

Feng, Shuaizhang; Zheng, Bingyong

Working Paper

Imperfect information, on-the-job training, and the employer size-wage puzzle: Theory and evidence

IZA Discussion Papers, No. 4998

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Feng, Shuaizhang; Zheng, Bingyong (2010) : Imperfect information, on-the-job training, and the employer size-wage puzzle: Theory and evidence, IZA Discussion Papers, No. 4998, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/36770>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 4998

**Imperfect Information, On-the-Job Training, and the
Employer Size-Wage Puzzle: Theory and Evidence**

Shuaizhang Feng
Bingyong Zheng

June 2010

Imperfect Information, On-the-Job Training, and the Employer Size-Wage Puzzle: Theory and Evidence

Shuaizhang Feng

*Princeton University
and IZA*

Bingyong Zheng

Shanghai University of Finance and Economics

Discussion Paper No. 4998

June 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Imperfect Information, On-the-Job Training, and the Employer Size-Wage Puzzle: Theory and Evidence^{*}

This paper develops a two-period labor market model with imperfect information and on-the-job training, and uses data from National Longitudinal Survey of Youth 1979 Cohorts (NLSY79) to test its predictions. We find that training does not explain the positive relationship between employer size and wage. In addition, for industries that display size-wage premium, workers in large establishments are more likely to receive on-the-job training but their return to training is smaller. Our theory, substantiated by the new empirical evidence, suggests that it is not large firms, per se, but firms that hire better workers who are paying a wage premium.

JEL Classification: D83, J31

Keywords: imperfect information, sorting, on-the-job training, size-wage premium

Corresponding author:

Shuaizhang Feng
411A Robertson Hall
Woodrow Wilson School
Princeton University
Princeton, NJ 08544
USA
E-mail: shuaizha@princeton.edu

^{*} We thank John Abowd, Francine Blau, Charlie Brown, Hongbin Cai, Yuyu Chen, Hank Farber, Peter Gottschalk, Larry Kahn, Alan Krueger, Lars Lefgren, Alan Manning, Bruce Meyer, Jesse Rothstein and seminar participants at Cornell University ILR, George Washington University, Peking University and Princeton University IRS Labor Lunch for comments. All remaining errors are our own.

1 Introduction

Employer size-wage premium is a well-known puzzle in labor economics.¹ It is perplexing that larger firms pay observationally equivalent workers higher wages than their smaller counterparts, which seems to violate the law of one price. Our inability to explain the size-wage relationship suggests that something important is missing in the understanding of labor market.

Recently, several new hypotheses have been proposed to explain the positive relationship between employer size and wage. These include the training hypotheses (e.g., Troske 1999, Zabojnik and Bernhardt 2001 and Oi and Idson 1999), on-the-job search model (e.g., Burdett and Mortensen 1998), and the imperfect information model (Feng and Zheng 2009). All of these theories generate positive size-wage relationship. Therefore, to examine their respective explanatory powers, it is necessary to explore additional empirically testable implications.

To do so, this paper first develops a two-period labor market model incorporating both employer learning in an environment of imperfect information and workers' human capital accumulation on the job. While the static model of Feng and Zheng (2009) generates both size-wage premium and (imperfect) positive sorting of worker type with firm size, to subject the idea of imperfect information to strict empirical scrutiny, a dynamic version of the model is called for. Over time, one may expect the size-wage premium to decline as information is gradually revealed. But due to training and other forms of human capital accumulations, heterogeneities in productivity among individual workers may increase, resulting larger size-wage premium. An empirically tractable model thus has to include both elements.

In the two-period model developed here, the first period is similar to the one-period model of Feng and Zheng (2009). The competitive labor market consists of heterogeneous workers and ex ante identical firms with only partial knowledge of worker type. The model has two types of equilibria, those that display positive size wage relationship and those that do not. In any equilibrium, firms posting higher wages hire

¹See e.g., Brown and Medoff (1989) and our literature review in Section 2.

workers who are more productive on average. As firms do not know worker's true type for sure, workers of the same productivity are paid differently if they are hired by different firms. Thus there exists wage dispersion conditional on worker type.

In the second period, incumbent firms evaluate worker's performance in the first period, and update beliefs about worker productivity. They then provide firm-specific training to workers they think are of high ability, but provide no training to those they believe are of low ability. All workers are paid according to their expected pre-training productivity. In equilibria that display positive size-wage premium, large firms train a higher proportion of workers, but the size-wage differential for trained workers is also smaller than that for untrained workers, i.e., return to training is smaller in large firms. In contrast, in equilibria that do not display positive size wage premium, neither of the two predictions holds.

We next conduct empirical analyses using data from the National Longitudinal Survey of Youth 1979 Cohorts (NLSY79), which contains extensive training information. As a preliminary test we examine the existence of size-wage premium for each industry individually. The preliminary analysis shows that the size-wage relationship is not uniform. Without controlling for unobserved personal heterogeneities, we find that some industries do not display positive size-wage relationship, and a few even have a negative relationship. Consistent with the theoretical model, for industries that do not display positive size-wage premium, there is no evidence of positive sorting between firm size and worker characteristics such as education and Armed Force Qualification Test (AFQT) score.

The main test focuses on industries that do display significant positive size-wage premium. It shows that differences in training cannot explain the existence of size-wage premium. In addition, return to training is lower in large establishments than in small establishments. With our preferred fixed effects specification, we find that return to training is lower by 7.5% in establishments with at least 500 employees than in smaller ones.² The probability of receiving training is also higher in large

²We have also used 250 and 100 as the cutoff points for large establishments, as well as natural log

establishments. These findings are consistent with the theoretical predictions.

As a falsification test, we also examine industries that do not show positive size-wage premium. After controlling for training and its interaction term with firm size, we find no evidence that return to training is smaller in large establishments. With control on unobserved person effect, the probability of receiving training is not higher in large establishments either. Our falsification test therefore indicates a link between lower return to training in large establishments and the existence of size-wage premium and sorting.

To summarize, our theory, as well as the empirical evidence suggests that it is not large firms per se, but firms hiring better workers who are paying a wage premium. We believe the same intuition sheds light on other forms of wage dispersion, such as inter-industry wage differential (e.g., Krueger and Summers 1988) and exporter-wage premium (e.g., Bernard and Jensen 1995, 1997). The model we have here can be easily extended to show that high productivity industries or exporters offer higher wages to attract more productive workers, but also hire some less productive workers due to imperfect knowledge of worker types. This generates both wage dispersion and sorting of workers across different industries, or between exporting and non-exporting firms.

The rest of the paper is organized as follows. Section 2 reviews related literatures. Section 3 presents the theoretical model. Section 4 examines empirical evidence from the NLSY79 data. The relevance of our new empirical findings to alternative theories is examined in section 5. Section 6 concludes. Appendix A collects proofs of the main results. Appendix B presents results of the falsification test.

of establishment size as a continuous measure. Results are robust to these alternative specifications.

2 Literature Review

2.1 Employer Size-Wage Effect

Moore (1911) first documented wage gains associated with working in large firms or establishments using data from Italian textile mills. Similar findings were subsequently reported for U.S. (see Lester 1967, Mellow 1982, Personick and Barsky 1982, Brown, Hamilton and Medoff 1990, Idson and Feaster 1990), the European countries (see Main and Reilly 1993, Abowd et al. 1999), and also for developing countries (see Velenchik 1997, Schaffner 1998, and Söderbom et al. 2005). The magnitude of size-wage premium is substantial. To put it 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.”

Many explanations have been advanced to rationalize the positive size-wage relationship within the perfect competition framework. Brown and Medoff (1989) have empirically examined six potential explanations, but only found modest support for the worker heterogeneity hypothesis. The five other explanations considered are: large firms offer inferior working conditions; large firms make greater use of high wages to forestall unionization; large firms have stronger ability to pay high wages; large firms face smaller pools of applicants relative to vacancies; large firms are less able to monitor their workers. In their conclusion, Brown and Medoff (1989) also hypothesize that the employer size-wage relationship may actually be a relationship between firm age and wage since large firms are usually old firms. However, subsequently Brown and Medoff (2003) have showed the firm age-wage relationship is explained by observed worker characteristics (see also Appendix A of Oi and Idson 1999).

Recently, some researchers have suggested size-wage premium is a result of productivity differences between workers in large firms and small ones that have not

been adequately controlled for in previous studies. Troske (1999) argues that both large employers and their employees are more likely to invest in firm-specific human capital. Thus, workers in large firms or establishments are paid more because they receive more firm-specific training. Similarly, Zabojnik and Bernhardt (2001) have showed corporate tournaments can induce workers in large firms to accumulate more general human capital through training. The productivity hypothesis in Oi and Idson (1999) is similar in spirit - workers in large firms are more productive because the production processes are organized differently in large and small firms. As productivity differences associated with size are endogenously determined, fixed effect approach is not adequate for resolving the issue.³ Rather, it is necessary to control for these differences, such as training workers received, with suitable data sets. Because of data limitation, this has not been done in previous studies.

Another prominent explanation of size-wage premium is the on-the-job search model of Burdett and Moretensen (1998). According to them, workers do not fully know potential jobs, thus have to either wait for a time period or incur a direct cost to sample from the pool of job offers. Workers search randomly and gradually move from low paying jobs to high paying jobs. In equilibrium, identical firms offer different wages. Those offering higher wages attract more workers but at the same time realize lower per worker profit.

2.2 Return to On-the-job Training

On-the-job training is considered as one of the most important channels through which workers accumulate human capital. Becker (1964) makes the critical distinction between general training and firm-specific training. General training increases workers' productivity in all firms, thus can only be financed by workers themselves. In contrast, firm-specific training only increases worker's productivity at the incumbent firm, and is financed by workers and firms jointly.

