A Structural Model of Demand for Apprentices*

Samuel Mühlemann
University of Berne (samuel.muehlemann@vwi.unibe.ch)

Jürg Schweri
Swiss Pedagogical Institute for Vocational Education (juerg.schweri@bbt.admin.ch)

Rainer Winkelmann
University of Zurich, CEPR & IZA (winkelmann@sts.unizh.ch)

Stefan C. Wolter
Swiss Coordination Centre for Research in Education, University of Berne,
CESifo & IZA (stefan.wolter@vwi.unibe.ch)

February 2005

Abstract

It is a widely held opinion that apprenticeship training represents a net investment for training firms, and that therefore firms only train if they have the possibility to recoup these investments after the training period. A recent study using a new firm-level dataset for Switzerland showed, however, that for 60 percent of the firms, the apprenticeship training itself does not result in net cost. In this context it seems important to examine the question whether the potential net cost of training (during the training period) are a major determinant for the demand for apprentices. Different count data models, in particular hurdle models, are used to estimate the effect of net cost on the demand for apprentices. The results show that the net cost have a significant impact on the training decision but no significant influence on the demand for apprentices, once the firm has decided to train. For policy purposes, these results indicate that subsidies for firms that already train apprentices would not boost the demand for apprentices.
JEL Classification: J24, C25
Keywords: Apprenticeship training, count data, probit-Poisson-log-normal model, Switzerland

*The study is based on two surveys financed by the Commission for Technology and Innovation (CTI credit 4289.1 BFS and 5630.1 BFS) with the help of the Swiss Federal Statistical Office and carried out in tandem with a similar survey in Germany conducted by the Federal Institute of Vocational Training (Bundesinstitut für Berufsbildung) in Bonn.

1 Introduction

Every year, almost 70,000 or more than half of the Swiss youngsters who complete their compulsory schooling choose to embark on what is called the dual education system, that is, a training programme combining vocational education at school with training in and work for a company. In order to guarantee each new cohort of school leavers that there is a sufficient number of firms willing to offer training opportunities, it is important to know which factors affect the demand for apprentices. From an economic point of view, it seems obvious that a firm would hire the more apprentices the higher the net benefit of apprenticeship training amounts to. So far, most studies suggested that apprenticeship training results in net cost during the apprenticeship (e.g. Beicht et al. 2004). This would imply that companies need to be able to recoup these net cost by employing graduated apprentices as skilled workers and paying them a wage lower than the value of their marginal productivity, because otherwise they would not offer an apprenticeship programme.

The most recent study of the Swiss apprenticeship system shows, unlike other studies, that about two thirds of the training firms find it profitable to train apprentices (Schweri et al. 2003). This suggests that for the majority of training firms, the possibility to recoup the investment during the apprenticeship training period is an important factor explaining their availability for training. That not every single company is able to offer apprenticeship posts under the same favourable conditions was subsequently pointed out by Wolter et al. (2003), who showed that the potential net cost of non-training firms would be positive and considerably higher than those of training companies, if they were to engage in training apprentices. These observations lead to the hypothesis that, at least in Switzerland where the high flexibility of the labour market puts some restrictions on the possibilities for firms to recoup training expenses after the training period has ended, the net cost of training during the training period is a decisive argument in favour or against the training of apprentices.

This paper extends the analyses made so far in testing directly whether the (potential) net cost of training during the training period explain the training probability of firms. Furthermore, it is estimated whether the net cost of training also affect the demand for apprentices, once the decision to train has been made. We use different types of count data models to account for the specific structure underlying the demand for apprentices.

The paper fills a gap in the training literature, as all previous studies (see section 3) estimating the demand for apprentices with firm-level data were not able to include data on the cost and benefit of the apprenticeship training in their analyses because of the lack of these data. The only exception known to the authors is a study by Niederalt et al. (2001), using, however, a very small and non-representative sample of firms.

The paper is organized as follows: Section 2 discusses briefly the apprenticeship system in Switzerland. Section 3 summarizes previous studies and motivates the hypothesis tested in this paper. Section 4 introduces the data and the sample design. In Section 5, (potential) net cost are estimated using a maximum likelihood selection model. Then, the effect of net cost on the demand for apprentices is estimated using different types of count data models. Section 6 concludes.

2 The apprenticeship system in Switzerland

The apprenticeship system is the route chosen by about 60 percent of the Swiss youngsters at upper secondary level. More than 180,000 adolescents are currently embarked on what is called the dual education system, that is, a training programme combining vocational education at school with training in and work for a company. Almost half of the remaining 40 percent of young people who complete compulsory education go on to attend grammar school (Gymnasium) to prepare them for university and a more academic career. The remainder opt either for other entirely school-based forms of education or pursues no form of post-compulsory education, placing Switzerland ahead of other OECD countries in terms of the percentage of the over-16 population attending any form of non-compulsory schooling.

Vocational training in a dual-education program usually lasts three to four years. Drop-out rates were fairly low in the past. Based on the responses of training firms in the survey used for this article, the drop out rates were on average 5 percent. Apprentices graduate with a diploma recognized throughout Switzerland attesting that the apprentice has a professional qualification. The quality of the training provided in Switzerland, which combines school lessons (1-2 days a week) with on-the-job training in a firm under the supervision of certified staff, is recognized internationally as meeting top standards. International comparisons show, in terms of scholastic and professional qualifications, that Swiss apprentices are more than a match for their upper secondary level peers attending school full-time (see e.g. Bierhoff and Prais 1997).

The employment period ends automatically on completion of training. Any extension of the employment period must be negotiated in a separate contract. Switzerland differs in this respect from some other countries where apprentices are protected from dismissal for a period of time after completing their training. Mobility is fairly high among young people who complete their apprenticeship, with only 36 percent still working at their original training site one year on (see Schweri et al. 2003).[1]

3 Previous studies

There have been several studies on the demand for apprentices using firm-level data in the last few years, but most of them did not include data on the cost-benefit ratio of apprenticeship training on the firm-level. Some of the studies described below used proxies for the cost-benefit ratio instead, others have simply ignored this possibility.

Harhoff and Kane (1997) found for Germany that "firms are much more willing to train when there are fewer firms around to poach their trainees" (p.184). They also showed that firms are less likely to train or train less apprentices if the work force in counties, or counties in commuting distance, is large.

