# Why has the share of training firms declined in Switzerland?*

*Juerg Schweri and Barbara Mueller*

The Swiss mass apprenticeship system is market based. The share of training firms as a percentage of all firms is therefore an indicator which receives much public attention. The share of training firms declined markedly from 1985 to 2001, dropping from 24.7 to 17.6 %. This has often been interpreted as a sign of firms' decreasing willingness to train apprentices. We use data from the firm census to assess whether the decline in the share of training firms can instead be explained by a range of independent variables. Besides firm characteristics and regional variables, we include supply-side factors such as demographic developments in the relevant age cohorts, which have been ignored in the empirical literature so far. Using pooled probit models, fixed-effects models and a Blinder-Oaxaca type decomposition, we are able to explain the variation in the share of training firms over time to a large extent. The main reasons for the decrease are increasing numbers of very small firms, shifts in industry composition, a reduction in the number of young people and an increasing share of young people going to grammar school. We discuss these developments in turn and conclude that they do not, in our opinion, provide sufficient reasons for state interventions in the apprenticeship market.

* This paper was released for publication in April 2007.

## Contents

## 1 Introduction[1]

The literature on training, especially on apprenticeship training as it can be found in German-speaking countries, has increased steadily in the last decade. New theoretical models concerning firms' motivations to train (Acemoglu and Pischke 1998 and 1999, Leuven 2005) are sometimes referred to as "new training literature" (Gerfin 2004). Despite this literature, our empirical knowledge about apprenticeship training systems is still very limited. Firstly, convincing empirical tests of the new theoretical models are still rare. Secondly, the longitudinal dynamics of the system have seldom been investigated. In this article, we focus on the second topic.

In Switzerland, apprenticeships are followed by almost two thirds of a cohort and constitute the main pathway of young people into the labour market. The fluctuation on the apprenticeship market leads to intense discussions in politics and the media; similar discussions are underway in Germany and Austria, too.

The quantitative importance of problems in the apprenticeship market is controversial. The most important question from a policy point of view is whether the system is able to integrate as many young people aged between 15 and 19 as possible in post-compulsory education or training. Unfortunately, the precise number of young people who look for an apprenticeship place but cannot find one is unknown. There are two relevant surveys which provide partial answers: the "Lehrstellenbarometer" (apprenticeship barometer) is an annual cross-section survey that gives an indication of how many young people are still looking for an apprenticeship place in April (1st survey) and in August (2nd survey). This survey has been carried out since 1997. Of those young people looking for an apprenticeship place at the beginning of the year, around 12 to 15 percent had only received a place in a one-year preparatory course or had found no solution at all by August. This share remained constant between 1997 and 2001, the time span we are analysing (OPET 2001). A longitudinal youth survey found that 5 percent of the young people interviewed either had not started or had dropped out of any post-compulsory education or training two years after finishing compulsory schooling (FSO/TREE 2003).

Although these figures seem quite low in comparison with other countries[2], they still indicate that there remains a segment of young people who cannot be integrated into the apprenticeship system.[3] Public discussions are therefore intense and a huge number of political initiatives have been launched at federal and cantonal level, many of them demanding state intervention in the apprenticeship market. Many different "solutions" are being discussed, such as firm subsidies, tax relief or the expansion of full-time schooling. One important reason put forward in favour of such policy interventions is that firms do not train enough apprentices and, in particular, that their training propensity has declined over time. According to this argument, some young people do not find apprenticeship places because firms' willingness to train has decreased. This decline is diagnosed from a simple descriptive indicator, the share of training firms in the economy.

We focus on this argument and show that the causes of the decline in this indicator have to be analysed before far-reaching conclusions can be drawn. To this end, we analyse firms' decisions to train apprentices and try to explain the variation over time in the share of training firms using different groups of explanatory variables. An important group of variables that has been largely ignored in political discussions as well as in empirical research is that of supply-side factors. In Switzerland, most young people start an apprenticeship directly after compulsory schooling. Therefore, the number of young people looking for apprenticeship places is strongly influenced by demographic development. As we will show, supply-side factors have sizeable effects on the training probability of firms and its variation over time. Taking supply-side factors into account, conclusions drawn from the empirical analysis will change substantially.

Standard human capital theory predicts that firms do not pay for general training (Becker 1962). The reason is that firms cannot extract a rent from skilled workers when labour markets are competitive, so they cannot recoup an investment during the apprenticeship period. The new training literature analyses frictional labour markets instead, which provide a rationale for firm investment in general training. Labour market frictions lead to a compressed wage structure (Acemoglu and Pischke 1998 and 1999), which means that skilled workers can be paid below their marginal product. The rent accrued from this gap grows with workers' training, giving firms an incentive to invest in the training of their

[1] We thank Carsten Kuechler, Jens Mohrenweiser, Fulvio Mulatero and two anonymous referees for helpful comments. Any errors are ours.

[2] The youth unemployment rate of people below the age of 25 was about 3 % percent in 1999 (Weber 2004), lower than in any of the countries considered in the overview article by Ryan (2001): France, Germany, Japan, Netherlands, Sweden, UK, USA.

[3] It has to be considered, though, that the aim of post-compulsory education for everyone is very recent.

employees as long as the marginal cost of training is lower than the marginal rent. The training provided in frictional labour markets will, however, remain below the social optimum that is achieved in competitive labour markets where individuals pay for their training.

