameters are biased for the whole sample, they are valid at least for the subpopulation of job changers and thus, remain informative. Model (5) considers unobserved job effects, thus is not directly subject to this criticism.

2.4 Do large establishments train more?

Previous studies (e.g., Bishop, 1997) have already found that large establishments train a higher proportion of their workers. In this subsection, we conduct the empirical examination ourselves using NLSY79 data, controlling for more detailed information such as AFQT.

$$TR^*_{it} = \beta Z_{it} + v_{it} \tag{6}$$

$$TR_{it} = \begin{cases} 1 & \text{if} \quad TR^*_{it} \geq 0 \\ 0 & \quad\text{otherwise} \end{cases} \tag{7}$$

In equations (6) and (7), TR^*_{it} is a latent variable that indicates the propensity to receive training, Z_{it} is a set of explanatory variables that include establishment size dummy, and v_{it} is the error term. In the simple Logit specification, $v_{it} = \epsilon_{it}$, while in the panel specifications, $v_{it} = C_i + \epsilon_{it}$. C_i is the unobserved person effect and ϵ_{it} is a logistic error term. Table 8 reports odds ratios for Logit, random effects Logit and fixed effects Logit specifications. Overall the results are consistent with the existing literature. Workers with higher education levels, longer tenure and more labor market experiences are more likely to receive on-the-job training. Controlling for other factors, workers in large firms are also significantly more likely to receive on-the-job training. The odds ratio is 2.2 and 2.3 in Logit and random effects Logit, and 1.7 in fixed effects Logit.

2.5 Discussions of existing theories

In light of our new empirical findings, we discuss here some recent explanations for the size-wage premium that has not been examined in Brown and Medoff (1989).[5] These include the on-the-job search model by Burdett and Mortensen (1998), the productivity hypothesis by Oi and Idson (1999), and the training explanations due to Troske (1999) and Zabojnik and Bernhardt (2001).

Burdett and Mortensen (1998) consider job search frictions in an economy with ex ante identical workers and firms.[6] As workers do not know wages offered by individual firms, they randomly sample a job from the wage offer distribution at a time. In equilibrium, workers take any job they sampled while unemployed, and jobs with higher pay than his current wage while employed, firms optimally choose to offer different wages and hire all job applicants. As firms posting higher wages will attract more workers, the equilibrium wage offers are positively correlated with firm sizes.

In the Burdett and Mortensen (1998) model, firms face a labor supply curve that is not perfectly elastic and each has a certain degree of monopsony power over workers (Green et al., 1996). Wage dispersion is positively related to firms' monopsony power, which is determined by the search frictions in the economy. As training usually contains a firm-specific component, trained workers are more attached to firms (see Lynch, 1992 and Bishop, 1997), which suggests that firms posses more monopsony power over trained workers than untrained ones. Consequently, we would expect the size-wage premium for trained workers to be larger under the search model. This clearly contradicts our empirical evidence.

According to the productivity hypothesis of Oi and Idson (1999), workers in large firms (establishments) are paid more simply because they are more productive. The central idea is that the production process in large and small firms are organized differently, so that similar workers might have different levels of productivity. For example, workers in large firms are required to work in teams, they work harder, use more capital, or have received more training. Thus, econometricians will observe a size-wage premium if productivity differences are not adequately controlled for, even using the fixed effects approach, as unobserved person effect is no longer constant. However, productivity hypothesis is not consistent with our new empirical finding with respect to return to training. Empirical studies such as Bishop (1997) have found that productivity increases for workers in large firms are usually bigger than increases in

[5]In their conclusion, Brown and Medoff (1989) also hypothesized that the employer size-wage relationship may actually be a relationship between firm age and wage since large firms are usually old firms (see also Oi and Idson, 1999). However, Brown and Medoff (2003) later found that the firm age-wage relationship is solely explained by observable worker characteristics.

[6]Other search models such as Albrecht and Axell (1984) and Gaumont et al. (2006) also generates wage dispersion but all workers are different ex ante.

smaller firms. In a competitive setting, this would translate into higher wage increases for trained workers in large firms.

Similarly, models have been proposed that attribute the observed size-wage premium to imperfect control of levels of training workers received in different sized firms. Troske (1999) argues that large firms not only hire more skilled workers, but also "produce" more skilled workers through training. Hu (2003) studies the hiring decisions of large firms and also hypothesizes that firm-specific human capital might explain the size-wage effect. Similarly, Zabojnik and Bernhardt (2001) develop a model in which incentives for workers to accumulate general human capital are provided by corporate tournaments. Due to differences in corporate structures and prize sizes, workers in large firms are induced to accumulate more general human capital and are thus paid more.

These training models predict no size-wage premium for untrained workers, which is grossly inconsistent with our empirical finding. The specific training model of Troske (1999) also implies that starting wages for workers in small firms should be *higher* than those in large firms, as workers and firms share the costs and benefits of specific training (Becker, 1975). Our data, on the contrary, suggest that large firms pay more even for job market entrants than small firms.

3 A Model of Labor Market Imperfect Information and On-the-job training

The model analyzed here considers a two-period competitive labor market with no barriers to entry and exit. Firms are initially identical, have the same production technology, and produce the same output. The output is traded at a competitive market at a price P which we normalize to 1. Firms maximize total profit and have a discount rate of zero.

There is a continuum of workers that can be sorted into two types: high type and low type. High type (also referred to as type H) workers account for proportion α of the population, and low type (also referred to as type L) workers account for proportion $1-\alpha$ of the population. A high type worker can produce some output if employed, while a low type worker will never be able to produce any output. Workers enter the labor market in the first period and retire at the end of the second period.

The proportion of the two types is common knowledge and known to all firms. However, an individual worker's type is unknown to both the individual and firms initially.[7] Hence, information is symmetric at the beginning, though firms and workers may learn about workers'

[7]Making an alternative assumption, namely allowing workers to know their own types, does not affect the main result in this model, but makes the analysis significantly more complicated.

types differently after some labor market experience. This is similar to a large literature on asymmetric learning in labor market, such as Waldman (1996), Gibbons and Waldman (1999), Golan (2005), and Schönberg (2007).

Applying for a job is costless; and workers may apply to all firms each period. Workers are risk-neutral and maximize wages received. They have a discount rate of zero. There is no cost associated with change of employer and hiring/firing workers, and thus no benefit to long-term contracts. Each period, workers will work for the firm that offers the highest wage, provided the highest wage is greater than their reservation wage $r(t)$. In case several firms are tied at the highest wage, workers will randomly choose one firm to work, except when one of these firms is their current employer.

For simplicity, let $r(t) = r$ for $t = 1, 2$, i.e., workers have the same reservation wage each period. Offering a wage $w < r$ is equivalent to no offer at all, as no worker will accept a wage less than the reservation wage. Clearly, when the reservation wage is too high, no firm would hire. To prevent this from happening, we assume that

$$r < \alpha.$$

That is, a firm could make a profit by hiring all workers at their reservation wage r. Alternatively, if firms randomly hire workers at wage r, they would still make a profit.

Wages are paid in advance of production, and no piece rate wage contract can be used. If a worker accepts an offer, he stays with that firm for that period. Otherwise, he will be unemployed for that period. The timing of the game is summarized in Figure 1.

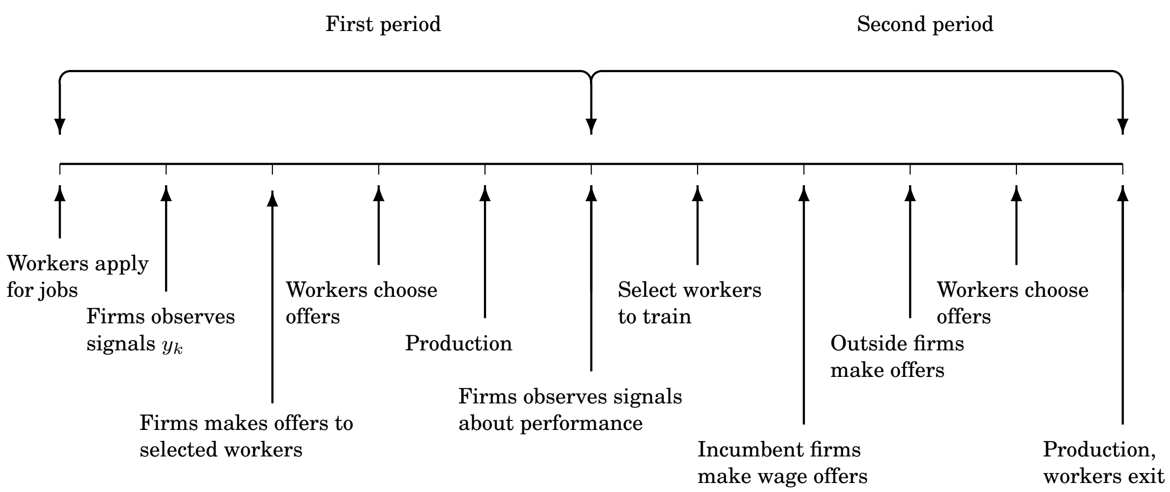

Figure 1: Timing of events

3.1 The first period

The first period starts with workers applying to all firms. While not knowing workers' types, firms each receive a private signal $y_k \in \{h, l\}$ as an imperfect indicator of workers' types. Conditional on a worker being of high type, firm k observes signal "h" with probability $\beta_H > 1/2$, and observes "l" with probability $1 - \beta_H$. Conditional on a worker being of low type, firm k observes "l" with probability $\beta_L > 1/2$ and observes "h" with probability $1 - \beta_L$. For simplicity, we assume private signals received by different firms are independent, and a worker recognized as low type by one firm may be taken as high type by another.