Becker's theory has been challenged by the empirical finding that, in many cases,

³See Gibbons and Katz (1992) for a similar argument.

firms do pay for general training (e.g., Krueger 1993, Autor 2001). New models have been developed explicitly incorporating labor market imperfections which compress the wage structure and allow firms to benefit from offering technically general training to their workers (Acemoglu and Pischke 1999). Therefore, the predictions on how general and firm-specific training affect wages are less clear-cut as many general trainings have become *de facto* firm-specific due to frictions in the labor market.

While economists have long been interested in knowing the magnitude of return to training, i.e., the effect of training on wage rates, for both theoretical and policy reasons, getting an estimate remains an empirical challenge. Unlike full-time schooling, training happens after workers enter the labor market, and is difficult to measure.⁴

Earlier empirical studies on return to training include Mincer (1983; 1988), Brown (1983; 1989), and Lillard and Tan (1992). However, all these studies use some proxies, not precise measure of actual training workers received. Thus, we only review later studies based on NLSY79.⁵ None of the studies, however, have examined whether returns to training differ for workers in different-sized firms or establishments.⁶

Lynch (1992) uses the 1979-1986 waves of NLSY79 to study the impact of training on wages. Her finding shows that on-the-job training significantly increases wages. Using later waves of NLSY79, Veum (1995) has found that while the incidence of on-the-job training is positively associated with higher wage level and wage growth, training duration has no effect on either. Using all waves of NLSY79 data up to 2000, Frazis and Loewenstein (2005) have found training has a large effect on wages, several times higher than the return to formal education, even after differences in wage growth, promotion, and measurement errors have been controlled for. They have also showed that using a single dummy for training incidence produces almost the same fit and median effect as including both training incidence and durations

⁴See Barron, Berger and Black, (1997) for a discussion of informal training, for example.

⁵Frazis and Loewenstein (2005) provide the most recent survey of empirical findings on on-the-job training from the NLSY79.

⁶Bishop (1997) provided some evidence that productivity gains from training for workers in large establishments are substantially larger than their counterparts in small establishments.

(Frazis and Loewenstein 2005, Table 2).

3 Theoretical model

In this section, we extend the cherry picking model of Feng and Zheng (2009) to incorporate information updating and on-the-job training. The economy consists of infinite number of ex ante identical firms, and a continuum of workers, α proportion being of high productivity type (type “ H ”) and $1 - \alpha$ being of low productivity type (type “ L ”). A high type worker can produce one unit of output if employed, while a low type worker can not produce any output.

Worker types are unknown to all parties initially, but firms and workers learn worker type over time.⁷ 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).

Workers are risk-neutral and maximize total wages received without discounting the second period wage. There is no cost associated with job application on the worker’s side and with hiring/firing on the firm’s side. As we do not allow for long term contracts, in each period, workers work for the firm that offers the highest wage, provided the highest wage is greater than their reservation wage $r(t)$, where $r(t) \in (0, \alpha)$. For simplicity, let $r(t) = r$ for $t = 1, 2$, i.e., they have the same reservation wage in both periods.

3.1 The first period

At the beginning of the first period, workers apply to all firms. While not knowing a worker’s type for sure, each firm makes a private assessment on job applicants,

⁷Allowing workers to know their own types does not change our main results, but will make the analysis much more complicated.

labeling a job applicant as either “*h*” or “*l*” type.⁸ A high type worker will be labeled as “*h*” type with probability β_H and as “*l*” type with probability $1 - \beta_H$. A low type worker will be labeled as “*l*” type with probability β_L and as “*h*” type with probability $1 - \beta_L$. Private assessments by different firms are independent, thus a worker labeled as “*l*” type by one firm may be taken as “*h*” type by another. For simplicity we let $\beta_H = \beta_L = \beta$, where $\beta \in (1/2, 1)$.

The hiring process is sequential, consisting of many rounds of auctions in which firms compete for the right to make offers. Each round, firms that have not hired post wages they are willing to offer to some workers in the remaining pool of applicants. The one with the highest wage wins the right to hire. In case several firms tie at the highest wage, one firm is randomly chosen as the winner. The winning firm then makes job offers to a positive measure of workers in the remaining pool at its posted wage. Workers who receive an offer then decide whether to accept it. After that the game moves on to the next round with the same process repeated for firms who have not hired and for workers who have not accepted any offer. Job market closes when no remaining firms finds it profitable to post a wage that will be accepted. Production then begins and firms realize their first period profit.

3.2 The second period

At the beginning of the second period, incumbent firms privately evaluate worker’s performance in the first period. Because of observational errors, the performance measure will not perfectly reflect a worker’s true performance. Conditional on a worker being of high type and thus performs well, an incumbent firm has a good evaluation about his performance, denoted as “*g*” with probability ρ_H , and has a bad evaluation, denoted as “*b*” with probability $1 - \rho_H$. Conditional on a worker being of low type and thus does not perform, the incumbent firm has a “*g*” evaluation with probability $1 - \rho_L$ and “*b*” with probability ρ_L . For simplicity we assume that $\rho_H = \rho_L = \rho \in (1/2, 1)$.

⁸For example, firms may gain some information on job applicants from their job application materials and interviews.

Given the performance evaluations, the incumbent firm then updates its belief about worker types using Bayes rule. Thus, if firm k 's prior belief of a worker being of high type is H_k , its posterior belief μ_k^b after a “b” evaluation will be

$$\mu_k^b = \frac{H_k(1 - \rho)}{H_k(1 - \rho) + (1 - H_k)\rho} < H_k; \quad (1)$$

and its posterior belief μ_k^g after a “g” evaluation will be

$$\mu_k^g = \frac{H_k\rho}{H_k\rho + (1 - H_k)(1 - \rho)} > H_k. \quad (2)$$

With this information, the incumbent firm makes wage offers to current employees and provide firm-specific training to selected workers. Training increases high type worker's productivity but not the low type's. After training, a high type worker can produce $q > 1$ unit of output, instead of one without training. A low type worker's output will be zero irrespective of his training status.⁹ Training is firm-specific and thus, training received at one firm does not increase worker's output at another.¹⁰ It is also costly: firms incur a cost of c to train one worker. Clearly, if the per worker training cost is too high, for example, $c > q - 1$, it would not be profitable to train any worker. On the other hand, if it is too low, for example, $c = 0$, it would be profitable to train all workers. To ensure training decision is nontrivial, we henceforth make the following assumption:

Assumption 1.

$$\frac{\rho(q - 1)\alpha\beta}{\rho\alpha\beta + (1 - \rho)(1 - \alpha)(1 - \beta)[\beta/(1 - \beta)]^{\kappa-1}} > c > \frac{\alpha\beta(1 - \rho)(q - 1)}{\alpha\beta(1 - \rho) + (1 - \alpha)(1 - \beta)\rho}.$$

⁹The assumption that training and worker ability are complements is a standard one in the literature, see e.g. Acemoglu and Pischke (1998).

¹⁰It is true that many training programs are general in the sense that skills taught are also useful in other firms. However, as noted in Acemoglu and Pischke (1999), in a non-competitive environment, labor market imperfections turn general skills into *de facto* specific skills. One can thus assume general training but incorporate some additional labor market frictions, such as matching and search, in the model to get similar results as we have here. Alternatively, Lazear (2010) provides a skill-weights interpretation to firm specific human capital.

That is, there are at least $\kappa > 0$ number of firms who would find it profitable to train workers with good performance evaluations, but no firm finds it profitable to train those with bad evaluations.

Outside firms have no direct contact with workers not currently working with them, and gain no extra information about those workers. Learning is asymmetric here, as in many previous studies, for example, Schönberg (2007) and Golan (2005, 2006).¹¹ The only difference is that learning is imperfect for the incumbent firm in our model, while it is perfect in theirs. 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 $\rho = 1$.

On the other hand, outside firms observe a worker's training status and wage offer from the incumbent firm in the second period, as well as the identity of firms (hence the average productivities of their workers) they are affiliated with.¹² Based on this information, outside firms can also make an offer to them. Workers choose the firm that offers the highest wage, and stay with the current employer if its offer ties with the highest outside offer. Production then starts, and firms realize their profits. After that, the game ends with all workers exiting the labor market.

3.3 Main results

Like the one period model of Feng and Zheng (2009), the two-period model has multiple equilibria. While wage dispersion is a distinct feature of all equilibria, not all exhibit positive size-wage differential. This is so as in any equilibrium, firms expect zero profit from each hire and can hire any portion of workers they think are of high type. In some equilibria, a firm that hires earlier (paying higher wages) will have a smaller workforce than one that hires later (paying lower wages). As our main

¹¹Alternatively, assuming outside firms gain the same information as incumbent firms, i.e., there is symmetric learning, does not change our results. This is because in our setup incumbent firms move first and are not allowed to make counter offers.

¹²We assume firms have short memory and do not remember the first-period assessments they may have on workers who have applied to but were not hired by them.

concern is the size wage premium puzzle, we will focus on equilibria that exhibit positive size-wage relationship. We therefore make the following assumption:

Assumption 2. *Firms treat observationally identical workers the same way.*

This assumption requires a firm to make the same offer to all workers who are observationally equivalent, i.e., those with the same expected productivity from the firm's perspective. This helps to eliminate the type of equilibria in which firms offering higher wages may hire a small portion of workers they think are of high type while those offering lower wages hire a large proportion.

Proposition 1. *Under assumptions 1 and 2, there exists 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 with a good performance evaluation receive firm-specific training.*
- 3. Trained workers receive higher wages than untrained workers in the same firm.*
- 4. Controlling for training status, large firms still pay more to workers than small firms.*

In the first period, firms make two choices at the time of hiring: to post a wage, and decide whom to hire after winning the right to hire. They face two observationally different groups of workers: those labeled as “ h ” type and those labeled as “ l ” type. Because of competition between ex ante identical firms, in equilibrium, winning wage in any round is equal to the sum of average productivity of “ h ” type workers and expected second-period per worker profit from those workers. Consequently, only workers labeled as “ h ” type will be hired. Under Assumption 2, each firm hires every worker labeled as “ h ” type. In this case, the average productivity of workers at firm k

who hires in the k -th round is:

$$H_k = \frac{\alpha\beta(1-\beta)^{k-1}}{\alpha\beta(1-\beta)^{k-1} + (1-\alpha)\beta^{k-1}(1-\beta)}. \quad (3)$$

The total measure of workers at firm k equals $N_k = \alpha(1-\beta)^{k-1}\beta + (1-\alpha)\beta^{k-1}(1-\beta)$. Since both H_k and N_k are decreasing in k , there is a positive relationship between firm size and wage.