Since labour market frictions (information asymmetries, transaction costs, dismissal laws, collective agreements, unionisation) are likely to exist to some degree in reality, firms have a choice about how to train apprentices: they can either pay moderate salaries and place the focus on the productive work of apprentices such that all expenses are covered during the training period. Then there is no necessity to keep on apprentices as skilled workers after the apprenticeship period. Or firms decide to invest in apprentices during the apprenticeship period in which case they have to make sure that enough apprentices remain with the firm as skilled workers in order to extract rents that cover the costs. The latter strategy is, however, hardly feasible for small firms which do not want to grow, since most of them do not have jobs in which to keep the apprentices after the training period.

There is a limited number of studies that attempt to test the premises of the new training literature (Acemoglu and Pischke 1998, Dustmann and Schoenberg 2004, Beckmann 2002). Most of the studies compare the wages of apprentices who stayed with their training firm (stayers) with those who left (movers). Depending on the exact assumptions about the labour market frictions, different hypotheses about movers' and stayers' wages emerge. The mover-stayer literature is, however, far from conclusive (see Clark 2002). Another strand of empirical literature has tried to explain the determinants of firms' training activity directly (without testing the theories mentioned above, though). Neubaeumer and Bellmann (1999) and Franz and Zimmermann (2000) used cross-sectional data for Germany. Muehlemann et al. (2007) estimated the demand for apprentices with a cross-sectional data set containing information on the cost and benefit of apprenticeship training for Swiss firms. Stoeger and Winter-Ebmer (2001) analysed longitudinal data for Austria and found a negative time trend in training activity. They did not include supply-side factors, however. Muehlemann and Wolter (2007) include demography in a cross-sectional analysis for Switzerland exploiting regional variation in firms' training activity.

In this paper, we use a longitudinal data set to analyse whether the change in the aggregate training activity of Swiss firms can be explained by demand-side variables (for which effects have been found in the above-mentioned papers) as well as supply-side variables. The paper is organised as follows: section 2 introduces the Swiss apprenticeship system. The estimation strategy is presented in section 3. Section 4 describes the data used for the analyses and presents the hypotheses to be tested. Section 5 contains the empirical results, section 6 concludes the paper with a discussion.

## 2 The Swiss apprenticeship training system

The apprenticeship system is the route chosen by most young Swiss people at upper secondary level. Around 60 % of young people who complete their compulsory schooling choose to embark on what is called the dual training 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 % 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 (just over 20 %) opt either for other entirely school-based forms of education or (less than 10 % of a cohort of 16-year-olds) pursue no form of post-compulsory education. This ranks Switzerland ahead of other OECD countries in terms of the percentage of the over-16 population attending school.

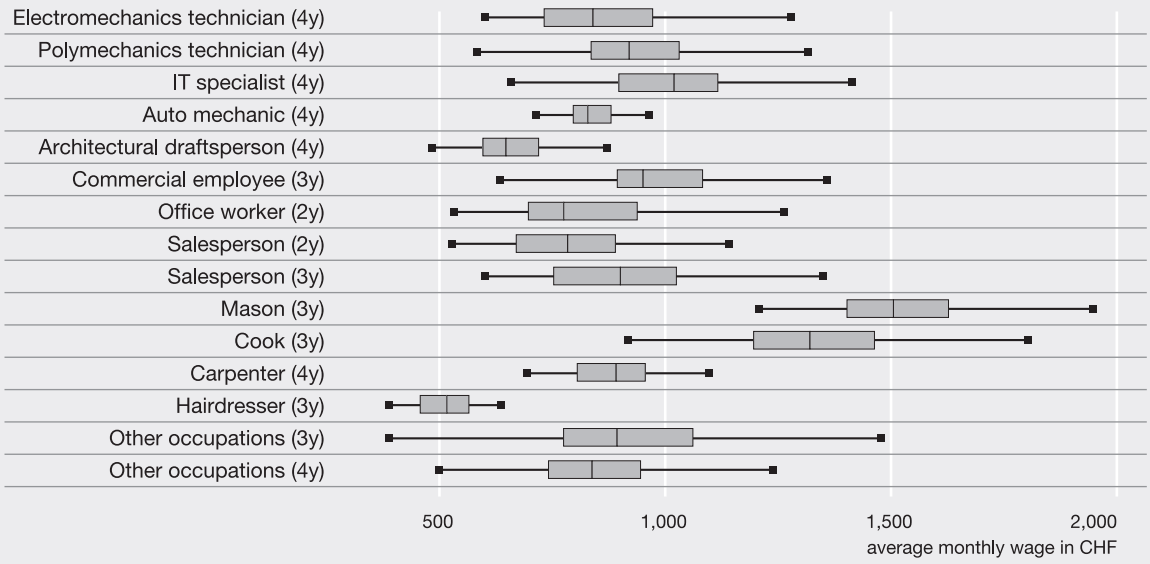
The apprenticeship system is market based: young people have no guarantee of receiving an apprenticeship place, nor are firms obliged to train apprentices. The apprenticeship market can therefore be seen as a sub-market of the labour market. Although employers' organisations often issue salary recommendations for apprentices, apprentices' salaries are determined by the employing company and not regulated by law or on the basis of collective agreements ("Tarifvertrag") between trade union federations and employers' federations.

It is crucial for the justification of our estimation strategy that the apprenticeship market works as a market and is not dominated by state or associations' regulations (e. g. by means of collusion).[4] Therefore, we present data on the wage variance on the apprenticeship market. We use a cross-sectional data set from Schweri et al. (2003) where 2352 Swiss

[4] Apart from the definition of occupations, certificates and the length of apprenticeships, of course (concerning this kind of basic regulation necessary to establish a mass apprenticeship system, see Acemoglu and Pischke 2000 and Malcomson et al. 2003).