We let $\beta_i = \beta$ for $i = H, L$,[8] and assume that $\beta \in (1/2, 1)$. When $\beta = 1/2$, firms have no ability to differentiate high type workers from low type workers. When $\beta = 1$, firms can perfectly distinguish high type from low type workers. When $1 > \beta > 1/2$, firms have some but less than perfect ability to distinguish, and this is the case we focus on.

Given the signals received about workers' types, firms simultaneously make wage offers to those they want to hire. While some workers may receive multiple offers, they can only accept one. The wages workers receive in the first period are verifiable and will be known to firms in the second period, however, offers they receive are not verifiable and remains workers' private information. In this sense, learning is asymmetric between firms and workers, as workers may acquire more information about their own types after some market experience.

Having received all the wage offers, workers will choose a firm that offers the highest wage, provided the highest offer is greater than his reservation wage r. After that, production begins, and firms realize their profit.

3.2 The second period

At the beginning of this period, firms observe workers' performance in the previous period. However, due to technical limitation, they do not observe workers' output directly. Instead, they observe a signal ω_j from each worker as an imperfect indicator of his performance and type.

Conditional on a worker being of high type, the incumbent firm observes a good signal "g" with probability ρ_H and observes a bad signal "b" with probability $1 - \rho_H$. Conditional on a worker being of low type, the incumbent firm observes "g" with probability ρ_L and observes a "b" with probability $1 - \rho_L$. For simplicity we assume $\rho_H = 1 - \rho_L = \rho > 1/2$; the probability of mistaking a high type for low type and the probability of mistaking a low type for high type are the same.

Given the signals received, the incumbent firm updates its belief about workers' types

[8] This assumption is made for simplicity and relaxing it will not affect our result.

using Bayes rule. If a firm k has a prior belief H_k of a worker being of high type, Bayes rule requires that its posterior belief be μ_0 if it receives a bad signal "b", where

$$\mu_0 = \frac{H_k(1-\rho)}{H_k(1-\rho) + (1-H_k)\rho} < H_k.$$

On the other hand, if it receives a good signal "g", the posterior belief μ_1 should be

$$\mu_1 = \frac{H_k\rho}{H_k\rho + (1-H_k)(1-\rho} > H_k.$$

Having acquired information on workers' productivities, incumbent firms can provide a firm-specific training to some employees. Training increases high type workers' productivity but not of the low type workers. After training, a high type worker can produce $q > 1$ unit of output in the second period, while he can only produce one unit without training. A low type worker's output, nonetheless, equals zero with or without training. Because of the firm specificity, training received at firm j does not increase workers' output at firm k ($k \neq j$). Training is costly, and it takes the incumbent firm a cost of c, $c > 0$, to train one worker. When the per workers training cost is high, for example, $c > (q-1)$, it would not be profitable to train any worker. On the other hand, when the cost equals zero, it is profitable to train every worker, unless firms know the worker involved is low type with certainty. Without loss of generality, we restrict our attention to situations where at least two firms provide training to workers from whom they receive a good signal "g", and no firm provides training to those from whom they receive a bad signal "b".

Assumption 1. *The parameter ρ, c are such that*

$$\frac{\alpha\rho(q-1)}{\alpha\rho + (1-\alpha)(1-\rho)} > c > \frac{\alpha\beta(1-\rho)(q-1)}{\alpha\beta(1-\rho) + (1-\alpha)(1-\beta)\rho}.$$

Meanwhile, outside firms, though have no direct contact with the workers, can also observes a private signal about each worker's performance. However, outside firms can only observe a coarser version of information received by the incumbent. Specifically, conditional on the incumbent firm observing a signal "g" ("b" respectively), outside firms observe a signal $\tilde{g}$ ($\tilde{b}$ respectively) with probability $\gamma \in (1/2, 1)$, and observe $\tilde{b}$ ($\tilde{g}$ respectively) with probability $1-\gamma$. One interpretation of this assumption is that the signal each outside firm observes is some private distortion of a random variable (incumbent firm's signal) that imperfectly indicates a worker's type. For example, the incumbent firm may have an evaluation of performance of each employee, which imperfectly indicate his true types. Outside firms may learn about those evaluations through some indirect channels, which makes their information less accurate than that of the incumbent firm. We will investigate the case when $1/2 < \rho\gamma < \rho$. That is, both the incumbent and outside firms have extra information to help them better judge workers' type, but the information outside firms have is less accurate than that of the incumbent firm.

The information structure is very close to Schönberg (2007) and Golan (2005 2006). However, we differ from them by not allowing the incumbent firm to perfectly learn workers' true types after the first period. Learning is imperfect for the incumbent firm as well as for outside firms here, whereas learning is perfect for the incumbent firm in their models. Of course, if firms and worker are to interact for many periods, incumbent firms' belief may gradually converge to their true types, which corresponds to a special case of our model when $\rho = 1$.

After observing information about workers' types, firms start to make wage offers. The incumbent firm moves first by offering each employee the chance to renew contracts with wage w. Having observed the offers made by the incumbent firm, outside firms can make offers to workers. Before making offers, outside firms know the wages each worker received in the previous period as well as his training status. Workers choose the firm that offer the highest wage, but stay with the incumbent firm if its offer is at least as high as their outside offers. This implies that to retain workers, incumbent firms need to offer their employees at least as much as their outside offers.

After workers and firms are matched, production begins. A trained type H worker who stays with his previous employer produces $q > 1$ unit of output, and a trained type H worker who changes employers, so as an untrained type H worker, produces one unit of output. A type L worker produces no output, irrespective of training status. When production is finished, firms realize their profits and workers exit the market.

3.3 Equilibrium concept

The solution concept we use is competitive equilibrium. A competitive equilibrium is characterized by the number of firms K, an equilibrium wage offer distribution $F(W)$ and firms' training decision such that the following conditions are satisfied

(a) Workers maximize wages giving the equilibrium wage offer;

(b) Firms make wage offers and train decisions to maximize profit given their belief about workers' type and competing firms' choices;

(c) Firms' belief about workers' type is derived from their prior belief using Bayes rule;

(d) Firms each make zero total profit.

Note that even if firms each make zero total profit overall, they may have a positive or negative profit in any period. For example, firms' may make a positive profit in the second period, when the benefit from firm-specific training more than offsets the cost of training.

4 Main result

In this model, workers are free to change employers and firms are free to fire workers. No long-term contracts can be beneficial either to firms or workers, so there is no loss of generality in restricting attention to spot-market contracting. In what follows we first show that there exists an equilibrium that generates size-wage premium.

Proposition 1. *There exists an equilibrium that has the following properties*

1. *In the first period, different firms pay different wages, with firms paying higher wages hiring more workers.*

2. *In the second period, only workers from whom a good signal "g" is observed receive firm-specific training.*

3. *Trained workers receive higher wages than untrained workers.*

For our purpose, we shall call the aforementioned equilibrium the equilibrium with size-wage premium. To show the existence of the equilibrium with size-wage premium, we use a backward induction approach, starting from the second period and then going backward to the first.

In the second period, incumbent firm receives signals from workers, and update beliefs about workers' types. They then decide whether to provide costly training to workers, and make wages offers after that. Having observed the wages offers of the incumbent firm, outside firms can then make offers to workers. At this point, outside firms have also acquired new information on workers' types and observed their training status.

As training is firm specific, the maximum outside offer a worker expects to get can not exceed his pre-training expected productivity. Hence, the incumbent firm will not offer a worker more than his pre-training expected productivity. Consequently, the incumbent firm alone captures the surplus from training, and training decision does not affect wage offers made by the incumbent firm. Hence, we can analyze firms' wage offers without worrying about training decisions.