Because training is firm-specific, firms will get all the surplus Π_k^2 from training. Zero profit condition therefore implies the first-period wage at firm k equals $H_k + \pi_k^2$, where $\pi_k^2 = \Pi_k^2/N_k \geq 0$ is firm k 's expected per worker profit from the second period.

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 labels as “ h ” type falls below the reservation wage r . This implies that in equilibrium, the number of hiring firms K is determined by the condition:

$$H_K + \pi_K^2 \geq r > H_{K+1} + \pi_{K+1}^2. \quad (4)$$

While incumbent firm has an informational advantage over outside firms, this does not translate into any monopsony power. This is so as outside firms know the average productivity of the incumbent's labor force, and can make offers after observing the its offer. Should the incumbent firm's wage offers are on average less than the expected pre-training productivity, outside firms can profitably outbid the incumbent and attract its workers away. In this sense, outside firms have a strategical advantage over the incumbent firm even though the latter has an informational advantage. This helps level the playing field for the two sides and results in zero pre-training profit for the incumbent firm in the second period.¹³

As incumbent firms pay workers according to their expected pre-training productivities, and training is firm-specific, it is profitable to train a worker as long as the

¹³Of 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 will not be the case.

expected increase in productivity more than offsets the training cost. In equilibria with positive size-wage differential, large firms have a higher proportion of productive workers; they therefore also train more of their workers.

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

Note 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 . Hence, the proportion of trained workers is higher in larger firms than in smaller firms.

Corollary 2. *Log wage differential is smaller for trained workers than for untrained workers, i.e., return to training is smaller in large firms.*

This result says that even though trained workers at a larger firm get higher pay than their smaller firm counterparts, the log wage differential is smaller for trained workers than for untrained workers. Empirically this would imply a lower return to training for workers at larger firms. The intuition for this result is as follows. Initially firms pay workers the average productivity of all workers. Because of cherry-picking, the average productivity at larger firms is higher than that at smaller firms. In the second period, as firms learn more about their workers, they update beliefs about worker productivity and adjust wages accordingly. This results in a decline of wage differentials associated with employer size for those workers viewed as high type. The declined wage differential for trained workers reflects learning by employers who only have imperfect knowledge of workers' productivity at the time of hiring.

REMARK: Note that without assumption 2, there exist equilibria that do not exhibit size-wage premium, as firms can hire any proportion of workers labeled as “h” type. In those equilibria, neither implications, corollary 1 and corollary 2 may hold.

For example, let firm i and j hires in the k -th and $(k + 1)$ -th round, respectively. Clearly $w_i > w_j$. If firm i hires only a very small portion of workers labeled as “h”,

while firm j hires all labeled as “ h ”, then it is possible that the measure of workers hired by firm i (N_i) is smaller than that of firm j (N_j). In this case, there is a negative relationship between employer size and wage. Also, the proportion of trained workers for firm i will be larger than in firm j , as the proportion of trainable (high type) workers is larger in firm i . Similarly, return to training is lower in firm i , which is now the smaller firm. Therefore, neither predictions holds when size-wage premium does not exist.

These predictions suggest that we can compare industries that do display size-wage premium with those that do not. If our explanation of the size-wage differential is correct, we would expect to observe the patterns stated in the two corollaries to appear only in industries that display size-wage premium, but not those that do not.

4 Empirical Results

4.1 Data

To test the predictions of the theoretical model, we conduct empirical analyses using data from National Longitudinal Survey of Youth 1979 Cohorts (NLSY79). Respondents in NLSY79 were interviewed annually from 1979 to 1994, and biannually since 1994. NLSY79 includes a nationally representative cross-sectional sample, a supplemental sample, and a military sample. In this paper, we restrict our analyses to the cross-sectional sample, and focus on the period 1986-2000, as there were no establishment size information for the period of 1981-1985, and some survey questions were restructured after 2000. To get a relatively homogeneous group of individuals, we consider only white males working full time (defined as 35 hours work per week and above) at nonunion jobs. Self-employed individuals are also excluded.

NLSY79 provides detailed training information. Like previous studies (e.g., Frazis and Loewenstein 2005), we focus on completed on-the-job training spells,¹⁴ and use

¹⁴A training spell is defined as on-the-job 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

a training dummy (training incidence) rather than actual training hours in order to be consistent with the theoretical model. However, our results remain unchanged if actual training hours are also included.¹⁵

We use establishment size information, taken from the question “number of employees at location of current job” or “number of employees at location at respondent’s job number 1”. Firm sizes were not available from the survey,¹⁶ but previous studies have shown that firm size and establishment size are highly correlated (Brown and Medoff, 1989). We follow the literature and use 500 employees as the cutoff point to define establishments as either large or small. Nevertheless, none of our qualitative conclusions changes when alternative cutoff points (250 and 100) or the natural log of establishment size are used.¹⁷

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 or less than \$1. In most years the loss of sample size is less than 2%. For each sample, the following information is recorded: year of tenure, total labor market experience, union status on the job, industry, occupation, part-time status, establishment size, local unemployment rate, on-the-job training incidence and hours, and off-the-job training incidence and hours. Variables such as education, marriage status, and AFQT score are also included.

by employer (type=9)” during the 1988-2000 surveys. The 1987 wave of NLSY79 does not contain any training questions. Instead, training happened in 1987 were recorded in the 1988 survey.

¹⁵Veum (1995) and Frazis and Loewenstein (2005) all find that when training incidence is controlled for, hours of training is no longer statistically significant.

¹⁶For workplaces within the firm, but other than where the employee is working from, NLSY79 only asked for whether the number of employees is above 1,000.

¹⁷These results are available upon request.

4.2 Preliminary analyses

As a first step, we use the NLSY79 data to reproduce the size-wage effect documented in the literature, without controlling for training status. To do so, we run OLS regressions on large establishment size dummy (500+) and other control variables for each industry separately. The results, as reported in Table 1, show that size-wage effects vary substantially by industry. For our main empirical analyses, we only use industries that show a positive and significant size-wage premium, including manufacturing, transportation, communication and other public utilities, wholesale and retail trade, business and repair services, and public administration.¹⁸ These industries account for 63% of the total sample. Other industries as reported in Panel B of Table 1 are used in falsification tests in Appendix B.¹⁹

There is some evidence that the existence of positive size-wage premium is closely related to sorting among worker characteristics. Table 1 shows that most industries in Panel A display positive sorting on college education and AFQT score, i.e., on average workers in large establishments are more likely to have at least college education and have higher AFQT scores. In contrast, those industries shown in Panel B of Table 1 typically display lower degrees of positive sorting. For two industries, health services and entertainment and recreation services, there is even negative sorting on both college education and AFQT. Not surprisingly, both industries also show negative size-wage relationship. This is consistent with our theoretical predictions: there is neither size-wage premium nor positive sorting on worker type in equilibria in which firms do not treat identical workers similarly.

Sample summary statistics are given in Table 2. Sample size is 9,314, with 1,644 individuals. The sample mean of the natural log of hourly wage (in cents) is 6.71, which corresponds to a real wage rate of \$8.2 in 1982-84 dollar. In terms of education,

¹⁸We do so as our focus is to explore the size-wage puzzle. Also, our theory suggests the existence of multiple equilibria with and without positive size-wage relationship.

¹⁹Results are robust to minor changes in industry composition, such as including mining in our main analyses.

56% of the sample has a high school diploma or less, while 19% has at least a Bachelor's degree. On average, workers have 4.6 years of tenure with current employers and 11.7 years of total labor market experience. 83% of the sample works in establishments with less than 500 employees, while the rest works in large ones. Average wage rate at large establishments is substantially higher. Workers at large establishments earn 0.33 log points more than those at small establishments. In addition, workers at large establishments tend to be more productive, with more years of schooling, higher AFQT scores, longer tenure with current employers and more on-the-job training.

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

$$LNW_{it} = \gamma L_EST_{it} + X'_{it}\beta + \varepsilon_{it}. \quad (5)$$

Here, the subscript i stands for worker and t for year. The dependent variable LNW is the log hourly wage, L_EST is a dummy for large establishment, and X 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, and industry dummies.

We also use a panel data model with additional control on unobserved person effects,

$$LNW_{it} = \gamma L_EST_{it} + X'_{it}\beta + \phi_i + \varepsilon_{it}. \quad (6)$$

Here ϕ is the unobserved person effects (indexed by i) that do not change for a given worker. The model is estimated using both random effects and fixed effects approaches.

Table 3 reports estimated coefficients as well as robust standard errors clustered at the person level. We see that as usual, 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 13.3% more than those in small ones. When person effects are controlled, the estimated coefficient is reduced to

7.2% in the random effects specification, and to 5.4% in the fixed effects specification. Nevertheless, they are all statistically significant at 1% level. Thus, even though a substantial proportion of the size-wage effect is due to worker heterogeneity, the remaining part is still statistically significant and practically large. This is not different from findings of previous studies (e.g., Brown and Medoff 1989).

4.3 Main results

We now introduce on-the-job training into the wage equation and allow for different returns to training in different-sized establishments,

$$LNW_{it} = \alpha TR_{it} + \gamma L_EST_{it} + \theta TRL_{it} + X'_{it}\beta + \varepsilon_{it}. \quad (7)$$

The dummy variable $TR_{it} = 1$ if worker i has finished at least one on-the-job training spell on the current job by year t , $TR_{it} = 0$ otherwise. TRL_{it} is the interaction term of TR_{it} with L_EST_{it} . The parameter of primary interest here is θ , as we want to know whether returns to training differ for establishments of different size.