Figure 1

**Boxplots of monthly wages (average across apprenticeship years, CHF) in selected occupations; duration of apprenticeship period in brackets**




*Data Source: Schweri et al. 2003, own calculations; N > 35 cases for all occupations; outside values excluded from graph.*

firms were asked about the cost and benefit of their apprenticeship training programs (see also Wolter et al. 2006). Figure 1 shows the wage variance between the most important occupations in terms of numbers of apprentices. Although wages do not vary within one firm for apprentices in the same occupation, figure 1 exhibits a high degree of wage variance within as well as between occupations, as can be expected in a free market where wages adapt to differing conditions and scarcities between occupations, industries, regions and individual firms.[5]

Vocational training in a dual-training programme usually lasts three to four years. A few of the approximately 250 occupations (mostly in the retail sector) permitted an apprenticeship period of just two years in the past. Firms report fairly low drop-out rates of around 5 % (Schweri et al. 2003). Apprentices graduate with a diploma recognized throughout Switzerland attesting that the apprentice has a vocational qualification. After or during an apprenticeship, a qualification called a "Berufsmatur" (professional baccalaureate) may be acquired, which additionally entitles the apprentice to begin third-level education at a university of applied sciences leading to a Bachelor degree. 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.[6]

The employment period ends automatically on completion of training. Any extension of the employment period (making the apprentice a fully-fledged employee) must be negotiated in a separate contract. Mobility is fairly high among young people who complete their apprenticeships, with only 36 % still working at their original training site one year on (Schweri et al. 2003).

Most firms in Switzerland succeed in covering all of the costs related to training with the apprentices' productive work during the apprenticeship period

[5] Mason apprentices earn exceptionally high wages since skilled masons earn relatively high wages as well. The opposite is true for hairdressers – the small median wage also explains why the dispersion looks rather small in absolute terms. The auto mechanic occupation seems to be an exception to the rule of sizeable wage dispersion even within occupations.

[6] Few descriptions of the Swiss vocational education and training system exist in English. For example, the international review of the vocational education literature by Ryan (1998) does not mention Switzerland. For a qualitative comparison of vocational education and training in Switzerland and the UK, see Bierhoff and Prais (1997).

(Schweri et al. 2003). For training firms, apprenticeships result in (small) net gains for the firm on average. This result is in line with Becker's classical prediction that apprentices will pay for general training through a low training wage (Becker 1962). The models of the new training literature, according to which firms pay for general training due to compressed wage structures, seem less important to describe actual firm behaviour in Switzerland. But even in Switzerland the result of net gains from training would not hold for non-training firms if they were to take up apprenticeship training, as has been shown in Wolter et al. (2006). The decision to train is at least partly based on the net cost or net gain that firms expect from training apprentices (Muehlemann et al. 2007). This is another piece of evidence showing that the Swiss apprenticeship system is shaped by economic market forces. The situation in Germany differs in this respect: although the institutional features of the Swiss apprenticeship system are very similar to those in Germany, the functioning of the systems seems to be less comparable. In Germany, the productive contribution of apprentices amounts to only half of the gross cost of training, resulting in substantial net cost for an average apprenticeship (Beicht et al. 2004).

A possible explanation for these differences is that the Swiss labour market is much less regulated than the German one, forcing firms to make productive use of the apprentices to cover costs. By contrast, many large German firms can afford to train apprentices without integrating them in the production process since their mobility after the apprenticeship period is reduced by labour market regulations. If labour markets are frictional, firms can recoup their training investment by keeping their apprentices as skilled workers (see introduction). Indeed, more than 50 % of German apprentices (as opposed to 36 % for Switzerland) are still employed by their original training firm one year on (Beicht et al. 2004). These differences suggest that frictional labour markets are an important explanation for firms' training propensity in Germany, while the Swiss apprenticeship system relies more on market forces.[7]

[7] This is not to say that Swiss labour markets are entirely frictionless or that the concept of a compressed wage structure is useless for Switzerland. Actually, one third of Swiss apprentices cause net costs for their training firms that have to be recouped somehow (Schweri et al. 2003). But in Germany, the vast majority of firms faces net costs during the apprenticeship period, and trade unions, dismissal laws and other labour market institutions (see Soskice 1994 and Franz and Soskice 1995) introduce frictions that are largely absent in Switzerland.

## 3 Estimation strategy

As discussed in the previous section, the Swiss apprenticeship market relies strongly on market forces. The number of apprenticeship places in the economy is thus determined by supply and demand on the apprenticeship market. The share of training firms is a result of the market outcome. To test whether firms' propensity to train has declined, we want to estimate firms' demand for apprentices. With the firm-level data we describe in section 4, we would therefore like to estimate a firm demand function:

$$A_{it}^{d} = \gamma_t + \gamma_w w_{it} + X_{it}^{d}\gamma_d + \mu_{it} \quad (1)$$

$A_{it}^{d}$ denotes the number of apprentices demanded by firm $i$ in period $t$, which depends on wage $w$ and demand-side factors $X$, e. g. firm characteristics. For now, we assume the effect of the independent variables to be constant over time, which is why the respective coefficients do not have a subscript $t$.