When training does not affect wage determination, the incumbent firm and outside firms are strategically symmetric, except that the incumbent firm has more accurate information on workers' performance and productivities. At first sight, this seems to give the incumbent firm an advantage over outside firms, and it may use this informational advantage to make extra profit, for example, paying some workers less than their pre-training expected productivity. This turns out not to be the case, as we show in the following result.

Claim 1. *Suppose firm k hires workers with average productivity of H_k in the first period. In the second period, firm k will offer*

$$w_k^g(H_k) = \frac{H_k\rho}{H_k\rho + (1-H_k)(1-\rho)}$$

to those from whom it receives a good signal "g", and will offer

$$w_k^b(H_k) = \frac{H_k(1-\rho)}{H_k(1-\rho) + (1-H_k)\rho}$$

to those from whom it receives a bad signal "b".

Thus, no firm has monopsony power in this market; firms always pay a worker his pre-training expected productivity. This is so as outside firms know the average productivity of the incumbent firm, and competition forces pre-training profit to zero in the second period. Of course, this result holds only if the incumbent firm can not make any counter offers. When the incumbent can make counter offers, as in Golan (2005) and Baron et al. (2006), this may not be the case.

Claim 1 shows that in the second period, incumbent firms offer employees their pre-training expected productivity. As a result, no worker changes employer in the second period. However, this is not to say that every worker would stay with his current employer in the second period. In fact, job separation may occur when some workers, i.e., untrained workers at some firms, find wages they can get fall below their reservation wage r in the second period. We will come back to this point after we characterize wage distribution in the first period.

The incumbent firm knows that it will pay each group of workers their pre-training expected productivities subsequently. Because of the firm specificity of training, it is profitable to train a worker as long as the expected increase in productivity more than offsets the training cost. Depending on parameter combinations, firms may find it profitable to train all workers regardless of signals received, train only those from whom a good signal "g" has been received, or provide no training at all. Let n_k (m_k respectively) be the measure of high type workers (low type workers respectively) employed by firm k, and let N_k be the total measure of workers employed, $N_k = n_k + m_k$. Thus, the pre-training expected productivity of workers at firm k would be $H_k = n_k/N_k$. In follows we present a preliminary result on firms' training decisions.

Lemma 1. *Firm k provides training to all its employees if and only if*

$$(1-\rho)(q-1)n_k - [(1-\rho)n_k + \rho m_k]c \geq 0, \tag{8}$$

provides training only to those from whom it receives a signal "g" if and only if (8) is violated and

$$\rho(q-1)n_k - [\rho n_k + (1-\rho)m_k]c \geq 0. \tag{9}$$

It provides no training to employees if and only if (9) is violated.

Note that as $\rho > 1/2$, condition (8) implies condition (9), but no the reverse; if it not profitable to train workers from whom firm k receives a good signal “g”, it will not be profitable to train those from whom firm k receives a bad signal “b”. The intuition is clear. Training cost per worker is the same for both groups of workers, those with a signal “g” and those with a signal “b”. However, those with a signal “g” are more likely to be of high type workers and therefore, the expected benefit from training is greater. It is more profitable to train workers with a good signal than those with a bad signal. Consequently, firm k trains both groups of workers if and only if the expected surplus from training workers with a signal “b” exceeds the cost of training them. When condition (8) fails but condition (9) holds, it is profitable to train those with a good signal but not profitable to train those with a bad signal. When condition (9) fails, training cost is so high that it is not profitable to train even good workers. No training will be provided in this case.

The previous result implies that firm k’ total profit in the second period equals Π_k if $\Pi_k > 0$, and zero otherwise. Note that

$$\Pi_k = n_k(q-1) - N_k c \tag{10}$$

if condition (8) is satisfied, but

$$\Pi_k = \rho(q-1)n_k - [\rho n_k + (1-\rho)m_k]c \tag{11}$$

if condition (9) holds but not condition (8) fails. In both cases, the average profit function π_k will be an increasing function of H_k, the average productivity of workers at firm k. To see that, we note that in the first case, average profit π_k equals

$$\pi_k(H_k) = \frac{\Pi_k}{N_k} = \rho(q-1)H_k - c, \tag{12}$$

while in the second case,

$$\pi_k(H_k) = \frac{\Pi_k}{N_k} = \rho(q-1)H_k - [\rho H_k - (1-\rho)(1-H_k)]c. \tag{13}$$

Clearly, π_k is an increasing function of H_k.

Claim 1 and Lemma 1 summarize firms’ choices in the second period. Incumbent firms pay each worker his pre-training expected productivity, and train a worker if and only if the expected surplus from training exceeds the cost c. After characterizing the first period wage distribution, we will come back for more on firms’ training decisions. However, for time being, we will leave it as it is and go backward to determine firms’ wage offers in the first period.

In the first period, firms do not know any worker’s type for sure. However, they have some ability to tell high type workers from low type workers. Therefore, they are not going to offer the same wage to every worker that applies. A further complication is that a worker may receive different wage offers from different firms. This is so as firms' ability to differentiate high type from low type workers are independent across firms, so one firm may think a worker as high type while another takes him as low type, regardless of his true type. Fortunately, the following two results significantly simply our analysis.

Claim 2. *In equilibrium, no two firms offer the same wage in the first period.*

The claim shows that, in equilibrium, firms always offer different wages, as offering the same wage is not a best response given competing firms' offers. Hence, the equilibrium wage distribution can not be degenerated when there are more one firms active in equilibrium, which we will show shortly.

Suppose now there are K firms active in the market, denote the set of wages offered in the first period by firm j by $W_j \equiv \{w_j^1, w_j^2, \cdots, \}$, and let $w_j = \max W_j$. That is, w_j is the highest wage offered by firm j in the first period. Without loss of generality, we also order firms by the ranking of highest wages they offer. That is, firm 1 offers the highest wage in the market, firm 2 offers the second highest wage, etc. It happens that in equilibrium, a firm only offers one wage to workers they hope to attract.

Claim 3. *No firms offer more than one wage in the first period.*

Thus, in the first period, firm 1 offers w_1, the highest wage and hires all it takes as high type workers. Because high type workers are more likely to be recognized as high type, firm 1 ends up hiring more than α proportion of high type and less than $1-\alpha$ proportion of low type workers, $H_1 > \alpha$.[9] This essentially leaves firm 2 with a pool of workers with less than α proportion of high type workers to choose from, as no worker who receives offer from firm 1 would accept firm 2 or any other firms' offers. Consequently, the average productivity of employees at firm 2 will be lower than that of firm 1, $H_2 < H_1$. Similar reasoning implies that for any two firms k and $k+1$, $H_{k+1} < H_k$. Since no workers getting offer from firm k would accept offers from firms $j > k$, we can treat the hiring process as a sequential move game in which firms take turns to select workers, with those offering higher wages moving earlier than those offering lower wages.

As firm 1 is the first to move, it can hire all workers it recognizes as high type, which implies that the average productivity of workers at firm 1 is

$$H_1 = \frac{\alpha\beta}{\alpha\beta + (1-\alpha)(1-\beta)}.$$

In general, the average productivity of workers at firm k would be

$$H_k = \frac{\alpha\beta(1-\beta)^{k-1}}{\alpha\beta(1-\beta)^{k-1} + (1-\alpha)\beta^{k-1}(1-\beta)}.$$

[9]Alternatively, H_1 can be taken as the average productivity of workers at firm 1.

The measure of workers firm k hires equals

$$N_k = \alpha(1-\beta)^{k-1}\beta + (1-\alpha)\beta^{k-1}(1-\beta).$$

Clearly, both the average productivity H_k and the measure of workers N_k hired by firm k decrease as k increases.

After firms $1, 2, \cdots, k-1, k$ have hired, the measure of workers remains to be hired will be

$$R_k = \alpha(1-\beta)^k + (1-\alpha)\beta^k,$$

and the average productivity of the remaining workers will be

$$S_k = \frac{\alpha(1-\beta)^k}{\alpha(1-\beta)^k + (1-\alpha)\beta^k} = \frac{\alpha[(1-\beta)/\beta]^k}{\alpha[(1-\beta)/\beta]^k + (1-\alpha)}.$$

Note that both R_k and S_k are strictly decreasing in k. As more and more firms have hired, the proportion of high type workers in the remaining pool decreases, so does the measure of unemployed workers. At certain point, it becomes unprofitable for another firm to hire, as the expected benefit from hiring a worker the firm recognizes as high type drops below the minimum wage r firm needs to pay. This implies that the equilibrium number of firms K is determined by the condition

$$H_k + \pi_K(H_K) \geq r > H_{K+1} + \pi_{K+1}(H_{K+1}). \tag{14}$$

In above we have showed that $\pi_k(H_k)$ is weakly increasing in H_k. The fact that H_k strictly decreases in k indicates $\pi_k(H_k)$ is a decreasing function of k. Hence, the number of firms K is uniquely determined. And firm K offers a wage of w_K to the workers it intends to hire, where

$$w_K = \max\{H_K + \pi_K(H_K), r\}, \tag{15}$$

As firm K is the last one to hire, one may wonder whether firm K may just offer r, the minimum wage to workers it intends to hire and earn a positive profit when $H_k + \pi_K > r$ and $H_{K+1} + \pi_{K+1} < r$. This can not happen, as an outside firm can profitable enter the market by offering $w = r + \varepsilon$ to any worker it recognizes as type H but not hired by firm $j \leq K-1$. Consequently, the potential of further entry ensures all firms makes zero total profit in equilibrium.