The first column of Table 4 gives the OLS results. The coefficient on L_EST is 14%, which is close to that in column 1 of Table 3, but now represents the wage difference between untrained workers in large and small establishments. Return to training is 8.2% in small establishments, while only 3.2% in large establishments. The coefficient of TRL equals -5% and is statistically significant at the 10% level.

One could argue that unobserved worker heterogeneity would bias the OLS estimates. Suppose that the omitted person effect is positively related to whether a person is trained or not (TR), and whether a person is employed in a large establishment (L_EST), then the OLS estimates of α and γ would be upward biased. However, a priori whether the estimate of θ is biased, or if biased, the direction of the bias is hard to know. To control for unobserved personal heterogeneities, we consider the following panel model:

$$LNW_{it} = \alpha TR_{it} + \gamma L_EST_{it} + \theta TRL_{it} + X'_{it}\beta + \phi_i + \varepsilon_{it}, \quad (8)$$

where ϕ is the unobserved person effect. The model is estimated using both random and fixed effects approaches, with results given in columns 2 and 3 in Table 4, respectively. In both cases, the coefficients for large establishment are substantially smaller than that in OLS. Return to training in small establishments is 5.3% in the random effect specification and 4.6% in the fixed effects specification, suggesting that workers are selected for training based on unobservable characteristics. Nevertheless, the difference between returns to training for small and large establishments becomes even larger, with point estimates at -6.6% and -7.5% in random and fixed effects specifications, respectively. Both are statistically significant at the 1% level. This indicates that return to training is substantially lower in large establishments even after person effects are taken into account.

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

$$LNW_{it} = \alpha TR_{it} + \gamma L_EST_{it} + \theta TRL_{it} + X'_{it}\beta + \phi_i + \eta_{J(it)} + \varepsilon_{it}, \quad (9)$$

where η indicates 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 are a good “match”, the productivity level would be higher than otherwise. This is in addition to the person effect captured by ϕ .

We estimate the model in two different ways: random effects at the job level and two-level mixed effects at both the person and job levels.²⁰ We do not consider the fixed effects approach as variations in establishment size within a job are primarily due to measurement errors.

Results are reported in the last two columns of Table 4. Again, return to training is smaller in large establishments by 5% and this difference is statistically significant

²⁰This approach is similar to the random effect specification, as error terms are still assumed to be orthogonal to explanatory variables. However, it allows for two levels (person and job spell) of unobserved effects. The model is estimated using XTMIXED in STATA. For more details, see Searle, Casella and McCulloch (1992).

at the 5% level. For the two-level mixed model, return to training is 3.3% in small establishments. The coefficient on the interaction term of training and establishment size is -5.3% and significant at the 1% level.

Our preferred specification is the fixed effects approach reported in column 3 of Table 4, which uses only within variations and is consistent even if person effects are correlated with the explanatory variables. There is strong evidence that return to training is lower in large establishments than in small ones based on the fixed effects results. The coefficient on *TRL* is -7.5% and significant at the 1% level with standard errors corrected for heteroscedasticity and clustered at the person level. One might worry about the measurement error problem. But in most cases measurement error would attenuate the coefficient estimate to zero. Overall we find strong empirical support for the theoretical predictions.

Table 5 reports results on probability of receiving training. We have run OLS, Logit, random effects Logit and fixed effects regressions, using training incidence as the dummy dependent variable.²¹ Average marginal effects are reported for Logit and random effects Logit in Table 5 so that results are all directly comparable. Our preferred estimate is again the fixed effects estimate, as person effects are allowed to be correlated with explanatory variables.²² Based on the fixed effects results reported in column 4 of Table 5, the probability of receiving on-the-job training for workers in large establishments is about 5% higher than their counterparts in small establishments. The coefficient is also significant at the 1% level based on robust standard errors clustered at the person level. The result is consistent with corollary 1, which states that large establishments train a higher proportion of workers.

²¹We could not perform fixed effect logit estimation, because the maximum likelihood function fail to converge in STATA.

²²In limited dependent variables models, linear probability specifications still estimate the average treatment effect correctly without making underlying latent variable assumptions, see discussions in Section 3.4.2 in Angrist and Pischke (2008).

4.4 Robustness checks

The above empirical test suggests that return to training is smaller in large establishments than in small ones, with the difference statistically significant and quantitatively important. Here we consider several robustness checks, and report the results in Table 6. Basic specifications are the same as in Table 4.

First, we include off-the-job training information. Although the focus of this paper is on-the-job training, it is useful to check whether adding off-the-job training information makes any difference. We aggregate all training spells not counted as on-the-job training into an off-the-job training variable. We also allow tenure effects to be different in large and small establishments, to capture possible different impacts of informal training not recorded in the data. As the results in panel A of Table 6 shows, there is no substantive changes about on-the-job training from results reported before. Based on the preferred fixed effects specification (column 2 of Table 6), the coefficient on the interaction term of on-the-job training and large establishments is -7.3% and significant at the 1% level. Other specifications give similar results. In addition, wage return to off-the-job training is essentially zero once individual heterogeneity is controlled for, and the interaction term of off-the-job training is not significantly different from zero either. The coefficient for large establishment remains large (7.6% based on the fixed effects specification) and highly significant in all specifications even after we control for all possibly differential impacts of formal and informal training.

Second, we consider the possible impact of measurement error in establishment size, which might be particularly problematic for fixed effects specifications. Using UK data, Manning (2003) finds substantial measurement error about establishment size in self-reported employee data. However, as first differencing attenuates the magnitude of all coefficients, the difference in returns to training between small and large establishments would be even larger if there were no measurement errors.²³ In

²³Barron, Berger and Black (1997) suggest that measurement errors in training variables are unrelated to establishment size.

panel B of Table 6, we exclude observations with establishment size falling between 475 and 525 and run the same set of regressions. In doing so we avoid using size variations that are near the 500 cutoff point which is liable to misreporting. Again, the results are essentially the same as those in Table 4; the coefficient on *TRL* is -7.2% and significant at the 1% level based on the fixed effects approach.

Next, we restrict the sample to the period of 1987-2000 and use only employer-paid on-the-job training spells. Training information is collected differently before and after 1987.²⁴ Also, NLSY79 allows us to identify who paid for on-the-job training in the 1987-2000 surveys. During this period of time, on average, over 90% of all on-the-job training spells are paid by employers (calculated based on person-year observations), and large establishments paid a higher percentage of training spells than smaller ones did. It is plausible that on-the-job training sessions paid and not paid by employers might have different contents and effects on wages.²⁵ In fact, Holtmann and Idson (1991) suggest that workers in large firms are paid more because they are more likely to receive employer-paid training, which implies that their starting wages are lower but subsequent wage profiles are steeper. Our results as shown in Panel C of Table 6 suggest that controlling for difference in proportions of employer-paid training does not change the main findings.

Lastly, to ensure our results are not driven by inexperienced workers who might have different levels of attachments to large and small establishments, we use workers with at least 5 years' labor market experience. Once again, results in panel D of Table 6 show no substantive changes from the main results. Coefficient on *TRL* is -8.2% and significant at the 1% level based on the fixed effects approach.

²⁴Despite 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 and up to two training spells that was still ongoing at last interview. 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 the latter period.

²⁵It is also likely that some of the on-the-job training spells are misclassified. Focusing on only employer-paid on-the-job training thus mitigates possible impacts from misclassifications.

5 Discussions

In above, we find that in industries displaying size-wage premium, workers in large establishments are more likely to receive on-the-job training, but enjoy lower return to training. In Appendix B, we conduct falsification tests and find that those empirical regularities do not hold for industries not displaying size-wage premium. This suggests that our main empirical findings cannot be explained by some uncontrolled factors that may differ by establishment size.²⁶ There is also some evidence that sorting on education and ability (measured by AFQT here) is related to the existence of size-wage premium (Table 1). Overall, the empirical evidence strongly support our theoretical predictions.

In light of the empirical evidence reported before, in particular the one on return to training, we discuss other explanations on size-wage premium. We focus on two alternative sets of theories, the training hypotheses and the on-the-job search model. However, we do not discuss explanations that have been rejected in previous studies such as Brown and Medoff (1989) and Troske (1999).

To start, we consider the firm-specific training hypothesis of Troske (1999). According to Troske (1999), large firms not only hire more skilled workers, but also “produce” more skilled workers by providing more firm-specific training. Hu (2003) studies the hiring decisions of large firms and then hypothesizes that firm-specific human capital might explain the size-wage effect. She also notes that it is hard to test this hypothesis directly, as data on firm-specific human capital investment and productivity are typically unavailable. The theoretic underpinning of this argument is Becker’s original analysis of firm-specific training. In Becker’s analysis, workers and firms share the costs and benefits of specific training. If workers in large firms receive more training, their starting wages would be lower but subsequent wage growth would be higher. Thus, workers in large establishments who have not received any training would be paid the same as their small establishment counterparts, or even

²⁶For example, large establishments may find it less costly to provide training to workers, and may provide more firm-specific training.

lower. Our empirical results, however, show positive size-wage relationship even for untrained workers.

The general human capital hypothesis of Zabojnik and Bernhardt (2001) is similar to the specific training hypothesis and does not imply a size-wage premium for workers who do not receive any training, which is not supported by the empirical evidence from the NLSY data. Further, it can not explain the finding of lower return to training in large establishments either.

Next we consider the productivity hypothesis of Oi and Idson (1999). According to them, production process in large and small firms are organized differently, so that similar workers might have different levels of productivity at different firms, which is consistent with the observed size-wage differential. The explanation is different from a simple selection story that could in principal be tested with control on constant person effects, as productivity differences are endogenous. Although one can never control for productivity differences perfectly, productivity hypothesis alone cannot explain why return to training is lower in large establishments. As training and worker productivity are complements, other things equal, one would expect productivity gain for workers in large firms to be bigger. This has been confirmed empirically by Bishop (1997), who reported that increases in productivity are bigger for workers who receive training in larger establishments.²⁷ In a competitive setting this would translate into higher wage increases for trained workers in large establishments, not the opposite. Thus productivity differences alone cannot be the whole answer for the size-wage puzzle.