In order to identify this classical firm demand function, one has to deal with the problem of the simultaneous determination of the observable combinations of $\{A_{it}, w_{it}\}$ in equilibrium. The supply side is represented by:

$$A_{it}^{s} = \delta_t + \delta_w w_{it} + X_{it}^{s}\delta_s + \nu_{it} \quad (2)$$

Since $A_{it}^{d} = A_{it}^{s}$ in equilibrium, $w_{it}$ is endogenous in equation (1). The classical solution is to use supply shifters $X_{it}^{s}$ as instruments for $w_{it}$ in the demand equation (1). Our data set (see section 4) provides us with supply shifters, but not with wage data. We therefore cannot estimate the structural equation (1) and instead estimate the reduced form equation where the endogenous variable $A_{it}$ is regressed on all available exogenous variables:

$$A_{it} = \beta_t + X_{it}^{s}\beta_s + X_{it}^{d}\beta_d + \varepsilon_{it} \quad (3)$$

Our main interest lies in the intercepts and the year dummies respectively. The allegation discussed in the media (see the introduction) is that the firms' willingness to train has declined over time. Assuming that no relevant time-variant variables have been omitted from the $X$ matrices, this corresponds to the hypothesis that $\beta_t$ is declining for higher t.

Supply shifters will be the share of 16-year-olds in a region and the share of grammar school pupils in a region. Both variables vary between regions as well as over time and have an influence on the number of young people looking for apprenticeship places.

We are not aware of any literature on the German-speaking apprenticeship systems that has included supply-side factors in the analysis. Our hypothesis is that an increase in the number of young people or a decrease in the share of grammar school pupils will lead to more training places: since most new apprentices are around 16 years old, an increase in their number leads to a shift in the supply curve. In equilibrium, wages will fall and more training contracts will be concluded.

We are especially interested to see whether the inclusion of supply-side factors significantly reduces the unexplained differences between time periods and thus changes our interpretation concerning the above-mentioned hypothesis on the differences between the $\beta_t$. The hypotheses on the effect of the demand-side independent variables, i.e. on the $\beta_d$, are discussed in the next section where we present the data set.

Equation (3) can in principle be estimated by pooled OLS. Another problem, however, is that the decision of a firm to train one apprentice instead of zero (i.e., to become a training firm) might be different from the decision to train six instead of five apprentices. One reason for this could be fixed entry costs when initiating training for the first time. The literature therefore typically uses two-step models where the first step is to analyse a dummy variable for training firms:

$$I_{it} = \begin{cases} 1 \text{ if } A_{it} > 0 \\ 0 \text{ if } A_{it} = 0 \end{cases}$$

Neubaeumer and Bellmann (1999) perform a probit estimation of $I_{it}$ followed by an OLS estimation of $A_{it}$ using only the training firms. Franz and Zimmermann (2000) as well as Stoeger and Winter-Ebmer (2001) use the probit as the first step of a Heckman two-step estimation. In the second step, the number of apprentices in training firms is analysed, taking into account the self-selection of training firms in the first step. For the hurdle or count data modelling approach see Muehlemann et al. (2007).

In this paper, we concentrate on the probit estimation of $I_{it}$. As we show in the next section, most of the variation over time stems from firms' changing training propensities and not from changes in the number of apprentices trained by training firms. The number of apprentices trained, given that a firm trains, has hardly changed.[8]

Unobserved heterogeneity is an obvious problem in our estimations. As discussed in section 4, only a limited number of independent variables are available. Many firm characteristics might influence the training decision but are not observed in our data. If these unobserved variables are correlated with the independent variables, the estimation is biased.[9] Assuming that the unobserved firm characteristics are constant over time, we can time-demean all variables in equation (3)[10] for each firm and thus get rid of the firm fixed effects. Because of the dummy dependent variable, we estimate a fixed-effects model based on a linear probability model.[11]

Another problem of equation (3) is that the apprenticeship market might not be homogenous. There are about 250 different occupations which require very different levels and combinations of cognitive and manual skills (see section 2). Demography might, for example, have differential effects for different occupations, which are distributed unevenly across the sectors of the economy. We therefore test for interaction effects between the variables of the two $X$ vectors representing the supply and the demand side.

After discussing the results of the pooled and fixed-effects models, we go on to assess the relative importance of the different variables for explaining the decrease in the share of training firms from 1985 to 2001. To this end, we use the decomposition introduced by Oaxaca (1973) and Blinder (1973)[12]:

$$\begin{aligned} \bar{I}^{01} - \bar{I}^{85} &= \hat{\beta}^{01}\bar{X}^{01} - \hat{\beta}^{85}\bar{X}^{85} + \hat{\beta}^{85}\bar{X}^{01} - \hat{\beta}^{85}\bar{X}^{01} \\ \bar{I}^{01} - \bar{I}^{85} &= (\hat{\beta}^{01} - \hat{\beta}^{85})\bar{X}^{01} + (\bar{X}^{01} - \bar{X}^{85})\hat{\beta}^{85} \end{aligned} \quad (4)$$

The decrease in the share of training firms is thus decomposed into one part explained by changes in the distribution of independent variables (second part of equation 4) and another part that cannot be explained and is attributed to a change in the coefficients (first part of equation 4). The decomposition

[8] Estimation results for the share of apprentices among all employees in a firm have been published in Mueller and Schweri (2006). The results do not alter our main conclusions, we do not, for example, find a time trend in these estimations, either.

[9] The outcome of a Hausman test suggests that this correlation does exist (p = 0.0000) and that a fixed-effects model should therefore be preferred over an (inconsistent) random effects model.

[10] With slight modifications in formula (3): we will use the training propensity $I_{it}$ instead of $A_{it}$ as the dependent variable, and non-linear models instead of OLS.

[11] The alternative is to estimate a conditional logit fixed-effects model (see e.g. Wooldridge 2002). In this model, only the observations that show a change in the dependent variable over time are included. Since many firms in our sample either always train or never train, we would lose about 80 percent of all observations.