Claim 4. *In equilibrium, the number of active firms K is decreasing in workers' reservation wage r, and equilibrium unemployment is increasing in r.*

From previous discussion we know that among the unemployed, the proportion of high ability workers is $S_{K(r)}$. Since S_k strictly decreases in k, the measure of unemployed productive workers will increase as the minimum wage r increases. This implication of our model is consistent with the traditional wisdom, and is in sharp contrast with that of Burdett and Mortensen (1998). In Burdett and Mortensen, an increase in minimum wage decreases inefficient unemployment due to search friction, and thus, "*employment increases with the minimum wage* even though atomistic wage competition, not classic monopsony in the formal sense of one buyer, characterizes the market structure."

Provided the reservation wage r is sufficiently small, the equilibrium wage offer in period one is not degenerate; there will be more than one firm, with each firm offering a different wage.

Claim 5. *Under the assumption that $r < \alpha$, there are at least two firms active in the market, i.e., $K \geq 2$.*

This result follows immediately from the assumption $r < \alpha$, as

$$w_2 \geq H_2 = \alpha > r,$$

and condition (15) implies that $K \geq 2$.

In above we have already given the conditions for a firm to train all employees, train only those from whom firm has a received a good signal "g", and provide no training to any of their employees. Now that we have determined the wage and firm size distribution in the first period, we can further pin down firms' training decisions in the second period.

Claim 6. *Under Assumption 1, at least two firms, firm 1 and 2, will provide training to the group of workers from whom a good signal "g" is received, no firm provides training to the group of workers from whom a bad signal "b" is received.*

Consequently, in the second period, only good workers, those from whom firms have received a good signal "g" will be trained. Given the above results, we are now ready to prove Proposition 1.

Proof of Proposition 1. Claim 1 though Claim 6 gives one equilibrium in which workers maximize wages in both periods, and firms maximize profits in wages offers and training decisions. In this equilibrium, workers maximize wages and firms maximize profit given beliefs about workers' types. Firms' beliefs are updated given information acquired using Bayes rule. Thus, neither firms nor workers has incentives to choose differently. In addition, each firm makes zero total profit. This confirms that the equilibrium outlines is indeed a competitive equilibrium of this model. In the equilibrium, in the first period, each firm k with size N_k pays wages $w_k = H_k + \pi(H_k)$, where

$$H_k = \frac{\alpha\beta(1-\beta)^{k-1}}{\alpha\beta(1-\beta)^{k-1} + (1-\alpha)\beta^{k-1}(1-\beta)}$$

and $\pi(\cdot)$ is an increasing function of H_k.

The first period wage offers is decreasing in k, with firm 1 offering the highest wage w_1 and firm K the lowest wage w_K. Given the wage distribution $w_1 > w_2 > \cdots > w_K$, the measure of workers firm k hires equals

$$N_k = \alpha(1-\beta)^{k-1}\beta + (1-\alpha)\beta^{k-1}\beta,$$

which decreases as k increases. Hence, in the first period, larger sized firms pay higher wages to workers of the same type than smaller size firms. This is also true after we control for workers' training status; trained workers (untrained workers respectively) at larger size firms get higher pay than trained workers (untrained workers respectively) at smaller size firms.

In the second period, firms only provide training to workers from whom a good signal "g" is received, as Claim 6 shows. In addition, Claim 1 shows that at any firm k, trained workers receive $w_k^g(H_k)$ and untrained workers receive $w_k^b(H_k)$. Clearly, $w_k^g(H_k) > w_k^b(H_k)$, trained workers receive higher wages than untrained workers. □

We summarize our findings as follows. In the first period, each firm posts a different wage, and the wage posted by firm k is $w_k = H_k + \pi(H_k)$, which strictly decreases in k. Hence, firm 1 post the highest wage and firm K offers the lowest. Each firm makes offer only to those it wants of hire, i.e., workers from whom it receives a signal "h". As workers maximize wages they receive, firms 1 can get all the workers it intends to hire, while the other firms only hire from the remaining pool of workers they extend offers. While firms make offers simultaneously, job matching ends up much like a sequential selection process, with firm 1 being the first one to choose and firm K being the last one to choose. As a result, firm 1 ends up hiring more workers than all the other firms, and firm K ends up hiring the least. In particular, firm k hires N_k measure of workers, where

$$N_k = \alpha(1-\beta)^{k-1}\beta + (1-\alpha)\beta^{k-1}\beta.$$

The number of active firms K is uniquely determined by the condition (14). The measure of workers not employed by any firm is

$$R_K = \alpha(1-\beta)^K + (1-\alpha)\beta^K.$$

Among them, $\alpha(1-\beta)^K$ measure is of high type.

In the second period, incumbent firms provide training only to those workers from whom a good signal "g" has been received, and offer each worker a wage equal to his expected productivity. As a result, no workers change employers in equilibrium, however, job separation may occur. Some workers, i.e., untrained workers at some firms, may drop out of the market when the wage they expect to get falls below their reservation wage in this period. When this occurs at a firm k, the measure of workers who leave the firm in the second period equals $\alpha(1-\beta)^{k-1}\beta(1-\rho)+(1-\alpha)\beta^{k-1}(1-\beta)\rho$. In addition, since H_k decreases as k increases, it follows immediately from Claim 1 that both $w_k^g(H_k)$ and $w_k^b(H_k)$ decrease in k. That is, after we controlling for workers' training status, larger firms pay higher wages to workers of the same type than smaller firms.

The fact that only workers with a good signal are trained has several implications. The first implication is that larger firms train disproportionally more workers than smaller firms.

Corollary 1. *The proportion of trained workers increases in firm size N_k and decreases in k.*

This result follows from the fact that firm size, N_k and the average quality of workers, H_k, both decrease in k. But the proportion of trained workers at firm k, $\rho H_k+(1-\rho)(1-H_K)$, is increasing in H_k and thus decreasing in k. This explains the empirical finding that larger firms train disproportionally more of their employees.

Another implication is that wage differentials between firms of different sizes will be smaller for trained workers than for untrained workers. In selecting workers to train, firms will try to identify productive workers to train. Because of this additional selection process, differences in average productivities for trained workers between firms of different sizes decreases. This helps to reduce wage differentials for trained workers.

Proposition 2. *Log wage differential is smaller for trained workers than for untrained workers.*

Hence, even though trained workers at a larger firm gets higher pay than trained workers at a smaller firm, the wage differential is smaller than the wage differential between untrained workers at the two firms. And the return to training is smaller at the larger firm than at the smaller firm. Previous researches, for example, the Burdett and Mortensen (1998) model, attribute the size-wage premium to frictions other than difference in workers' qualities. As a result, though being able to generate size-wage premium as an equilibrium outcome, they fail to reconcile size-wage premium with empirical finding on wage differential for trained workers.

In above we have identified one type of equilibrium, the equilibrium with size-wage premium. Note that the equilibrium with size-wage premium is not a single equilibrium and renaming of firms does not change any properties of the equilibrium. A question naturally arise, are there any other types of equilibrium ? The answer is no, as the following result shows.

Proposition 3. *There exists no equilibrium other than the equilibrium with size-wage premium.*

Therefore, the equilibrium is unique up to renaming of the firms. This result is not surprising, as Claim 2 and Claim 3 already show that in any equilibrium, firms post different wages and each firm post one wage only. This ensures that no other equilibrium can exist.

Thus, our model also generate size-wage premium. However, unlike the equilibrium search model, the size-wage premium comes mainly from differences in average productivities across firms. Workers of the same ability may be paid differently because they are pooled with workers of different average productivities.

Consequently, as more information becomes available to help firms better sort workers in the second period, the wage differential between firms of different sizes will decrease; return to training is smaller for larger firms than for small firms. Hence, by emphasizing quality difference to explain the size-wage premium, our model helps reconcile the new findings with the other two stylized facts documented in the literature.

5 Conclusions

In this paper we make two contributions to the literature of size-wage premium. First, using NLSY79 data, we show that large establishments pay more to their workers than small establishments, and train a higher proportion of their workforce, thus confirm findings from previous studies. In addition, we find that return to training for workers in large establishments is lower than those in small establishments. This empirical finding is new and fairly robust across different specifications. No existing theories could explain these three stylized facts simultaneously.