Neubäumer and Bellmann (1999) also examined German firms without data on cost and benefit. Their data set contained, however, qualitative assessments by the non-training firms of the factors affecting the training decision. Their findings were, that about two thirds of the firms do not train because they do not meet the legal requirements for offering apprenticeship training. For

[1] In Germany, the corresponding figure is closer to 50 percent, see Winkelmann (1996) and Euwals and Winkelmann (2002).

the remaining third of the non-training companies, an unfavourable cost-benefit ratio was one of the main reasons not to train. Further, the analysis of the training intensity of training firms showed that the firm size is the major factor explaining the ratio of apprentices and total number of employees. Similar to Switzerland, the majority of all apprentices in Germany are trained by small firms with less than 50 employees. Wolter and Schweri (2002) found similar results for the training intensity amongst training firms in Switzerland, but in addition, the net cost of training also had a significant impact, even if other factors like the firmsize or the training profession were accounted for.[2] This result goes in line with the observation made by Neubäumer and Bellmann (1999) that the training intensity was particularly high in industries where previous studies (see Bardeleben et al. 1995) had shown low or even negative net costs for apprentices.[3]

Stöger and Winter-Ebmer (2001) found that the number of apprentices trained in Austria declined over time, but they were not able to find explanations for this time-trend. As in Germany and Switzerland, they observed that large firms are more likely to train apprentices, but the intensity of training is the highest for small firms.

In a different type of analysis, Fougère and Schwerdt (2002), comparing the French and the German apprenticeship system, analysed the determinants of firms' demand for apprentices. Using a production function approach, they found evidence that apprentices participate significantly in the production process only in medium size firms. They conclude from this result, that small and large firms train apprentices less for the motive of exploiting the value of their productivity, but rather because of their difficulty to find skilled workers on the external labour market.

Beckmann (2002) finally used a zero-inflated negative binominal model to test the implications of the theory proposed by Acemoglu and Pischke (1998, 1999a, 1999b) for firms from West and East Germany. He defined two regimes, one where firms will never train apprentices, and a second regime, in which firms can both train zero or a positive number of apprentices. He found evidence that a high degree of estimated wage compression encourages firms to train apprentices. His findings

[2] Niederalt et al. (2001) found similar results for a small sample of 27 Bavarian companies for which they had calculated the cost of apprenticeship training.

[3] An assessment of the German situation was repeated in an analysis made by Dietrich (2000).

apply only for companies for which he assumed positive net costs of apprenticeship training, as he excluded firms from his sample for which he assumed zero or negative net cost of training (companies with less than five employees). Such a reduction of the sample of analysed firms is in our view inappropriate to analyse the overall decision of firms to train apprentices, both in Germany and Switzerland. Furthermore, with the data used in this study, such assumptions are not necessary as the net cost of training can be included directly in the estimation of the demand for apprentices.

4 Data

The data used here is from two representative surveys conducted in Swiss firms in the year 2001 by the Centre for Research in Economics of Education at the University of Berne and the Swiss Federal Statistical Office (see Schweri et al, 2003 or Wolter and Schweri 2002). In one survey, training firms were asked about their cost and benefit of apprenticeship training. In the second survey, non-training firms were asked to fill in the identical questionnaire as training firms, but without the questions directly related to the costs and benefits of training.

The original dataset contains 2352 training firms and 2230 non-training firms, but the public sector has been excluded, because the profit-maximizing principle does not fully apply to such firms. Also excluded are firms that cannot make independent decisions about training because they are part of a larger enterprise. The dataset used in this paper embraces 1971 training firms and 1661 non-training firms. Table A (in the appendix) shows the means and standard deviations of the variables.

For training firms, net cost, i.e. cost minus benefits, were derived as follows: the main parts of cost are the wages of apprentices and the cost for the training personnel, which together add up to about 90 percent of total cost. The remainder are cost for material, infrastructure and other. The benefits are calculated by the type of work the apprentices perform. This benefit is broken down into production activities that would otherwise be performed by unskilled workers or skilled workers. While we can assume in the first case that the apprentice's performance has the same value as that of an unskilled worker, for the second case the value of the apprentice's performance is compared to that of a fully trained skilled worker.[4] Figure 1 (in the appendix) shows the distribution of net cost for the 1971 training firms. The average is negative (benefits exceed costs) at CHF -6,174. Net cost are negative for 60 percent of all training firms.

5 Econometric models and empirical analysis

In order to estimate the effect of net cost of training on the training decision, we require the potential net cost for currently non-training firms, which we naturally cannot observe. Hence, we need to estimate these counterfactual costs, and the first subsection shows how to do that. In the next subsection, we present different count data models that can be used to estimate the effect of predicted net cost on the demand for apprentices. A particular feature of the data is the high proportion of zeros, i.e., the large fraction of non-training firms. In a demand context, these observations represent a corner solution. Therefore, we concentrate on count data models for corner solutions, so-called "hurdle models". A useful property of these models in the present context is that they allow us to distinguish between cost elasticities at the extensive and intensive margins. The third subsection presents the results.

5.1 Estimation of potential net cost

The net cost of training are observed only for training firms. To use these costs to estimate the potential cost of training for currently non-training firms, we have to realize that training firms are not a random sample from the universe of all firms, and that training and non-training firms may differ systematically in observable and unobservable characteristics. In order to account for observable differences, we use a linear regression framework according to which

$$y_{1j} = x'_{1j}\beta_1 + \varepsilon_{1j} \qquad j = 1, \dots, 3632 \tag{1}$$

where y_{1j} are net cost and x_{1j} is a vector containing variables concerning firm size, number of skilled workers, industry, apprenticeship profession, region, ownership of the firm and a variable

[4] For more details on the cost and benefit model used in this study see Wolter et al. (2003) or Schweri et al. (2003).

indicating whether a firm would like to reduce the time an apprentice spends in vocational school.[5] We don't estimate a log-linear model because net cost can be positive as well as negative. If $\hat{\beta}_1$ is a consistent estimator for β_1, we can predict the net cost for non-training firms consistently as $\hat{y}^{nt}_{1j} = x^{nt'}_{1j}\hat{\beta}_1$.