[12] In fact, the decomposition can be used for any combination of the four points in time under scrutiny. We use the two extreme points in time (1985 and 2001) in the formulas to simplify the discussion.

can also be applied to non-linear models with dichotomous dependent variables. The decomposition can then be written as (Fairlie 2003):

$$\bar{I}^{01} - \bar{I}^{85} = \left[ \sum_{i=1}^{N^{01}} \frac{F(X^{01}\hat{\beta}^{01})}{N^{01}} - \sum_{i=1}^{N^{01}} \frac{F(X^{01}\hat{\beta}^{85})}{N^{01}} \right] + \left[ \sum_{i=1}^{N^{01}} \frac{F(X^{01}\hat{\beta}^{85})}{N^{01}} - \sum_{i=1}^{N^{85}} \frac{F(X^{85}\hat{\beta}^{85})}{N^{85}} \right] \quad (5)$$

The basic idea is the same as above. Equation (4) is not valid for non-linear models since $\bar{I} = F(X\beta)$ does not necessarily hold as in linear models. Equation (5) therefore uses the sample means of the firms' training probabilities $F(\cdot)$ to decompose the difference in the share of training firms between the two years into an unexplained part and a part explained by all of the independent variables together.

This technique makes it possible to further decompose the explained part and to identify the relative contribution of each independent variable to the total decrease. The empirical importance of supply-side factors can thus be compared with that of the demand-side factors. The linear decomposition for the single variables can easily be seen in the second part of equation (4): the changes in the independent variables are multiplied by the respective coefficients, the total explained change thus results from the sum of the changes caused by the individual independent variables. In non-linear models, the independent contribution of one variable depends on the value of the other variables through the respective non-linear function.[13] The contribution of a single variable to the difference in firms' training shares is therefore evaluated by holding the distribution of the other variables constant (see Fairlie 2003).

## 4 The data and hypotheses

Our basic data set consists of the firm census data of the Swiss Federal Statistical Office. We exploit a question about whether a firm had employed one or more apprentices.[14] So far, the firm census has been carried out in 1985, 1991, 1995, 1998, 2001 and 2005. The question about apprentices was not asked in 1991. The 2005 data were not yet available when this article was written. We can thus use the firm census data for the years 1985, 1995, 1998 and 2001. A major advantage of the data is that the firm census encompasses all Swiss firms. We can include in the analyses the full population of Swiss firms in the available years. Since we can also trace firms that existed in more than one of these years, we can construct a panel data set that does not suffer from attrition due to non-response. The survey frequency of one census every three to four years is not a problem for our purposes since apprenticeships last three to four years. The large gap between 1985 and 1995 is unfortunate, and more survey times (before 1985 and after 2001) would admittedly have enhanced the analytical possibilities. The advantage of a population data set also comes at the price of a reduced set of variables.[15]

Descriptive information on the training activities of firms is presented in figure 2. It shows how the share of training firms[16] has evolved over time (bright bars). There was a sharp decline between 1985 and 1995, then the share of training firms grew again. Between 1985 and 2001, the indicator still shows a decline of 7.1 percentage points, which led to public debates about firms' willingness to train apprentices as mentioned in the introduction.

The mean share of apprentices[17] in the training firms (grey bars) has hardly changed over time. The variation in the share of apprentices among the total of the employees in the economy (black bars) is caused by the variation in the share of training firms. The share of training firms variable will therefore be the dependent variable in our estimations.[18]

The following independent variables are available from the firm census data: firm size, industry, firm type (independent firm, headquarters or branch), region (i.e. canton) and district (rural or urban). Descriptives for these variables can be found in tables 4 and 5 in the appendix.

---

[13] We do not enter into the details of the technique that are discussed in Fairlie (2003). E.g., equation (5) does not hold exactly for probit models, but typically holds very closely. Furthermore, the equation has to be adapted for $N^{85} \neq N^{01}$.

[14] The formulation of the question was "Wie viele Lehrlinge beschäftigt diese Arbeitsstätte?" (How many apprentices does this establishment employ?). An explanation specified who exactly counts as an apprentice: "Als Lehrling gilt, wer aufgrund eines Lehrvertrags einen dem Berufsbildungsgesetz oder entsprechenden kantonalen Gesetzen unterstellten Beruf erlernt. Ebenfalls als Lehrlinge zu zählen sind die Personen, die eine Anlehre absolvieren." (Anyone who, on the basis of an apprenticeship contract, is learning an occupation subordinated to the Vocational Training Act or corresponding cantonal laws counts as an apprentice. People who complete a basic training (Anlehre) are also to be counted as apprentices)

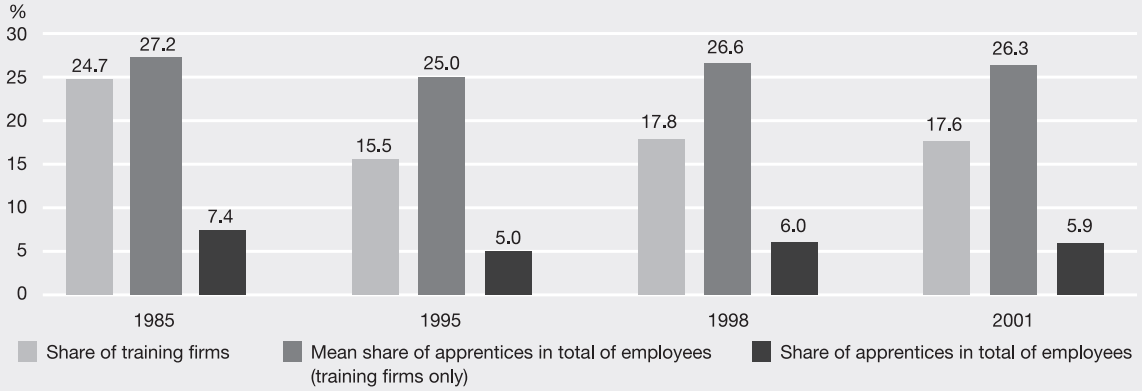
[15] In Switzerland, large firm panels comparable with the IAB establishment panel for Germany are not available. Employer-employee matched data sets including a wide variety of variables on firms as well as employees are also not known.