Second, we develop a two-period model of labor market with heterogeneous workers and imperfect observability of worker type. The model generates implications that are consistent with all our empirical findings. Thus, we believe this paper has offered an alternative way of understanding the puzzling size-wage premium phenomenon in the labor market.

References

Abowd, J., F. Kramarz, and D. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*, 251–333. 1

Albæ k, K., M. Arai, R. Asplund, E. Barth, and E. Madsen (1998). Employer size wage effects in the nordic countries. *Labour Economics 5*(4), 425–448. 1

Albrecht, J. and B. Axell (1984). An equilibrium model of search unemployment. *Journal of Political Economy 92*, 824–840. 10

Albrecht, J., P. A. Gautier, and S. Vroman (2006). Equilibrium directed search with multiple applications. *Review of Economic Studies 73*(4), 869–891. 3

Baron, J. M., M. C. Berger, and D. A. Black (2006). Selective counteroffers. *Journal of Labor Economics 24*, 385–409. 17

Barron, J., M. Berger, and D. Black (1997). How well do we measure training? *Journal of Labor Economics 15*, 507–528. 4, 8

Becker, G. (1975). *Human Capital*. New York: Columbia University Press. 11

Bishop, J. H. (1997). What we know about employer-provided training: a review of the literature. *Research in Labor Economics 16*, 19–87. 9, 10

Brown, C., J. Hamilton, and J. Medoff (1990). *Employers large and small*. Harvard University Press, Cambridge MA. 1

Brown, C. and J. Medoff (1989). The employer size-wage effect. *Journal of Political Economy 97*, 1027–1059. 1, 2, 3, 4, 6, 10

Brown, C. and J. Medoff (2003). Firm age and wages. *Journal of Labor Economics 21*(3), 677–697. 10

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review 39*, 257–273. 2, 10, 21, 23

Frazis, H. and M. Lowenstein (2005). Reexamining the returns to training: Functional form, magnitude, and interpretation. *Journal of Human Resources 40*(2), 453–476. 3

Gaumont, D., M. Schindler, and R. Wright (2006). Alternative theories of wage dispersion. *European Economic Review 50*, 831–848. 10

Gibbons, R. and L. Katz (1992). Does unmeasured ability explain inter-industry wage differentials? *Review of Economic Studies 59*(3), 515–535. 2, 9

Gibbons, R. and M. Waldman (1999). A theory of wage and promotion dynamics inside firms. *Quarterly Journal of Economics 114*, 1321–58. 12

Golan, L. (2005). Counteroffers and efficiency in labor markets with asymmetric information. *Journal of Labor Economics 23*, 373–93. 12, 15, 17

Golan, L. (2006). Wage signaling: a dynamic model of intra-firm bargaining and asymmetric information. working paper. 15

Green, F., S. Machin, and A. Manning (1996). The employer size-wage effect: Can dynamic monopsony provide an explanation? *Oxford Economic Papers, New Series 48*(3), 433–455. 1, 10

Hu, L. (2003). The hiring decisions and compensation structures of large firms. *Industrial and Labor Relations Review 56*(4), 663–681. 11

Idson, T. and D. Feaster (1990). A selectivity model of employer-size wage differentials. *Journal of Labor Economics 8*(1), 99–122. 1

Lallemand, T., R. Plasman, and F. Rycx (2005). The establishment-size wage premium: evidence from european countries. IZA Discussion paper No. 1569. 5

Lester, R. (1967). Pay differentials by size of establishment. *Industrial Relations 7*, 57–67. 1

Lynch, L. (1992). Private-sector training and the earnings of young workers. *American Economic Review 82*(1), 299–312. 3, 10

Main, B. G. and B. Reilly (1993). The employer size-wage gap, evidece from britain. *Economica 60*(238), 125–142. 1

Mellow, W. (1982). Employer size and wages. *Review of Economics and Statistics 64*, 495–501. 1

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy 105*(2), 385–411. 3

Moore, H. (1911). *Laws of Wages*. Macmillan, New York, reprinted 1967. 1

Neal, D. and W. R. Johnson (1996). The role of pre-market factors in black-white wage differences. *Journal of Political Economy 104*(5), 869–895. 3

Oi, W. and T. Idson (1999). Firm size and wages. In O. Ashenfelter and D. Card (Eds.), *Handbook of labor economics*, Volume 3, pp. 2165–2215. 10

Personick, M. and C. Barsky (1982). White-collar pay levels linked to corporate work force size. *Monthly Labor Review 105*, 23–28. 1

Schaffner, J. (1998). Premiums to employment in larger establishments: evidence from peru. *Journal of Development Economics 55*(1), 81–113. 1

Schönberg, U. (2007). Testing for asymmetric employment learning. *Journal of Labor Economics 25*(4), 651–691. 12, 15

Shi, S. (2002). A directed search model of inequality with heterogeneous skills and skill-biased technology. *Review of Economic Studies 69*(2), 467–491. 3

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy 113*(5), 996–1025. 3

Söderbom, M., F. Teal, and A. Wambugu (2005). Unobserved heterogeneity and the relation between earnings and firm size: evidence from two developing countries. *Economics Letters 87*(2), 153–159. 1

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics 87*, 355–374. 2

Stigler, G. (1962). Information in the labor market. *Journal of Political Economy 70*, 94–105. 2

Troske, K. (1999). Evidence on the employer size-wage premium from worker-establishment matched data. *The Review of Economics and Statistics 81*(1), 15–26. 1, 2, 10, 11

Velenchik, A. (1997). Government intervention, efficiency wages, and the employer size wage effect in zimbabwe. *Journal of Development Economics 53*(2), 305–338. 1

Veum, J. (1995). Sources of training and their impact on wages. *Industrial and Labor Relations Review 48*(4), 812–826. 3, 4

Waldman, M. (1996). Asymmetric learning and the wage/productivity relationship. *Journal of Economic Behavavior and Organization 31*, 419–29. 12

Winter-Ebmer, R. and J. Zweimüller (1999). Firm-size wage differential in switerland: evidence from job-changers. *American Economic Review, Papers and Proceedings 89*(2), 89–93. 1

Zabojnik, J. and D. Bernhardt (2001). Corporate tournaments, human capital acquisition, and the firm size-wage relation. *Review of Economic Studies 68*(3), 693–716. 2, 10, 11

Appendix

Proof of Claim 1. We prove this result by showing that the incumbent firm can not make any profit except that from firm specific training. As the decision of training does not affect wage offers, we simplify the proof a bit by pretending there is no training at all. That is, for proof of this result, we can look at a special case where $q = 1$ and $c = 0$.

In this case, firm k could not make any wage offers that result in positive profit in the second period. To see that, we first suppose firm k offers worker i an wage $w_k^i(2)$ such that $\sum_{i\in k} w_k^i(2) < \sum_{i\in k} H_k$, where $i \in k$ indicates sum over firm k's employees, then an outside firm could make a profit by attracting all of k's employee. This could be done even if the outside firm has acquired no extra information in the second period. Of course, it could do better with the additional information.

Since firm k earns zero profit, could it offer the same wage $w_k^2 = H_k$ to all workers in the second period, irrespective of signals received? Suppose that is the case, an outside firm j could make a profit by offering $w_k^2 + \varepsilon$ to those from whom firm j observes a signal "$\tilde{g}$". Consequently, the expected productivity of the remaining pool for firm k would be strictly less than H_k which leads to a loss in this period. Hence we conclude that firm k offers each employee her pre-training expected productivity in the second period.

If a worker is hired by firm k in period one, the chance that he is type H equals H_k. Thus the posterior belief that he is of type H is

$$\mu_1(H_k) = \frac{H_k\rho}{H_k\rho + (1-H_k)(1-\rho)} > H_k$$

if he produces a good signal in period one, and is

$$\mu_0(H_k) = \frac{H_k(1-\rho)}{H_k(1-\rho) + (1-H_k)\rho} < H_k$$

otherwise. Zero profit condition implies the second period wage offer should be $w_k^g = \mu_1(H_k)$, and $w_k^b = \mu_0(H_k)$ respectively. □

Proof of Lemma 1. As training is firm specific, the incumbent firm alone will capture the surplus from training. With no training, firm k's profit in the second period would be

$$\Pi_k = n_k - [\rho n_k + (1-\rho)m_k]w_k^g - [(1-\rho)n_k + \rho m_k]w_k^b = 0,$$

where w_k^g and w_k^b are wages paid to those with a good signal and bad signal respectively.

If firm k provides training only to workers with a signal "g", its profit would be

$$\Pi_k = \rho q n_k - [\rho n_k + (1-\rho)m_k]c - [\rho n_k + (1-\rho)m_k]w_k^g - [(1-\rho)n_k + \rho m_k]w_k^b,$$

which is greater than zero if and only if the condition in (9) is satisfied. That is, the expected surplus from training must exceeds the cost of training.