However, direct estimation of (1) by ordinary least squares using training firms only leads to selection bias as long as training and non-training firms differ in unobservables as well observables. Let

$$y_{2j} = \begin{cases} 1 \text{ if firm } j \text{ trains} \\ 0 \text{ if firm } j \text{ does not train} \end{cases} \tag{2}$$

Selection bias arises if $\mathrm{E}(\varepsilon_{1j}|y_{2j} = 1) \neq \mathrm{E}(\varepsilon_{1j}|y_{2j} = 0)$. We follow here the standard Heckman formulation with a latent continuous variable for the training decision:

$$y_{2j} = \mathbb{1}[x'_{2j}\beta_2 + \varepsilon_{2j} < 0] \tag{3}$$

This is a reduced form version of the firms' profit maximization condition, according to which firms train as long as the marginal benefit of doing so is positive. The benefit of training has two components. First, an immediate benefit arises if the net cost of training are negative. Secondly, a benefit can arise in later periods, for example because the firm needs skilled workers and it may be difficult to recruit such workers in the outside labour market. This reasoning implies that x_{2j} should include all determinants of net cost x_{1j}. These variables obviously affect the decision to train, although not necessarily through their influence on net cost only. They may also have an indirect effect, by affecting other aspects of the training decision. In addition x_{2j} includes a variable measuring the tightness of the labor market for skilled workers (the exclusion restriction). This variable, a dummy indicating whether a firm has difficulties finding skilled workers on the labour market, does not influence the net cost of the apprenticeship training but it is expected to have a significant impact on the firm's decision to train when the net cost of the apprenticeship training alone would be positive.

[5] The last variable will be used later on to identify the cost elasticity in a structural demand model. It is assumed to affect demand only through net cost.

If we moreover assume that

$$(\varepsilon_{1j},\, \varepsilon_{2j}) \sim Normal\left(0, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}\right) \tag{4}$$

we obtain that

$$\mathrm{E}(y_{1j}|x_{1j}, x_{2j}, y_{2j} = 1) = x_{1j}'\beta_1 - \sigma_{12}\frac{\phi(x_{2j}'\beta_2)}{\Phi(x_{2j}'\beta_2)} \tag{5}$$

for training firms, and

$$\mathrm{E}(y_{1j}|x_{1j}, x_{2j}, y_{2j} = 0) = x_{1j}'\beta_1 + \sigma_{12}\frac{\phi(x_{2j}'\beta_2)}{1 - \Phi(x_{2j}'\beta_2)} \tag{6}$$

for non-training firms, where σ_{12} denotes the covariance between the error in the net cost equation and the error in the training equation. If $\rho = \sigma_{12}/\sigma_1 \neq 0$, then the expected value of the error term is $\mathrm{E}(\varepsilon_{1j}|x_{1j}, y_{2j} = 1) \neq 0$, and the self-selection problem results in biased estimates in the standard regression using training firms only. A positive ρ implies that the higher the unobserved net cost component ε_{1j}, the less likely is the training condition (3) satisfied, i.e., the lower the probability of training. The net cost of training firms then underestimate the average net cost, as seen in (5).

To solve the problem, a maximum likelihood estimation procedure can be applied, where the contribution of firm j to the log likelihood function (see Wooldridge 2002) is

$$\begin{aligned} L_j &= (1 - y_{2j}) \ln[1 - \Phi(x_{2j}'\beta_2)] \\ &\quad + y_{2j}\left\{\ln\Phi\left(\frac{x_{2j}'\beta_2 + \sigma_{12}\sigma_1^{-2}(y_{1j} - x_{1j}'\beta_1)}{\sqrt{1 - \sigma_{12}^2\sigma_1^{-2}}}\right) + \ln\phi\left(\frac{y_{1j} - x_{1j}'\beta_1}{\sigma_1}\right) - \ln(\sigma_1)\right\} \end{aligned}$$

The estimation results are given in Table 1. The estimated ρ is positive indeed, so that training firms self-select based on absolute cost advantage. The point estimate is 0.5, and a Wald test for the null hypothesis $H_0 : \rho = 0$ has a p-value of 0.0013, so that the null is rejected.

Table 1: MLE selection model, Dependent variable: Net cost of training

	Net Cost		Training yes/no	
	Coeff.	Std. Err.	Coeff.	Std. Err.
Industrial sector	-1510.6	(2687.3)	0.145†	(0.066)
Construction sector	-7799.4†	(3579.8)	0.329†	(0.087)
French part of Switzerland	1733.4	(2580.1)	-0.240†	(0.060)
Italian part of Switzerland	8479.1	(5750.7)	0.044	(0.144)
Profession:				
Commercial employee	-12184.5†	(2705.4)	-0.162†	(0.065)
Electromechanics technician	-50675.5†	(6150.5)	0.101	(0.174)
Polymechanics technician	24259.5†	(4781.7)	-0.380†	(0.135)
Cook	6976.6	(5686.0)	-0.918	(0.136)
IT specialist	10968.9	(5764.4)	-0.508†	(0.138)
Mason	-8595.4	(6289.2)	-0.653†	(0.168)
Architectural draftsperson	-17458.1†	(5779.9)	0.542†	(0.136)
Salesperson (2 year)	-4930.8	(5735.7)	-0.874	(0.133)
Salesperson (3 year)	-7171.0	(7296.3)	0.208	(0.178)
Carpenter	-22790.2†	(6462.1)	0.377†	(0.175)
Auto mechanic	4007.9	(7396.1)	0.199	(0.184)
Hairdresser	-25836.7†	(7627.9)	1.051†	(0.201)
Office worker	-5755.4	(7861.2)	-0.421†	(0.154)
Assistant in a doctor's office	-10143.9	(7498.6)	0.506†	(0.163)
Automation technician	20615.2†	(7777.4)	0.574†	(0.293)
Electronics technician	24008.1†	(8739.7)	-0.190	(0.264)
Structural draftsperson	-34250.5†	(7904.6)	0.725†	(0.232)
Firm-characteristics:				
Foreign-owned company	5617.6	(3212.0)	-0.615†	(0.074)
Firm size: 4-9 employees	-3631.6	(3561.6)	0.168†	(0.069)
Firm size: 10-49 employees	-6742.8	(4148.2)	0.362†	(0.085)
Firm size: 50-100 employees	-8341.6	(5029.0)	0.862†	(0.109)
Firm size: >100 employees	-4287.8	(5499.7)	1.104†	(0.121)
Reduction in school days	8327.6†	(3753.6)	-0.368†	(0.082)
Number of skilled workers (ln)	-2498.7†	(1095.1)	0.317†	(0.031)
Difficulties in finding qualified labor			0.303†	(0.049)
Constant	24676.6	(5606.5)	-0.732	(0.066)
ρ	0.503	(0.079)		
σ_1	42580.8	(1152.2)		
Log-Likelihood	-25551.2			
Observations	1971		3632	

†Effect is significant at the 5 percent level. The reference site is a Swiss-owned company located in the German-speaking part of Switzerland, has more than 100 employees, and trains apprentices in the category "Other occupations" in the service sector.