Therefore, the empirical findings in this paper suggest that one cannot easily rationalize the existence of size-wage premium using differences in training and other unobserved and possibly size-dependent worker heterogeneities. One has to go beyond the perfect competition paradigm and consider labor market imperfections directly.

²⁷The complementarity between training and worker productivity is also demonstrated by the empirical finding that firms usually select better workers to receive on-the-job training (Lynch, 1992, Lillard and Tan, 1992).

The most prominent “non-neo-classical” theory of size-wage premium is the on-the-job search model of Burdett and Mortensen (1998). Although both search and our imperfect information explanations rely on some form of labor market imperfections, the underlying intuition is quite different. While search explanation suggests that large firms pay more in order to attract more workers, our imperfect information story postulates that large firms pay more in order to attract better workers. Search model has not taken account of the correlation of sorting and size-wage premium across industries. Further, it can not explain the negative size-wage differential.

So far, search models have not provided a direct answer to what would happen to return to training in different sized firms or establishments. For example, Quercioli (2005) has extended the Burdett-Mortensen model by allowing firms to provide specific training. But workers in the same firm receive the same amount of training in equilibrium. Thus, it is impossible to compare wages for workers with and without training. Overall, it is not obvious that search frictions alone would generate the empirical findings reported in this paper.

6 Conclusion

This paper extends the Feng and Zheng (2009) model, which has proposed a new imperfect information explanation to the size-wage premium puzzle. We incorporate employer learning and on-the-job training in a two-period setting, and generate two types of equilibria, one with size-wage premium and the other without. In all equilibria with size-wage premium, there is also positive sorting of worker types with firm size. Large employers train a higher proportion of workers but return to training at large firms is lower. This is not true in equilibria without size-wage premium.

We then use data from NLSY79 to test predictions of the model. We find that training does not explain the relationship between employer size and wage. The empirical results also show that, for industries that display size-wage premium, workers in large establishments have a higher probability of receiving on-the-job training but

lower return to training. Consistent with the theory, these empirical regularities do not appear in industries that do not display size-wage premium.

Our theory suggests that it is not necessarily large firms, but firms that hire better workers, that are paying a wage premium. We believe this sheds light on other forms of wage dispersion, such as inter-industry wage differentials and exporter-wage premium. Future research should provide more direct empirical evidence on the relationship between sorting and wage premium.

References

- Abowd, J., F. Kramarz, and D. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67, 251–333.
- Acemoglu, D. and J. Pischke (1998). Why do firms train? theory and evidence. *Quarterly Journal of Economics* 113, 79–119.
- Acemoglu, D. and J. Pischke (1999). Beyond becker: Training in imperfect labour markets. *Economic Journal* 109, 112–142.
- Angrist, J. and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Autor, D. (2001). Why do temporary help firms provide free general skills training. *Quarterly Journal of Economics* 116, 1409–1448.
- Baron, J. M., M. C. Berger, and D. A. Black (2006). Selective counteroffers. *Journal of Labor Economics* 24, 385–409.
- Barron, J., M. Berger, and D. Black (1997). How well do we measure training? *Journal of Labor Economics* 15, 507–528.
- Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*. New York: NBER.

- Bernard, A. B. and B. J. Jensen (1995). Exporters, jobs, and wages in us manufacturing: 1976-87. *Brookings Papers on Economic Activity: Microeconomics*, 67-112.
- Bernard, A. B. and B. J. Jensen (1997). Exporters, skill upgrading and the wage gap. *Journal of International Economics* 42, 3–31.
- Bishop, J. H. (1997). What we know about employer-provided training: a review of the literature. *Research in Labor Economics* 16, 19–87.
- Brown, C., J. Hamilton, and J. Medoff (1990). *Employers large and small*. Cambridge, MA: Harvard University Press.
- Brown, C. and J. Medoff (1989). The employer size-wage effect. *Journal of Political Economy* 97, 1027–1059.
- Brown, C. and J. Medoff (2003). Firm age and wages. *Journal of Labor Economics* 21(3), 677–697.
- Brown, J. (1989). Why do wages increase with tenure? *American Economic Review* 79, 971–991.
- Brown, J. N. (1983). Are those paid more really no more productive? measuring the relative performance of tenure as on-the-job training in explaining wage growth. Princeton Industrial Relations Papers.
- Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review* 39, 257–273.
- Feng, S. and B. Zheng (2009). Cherry-picking in labor market with imperfect information. IZA Discussion paper NO. 4309.
- Frazis, H. and M. Lowenstein (2005). Reexamining the returns to training: Functional form, magnitude, and interpretation. *Journal of Human Resources* 40(2), 453–476.

- Gibbons, R. and L. Katz (1992). Does unmeasured ability explain inter-industry wage differentials? *Review of Economic Studies* 59(3), 515–535.
- Gibbons, R. and M. Waldman (1999). A theory of wage and promotion dynamics inside firms. *Quarterly Journal of Economics* 114, 1321–58.
- Golan, L. (2005). Counteroffers and efficiency in labor markets with asymmetric information. *Journal of Labor Economics* 23, 373–93.
- Golan, L. (2006). Wage signaling: a dynamic model of intra-firm bargaining and asymmetric information. working paper.
- Holtmann, A. and T. Idson (1991). Employer size and on-the-job training decisions. *Southern Economic Journal* 58(2), 339–355.
- Idson, T. and D. Feaster (1990). A selectivity model of employer-size wage differentials. *Journal of Labor Economics* 8(1), 99–122.
- Krueger, A. B. (1993). How computers have changed the wage structure: evidence from microdata, 1984-1989. *Quarterly Journal of Economics* 108, 33–60.
- Krueger, A. B. and L. H. Summers (1988). Efficiency wages and the inter-industry wage structure. *Econometrica* 56, 259–93.
- Lazear, E. P. (2010). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy*, forthcoming.
- Lester, R. (1967). Pay differentials by size of establishment. *Industrial Relations* 7, 57–67.
- Lillard, L. A. and H. W. Tan (1992). Private sector training: Who gets it and what are its effects. *Research in Labor Economics* 13, 1–62.
- Lynch, L. (1992). Private-sector training and the earnings of young workers. *American Economic Review* 82(1), 299–312.

- Main, B. G. and B. Reilly (1993). The employer size-wage gap, evidence from Britain. *Economica* 60(238), 125–142.
- Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton, NJ: Princeton University Press.
- Mellow, W. (1982). Employer size and wages. *Review of Economics and Statistics* 64, 495–501.
- Mincer, J. (1983). Union effects: Wages, turnover and job training. *Research in Labor Economics* 5, 217–52.
- Mincer, J. (1988). Job training, wage growth and labor turnover. NBER working paper No. 2090.
- Moore, H. (1911). *Laws of Wages*. Macmillan, New York, reprinted 1967.
- 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. North-Holland.
- Personick, M. and C. Barsky (1982). White-collar pay levels linked to corporate work force size. *Monthly Labor Review* 105, 23–28.
- Quercioli, E. (2005). Training, turnover, and search. *International Economic Review* 46, 133–43.
- Schaffner, J. (1998). Premiums to employment in larger establishments: evidence from Peru. *Journal of Development Economics* 55(1), 81–113.
- Schönberg, U. (2007). Testing for asymmetric employment learning. *Journal of Labor Economics* 25(4), 651–691.
- Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components*. New York, NY: John Wiley and Sons.

- 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.
- 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.
- Velenchik, A. (1997). Government intervention, efficiency wages, and the employer size wage effect in zimbabwe. *Journal of Development Economics* 53(2), 305–338.
- Veum, J. (1995). Sources of training and their impact on wages. *Industrial and Labor Relations Review* 48(4), 812–826.
- Waldman, M. (1996). Asymmetric learning and the wage/productivity relationship. *Journal of Economic Behavior and Organization* 31, 419–29.
- 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.

Table 1: Preliminary evidence on size-wage premium by industry

Industry	Sample				Large Estab.		Small Estab.	
	Large Estab.	R^2	Size	College	AFQT	College	AFQT	
A: Industries displaying significant positive size-wage premium								
Manufacturing	0.144*** (0.021)	0.50	3564	0.35	64	0.18	55	
Transport., Commun., & Other Public Util.	0.107*** (0.042)	0.36	1087	0.33	64	0.11	49	
Wholesale Trade	0.258*** (0.059)	0.42	605	0.41	62	0.17	56	
Retail Trade	0.084** (0.042)	0.31	2059	0.27	56	0.11	54	
Business & Repair	0.124*** (0.047)	0.44	1385	0.44	72	0.18	55	
Public Administration	0.128*** (0.049)	0.46	614	0.34	69	0.2	59	
B: Industries displaying no significant positive size-wage premium								
Agriculture, Forestry, and Fisheries	-0.108 (0.123)	0.31	559	0.16	54	0.14	47	
Mining	0.104 (0.097)	0.48	164	0.23	59	0.11	38	
Construction	0.054 (0.043)	0.24	1938	0.04	43	0.04	41	
Finance, Insurance and Real Estate	0.019 (0.056)	0.50	855	0.57	83	0.44	69	
Personal services	0.086 (0.116)	0.37	246	0.12	50	0.15	46	
Entertainment and Recreation Services	-0.020 (0.113)	0.38	249	0.18	53	0.27	57	
Education Services	0.070 (0.068)	0.32	424	0.72	81	0.52	68	
Health Services	-0.188*** (0.071)	0.45	447	0.37	66	0.49	72	
Other Profess. Service	0.015 (0.075)	0.27	610	0.55	82	0.52	79	

Note: Dependent variable is log wage. Other control variables are also included.

S.E. robust to heteroscedasticity and clustered at the person level. ***, ** and * stand for significance at 1%, 5%, and 10% levels, respectively.