Similarly, for firm k to provide training to those with a signal "b", the expected surplus from training this group of workers must exceed the cost, i.e.,

$$(1-\rho)n_k(q-1) - [(1-\rho)n_k + \rho m_k]c \geq 0.$$

Note that this is equivalent to

$$(1-\rho)\{(q-1)n_k - [n_k + \frac{1-\rho}{\rho} m_k]c\} \geq 0,$$

which indicates

$$\rho\{(q-1)n_k - [n_k + \frac{1-\rho}{\rho} m_k]c\} = \rho(q-1)n_k - [\rho n_k + (1-\rho)m_k]c \geq 0$$

as $\rho > 1/2$. Hence if it is profitable for firm k to train workers with a bad signal "b", it would be profitable to train all workers. $\square$

Proof of Claim 2. Suppose two firms, firm k and firm j offer the same wage w. Given the pool of workers available to the two firms, firms make offers to workers they reckon as type H workers. Let n_k (respectively n_j) be measure of type H workers and m_k (respectively m_j) be measure of type L workers who will receive firm k's (respectively firm j's) offer. Given the information that the two firms have similar information technology β and they offer the same wage, it is clear that $n_k = n_j$ and $m_k = m_j$. Also let n and m be the measure of type H and type L workers that receive offers from both firms. As the probabilities β are independent, clearly $n < n_k$ and $m < m_k$. As workers randomly choose a firm to work for, firm k will attract $n_k - n/2$ measure of type H and $m_k - m/2$ measure of type L workers. Zero profit condition implies that wage w offered by the the two firms must be

$$w = \frac{(n_k - n/2)}{n_k - n/2 + m_k - m/2} = \frac{(n_j - n/2)}{n_j - n/2 + m_j - m/2}.$$

However, in this case, firm k could profit by offering a higher wage $w + \varepsilon$. Doing so, its profit would be

$$\Pi'_k(w+\varepsilon) = \left[\left(\frac{n_k}{n_k + m_k} - \frac{(n_k - n/2)}{n_k - n/2 + m_k - m/2}\right) - \varepsilon\right](n_k + m_k).$$

If we let $\tilde{N}^H$, $\tilde{N}^l$ be the measure of type H, type L workers in the pool of workers firms k and j hires from, then we have

$$n_k = \tilde{N}^H \beta, \quad n = \tilde{N}^H \beta^2,$$

$$m_k = \tilde{N}^L (1-\beta), \quad m = \tilde{N}^L (1-\beta)^2.$$

The condition $\beta > 1/2 > 1 - \beta$ indicates

$$\frac{n/2}{n/2 + m/2} = \frac{\tilde{N}^H \beta^2}{\tilde{N}^H \beta^2 + \tilde{N}^L (1-\beta)^2} > \frac{\tilde{N}^H \beta}{\tilde{N}^H \beta + \tilde{N}^L (1-\beta)} = \frac{n_k}{n_k + m_k}.$$

So we have

$$\frac{n_k - n/2}{n_k + m_k - n/2 - m/2} < \frac{n_k}{n_k + m_k}.$$

Hence, by offering a wage $w+\varepsilon$, firm k's profit strictly increase, $\Pi(w+\varepsilon) > 0 > \Pi(w)$; two firms offering the same wage w can not be part of any equilibrium. The same argument implies that more than two firms offering the same wage can not be part of any equilibrium $\square$

Proof of Claim 3. Since firms 1 has the highest wage, any worker who receives its offer will accept the offer. Because of imperfect information, firm 1 may recognize a productive worker as type H with probability β and take a unproductive worker as type H with probability $1-\beta$. Given the proportion of the two types in the population, Bayes rule requires that the probability of workers accepting firm 1's offer being of type H to be

$$H_1 = \frac{\alpha\beta}{\alpha\beta + (1-\alpha)(1-\beta)}.$$

Zero-profit condition implies that the first period wage equals

$$w_1(H_1) = H_1 + \pi(H_1),$$

where $\pi(H_1)$ is firm 1's expected profit per worker in the second period. The probability that a randomly chosen worker not hired by firm 1 is of type H equals

$$S_1 = \frac{\alpha(1-\beta)}{\alpha(1-\beta) + (1-\alpha)\beta}.$$

Now suppose firm 1 also offer another wage w_1', $w_1' \neq w^1$, to some other workers it recognized as type L. Note that given firm 1's belief, the highest wage w_1^2 it is willing to offer to any workers not getting w_1 is $S_1 + \pi(S_1)$, where $\pi(S_1)$ is the profit it expects to get from each of these workers in the second period.

However, offering $w_1^2 = S_1 + \pi(S_1)$ is not profitable as another firm, firm 2 can offer $w_2 = H_2 + \pi(H_2)$ to any workers it recognizes as type H but not getting the offer w_1, where

$$H_2 = \frac{\alpha\beta(1-\beta)}{\alpha\beta(1-\beta) + (1-\alpha)\beta(1-\beta)} > S_1.$$

As $\pi(\cdot)$ is a weakly increasing function, $\pi(H_2) \geq \pi(S_1)$ and $w_2 > w_1^2$. This indicates that any worker getting offer w_1^2 and w_2 would accept firm 2's offer of w_1. As a result, the average productivity of the group of workers firm 1 hires at wage w_1^2 is no greater than S_2 where

$$S_2 = \frac{\alpha(1-\beta)^2}{\alpha(1-\beta)^2 + (1-\alpha)\beta^2} = \frac{\alpha[(1-\beta)/\beta]^2}{\alpha[(1-\beta)/\beta]^2 + (1-\alpha)}.$$

As $S^2 < S^1$, firm 1 would suffer a loss from offering w_1^2.

Similar argument implies that it is not profitable for firm 1 to offer any wage $w_1' \geq r$ as some other firm k may offer a wage just high enough to draw some type H workers from the pool of workers it intends to hire, which leaves firm 1 with a loss. For the same reasoning, the other firms, firm $2, 3, \cdots$ also make only one offer in any equilibrium. $\square$

Proof of Claim 4. The first part of the proposition follows immediately from the fact that H_k is decreasing in k and that K is determined by the condition

$$H_k + \pi_K \geq r > H_{K+1} + \pi_{K+1}.$$

To show the second part, we note that the measure of workers not hired by firm $j \leq k$ equals

$$R_k = \alpha(1-\beta)^k + (1-\alpha)\beta^k.$$

And the measure of unemployed workers equals R_K. As K decreases in r, R_K increases in r. Hence we conclude that equilibrium unemployment strictly increases in the reservation wage r. $\square$

Proof of Claim 5. To prove the first part, we first note that for firm k,

$$n_k = \alpha\beta(1-\beta)^{k-1}, \qquad m^k = (1-\alpha)\beta^{k-1}(1-\beta).$$

By the condition in (9), firm k would train those with a signal "g' if and only if

$$\Pi_k^g = \rho\alpha\beta(1-\beta)^{k-1}(q-1) - [\rho\alpha\beta(1-\beta)^{k-1} + (1-\rho)(1-\alpha)\beta^{k-1}(1-\beta)]c \geq 0.$$

As $\beta > 1/2$, Π_k^g is decreasing in k; if it is profitable for firm $k \geq 2$ to train workers with a signal "g", it is must be profitable for firm $k-1$. Assumption 1 indicates

$$\Pi_2 = \beta(1-\beta)[\rho\alpha + (1-\rho)(1-\alpha)]\left\{\frac{\rho\alpha}{\rho\alpha + (1-\rho)(1-\alpha)}\right\} > 0. \tag{A.1}$$

That is, at least two firms, firm 1 and firm 2 will provide training to some workers.