If we use the selectivity-corrected net cost estimates in column 1 of Table 1, we find that the predicted net cost of an average apprenticeship program of a randomly drawn firm in Switzerland amounts to about CHF 13,608.[6] Of course, there is considerable variation across firms, as seen in Figure 2 (in the appendix) which shows a histogram of the predicted net cost of training and non-training firms. The predictions are now based on (1) and thus conditional on x_{1j} but unconditional on the training decision. As expected, predicted net cost tend to be higher for non-training firms than for training firms - the averages are CHF 18,540 and CHF 9,453, respectively. However, there is considerable overlap in the two distributions. 5.45 percent of all non-training firms (about 8,090 firms), are predicted to have negative net cost and still do not train. At the point of time of the survey some 5,000 offered training posts were not filled because of a lack of either no or not appropriate applicants. Therefore we suppose that more than half of these non-training firms with negative net cost did not train because of this reason. The rest of these firms – according to the model, must have a positive unobserved cost component.

In the following analysis, the predicted net cost – unconditional on the training decision – are used as an explanatory variable in a structural model in order to estimate the cost elasticity of the demand for apprentices. Since the dependent variable, the number of apprentices, is a count, we discuss first appropriate econometric models for non-negative integer valued random variables.

5.2 Count data models for demand of apprentices

Let $n_j = 0, 1, 2, \ldots$ denote the number of apprentices employed by firm j and remember that $\hat{y}_{1j}$ are the predicted net cost (these can be both positive or negative). The main objects of interest are the cost elasticity of demand for apprentices at the extensive margin

$$\eta_1 = \frac{\partial P(n_j > 0)}{\partial \hat{y}_{1j}} \frac{\hat{y}_{1j}}{P(n_j > 0)},$$

[6] These results differ from the results in Wolter et al. (2003) because in this paper no weights are used. The use of weights is not required to obtain consistent estimates of the regression parameters although differences in small samples occur. From a qualitative point of view, the results remain the same with or without the use of weights. The weights have been dropped since some of the following count data models have not been implemented for use with weights.

and the cost elasticity at the intensive margin

$$\eta_2 = \frac{\partial \mathrm{E}(n_j|n_j > 0)}{\partial \hat{y}_{1j}} \frac{\hat{y}_{1j}}{\mathrm{E}(n_j|n_j > 0)}.$$

Since $\mathrm{E}(n_j) = P(n_j > 0)\mathrm{E}(n_j|n_j > 0)$ it follows that the overall cost elasticity equals $\eta_1 + \eta_2$. We will estimate these elasticities from count data models since n_j is a non negative integer. The Poisson regression model, for example, postulates that

$$P(n_j|\lambda_j) = \frac{\exp(-\lambda_j) n_j^{\lambda_j}}{n_j!}$$

where

$$\lambda_j = \exp(x_j'\delta + \gamma\hat{y}_{1j})$$

In this case, $P(n_j > 0) = 1 - \exp(-\lambda_j)$ and

$$\eta_1 = \left(\frac{\lambda_j \exp(-\lambda_j)}{1 - \exp(-\lambda_j)}\right) \gamma\hat{y}_{1j}$$

Similarly, $\mathrm{E}(n_j|n_j > 0) = \lambda_j/(1 - \exp(-\lambda_j))$ so that

$$\eta_2 = \left(1 - \frac{\lambda_j \exp(-\lambda_j)}{1 - \exp(-\lambda_j)}\right) \gamma\hat{y}_{1j}$$

We see immediately that $\eta_2 = \gamma\hat{y}_{1j} - \eta_1$. This tight functional relationship between η_1 and η_2 is a direct consequence of the assumption that the same distributional model (and the same set of parameters) generates zeros and positive counts. This assumption needs to be relaxed if one wants to test separately for the statistical significance of intrinsic and extrinsic cost elasticities – as we do. The econometric tools that allows one to perform such separate tests are the class of so-called hurdle models (Mullahy, 1986).

A hurdle model combines a binary model for the training decision with a truncated-at-one count data model for the number of apprentices employed by training firms. A reduced form binary training decision was already modelled before, in equation (3). In keeping with the previous model assumptions, the "structural" training decision equation is again of a standard probit form, this time including the predicted net cost $\hat{y}_{1j}$. In order to capture the indirect effect of the exogenous variables on the training decision, they are included separately, although one exclusion restriction is needed for identification. We assume that the regressors x_{3j} are the same as x_{2j} except that the variable “reduction in vocational school time is desired (yes/no)” is excluded.

The probability function of the counts is then

$$P(n_j|x_{3j}, \hat{y}_{1j}) = \begin{cases} 1 - \Phi(x'_{3j}\delta_1 + \gamma_1\hat{y}_{1j}) & n_j = 0 \\ \Phi(x'_{3j}\delta_1 + \gamma_1\hat{y}_{1j})\,\dfrac{f(n_j|x_{3j}, \hat{y}_{1j})}{1 - f(0|x_{3j}, \hat{y}_{1j})} & n_j = 1, 2, \ldots \end{cases} \tag{7}$$

where Φ is the cumulative distribution function of the standard normal distribution. Here, $f(n_j)$ can be any proper conditional distribution model for count data. The most prominent examples are the Poisson distribution and the negative binomial distribution (Winkelmann, 2003). Among these two, the negative binomial model is usually prefered (and the Poisson model is rejected by formal tests) because it allows for unobserved heterogeneity that is an important feature of almost all economic data. Assume that

$$\lambda_j = \exp(x'_{3j}\delta_2 + \gamma_2\hat{y}_{1j} + u_j) \tag{8}$$

where $\exp(u_j)$ is gamma distributed independently of x_{3j} and $\hat{y}_{1j}$. If $n_j|u_j$ is Poisson distributed, then the unconditional model n_j can be shown to be negative binomial.