[16] This share is computed as the number of training firms over the number of all firms. Here, as throughout the paper, only private firms are included.

[17] The share of apprentices in a firm is computed as the number of apprentices over the number of all employees (including apprentices) in a firm.

[18] Compare also footnote 8.

Figure 2
**Swiss firms' training activity 1985–2001**




*Source: Figure based on firm census data of the Federal Statistical Office.*

We have the following hypotheses on how these variables affect firms' training activity: larger firms are known to have a higher training propensity (Neubaeumer and Bellmann 1999, Franz and Zimmermann 2000, Stoeger and Winter-Ebmer 2001). This can be explained by two factors: first, larger firms are more likely to have enough suitable work to be able to use apprentices efficiently in the production process. Second, they are also more likely to have a vacancy for a skilled worker when the apprentice has finished his training, providing the opportunity to extract a rent from him (see section 1).

Different industries are shaped by different production technologies and different skill needs. The training strategies of firms will therefore differ between industries: in the construction industry, apprentices can typically work productively early on. Firms find it favourable to employ apprentices (instead of unskilled workers with higher wages) and should thus have a high training propensity. Services have less tradition in apprenticeship training than crafts and manufacturing and show less training activity according to the descriptive statistics. We also distinguish between traditional and modern services and manufacturing. Modern services and manufacturing might operate in a faster changing environment, which impedes the training of apprentices over several years. Single firms and headquarters probably have a wider array of activities, whereas specialised branches of a firm might show a smaller propensity to train. For Swiss regions it is well known that in the French and Italian speaking parts there are more full-time vocational schools and slightly fewer dual apprenticeship places. In rural districts the reputation effects of apprenticeships (with customers) might be more important, and poaching might be less of a problem. In rural districts, we therefore expect the share of training firms to be higher.

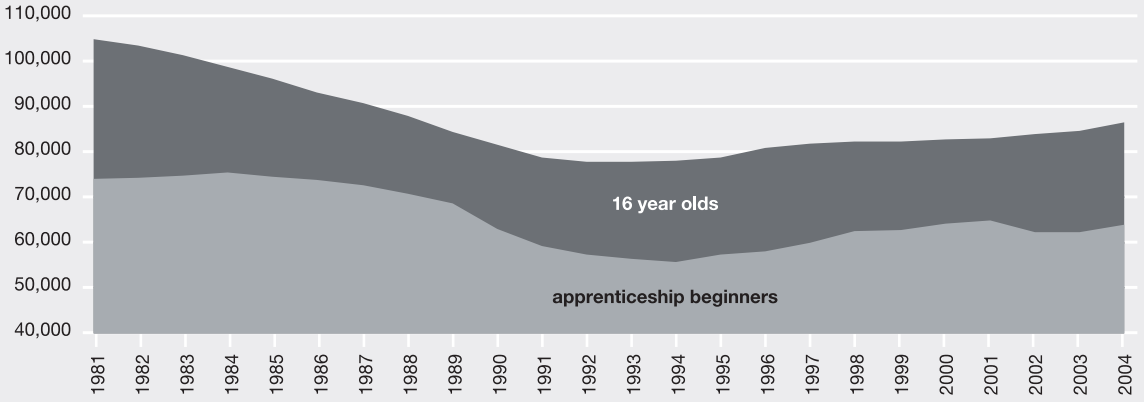
In order to include supply-side variables, we added two demographic variables: the share of 16-year-olds among the working-age population of a canton and the share of grammar school pupils among the 16-year-olds of a canton. These variables were taken from other data sets published by the Federal Statistical Office. In the appendix, a table of frequencies for all variables and all years can be found.

We use cantonal level variables since cantons form strong political entities in Switzerland. They vary markedly in the composition of their population. Moreover, most competencies in educational policies reside with cantons and not with the federal authorities. Therefore, the share of 16-year-olds as well as the share of grammar school pupils among 16-year-olds vary markedly between cantons.[19]

The rationale for including supply-side factors in the estimations was discussed in the previous section about the estimation strategy. Figure 3 shows that demography is very likely to have a strong influence on the variance in the outcomes of the apprenticeship market over time. The number of 16-year-olds showed a significant downward trend from the early eighties to the mid-nineties and rose again afterwards. With a small time lag, the number of people starting apprenticeships seems to mirror these

[19] The share of 16-year-olds varies between 1.57 percent and 3.28 percent between cantons in 1985. The mean share of 2.27 percent in 1985 dropped to 1.65 percent in 1995. The mean share of grammar school pupils in Switzerland rose from 18.8 percent in 1985 to 25.5 percent in 2001. The minimum share in 2001 was 13.2 percent, the maximum 62.1 percent.

Figure 3
**Development of the number of 16-year-olds and of people starting apprenticeships from 1981–2004**




*Data source: Swiss Federal Statistical Office.*

trends. The share of training firms might therefore be affected by the number of apprenticeship candidates available on the market.