To see that no firm provides training to those with a signal "b", we note that profit for firm k to train these workers would be

$$\Pi_k^b = (1-\rho)\alpha\beta(1-\beta)^{k-1}(q-1) - [(1-\rho)\alpha\beta(1-\beta)^{k-1} + \rho(1-\alpha)\beta^{k-1}(1-\beta)]c,$$

which is decreasing in k as $\beta > 1/2$, if it is not profitable form firm k to train these workers, it can not be profitable for firm $k+1$ to do so. Assumption 1 indicates

$$\Pi_1^b = \beta(1-\beta)[(1-\rho)\alpha\beta + \rho(1-\alpha)(1-\beta)]\left\{\frac{(1-\rho)\alpha\beta}{(1-\rho)\alpha\beta + \rho(1-\alpha)(1-\beta)}\right\} < 0.$$

That is, it is not profitable for firms to provide training to workers with a signal "b". $\square$

Proof of Proposition 2. Two firms, k and j, with k be the larger firm. Let the measure of workers firm k (firm j respectively) employs be N_k (N_j respectively). Also let n_k (n_j respectively) be the measure of type H workers firm k (firm j respectively) hires, and m_k (m_j respectively) be the measure of type L workers firm k (firm j respectively) hires. By assumption,

$$n_k > n_j, \quad n_k/N_k > n_j/N_j \Longrightarrow$$
$$\frac{n_k}{n_j} > \frac{n_k + m_k}{n_j + m_j} \Longrightarrow \frac{m_k}{m_j} < \frac{N_k}{N_j} < \frac{n_k}{n_j}.$$

The wage differential between workers at firm k and j in the first period is

$$\Phi_{kj} = \frac{\{n_k + (q-1)\rho n_k - [\rho n_k + (1-\rho)m_A]c\}/N_k}{\{n_j + (q-1)\rho n_j - [\rho n_j + (1-\rho)m_j]c\}/N_j}.$$

The wage differential between trained workers at firm k and j in the second period is

$$\Phi^t = \frac{\rho n_k/[\rho n_k + (1-\rho)m_k]}{\rho n_j/[\rho n_j + (1-\rho)m_j]} = \frac{n_k}{n_j} \cdot \frac{\rho n_j + (1-\rho)m_j}{\rho n_k + (1-\rho)m_k}.$$

Let

$$\Gamma(\rho) = \frac{\rho n_j + (1-\rho)m_j}{\rho n_k + (1-\rho)m_k}.$$

Note that $\Gamma' < 0$ and $\Gamma(1/2) = N_j/N_k$. By our assumption, $\rho > 1/2$, so we conclude $\Gamma < N_j/N_k$ and

$$\Phi^t < \frac{n_k/N_k}{n_j/N_j}.$$

On the other hand, we note that

$$\frac{n_k + (q-1)\rho n_k}{n_j + (q-1)\rho n_j} = \frac{n_k}{n_j}, \qquad \text{and} \quad \frac{[\rho n_k + (1-\rho)m_k]c}{[\rho n_j + (1-\rho)m_j]c} < \frac{n_k}{n_j}$$

because $n_k/n_j > m_k/m_j$. This implies that

$$\frac{n_k + (q-1)\rho n_k - [\rho n_k + (1-\rho)m_k]c}{n_j + (q-1)\rho n_j - [\rho n_j + (1-\rho)m_j]c} > \frac{n_k}{n_j},$$

and

$$\Phi > \frac{n_k/N_k}{n_j/N_j} > \Phi^t.$$

The wage differential is smaller in the second period for trained workers than the wage differential in the first period.

The wage differential between untrained workers at firm k and j in the second period is

$$\Phi^u = \frac{(1-\rho)n_k/[(1-\rho)n_k + \rho m_k]}{(1-\rho)n_j/[(1-\rho)n_j + \rho m_j]} = \frac{n_k}{n_j} \cdot \frac{(1-\rho)n_j + \rho m_j}{(1-\rho)n_k + \rho m_k}.$$

Because $1-\rho < 1/2$, $\Gamma(1-\rho) > N_j/N_K$, which implies

$$\Phi^u > \frac{n_k/N_k}{n_j/N_j} > \Phi^t.$$

To see that $\Gamma'(\rho) < 0$ we differentiate it with respect to ρ, which gives

$$\begin{aligned}
\Gamma' &= \frac{[\rho n_k + (1-\rho)m_k](n_j - m_j) - [\rho n_j + (1-\rho)m_j](n_k - m_k)}{[\rho n_k + (1-\rho)m_k]^2} \\
&= \frac{[\rho n_k n_j\{[1+\frac{1-\rho}{\rho}\frac{m_k}{n_k}](1-\frac{m_j}{n_j}) - [1+\frac{1-\rho}{\rho}\frac{m_j}{n_j}](1-\frac{m_k}{n_k})]\}}{[\rho n_k + (1-\rho)m_k]^2} \\
&= \frac{[\rho n_k n_j\{\frac{2-\rho}{\rho}[\frac{m_k}{n_k} - \frac{m_j}{n_j}]\}}{[\rho n_k + (1-\rho)m_k]^2},
\end{aligned}$$

which is strictly negative as $m_k/n_k < m_j/n_j$.

□

Proof of Proposition 3. This follows immediately from Claim 2 and 3, which shows that in any equilibrium, each firm offers a different wage and only one wage in the first period. This implies that firms offer higher wages hire more workers in the first period. Firms choices in the second period follow from Claim 1 and Lemma 1. □

Table 1: Variable Definitions

Variable	Description
Y	Natural logarithm of real wage, deflated by CPI-U (1982–$1984 = 100$)
ED_L12	Less than 12 years of schooling (less than high school)
ED_12	Equal to 12 years of schooling (high school graduates)
ED_1315	Between 13 to 15 years of schooling (some college)
ED_16	Equal to 16 years of schooling (college graduates)
ED_G16	More than 16 years of schooling (More than college graduates)
MARR_NM	A dummy = 1 if the person is never married
MARR_SP	A dummy = 1 if the person is married with spouse present
MARR_OT	A dummy = 1 if the person is married with spouse not present
UNEMP	Local unemployment rate categories: 1: $< 3\%$; 2: 3%–5.9%; 3: 6%–8.9%; 4: 9%–11.9%; 5: 12–14.9%; 6: $\geq 15\%$
NE	A dummy = 1 if the work is in the Northeast region
NC	A dummy = 1 if the work is in the North Central region
WEST	A dummy = 1 if the work is in the West region
SOUTH	A dummy = 1 if the work is in the South region
AFQT	Armed forces qualification test percentile score
AFQTSQ/100	Square of AFQT divided by 100
TENURE	Total years of tenure with current employer
TENURESQ/100	Square of TENURE divided by 100
EXP	Total years of labor market experience
EXPSQ/100	Square of EXP divided by 100
S_EST	A dummy = 1 if in an establishment with less than 500 employees, 0 otherwise
L_EST	A dummy = 1 if in an establishment with at least 500 employees, 0 otherwise
TR	A dummy = 1 if completed at least one on-the-job training spell with current employer, 0 otherwise
TRS	Interaction term of TR with S_EST
TRL	Interaction term of TR with L_EST
OFF_TR	A dummy = 1 if completed at least one off-the-job training spell with current employer, 0 otherwise
L_TENURE	Interaction term of TENURE with L_EST

See next page

TABLE 1. Variable Definitions (Continued)

Variable	**Description**
L_TENURESQ/100	Interaction term of TENURESQ/100 with L_EST
IND1	A dummy = 1 if in the mining ind. (SIC: 47–57), 0 otherwise
IND2	A dummy = 1 if in the construction ind. (SIC: 67–77), 0 otherwise
IND3	A dummy = 1 if in the manufacturing ind. (SIC: 107–398), 0 otherwise
IND4	A dummy = 1 if in the public utilities ind. (SIC: 407–479), 0 otherwise
IND5	A dummy = 1 if in the wholesale trade ind. (SIC: 507–588), 0 otherwise
IND6	A dummy = 1 if in the financial services ind. (SIC: 707–718), 0 otherwise
IND7	A dummy = 1 if in the business and repair services ind (SIC: 727–759), 0 otherwise
IND8	A dummy = 1 if in the personal services ind. (SIC: 769–798), 0 otherwise
IND9	A dummy = 1 if in the public administration ind. (SIC: 907–937), 0 otherwise
WEIGHT	NLSY79 Sampling Weight

Table 2: Firm Size-Wage Relationship, Year 2002, by Industries

	Firm Size (Number of Employees)			
	0–19	**20–99**	**100–499**	**500+**
Agriculture, forestry, fishing, & hunting	26	27	N/A	N/A
Mining	39	43	50	58
Utilities	32	46	52	68
Construction	33	40	44	47
Manufacturing	31	35	37	44
Wholesale trade	39	41	43	51
Retail Trade	21	27	29	19
Transportation & Warehousing	28	30	32	40
Information	44	46	51	55
Finance & insurance	46	55	56	61
Real estate & rental & leasing	30	32	34	35
Professional, scientific, & technical services	44	54	58	56
Management of companies & enterprises	66	50	52	73
Administrative & support & waste management	29	27	23	25
Educational services	21	22	26	29
Health care & social assistance	39	33	27	34
Arts, entertainment, & recreation	35	19	34	24
Accommodation & food services	12	12	12	15
Other services	20	21	26	27
Total	**30**	**31**	**34**	**39**

Note. *Shaded industries are those with average wages not necessarily increase with firm size. Information from Small Business Administration website http://www.sba.gov/advo/research/data.html.*