However, it has been argued elsewhere (Winkelmann, 2004) that a more suitable model for incorporating unobserved heterogeneity is the *Poisson log-normal model*. This model assumes that conditional on u_j, n_j is Poisson distributed with parameter λ_j defined as in (8). Moreover, $u_j \sim Normal(0, \sigma_u^2)$. In previous empirical applications, the Poisson log-normal model usually outperformed the negative binomial model based on selection criteria for non-nested models. If the Poisson log-normal model is combined with a probit hurdle as in (7), we obtain the *probit-Poisson-log-normal model* (PPLN) with individual likelihood contribution

$$\begin{aligned} f(n_j) = {} & [1 - \Phi(x'_{3j}\delta_1 + \gamma_1\hat{y}_{1j})]^{\mathbb{1}(n_j=0)} \\ & \times \left[\Phi(x'_{3j}\delta_1 + \gamma_1\hat{y}_{1j}) \int_{-\infty}^{\infty} \frac{\exp(-\lambda_j(u_j))(\lambda_j(u_j))^{n_j}}{[1 - \exp(-\lambda_j(u_j))]n_j!} \frac{1}{\sigma_u}\phi(u_j/\sigma_u)\mathrm{d}u_j\right]^{\mathbb{1}(n_j>0)} \end{aligned} \tag{9}$$

where ϕ denoted the standard normal density. The parameters δ_1, δ_2, γ_1, γ_2 and σ_u^2 can be jointly estimated by maximum likelihood. A useful property of all standard hurdle models is that the log-likelihood function has two additive parts $L_1(\delta_1, \gamma_1)$ and $L_2(\delta_2, \gamma_2, \sigma_u^2)$ that can be maximized separately with respect to the corresponding parameters. In the probit Poisson log-normal model described here, $\hat{\delta}_1$ and $\hat{\gamma}_1$ are obtained from estimating a probit model with all data whereas the remaining parameters are obtained by maximizing a truncated-at-zero Poisson log-normal model using observations on training firms only. Although the integral in this second part has no analytical solution and thus requires numerical integration, this can be readily done using the Gauss-Hermite method.[7]

We can now return to the question of the elasticities at the extensive and intensive margins. For the extensive margin, the cost elasticity is readily calculated as

$$\eta_1 = \frac{\phi(x'_{3j}\delta_1 + \gamma_1 \hat{y}_{1j})}{\Phi(x'_{3j}\delta_1 + \gamma_1 \hat{y}_{1j})} \gamma_1 \hat{y}_{1j}$$

The demand elasticity for training firms cannot be calculated in closed form. However, it can be computed numerically.

5.3 Results

We will concentrate our discussion on the estimation results for the PPLN model. We also estimated a number of alternative available count data models, as mentioned earlier. Specification tests showed that the PPLN model fits the data best. First, the Poisson model has a log-likelihood of -8899.9, compared to a log-likelihood of -5426.7 of the negative binomial model. The two models are nested with one restriction, and a likelihood ratio test therefore clearly rejects the Poisson assumption of no unobserved heterogeneity. Next, the simple negative binomial model was tested against the negative binomial model with hurdle (log-likelihood -5332.7). With 31 restrictions, the likelihood ratio test statistic has a p-value of zero. This result means that in this application, it is important to estimate the parameters for the training decision separately from those for the quantity decision among training firms. Different mechanisms are at work, and imposing a single probability model for both extensive and intensive margins would lead to spurious interpretations.

[7] An extension of the model allows for correlation between u_j and the error in the probit part of the model, leading to the *Probit Poisson-log-normal model with correlated errors* (Winkelmann, 2004). In this case, the log-likelihood function can no longer be maximized separately.

Table 2: Demand for apprentices – Probit Poisson log-normal model

	Structural Probit (training yes/no)		Poisson log-normal (# apprentices 1+)	
	Coeff.	Std. Err.	Coeff.	Std. Err.
Industrial sector	0.072	(0.067)	-0.061	(0.063)
Construction sector	-0.013	(0.116)	-0.111	(0.118)
French part of Switzerland	-0.150†	(0.062)	-0.465†	(0.067)
Italian part of Switzerland	0.374†	(0.169)	-0.371†	(0.156)
Profession:				
Commercial employee	-0.723†	(0.134)	-0.263†	(0.129)
Polymechanics technician	0.667†	(0.280)	0.288	(0.250)
IT specialist	-0.056	(0.176)	-0.096	(0.153)
Cook	0.222	(0.153)	0.101	(0.150)
Electromechanics technician	-2.161†	(0.533)	1.016†	(0.492)
Mason	-1.032†	(0.190)	-0.168	(0.157)
Architectural draftsperson	-0.252	(0.216)	-0.140	(0.246)
Salesperson (2 year)	-0.305†	(0.141)	0.086	(0.144)
Auto mechanic	0.390†	(0.189)	0.648†	(0.157)
Carpenter	-0.635†	(0.286)	0.214	(0.265)
Salesperson (3 year)	-0.106	(0.190)	0.208	(0.180)
Office worker	-0.673†	(0.163)	-0.866†	(0.311)
Assistant in a doctor's office	0.053	(0.190)	-0.019	(0.291)
Structural draftsperson	-0.808†	(0.407)	0.127	(0.377)
Hairdresser	-0.065	(0.324)	0.886†	(0.305)
Automation technician	1.457†	(0.368)	0.576†	(0.247)
Electronics technician	0.877†	(0.353)	0.365	(0.275)
Firm-characteristics:				
Foreign-owned company	-0.373†	(0.091)	-0.044	(0.080)
Firm size: 4-9 employees	0.013	(0.078)	0.081	(0.133)
Firm size: 10-49 employees	0.063	(0.109)	0.346†	(0.149)
Firm size: 50-100 employees	0.500†	(0.137)	0.496†	(0.161)
Firm size: >100 employees	0.919†	(0.129)	0.823†	(0.152)
Number of skilled workers (ln)	0.208†	(0.040)	0.522†	(0.033)
Difficulties in finding qualified workers	0.291†	(0.050)	0.080	(0.049)
Net cost of training (in thousands)	-0.044†	(0.010)	0.001	(0.009)
Constant	0.366	(0.259)	-0.857	(0.269)
$\ln(\sigma_u)$			-0.589†	(0.036)
Log-Likelihood		-5262.1		
Observations	3632		1971	

†Effect is significant at the 5 percent level.