The share of 16-year-olds opting for grammar school accounts for shifts in the preferences for general vs. vocational education. An increase in the share of grammar school pupils, as observed especially between 1985 and 1995, will reduce the number of candidates on the apprenticeship market and is therefore expected to lower the overall share of training firms. Not including this variable might bias the time dummies in our estimations, which we want to reflect possible time trends in firm behaviour.[20]

## 5 Results

### *Pooled estimations*

Table 1 presents probit estimations for the full data set (all firms, all years) where the dependent variable is a dummy training variable. Marginal effects are computed at the mean of the explanatory variables. The first column contains a model that excludes supply-side factors. While firm characteristics, industry and regional dummies all have significant effects, they cannot fully explain the differences between the years. Controlling for these factors, the training propensity of firms was still 3.9 percentage points higher in 1985 than it was in 2001 (compared with 7.1 percentage points difference in the descriptive statistics, see figure 2).

The next column shows the results including supply-side factors. Demographic development does have a substantial significant effect, as can be seen from the following illustration: from 1985 to 1995, the share of sixteen-year-olds dropped from 2.27 % to 1.65 % (for all of Switzerland). Based on model 2 in table 1, this decrease would *ceteris paribus* have caused a drop in the share of training firms of 1.8 percentage points.[21] The share of grammar school pupils, however, did not have an effect on training propensity according to this estimation.

While most of the coefficients do not react strongly to the inclusion of these factors, the dummy for 1985 drops by more than half and is reduced from 3.9 to 1.8 percentage points. So the difference between 1985 and 2001 can now largely be explained by the independent variables. Although demography does not increase the goodness-of-fit, it does make an important difference for the conclusions that can be drawn concerning the development of firms' train-

[20] Including the share of grammar school pupils would be questionable if it was itself influenced by the outcome of the apprenticeship market. This would be the case if young people who do not get an apprenticeship place went to grammar school instead. This is unlikely, however: in the short run, grammar schools have a given infrastructure and stock of teachers. The number of grammar school places is therefore inelastic in the short run. Empirical evidence supports this: both the number and the share of grammar school pupils in the relevant age cohort change only gradually and do not show cyclical fluctuations. A study by the Federal Statistical Office which aims to predict the development of pupil numbers confirms that the number of grammar school pupils did not depend on business cycles in the past (FSO 2004).

[21] The respective probabilities for these values of the demography variable have been computed at the mean of all other variables.

Table 1

## Estimating firms' training propensity

(dependent variable: training firm = 1, non-training firm = 0)

| | Model 1 Probit | Model 2 Probit | Model 3 Probit | Model 4 LPM | Model 5 FE Panel | Model 6 FE Panel |
|---|---|---|---|---|---|---|
| **Firm size (Reference = <2 FTE)** | | | | | | |
| 2 FTE | 0.144** | 0.144** | 0.147** | 0.093** | 0.005** | 0.005** |
| 3–4 FTE | 0.263** | 0.263** | 0.249** | 0.186** | 0.010** | 0.010** |
| 5–9 FTE | 0.388** | 0.388** | 0.339** | 0.278** | 0.035** | 0.034** |
| 10–19 FTE | 0.496** | 0.496** | 0.417** | 0.348** | 0.082** | 0.081** |
| 20–49 FTE | 0.599** | 0.599** | 0.513** | 0.442** | 0.164** | 0.162** |
| 50–99 FTE | 0.685** | 0.685** | 0.606** | 0.543** | 0.262** | 0.259** |
| 100–149 FTE | 0.742** | 0.742** | 0.670** | 0.628** | 0.345** | 0.341** |
| 150–249 FTE | 0.770** | 0.770** | 0.708** | 0.677** | 0.405** | 0.401** |
| 250–499 FTE | 0.805** | 0.805** | 0.753** | 0.749** | 0.508** | 0.503** |
| 500–999 FTE | 0.821** | 0.821** | 0.779** | 0.789** | 0.593** | 0.587** |
| >1000 FTE | 0.831** | 0.830** | 0.789** | 0.810** | 0.670** | 0.664** |
| **Firm type (Reference group = independent firm)** | | | | | | |
| Corporate headquarters | 0.048** | 0.048** | 0.062** | 0.062** | | |
| Branch | –0.027** | –0.027** | –0.016** | –0.037** | | |
| **Industry dummies (45)** | Yes | Yes | No | Yes | No | No |
| **Industry group (Reference group = traditional services)** | | | | | | |
| Traditional manufacturing | | | 0.007** | | | |
| Modern manufacturing | | | -0.032** | | | |
| Construction | | | 0.046** | | | |
| Modern services | | | -0.018** | | | |
| **Rural district (Reference group = urban)** | 0.009** | 0.009** | 0.002 | 0.009** | | |
| **Demography** | | 0.031** | 0.028** | 0.058** | 0.044** | |
| x Traditional manufacturing | | | | | | 0.051** |
| x Modern manufacturing | | | | | | 0.048** |
| x Construction | | | | | | 0.058** |
| x Traditional services | | | | | | 0.041** |
| x Modern services | | | | | | 0.036** |
| **Share of grammar school pupils** | | | | | | |
| (Coeff * 100) | | –0.022 | –0.025 | 0.059* | –0.073** | –0.076** |
| **Regional dummies (26)** | Yes | Yes | Yes | Yes | No | No |
| **Year dummies (Reference group = 2001)** | | | | | | |
| Year 1985 | 0.039** | 0.018** | 0.026** | 0.012** | 0.019** | 0.019** |
| Year 1995 | –0.034** | –0.032** | –0.030** | –0.030** | –0.036** | –0.036** |
| Year 1998 | 0.000 | 0.000 | 0.002* | 0.001 | –0.004** | –0.004** |
| (Pseudo) R2 | 0.198 | 0.198 | 0.145 | 0.184 | 0.091 | 0.095 |
| N | 1,252,132 | 1,252,132 | 1,252,132 | 1,252,132 | 1,252,132 | 1,252,132 |