Table 3: Summary Statistics, Whole Sample

Variable	**# of Observations**	**Mean**	**Standard Deviation**	**Min**	**Max**
LNW	8972	6.77	0.49	4.61	9.10
ED_L12	9261	0.13	0.33	0	1
ED_12	9261	0.43	0.49	0	1
ED_1315	9261	0.23	0.42	0	1
ED_16	9261	0.19	0.39	0	1
ED_G16	9261	0.04	0.19	0	1
NE	9285	0.19	0.39	0	1
NC	9285	0.33	0.47	0	1
WEST	9285	0.15	0.36	0	1
SOUTH	9285	0.32	0.47	0	1
UNEMP	9148	2.59	0.88	1	6
MARR_NM	9317	0.28	0.45	0	1
MARR_SP	9317	0.60	0.49	0	1
MARR_OT	9317	0.12	0.33	0	1
AFQT	8919	56.25	28.29	1	99
TENURE	9155	4.26	4.26	0.02	24.92
EXP	9318	11.34	4.56	0.35	23.00
TR	9318	0.19	0.39	0	1
OFF_TR	9318	0.15	0.35	0	1
S_EST	9165	0.83	0.38	0	1
L_EST	9165	0.17	0.38	0	1

Table 4: Sample Means for the whole sample, small establishments and large establishments

Variable	**Whole Sample**	**Small Establishments**	**Large Establishments**
LNW	6.77	6.71	7.06
ED_L12	0.13	0.14	0.05
ED_12	0.43	0.46	0.28
ED_1315	0.23	0.22	0.24
ED_16	0.19	0.16	0.34
ED_G16	0.04	0.02	0.10
NE	0.19	0.18	0.23
NC	0.33	0.33	0.35
WEST	0.15	0.16	0.12
SOUTH	0.32	0.32	0.30
UNEMP	2.59	2.62	2.46
MARR_NM	0.28	0.28	0.23
MARR_SP	0.60	0.59	0.66
MARR_OT	0.12	0.12	0.10
AFQT	56.25	53.51	70.02
TENURE	4.26	4.09	5.22
EXP	11.34	11.26	11.74
TR	0.19	0.15	0.38
OFF_TR	0.15	0.13	0.22
S_EST	0.83	1.00	0.00
L_EST	0.17	0.00	1.00

Table 5: Establishment Size-Wage Effects: OLS, Random Effects, and Fixed Effects Estimates

	OLS	**Random Effects**	**Fixed Effects**
L_EST	0.146^{***}	0.066^{***}	0.039^{***}
	(0.012)	(0.012)	(0.012)
ED_L12	-0.066^{***}	-0.065^{**}	
	(0.015)	(0.029)	
ED_1315	0.174^{***}	0.161^{***}	
	(0.012)	(0.023)	
ED_16	0.378^{***}	0.368^{***}	
	(0.014)	(0.029)	
ED_G16	0.501^{***}	0.525^{***}	
	(0.025)	(0.052)	
AFQT	0.004^{***}	0.003^{**}	
	(0.001)	(0.001)	
AFQTSQ/100	-0.001	-0.001	
	(0.001)	(0.001)	
UNEMP	-0.031^{***}	-0.024^{***}	-0.022^{***}
	(0.005)	(0.004)	(0.005)
TENURE	0.033^{***}	0.029^{***}	0.025^{***}
	(0.003)	(0.002)	(0.002)
TENURESQ/100	-0.118^{***}	-0.124^{***}	-0.121^{***}
	(0.018)	(0.015)	(0.015)
EXP	0.027^{***}	0.030^{***}	0.032^{***}
	(0.005)	(0.004)	(0.004)
EXPSQ/100	-0.021	-0.017	-0.017
	(0.018)	(0.014)	(0.015)

Note. *Dependent variable is log hourly real wage. Number of observations is 8,130. The other explanatory variables are regional dummies, marriage dummies, and industry dummies. The base group workers are full-time white males with 12 years of schooling, work in a small-sized establishment.* *** *stands for significance at* 1% *level,* ** *stands for significance at* 5% *level,* * *stands for significance at* 10% *level.*

Table 6: Return to Training by Establishment Size: OLS and Person Effects Models

	OLS	**Random Effects**	**Fixed Effects**
L_EST	0.149***	0.064***	0.033*
	(0.023)	(0.018)	(0.019)
TR	0.068***	0.037***	0.030**
	(0.013)	(0.012)	(0.012)
TRL	−0.066**	−0.060***	−0.057**
	(0.026)	(0.022)	(0.023)
ED_L12	−0.066***	−0.065**	
	(0.015)	(0.029)	
ED_1315	0.171***	0.160***	
	(0.012)	(0.023)	
ED_16	0.374***	0.365***	
	(0.014)	(0.029)	
ED_G16	0.497***	0.524***	
	(0.025)	(0.052)	
AFQT	0.004***	0.003**	
	(0.001)	(0.001)	
AFQTSQ/100	−0.001	−0.001	
	(0.001)	(0.001)	
UNEMP	−0.031***	−0.024***	−0.022***
	(0.005)	(0.004)	(0.005)
TENURE	0.031***	0.027***	0.023***
	(0.003)	(0.003)	(0.003)
L_TENURE	0.005	0.007	0.008
	(0.007)	(0.006)	(0.006)
TENURESQ/100	−0.108***	−0.116***	−0.113***
	(0.02)	(0.016)	(0.017)
L_TENURESQ/100	−0.028	−0.033	−0.037
	(0.045)	(0.035)	(0.035)
EXP	0.026***	0.029***	0.032***
	(0.005)	(0.004)	(0.004)
EXPSQ/100	−0.019	−0.016	−0.016
	(0.018)	(0.014)	(0.015)

Table 7: Return to Training by Establishment Size: Person plus Job Match Effects Models

	Random Effects	**Fixed Effects**	**Two-level Mixed Model**
L_EST	0.072^{***}	-0.021	0.054^{***}
	(0.019)	(0.025)	(0.018)
TR	0.033^{***}	0.015	0.029^{**}
	(0.013)	(0.015)	(0.012)
TRL	-0.041^{*}	-0.031	-0.044^{**}
	(0.022)	(0.024)	(0.021)
ED_L12	-0.035^{*}		-0.056^{*}
	(0.020)		(0.029)
ED_1315	0.177^{***}		0.162^{***}
	(0.017)		(0.023)
ED_16	0.406^{***}		0.377^{***}
	(0.021)		(0.029)
ED_G16	0.561^{***}		0.532^{***}
	(0.040)		(0.052)
AFQT	0.004^{***}		0.004^{***}
	(0.001)		(0.001)
AFQTSQ/100	-0.001		-0.001
	(0.001)		(0.001)
UNEMP	-0.027^{***}	-0.020^{***}	-0.025^{***}
	(0.004)	(0.006)	(0.004)
TENURE	0.027^{***}	-0.091^{***}	0.025^{***}
	(0.003)	(0.027)	(0.003)
L_TENURE	0.004	0.012^{**}	0.005
	(0.005)	(0.006)	(0.005)
TENURESQ/100	-0.098^{***}	-0.059^{***}	-0.098^{***}
	(0.016)	(0.017)	(0.015)
L_TENURESQ/100	-0.031	-0.067^{**}	-0.033
	(0.031)	(0.032)	(0.030)
EXP	0.028^{***}	0.145^{***}	0.029^{***}
	(0.004)	(0.028)	(0.004)
EXPSQ/100	-0.026^{*}	-0.045^{**}	-0.022
	(0.016)	(0.021)	(0.015)

Table 8: Logit Models for Training Determination (Odds Ratios)

	Logit	**Random Effects Logit**	**Fixed Effects Logit**
L_EST	2.176^{***}	2.276^{***}	1.669^{***}
	(0.160)	(0.319)	(0.278)
ED_L12	0.665^{***}	0.395^{**}	
	(0.097)	(0.168)	
ED_1315	1.562^{***}	3.700^{***}	
	(0.130)	(0.935)	
ED_16	1.271^{***}	2.441^{**}	
	(0.120)	(0.859)	
ED_G16	1.750^{***}	6.741^{***}	
	(0.282)	(3.337)	
AFQT	1.021^{***}	1.027	
	(0.006)	(0.018)	
AFQTSQ/100	0.989^{**}	0.986	
	(0.005)	(0.017)	
UNEMP	0.961	0.968	0.993
	(0.036)	(0.066)	(0.075)
TENURE	1.250^{***}	1.509^{***}	1.526^{***}
	(0.026)	(0.054)	(0.067)
TENURESQ/100	0.451^{***}	0.398^{***}	0.807
	(0.055)	(0.083)	(0.249)
EXP	1.322^{***}	1.736^{***}	1.664^{***}
	(0.053)	(0.113)	(0.115)
EXPSQ/100	0.451^{***}	0.200^{***}	0.214^{***}
	(0.068)	(0.048)	(0.278)