After we settled for hurdle models, we finally compared the negative binomial model with hurdle and the PPLN. The two models are not nested. However, since the log-likelihood of the PPLN is -5262.1, and since it has one parameter less than the negative binomial hurdle model, it is clearly the better fitting hurdle model.

Table 2 shows all regression coefficients for the PPLN model. The first column gives the probit coefficients, whereas the third column shows the coefficients for the truncated-at-zero Poisson log-normal model. The main parameter of interest is the structural effect of net cost. The regression parameter is negative and statistically significant in the hurdle part but practically zero for the positives. This means that an increase in net cost reduces the probability of offering (any) training, but such an increase is unrelated to the number of apprentices a firm trains, once it has decided to train. A potential economic explanation for this pattern is that small firms who train a small number of apprentices have in general lower net cost of training than large firms training many apprentices. Although training is lucrative for small firms, their demand for apprentices is limited by an upper bound that depends on the number of skilled workers able to train apprentices. If they would extend their demand beyond that limit, the additional need for trainers (and additional infrastructure) would result in positive marginal net cost for an additional apprentice. Large firms with positive net cost of training are not sensitive to marginal variations, because the number of apprentices is determined by the number of future vacancies for skilled workers. In this situation, a reduction in the net cost of training would only translate into an additional demand for apprentices if the marginal net cost of training were smaller than the marginal benefit for the period after the apprenticeship. The latter depends on the probability that the firm is able to offer its former apprentice a job. This together leads to the result that once a firm has decided to train apprentices, the number of apprentices does not depend on marginal variations in the net costs of training. The effect of net cost is therefore entirely restricted to the extensive margin.

To put a quantitative meaning on the estimated coefficient of -0.044, we can compute the elasticity at the average values of the regressors. In this case, $\hat{\eta}_1 = -0.45$ - a one percent increase in the net cost reduces the probability of training by 0.45 percent. Obviously, these marginal elasticities should not be extrapolated too far, as the probit model is highly non-linear. Alternatively, we can consider absolute percentage point changes in the probability of training. For example, starting again from average values, an increase in net training cost by one thousand will reduce the probability of training by 1.8 percentage points. Similarly, an increase in net training cost by one standard deviation reduces the probability of training by 27.4 percentage points.

In order to summarize the main result of the analysis, we find that the elasticity at the intensive margin is zero, whereas the estimated elasticity at the extensive margin amounts to -0.45, an economically substantial effect. We conclude by mentioning some of the other results. Firm size, measured by the number of employees, and the number of skilled workers have the expexted positive effects on both the probability of training and on the number of apprentices among training firms. Foreign owned firms are less likely to train, as are firms in the French speaking part of Switzerland. The training firms in the French part in addition train fewer apprentices – for given firm size, net cost, etc.– than otherwise similar firms in the German speaking part of Switzerland. These results largely confirm those from previous studies.

5.4 A simulation for subsidies

In the year 2000, the year of reference of our survey, about 74'500 new apprentices were employed by Swiss firms, but not every school leaver interested in apprenticeship training was able to find a training position. In subsequent years, the gap between the number of school leavers interested in apprenticeship training and the number of available apprenticeship positions increased steadily due to the unfavorable general economic climate. Political initiatives tried to fight the imbalance in the apprenticeship market with the idea of subsidizing firms for their training. Taking into account our findings, the subsidy should be given to non-training firms, where it would have a large effect, but not to training firms, where it would have no effect at all. A cost subsidy to a non-training firm would reduce net cost of training and thereby increase the probability to offer a training position. The question remains how much such a program would cost.

In the year 2004, according to the "Lehrstellenbarometer 2004", some 8,000 interested school leavers did not succeed in finding a training position. With an average of 1.04 newly created apprenticeship posts per training firm, one would need to attract 7,666 new training firms, an increase in the proportion of training firms by 3.66 percentage points. As shown in the previous section, the training ratio increases by 1.8 percentage points for each reduction in cost by CHF 1,000. Thus, the targeted increase by 3.66 percentage points requires a subsidy of CHF 1,949 per apprentice.

Of course, it would be difficult in practice to assess whether a given firm would have offered no training post in the absence of a subsidy, and thus to discriminate between firms who do already train apprentices and new firms when deciding whom to offer the training subsidies. In the spirit of our model, one would need to have detailed information on the net cost of training for that firm. However, firms can be expected to respond strategically and overstate their true cost once such a scheme would be in place. Therefore all known political initiatives demanding subsidies for training start with the idea that each training post would be subsidized, and not only the "additional" ones. This windfall gain for "old" training companies creates as a consequence much higher total costs of the subsidy than a regime that could be targeted at the new companies only. As a consequence the subsidies alone would amount to more than CHF 17,000 per newly created apprenticeship position.

Table 3: Subsidy simulation

New apprenticeship positions	Total cost of subsidies (in Mio CHF)	Subsidy per newly created apprenticeship position (in CHF)	Total cost per apprenticeship position (in CHF)
8,000	139.7	17,463	62,463
1,000	15.8	15,758	60,758
10,000	179.5	17,950	62,950

The amount depends on the number of newly created apprenticeship positions but is basically driven by the total number of training firms. To illustrate this, a scenario with an additional 1,000 and another scenario with 10,000 additional new apprenticeship positions is shown in Table 3. To the costs of subsidies, the costs of school for an apprentice must be added, which amount on average to CHF 15,000 per year. The total cost of education (subsidy included) of an additionally trained apprentice would be more than CHF 60,000. For purposes of comparison, the total costs for the highest form of a full-time general education in Switzerland (Gymnasium) are on average CHF 58,500 for a 3-year period. Taking into account that the administration of such subsidies would generate additional costs of 10 to 20 percent of the total amount of subsidies, the "artificial" creation of new apprenticeship positions through subsidies seems already questionable from a cost point of view.

6 Concluding remarks

The paper has made both a methodological and a substantive contribution. On the methodological side, it is the first time that the firm's demand for apprentices is modelled in a structural framework. In this empirical framework, the main parameter of interest, the cost elasticity of demand, can be identified from observing cost data for training firms alone. The problem that costs cannot be observed for non-training firms is overcome by using predictions from a selectivity corrected cost equation. A hurdle count data model is then used to estimate the structural demand equation, so that cost elasticities at the extensive and intensive margins can be estimated separately. To estimate the model, we employed a unique firm-level dataset that includes training and currently non-training firms and provides detailed cost information for training firms.