Table 2: Sample Means and Standard Deviations

	Whole Sample	Large Establishments (500+)	Small Establishments
Log wage	6.71 (0.52)	6.98 (0.48)	6.65 (0.51)
Education			
- Less than 12 years	0.10	0.07	0.11
- 12 years	0.46	0.35	0.48
- 13-15 years	0.25	0.21	0.25
- 16 years	0.16	0.29	0.13
- More than 16 years	0.03	0.08	0.02
AFQT	56 (27.3)	64 (27.8)	54 (26.9)
Local Unemployment Rate	2.6 (0.90)	2.5 (0.79)	2.7 (0.92)
Job Tenure	4.6 (4.6)	5.8 (4.9)	4.4 (4.5)
Labor Market Experience	11.7 (4.7)	12.2 (4.9)	11.6 (4.7)
On-the-job Training (incidence)	0.2	0.36	0.17
Establishment size ≥ 500	0.17	1	0

Note: Total sample size is 9,314. Standard deviations are given in parentheses for non-dummy variables.

Table 3: Wage Regressions Not Controlling for Training

	OLS	Random Effects, Person level	Fixed Effects, Person Level
LE (500+)	0.133*** (0.016)	0.072*** (0.012)	0.054*** (0.013)
Education			
- Less than 12 years	-0.072*** (0.024)	-0.063*** (0.024)	NA
- 13-15 years	0.141*** (0.022)	0.135*** (0.022)	NA
- 16 years	0.272*** (0.031)	0.303*** (0.031)	NA
- More than 16 years	0.344*** (0.060)	0.399*** (0.058)	NA
AFQT	0.002*** (0.0004)	0.002*** (0.0004)	NA
UNEMP	-0.030*** (0.007)	-0.029*** (0.005)	-0.028*** (0.005)
Job Tenure	0.036*** (0.004)	0.034*** (0.003)	0.031*** (0.003)
Job Tenure Squared	-0.001*** (0.0002)	-0.002*** (0.0002)	-0.002*** (0.0002)
EXP	0.033*** (0.005)	0.041*** (0.005)	0.045*** (0.006)
EXP Squared	-0.0004** (0.0002)	-0.0006*** (0.0002)	-0.0007*** (0.0002)

Note 1: Regressions based on main sample shown in Table 2. Sample size is 9,314.

Note 2: Dependent variable is log wage. 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, who work in a small-sized establishment.

Note 3: Standard errors (S.E.) in parentheses adjusted for heteroscedasticity and clustered at the person level.

***, ** and * stand for statistical significance at 1%, 5%, and 10% levels.

LE: large establishment

UNEMP: local unemployment rate

EXP: labor market experience

Table 4: Wage Regressions Controlling for Training

	OLS	Random Effects, Person Level	Fixed Effects, Person Level	Random Effects, Job Level	Mixed Effects, Person and Job Levels
LE (500+)	0.140*** (0.017)	0.087*** (0.013)	0.072*** (0.015)	0.073*** (0.012)	0.068*** (0.013)
OJT	0.082*** (0.019)	0.053*** (0.013)	0.046*** (0.014)	0.032*** (0.012)	0.033*** (0.012)
LE x OJT	-0.050* (0.029)	-0.066*** (0.021)	-0.075*** (0.022)	-0.050** (0.020)	-0.053*** (0.020)
Education					
- Less than 12 years	-0.069*** (0.024)	-0.062*** (0.024)	NA	-0.053*** (0.018)	-0.059** (0.028)
- 13-15 years	0.137*** (0.021)	0.133*** (0.022)	NA	0.142*** (0.016)	0.131*** (0.021)
- 16 years	0.270*** (0.031)	0.302*** (0.031)	NA	0.327*** (0.023)	0.316*** (0.028)
- More than 16 years	0.341*** (0.061)	0.399*** (0.058)	NA	0.432*** (0.047)	0.422*** (0.052)
AFQT	0.002*** (0.0004)	0.002*** (0.0004)	NA	0.002*** (0.0003)	0.003*** (0.0004)
UNEMP	-0.031*** (0.007)	-0.029*** (0.005)	-0.028*** (0.005)	-0.025*** (0.004)	-0.024*** (0.005)
Job Tenure	0.034*** (0.004)	0.033*** (0.003)	0.031*** (0.003)	0.033*** (0.003)	0.031*** (0.002)
Job Tenure Squared	-0.001*** (0.0002)	-0.002*** (0.0002)	-0.002*** (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0001)
EXP	0.033*** (0.005)	0.040*** (0.005)	0.044*** (0.006)	0.032*** (0.004)	0.036*** (0.004)
EXP Squared	-0.0004** (0.0002)	-0.0006*** (0.0002)	-0.0006*** (0.0002)	-0.0003* (0.0002)	-0.0004*** (0.0002)

Note 1 & 2: Same as Note 1 & 2 for Table 3.

Note 3: S.E. for the first three columns adjusted for heteroscedasticity and clustered at the person level. S.E. for the fourth column adjusted for heteroscedasticity and clustered at job level. S. E. for the fifth column (mixed effects results) have been adjusted for heteroscedasticity but not clustered. ***, ** and * stand for significance at 1%, 5%, and 10% levels, respectively.

OJT: on-the-job training (incidence).

Table 5: Probability of Receiving On-the-job Training

	OLS	Logit	Random Effects Logit, Person Level	Fixed Effects, Person Level
LE (500+)	0.122*** (0.021)	0.095*** (0.015)	0.042*** (0.009)	0.051*** (0.017)
Education				
- Less than 12 years	-0.048** (0.021)	-0.098** (0.038)	-0.089*** (0.028)	NA
- 13-15 years	0.053** (0.022)	0.048** (0.020)	0.050*** (0.016)	NA
- 16 years	0.016 (0.029)	0.011 (0.025)	0.024 (0.021)	NA
- More than 16 years	0.039 (0.060)	0.029 (0.045)	0.037 (0.037)	NA
AFQT	0.001* (0.0003)	0.001* (0.0004)	0.001** (0.0003)	NA
UNEMP	0.004 (0.007)	0.001 (0.007)	-0.002 (0.004)	0.002 (0.006)
Job Tenure	0.037*** (0.004)	0.035*** (0.004)	0.026*** (0.002)	0.034*** (0.004)
Job Tenure Squared	-0.001*** (0.0003)	-0.001*** (0.0002)	-0.001*** (0.0001)	-0.001** (0.0002)
EXP	0.012*** (0.004)	0.028*** (0.006)	0.029*** (0.004)	0.030*** (0.005)
EXP Squared	-0.0002 (0.0002)	-0.001*** (0.0002)	-0.001*** (0.0001)	-0.0008*** (0.0002)

Note 1: Regressions based on the main sample shown in Table 2. Sample size is 9,685.

Note 2: Dependent variable is incidence of on-the-job training. Other explanatory variables are dummies for region, marriage and industry. Base group: full-time white males with 12 years of schooling, work in a small-sized establishment.

Note 3: For OLS, Logit and Fixed Effects results, S.E. adjusted for heteroscedasticity and clustered at the person level. Random Effects Logit S.E. not clustered. ***, ** and * stand for significance at 1%, 5%, and 10% levels, respectively.

Note 4: Results reported for Logit and Random Effects Logit are average marginal effects.

Table 6: Robustness Checks

	OLS	Random Effects, Person Level	Fixed Effects, Person Level	Random Effects, Job Level	Mixed Effects Person and Job Levels
A: Adding Off-the-job Training and Allowing Tenure Effects to be different					
LE	0.123*** (0.023)	0.084*** (0.019)	0.076*** (0.020)	0.083*** (0.020)	0.074*** (0.018)
OJT	0.075*** (0.019)	0.053*** (0.013)	0.047*** (0.014)	0.031** (0.012)	0.034*** (0.013)
LE x OJT	-0.050 (0.031)	-0.069*** (0.022)	-0.073*** (0.023)	-0.047** (0.019)	-0.054** (0.022)
OFT	0.069*** (0.023)	0.003 (0.017)	-0.014 (0.019)	0.003 (0.018)	-0.005 (0.014)
LE x OFT	-0.081** (0.036)	0.006 (0.027)	0.029 (0.029)	0.025 (0.024)	0.028 (0.026)
B: Excluding Those with Establishment Size between 475 and 525					
LE	0.133*** (0.018)	0.082*** (0.015)	0.067*** (0.017)	0.072*** (0.013)	0.067*** (0.014)
OJT	0.082*** (0.019)	0.053*** (0.013)	0.046*** (0.014)	0.031*** (0.012)	0.033*** (0.013)
LE x OJT	-0.044 (0.030)	-0.061*** (0.023)	-0.072*** (0.024)	-0.048** (0.022)	-0.050** (0.022)
C: Restricting to Post-1987 years and employer-paid on-the-job Training					
LE	0.142*** (0.018)	0.089*** (0.013)	0.075*** (0.015)	0.072*** (0.012)	0.070*** (0.013)
OJT	0.079*** (0.019)	0.047*** (0.013)	0.039*** (0.014)	0.036*** (0.012)	0.034*** (0.013)
LE x OJT	-0.053* (0.029)	-0.073*** (0.020)	-0.083*** (0.021)	-0.052*** (0.019)	-0.057*** (0.021)
D: Restricting to those with at least 5 years' labor market experience					
LE	0.146*** (0.017)	0.090*** (0.013)	0.074*** (0.015)	0.072*** (0.012)	0.071*** (0.014)
OJT	0.081*** (0.019)	0.047*** (0.013)	0.039*** (0.014)	0.031** (0.012)	0.030** (0.013)
LE x OJT	-0.053* (0.029)	-0.072*** (0.021)	-0.082*** (0.022)	-0.052*** (0.020)	-0.058*** (0.021)

Note 1: Regressions based on the main sample (Table 2). Sample sizes vary for each panel. Panel A includes interaction terms of tenure with size. Other explanatory variables included are the same as in Table 4.

Note 2 & 3: Same as Note 2 & 3 for Table 4.

OFT: Off-the-job training.