The results are quite striking. We find that the cost elasticity at the extensive margin, i.e. with regards to the probability whether to train or not to train, amounts to -0.45, an economically substantial effect. By contrast, the cost elasticity at the intensive margin, i.e. for the number of apprentices among training firms, is zero.

We close with the substantive conclusion that in order to increase the demand for apprentices, i.e. the number of apprenticeship positions offered each year, one would need to direct subsidies at non-training firms and exclude training firms. In this case, the required subsidy would be quite modest. If, however, such a discrimination is politically and practically infeasible, the costs for creating additional apprenticeship positions would be prohibitively high.

References

Acemoglu, Daron and Pischke, Jörn-Steffen (1998): Why do Firms train? Theory and Evidence, *Quarterly Journal of Economics*, 113, 79-119.

Acemoglu, Daron and Pischke, Jörn-Steffen (1999a): Beyond Becker. Training in Imperfect Labour Markets, *Economic Journal*, 109, F112-F142.

Acemoglu, Daron and Pischke, Jörn-Steffen (1999b): The Structure of Wages and Investment in General Training, *Journal of Political Economy*, 107, 539-572.

Beckmann, Michael (2002): Wage Compression and Firm-Sponsored Training in Germany: Empirical Evidence for the Acemoglu-Pischke Model from a Zero-inflated Count Data Model, *Applied Economics Quarterly (Konjunkturpolitik)*, 48 (3-4), 368-387.

Beicht, Ursula et al. (2004): Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland, Berichte zur beruflichen Bildung, Heft 264, Bonn: Bundesinstitut für Berufsbildung.

Bierhoff, Helvia and S. J. Prais (1997): *From school to productive work. Britain and Switzerland compared*. The National Institute of Economic and Social Research. Cambridge: Cambridge University Press.

Dietrich, Hans (2000): Betriebliches Ausbildungsverhalten im Kontext der betrieblichen Altersstruktur, in: R. George and O. Struck (Eds.) *Generationenaustausch im Unternehmen*, Rainer Hampp Verlag: München, 159-176.

Euwals, Rob and Winkelmann, Rainer (2002): Mobility after Apprenticeship - Evidence from Register Data, *Applied Economics Quarterly (Konjunkturpolitik)*, 48 (3-4), 256-278.

Fougère, Denis and Schwerdt, Wolfgang (2002): Are Apprentices Productive?, *Applied Economics Quarterly (Konjunkturpolitik)*, 48 (3-4), 317-346.

Franz, Wolfgang et al. (2000): Die betriebliche Ausbildungsbereitschaft im technologischen und demographischen Wandel, Schriftenreihe des ZEW, Band 46, Baden-Baden: Nomos Verlagsgesellschaft.

Harhoff, Dietmar and Kane, Thomas J. (1997): Is the German Apprenticeship System a Panacea for the U.S. Labor Market?, *Journal of Population Economics*, 100, 1147-1158.

Lehrstellenbarometer (2004): Lehrstellenbarometer August 2004, Link Institut Luzern im Auftrag des Bundesamtes für Berufsbildung und Technologie BBT, http://www.bbt.admin.ch/berufsbi/projekte/barometer/archiv/d/kurzbericht_aug_04_d.pdf.

Mullahy, John (1986): Specification and testing in some modified count data models, *Journal of Econometrics* 33, 341-365.

Niederalt, Michael et al. (2001): Betriebliches Ausbildungsverhalten zwischen Kosten-Nutzen-Kalkül und gesellschaftlicher Verantwortung, Friedrich-Alexander-Universität Erlangen-Nürnberg, Diskussionspapiere No. 7.

Neubäumer, R. and Bellmann, L. (1999): Ausbildungsintensität und Ausbildungsbeteiligung von Betrieben: Theoretische Erklärungen und empirische Ergebnisse auf der Basis des IAB-Betriebspanels 1997, in: Beer, D., Frick, B., Neubäumer, R. and Sesselmeier, W. (eds.): Die wirtschaftlichen Folgen von Aus- und Weiterbildung, Rainer Hampp Verlag: München und Mering, 9-41.

Schweri, Jürg et al. (2003): Kosten und Nutzen der Lehrlingsausbildung aus der Sicht Schweizer Betriebe, Beiträge zur Bildungsökonomie, Band 2, Chur und Zürich: Rüegger Verlag.

Stöger, Klaus and Winter-Ebmer, Rudolf (2001): Lehrlingsausbildung in "Osterreich: Welche Betriebe bilden Lehrlinge aus?, Johannes Kepler Universität Linz, Arbeitspapier Nr. 0110.

Winkelmann, Rainer (1996): Employment Prospects and Skill Acquisition of Apprenticeship-Trained Workers in Germany, *Industrial and Labor Relations Review*, 49(4), 658-672.

Winkelmann, Rainer (2003): *Econometrics Analysis of Count Data* 3rd edition, Heidelberg: Springer.

Winkelmann, Rainer (2004): Health Care Reform and the Number of Doctor Visits - An Econometric Analysis, *Journal of Applied Econometrics* 19, 455-472.

Wolter, Stefan C. and Schweri, Jürg (2002): The Cost and Benefit of Apprenticeship Training: The Swiss Case, *Applied Economics Quarterly (Konjunkturpolitik)*, 48 (3-4), 347-367.

Wolter, Stefan C., Mühlemann, Samuel and Schweri, Jürg (2003): Why some firms train apprentices and many others do not, IZA Discussion Paper No. 916.

Wooldridge, Jeffrey M. (2002): *Econometric analysis of cross section and panel data*, Cambridge, Massachusetts: The MIT Press.

Appendices

Table A: Sample descriptives
(N=3632; sampling weights have been used.)

Variable	Mean	Std. Dev.
Training firm	0.291	0.454
Number of apprentices	0.675	3.878
Firm size 1-3	0.331	0.471
Firm size 4-9	0.403	0.491
Firm size 10-49	0.224	0.417
Firm size 50-99	0.023	0.151
Firm size >100	0.019	0.135
ln (Number of skilled workers)	0.313	3.035
Sector: Service	0.679	0.467
Industry	0.136	0.343
Construction	0.119	0.324
German part of Switzerland	0.750	0.433
French part of Switzerland	0.222	0.416
Italian part of Switzerland	0.029	0.167
Foreign firm ownership	0.116	0.320
Difficulties in finding qualified labor	0.403	0.491
Reduction in school days	0.113	0.317
Commercial employee	0.177	0.381
Polymechanics technician	0.019	0.136
IT specialist	0.028	0.164
Cook	0.070	0.255
Electromechanics technician	0.020	0.139
Mason	0.025	0.157
Architectural draftsperson	0.030	0.171
Salesperson (2 years)	0.062	0.242
Salesperson (3 years)	0.026	0.159
Auto mechanic	0.020	0.141
Carpenter	0.025	0.155
Office worker	0.033	0.180
Assistant in a doctor's office	0.021	0.145
Structural draftsperson	0.010	0.100
Hairdresser	0.017	0.129
Automation technician	0.004	0.063
Electronics technician	0.004	0.064
Other professions	0.409	0.492

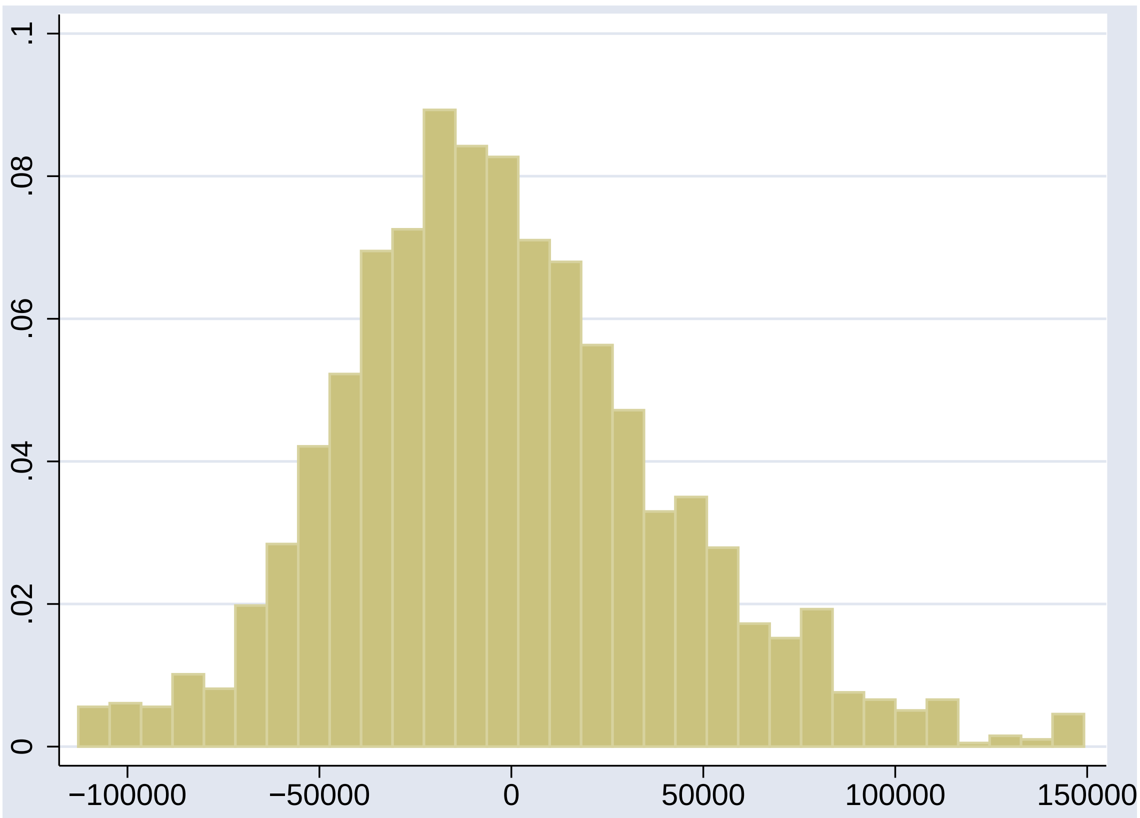

Figure 1: *Distribution of net cost for training firms*

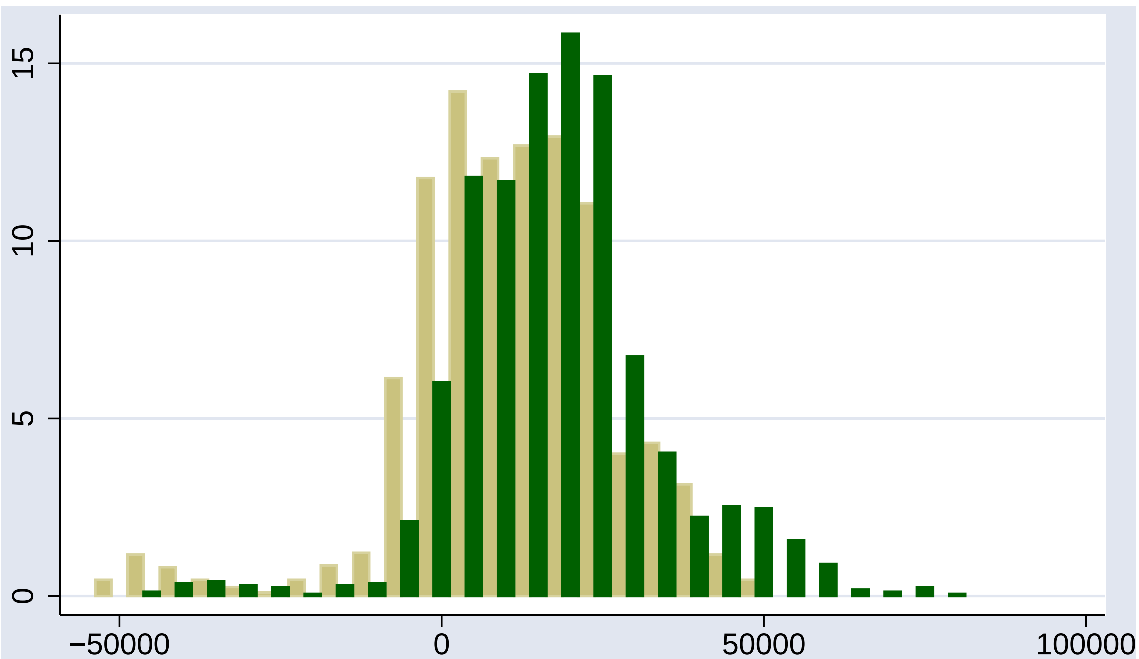

Figure 2: *Overlapping histograms for predicted net cost of training firms (light color) and non-training firms (dark color)*