Appendix A: Proofs of results in Section 3

In order to establish Proposition 1, we follow a backward induction approach, starting from the second period and then going back to the first period. In the second period, incumbent firms have private evaluations of workers' performance and update beliefs about their types accordingly. Given the updated belief, they then make wage offers to current employees and decide whom to train. Having observed workers' training status and wage offers received from the incumbent firms, outside firms can compete with incumbent firms for good workers. We begin by analyzing incumbent firm's wage offer and training decision in the second period.

Claim 1. *Incumbent firm makes zero pre-training profit from current employees in the second period.*

Proof. For workers with average pre-training productivity of H_k , incumbent firm k can distinguish between two subgroups after observing their performance in the first period. Bayes rule implies that the updated belief for workers it has a good evaluation is $\mu_k^g = H_k\rho/[H_k\rho + (1 - H_k)(1 - \rho)]$, and belief for those it has a bad evaluation is $\mu_k^b = H_k(1 - \rho)/[H_k(1 - \rho) + (1 - H_k)\rho]$.

Note first that firm k 's average wage offer to their workers in the second period cannot be lower than H_k . Suppose it is the case, then the firm would lose all its workers, as an outside firm can offer each worker ε more, and still be able to make a profit.

On the other hand, profit maximization implies that firm k 's average wage offer will not be higher than H_k . This is because it can always offer μ_k^g to those workers with an evaluation of "g" and offer μ_k^b to those workers with an evaluation of "b". Doing so makes its average wage offer equal H_k , while no outside firms would find it profitable to steal workers from firm k . □

Note that the key to the zero pre-training profit result lies in the strategic advantage of outside firms over the incumbent firm in wage offering process. Outside firms know the average productivity of workers at firm k , and also observe firm k 's

wage offers before making their own. This results in zero pre-training profit for the incumbent firm even though it has an informational advantage over outside firms.

Let N_k be the total measure of workers hired by firm k in the first period, including n_k high productivity and m_k low productivity workers. We now have the following claim regarding firms' training decisions in the second period.

Claim 2. *Firm k provides training to all its employees when training cost*

$$c \leq \underline{c} = \frac{(1 - \rho)(q - 1)n_k}{(1 - \rho)n_k + \rho m_k}, \quad (\text{A.1})$$

provides training only to those with a good performance evaluation when

$$\underline{c} < c \leq \bar{c} = \frac{\rho(q - 1)n_k}{\rho n_k + (1 - \rho)m_k}. \quad (\text{A.2})$$

It provides no training to employees when $c > \bar{c}$.

Proof. Claim 1 has already showed that firm k makes zero pre-training profit from its employees. Hence, the decision on whether to train a worker depends on the marginal benefit and marginal cost of training alone.

The benefit for firm k from offering training to workers with evaluation “ g ” is $(q - 1)\rho n_k$ and the cost is $[\rho n_k + (1 - \rho)m_k]c$. Firm k will provide training to this group of workers if and only if $(q - 1)\rho n_k \geq [\rho n_k + (1 - \rho)m_k]c$, i.e., $c \leq \bar{c} \equiv ((q - 1)\rho n_k)/(\rho n_k + (1 - \rho)m_k)$.

Similarly, firm k will provide training to those with a evaluation “ b ” if and only if the benefit from training this group of workers exceed the cost, i.e.,

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

This is equivalent to $c \leq \underline{c} \equiv (1 - \rho)n_k(q - 1)/((1 - \rho)n_k + \rho m_k)$.

Finally, using $1/2 < \rho < 1$, we have $\bar{c} \geq \underline{c}$. The results then follow. \square

Claim 3. *Firm k 's second period total profit (Π_k^2) divided by first period measure of workers (N_k) is weakly increasing in first period average productivity (H_k).*

Proof. Based on Claim 1, firm k 's profit in the second period should equal the surplus from offering training to workers, as wages on average equal to workers' pre-training productivities. Let Π_k^2 be firm k 's second period total profit, then

$$\Pi_k^2 = n_k(q - 1) - N_k c \quad (\text{A.3})$$

if training cost $c \leq \underline{c}$, which corresponds to the case that firm k offer training to all workers, and

$$\Pi_k^2 = \rho(q - 1)n_k - [\rho n_k + (1 - \rho)m_k]c \quad (\text{A.4})$$

if $\underline{c} < c \leq \bar{c}$, which corresponds to the case that firm only offer training to “ g ” workers. In both cases, the average profit function $\pi_k^2 = \Pi_k^2/N_k$ will be an increasing function of H_k . To see that, we note that in the first case,

$$\pi_k^2 = (q - 1)H_k - c, \quad (\text{A.5})$$

while in the second case,

$$\begin{aligned} \pi_k^2 &= \frac{\Pi_k^2}{N_k} = \rho(q - 1)H_k - [\rho H_k + (1 - \rho)(1 - H_k)]c \\ &= \rho[(q - 1) - c]H_k + (1 - \rho)H_k c - (1 - \rho)c. \end{aligned} \quad (\text{A.6})$$

When $c > \bar{c}$, firm k provides no training and the second period profit is zero. \square

We now go back to examine firms' decisions in the first period. The game is essentially the same as analyzed in Feng and Zheng (2009). Let the wage offered by firm k be w_k , and workers hired by firm k have average productivity H_k . The measure of workers firm k hires equals

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

As competition drives down firms' total profit from both periods to zero, the wage offer in the first period should equal to the first period average productivity plus second period per worker profit, i.e., $w_k = H_k + \pi_k^2$.

In equilibrium, the number of firms K is determined by the condition

$$H_K + \pi_K^2 \geq r > H_{K+1} + \pi_{K+1}^2. \quad (\text{A.7})$$

There exists such a finite K as $w_k = H_k + \pi_k^2$ goes to zero in the limit as k increases.

Given firms' first period decisions, we are now ready to pin down their training decisions in the second period, using Assumption 1.

Claim 4. *Under Assumption 1, firm 1,2,..., κ (with $\kappa < K$) will provide training to the group of workers with a good performance evaluation ("g"), no firm provides training to the group of workers with a bad performance evaluation ("b").*

Proof. We first show that under Assumption 1, firm 1,2,..., κ (with $\kappa < K$) will provide training to those with evaluation "g". Note that for firm k ,

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

By the condition in (A.2), firm k would train those with evaluation "g" if and only if

$$c \leq \frac{(q - 1)\rho n_k}{\rho n_k + (1 - \rho)m_k} = \frac{\rho(q - 1)\alpha\beta}{\rho\alpha\beta + (1 - \rho)(1 - \alpha)(1 - \beta)[\beta/(1 - \beta)]^{k-1}}.$$

As long as assumption 1 holds, the above equation will hold for any $k \leq \kappa$.

To show that no firm will provide training to those with evaluation "b", we note that firm k will not provide training to these workers if and only if

$$c > \frac{(1 - \rho)n_k(q - 1)}{(1 - \rho)n_k + \rho m_k} = \frac{(1 - \rho)(q - 1)\alpha\beta}{(1 - \rho)\alpha\beta + \rho(1 - \alpha)(1 - \beta)[\beta/(1 - \beta)]^{k-1}}.$$

The quantity on the right hand side is decreasing in k as $1/2 < \beta < 1$. If firm 1 does not train workers with evaluation "b", no other firms will do so. Therefore, $c > (1 - \rho)(q - 1)\alpha\beta/[(1 - \rho)\alpha\beta + \rho(1 - \alpha)(1 - \beta)]$ is sufficient for all firms not to train workers with a bad evaluation. \square

Claim 5. *If firm k provides training to some of its workers, then with average productivity of H_k , then in the second period, it will offer*

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

to those from whom it has a good performance evaluation, and offer

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

to those from whom it has a bad performance evaluation.

Proof. Claim 1 shows that firm k is indifferent between offering ω_k^g to those with a good evaluation, ω_k^b to those with a bad evaluation and any other wage schedule with an average of H_k . However, as outside firms observe workers' training status and wage offers, firm k can only offer ω_k^g to those trained workers (who have received a good evaluation), and ω_k^b to those workers that have received a bad evaluation. Should firm k offers any wage $w'_g < \omega_k^g$ to those it intends to train, outside firms can profitably offer them $w'_g + \epsilon$ to attract them away. Since firm k can not offer anything less than ω_k^g to those with a "g" evaluation, it will not offer anything more than ω_k^b to those with a "b" evaluation. \square

Note that for a firm ($k > \kappa$) that do not offer training to any of its workers, it is possible that they pay the same wage (which equals average productivity H_k) to all workers, regardless of the performance evaluations.

Proof of Proposition 1. In equilibrium, workers maximize total wages in both periods, and firms maximize profits by making wage offers and training decisions, conditional on their beliefs about worker type. Neither firms nor workers has incentives to choose differently. In addition, each firm makes zero total profit. In the first period, firm k with size N_k pays wages $w_k = H_k + \pi_k^2$, where

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

and π_k^2 is weakly increasing in H_k . Therefore, w_k decreases in k and different firms pay different wages in the first period.

Also, since we restrict firms to hire all workers they label as "h" type (assumption 2), the measure of workers firm k hires equals

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

which decreases as k increases. Hence, in the first period, larger sized firms pay higher wages to workers of the same type than smaller firms.

In the second period, firms only provide training to workers with a good performance evaluation. In addition, trained workers at firm k receive ω_k^g and untrained

workers receive ω_k^b . Clearly, $\omega_k^g > \omega_k^b$, trained workers receive higher wages than untrained workers.

Since H_k decreases as k increases, it follows immediately from Claim 5 that both ω_k^g and ω_k^b decrease in k . That is, controlling for workers' training status, larger firms still pay higher wages than smaller firms in the second period. \square

Proof of Corollary 1. Previously we have shown that H_k is decreasing in k . But the proportion of trained workers at firm k equals $\rho H_k + (1 - \rho)(1 - H_k)$. Since $\rho > 1/2 > 1 - \rho$, it immediately follows that $\rho H_k + (1 - \rho)(1 - H_k)$ is decreasing in k as well. \square

Proof of Corollary 2. Consider two firms, i and j , with i being the larger firm while j being the smaller one. Denote the probability of a worker at firm i (firm j) being of high type as H_i (H_j). Previous result implies that $H_i > H_j$. Log wage differential is smaller for trained workers than for untrained workers if and only if:

$$\frac{\rho H_i / [\rho H_i + (1 - \rho)(1 - H_i)]}{\rho H_j / [\rho H_j + (1 - \rho)(1 - H_j)]} < \frac{(1 - \rho) H_i / [(1 - \rho) H_i + \rho(1 - H_i)]}{(1 - \rho) H_j / [(1 - \rho) H_j + \rho(1 - H_j)]}. \quad (\text{A.8})$$

To see this, note the inequality implies

$$\ln \omega_i^g - \ln \omega_j^g < \ln \omega_i^b - \ln \omega_j^b \iff \ln \omega_i^g - \ln \omega_i^b < \ln \omega_j^g - \ln \omega_j^b.$$

As $H_i > H_j$, $H_j/H_i < (1 - H_j)/(1 - H_i)$. In addition, $\rho > 1/2 > 1 - \rho$, Hence, we have:

$$\frac{(1 - \rho) H_j + \rho(1 - H_j)}{(1 - \rho) H_i + \rho(1 - H_i)} > \frac{\rho H_j + (1 - \rho)(1 - H_j)}{\rho H_i + (1 - \rho)(1 - H_i)}.$$

Some transformation gives the following result:

$$\begin{aligned} & \frac{(1 - \rho) H_i}{(1 - \rho) H_j} \left[\frac{(1 - \rho) H_j + \rho(1 - H_j)}{(1 - \rho) H_i + \rho(1 - H_i)} \right] = \frac{(1 - \rho) H_i / [(1 - \rho) H_i + \rho(1 - H_i)]}{(1 - \rho) H_j / [(1 - \rho) H_j + \rho(1 - H_j)]} \\ & > \frac{\rho H_i}{\rho H_j} \left[\frac{\rho H_j + (1 - \rho)(1 - H_j)}{\rho H_i + (1 - \rho)(1 - H_i)} \right] = \frac{\rho H_i / [\rho H_i + (1 - \rho)(1 - H_i)]}{\rho H_j / [\rho H_j + (1 - \rho)(1 - H_j)]}. \end{aligned}$$

Hence we conclude that log wage differential is smaller for trained workers than for untrained workers. \square

Appendix B: Falsification tests

Despite the robustness checks, one can still come up with stories attributing the lower return to training in large establishments to omitted variables. For example, if NLSY79 captures less training spells in small establishments than in large ones, estimated wage return to training in small establishments could be upward biased.²⁸

Therefore, a useful exercise would be to examine industries which show no positive size-wage effects and are not used in the main analyses. Those industries are listed in Panel B of Table 1, including agriculture, forestry and fisheries; mining; construction; finance, insurance and real estate; personal services; entertainment and recreation services; education and other professional services. Table A1 presents results for wage regressions with no control on training, similar to Table 3. The coefficient on large establishment size is almost zero in the OLS specification, but becomes negative in random effects specification. In column 3 of Table A1, the fixed effects specification shows a negative size-wage premium of -6.2% that is significant at the 5% level. Thus those industries display very different size-wage relationships, as predicted by our theory. Coefficient estimates for other variables, such as education, AFQT, tenure, and labor market experience are similar in magnitude to those reported in Table 3.

Table A2 presents the results when on-the-job training information is included. Note the interaction term of large establishment and on-the-job training is always positive, although not statistically significant in some specifications. Hence, the result suggests that for industries with no positive size-wage premium, there is also no evidence that return to training is lower in large establishments.

Table A3 reports regression results on probability of receiving on-the-job training. The specifications are the same as those in Table 5. OLS, Logit, and Random effects Logit models all suggest that workers in large establishment are more likely to receive training, but the coefficients are significantly smaller in magnitude than in Table 5. Once we control for person effects that are possibly correlated with other explanatory

²⁸Large establishments may provide more formal training, and small establishments provide more informal training which is not captured in the NLSY data.

variables with the fixed effects approach (column 4 of Table A3), the probability of receiving training is the same in large and small establishments. This contrasts with our results from Table 5. Note that in both results shown in Table 5 and Table A3, we have not controlled for training cost. Large establishments may enjoy economies of scale in terms of offering formal training to its employees, but the effect should be the same for all industries.

Table A1: Wage Regressions Not Controlling for Training for Industries Not Displaying Size-Wage Premium

	OLS	Random Effects, Person Level	Fixed Effects, Person Level
LE (500+)	0.008 (0.026)	-0.038 (0.023)	-0.062** (0.025)
Education			
- Less than 12 years	-0.036 (0.034)	-0.033 (0.033)	NA
- 13-15 years	0.174*** (0.031)	0.188*** (0.032)	NA
- 16 years	0.259*** (0.041)	0.306*** (0.042)	NA
- More than 16 years	0.394*** (0.058)	0.436*** (0.063)	NA
AFQT	0.002*** (0.0006)	0.002*** (0.0005)	NA
UNEMP	-0.021** (0.010)	-0.016* (0.008)	-0.014 (0.010)
Job Tenure	0.031*** (0.006)	0.026*** (0.005)	0.022*** (0.005)
Job Tenure Squared	-0.001*** (0.0003)	-0.001*** (0.0003)	-0.001*** (0.0003)
EXP	0.024*** (0.008)	0.039*** (0.007)	0.047*** (0.009)
EXP Squared	-0.000 (0.0004)	-0.0004 (0.0003)	-0.0007** (0.0003)

Note 1: Regressions based on the sample listed in Panel B of Table 1. Sample size is 5,546.

Note 2: Same as Note 2 in Table 3.

Note 3: Same as Note 3 in Table 3.

Table A2: Wage Regressions Controlling for Training for Industries Not Displaying Size-Wage Premium

	OLS	Random Effects, Person Level	Fixed Effects, Person Level	Random Effects, Job Level	Mixed Effects, Person and Job Levels
LE (500+)	-0.016 (0.030)	-0.045* (0.026)	-0.063** (0.028)	-0.033* (0.020)	- 0.042** (0.020)
OJT	0.016 (0.030)	0.020 (0.025)	0.018 (0.029)	0.012 (0.021)	0.011 (0.023)
LE x OJT	0.087 (0.055)	0.029 (0.038)	0.002 (0.040)	0.078*** (0.030)	0.061 (0.038)
Education					
- Less than 12 years	-0.036 (0.034)	-0.033 (0.033)	NA	-0.049* (0.025)	-0.041 (0.036)
- 13-15 years	0.172*** (0.031)	0.187*** (0.032)	NA	0.187*** (0.025)	0.185*** (0.031)
- 16 years	0.261*** (0.041)	0.306*** (0.042)	NA	0.288*** (0.033)	0.304*** (0.040)
- More than 16 years	0.392*** (0.057)	0.435*** (0.063)	NA	0.458*** (0.049)	0.444*** (0.054)
AFQT	0.002*** (0.0005)	0.002*** (0.0005)	NA	0.002*** (0.0004)	0.002*** (0.0005)
UNEMP	-0.020** (0.010)	-0.016* (0.008)	-0.014 (0.010)	-0.022*** (0.007)	-0.019*** (0.007)
Job Tenure	0.030*** (0.006)	0.025*** (0.005)	0.021*** (0.005)	0.027*** (0.004)	0.025*** (0.004)
Job Tenure Squared	-0.001*** (0.0004)	-0.001*** (0.0003)	-0.001*** (0.0003)	-0.001*** (0.0003)	-0.001*** (0.0002)
EXP	0.024*** (0.008)	0.039*** (0.007)	0.046*** (0.009)	0.027*** (0.006)	0.033*** (0.006)
EXP Squared	0.00002 (0.0003)	-0.0004 (0.0003)	-0.0007** (0.0003)	-0.0001 (0.0003)	-0.0003 (0.0002)

Note 1: Regressions based on the sample listed in Panel B of Table 1. Sample size is 5,546.

Note 2: Same as Note 2 in Table 4.

Note 3: Same as Note 3 in Table 4.

Table A3: Probability of Receiving On-the-job Training for Industries Not Displaying Size-Wage Premium

	OLS	Logit	Random Effects Logit, Person Level	Fixed Effects, Person Level
LE (500+)	0.089*** (0.025)	0.063*** (0.017)	0.016** (0.008)	0.009 (0.019)
Education				
- Less than 12 years	-0.019 (0.021)	-0.060 (0.042)	-0.023 (0.020)	NA
- 13-15 years	0.037 (0.025)	0.026 (0.023)	0.022 (0.014)	NA
- 16 years	-0.030 (0.036)	-0.028 (0.031)	-0.003 (0.016)	NA
- More than 16 years	0.032 (0.048)	0.013 (0.036)	0.005 (0.022)	NA
AFQT	0.0002 (0.0004)	0.0002 (0.0005)	0.0003 (0.0002)	NA
UNEMP	-0.008 (0.008)	-0.011 (0.009)	-0.005 (0.003)	-0.0004 (0.007)
Job Tenure	0.025*** (0.004)	0.026*** (0.005)	0.012*** (0.002)	0.024*** (0.005)
Job Tenure Squared	-0.001*** (0.0003)	-0.001*** (0.0004)	-0.0004*** (0.0001)	-0.001*** (0.0003)
EXP	0.012*** (0.005)	0.020*** (0.007)	0.013*** (0.003)	0.018*** (0.006)
EXP Squared	-0.0003* (0.0002)	-0.001** (0.0003)	-0.0004*** (0.0001)	-0.0006** (0.0002)

Note 1: Regressions based on the sample shown in Panel B of Table 1. Sample size is 5,879.

Note 2: Same as Note 2 for Table 5.

Note 3: Same as Note 3 for Table 5.

Note 4: Same as Note 4 for Table 5.