*Notes: Probit results show marginal effects at the means of the other explanatory variables.*
** and ** denote significance at the 1 percent and 1 per mill levels respectively, and are based on robust standard errors accounting for heteroskedasticity (and clustering of firms in the pooled models).*
*FTE= full-time equivalents. The equivalent of a full-time worker: for example, two people who each work half time correspond to one FTE.*

ing propensity: An unexplained difference of 1.8 percentage points does not suggest a major change in firms' willingness to train. Furthermore, another factor might be responsible for the remaining differences, namely the business cycle. We do not have firm level data on the course of business, but it is striking that GDP growth had been high in the years before 1985 whereas it was particularly low in the early nineties. Without being able to test this hypothesis, it offers a plausible explanation for the differences. If the hypothesis is true, the dummies would mainly reflect business cycle effects.

Apparently, demography might be less important for the cross-section analysis of firms' training propensity, but for the longitudinal analysis it is crucial.

In model 3, the industry dummies have been reduced to five dummies. This estimation shows a clearly smaller goodness-of-fit; it is therefore only used to give some results concerning industries, otherwise the coefficients of model 2 are regarded as more reliable. In comparison with the traditional services (reference group), firms in the construction sector show a higher training propensity, whereas modern services and modern manufacturing exhibit lower training propensities. Further analysis with a dummy variable indicating new firms, i.e. firms that had not been observed in the earlier periods, show that these differences are mainly due to the fact that young firms cluster in the modern industries of the economy.[22] Young firms are less likely to train apprentices if there are fixed costs associated with engaging in training for the first time (mainly transaction costs and investment in training instructors).

### *Fixed-effects estimations*

The last two columns of table 1 present results of the fixed-effects estimations. Model 4 shows the underlying, pooled linear probability model. The marginal effects of the LPM model are very close to those of the probit model 2, with the notable exception of the share of grammar school pupils, which has a counterintuitive positive sign. The fixed-effects estimates in model 5, however, yield the expected negative sign for the share of grammar school pupils. The only other major change compared with the LPM model and the probit models occurs with firm size: the dummy coefficients decrease, that is the dummies for small firms show a massive decline in coefficient size. It seems that the firm-size coefficients absorb quite a lot of unobserved cross-sectional heterogeneity. These coefficient estimates should not be interpreted causally, however; the training propensity of growing or shrinking firms change on much smaller scales, as the fixed-effects estimates show.

The effect of the demography variable is confirmed in model 5, and the coefficient is even higher than in the pooled probit models (though slightly smaller than in the LPM model 4). The change in demography between 1985 and 1995, which would have caused a drop in the share of training firms of 1.8 percentage points according to model 2, now suggests a decrease of 2.7 percentage points.

Moreover, the year dummy estimates of the probit models also hold in the panel regression. The longitudinal conclusions of the pooled model are thus confirmed: the decrease in the share of training firms can be largely explained by a combination of demand-side, supply-side and regional factors. There is no evidence of a steady downward trend in firms' willingness to train since training propensity increased after 1995.

Model 6 introduces interaction effects between demography and the five aggregated industry dummies into the fixed-effects estimation. While time-fixed variables such as industry dummies cannot have their own partial effect on the change of the dependent variable over time, time-variant variables such as demography might for example have a differential effect in different industries. Indeed, column 6 shows that the effect of demography is stronger in the construction sector than in other industries.[23] This seems plausible since in the construction sector there is an excess of training places. The occupations offered by this sector seem to be less attractive to young people. On the other hand, these occupations are very profitable for the firms, since the main construction occupations such as mason and electrician generate high net returns for firms. Therefore, construction firms are most ready to employ (more) apprentices when there are more young people on the market. Though these results are interesting in themselves, the inclusion of the interaction effects had no impact on the other coefficients.

### *Decomposition analysis*

So far, we have seen that there remains a rather small, though significant difference between the years 1985 and 2001 as measured by a year dummy.

[22] The variable "new firm" is usually not included in the estimations because one loses the data for 1985: it is not known how old firms were in 1985 since there was no earlier firm census. The results of these analyses can be found in Mueller and Schweri 2006.

[23] All of the interaction coefficients are significantly different from each other.

Table 2

**Prediction of the share of training firms using data and cross-section probits for all periods with two different sets of independent variables**

(underlying probit estimation results in appendix, table 6)

| Variables in underlying estimations: regions and all firm characteristics | Actual share of training firms | Predicted share of training firms using … | | | |
|---|---|---|---|---|---|
| | | Coefficients 1985 | Coefficients 1995 | Coefficients 1998 | Coefficients 2001 |
| Data 1985 | 0.247 | 0.247 | 0.170 | 0.210 | 0.209 |
| Data 1995 | 0.155 | 0.231 | 0.155 | 0.191 | 0.190 |
| Data 1998 | 0.178 | 0.219 | 0.145 | 0.178 | 0.178 |
| Data 2001 | 0.176 | 0.219 | 0.144 | 0.177 | 0.176 |
| | Actual change 1985–2001 | Explained difference 1985–2001 | | | |
| | –0.071 | –0.028 | –0.026 | –0.033 | –0.033 |

| Variables in underlying estimations: all (regions and all firm characteristics plus supply-side factors) | Actual share of training firms | Predicted share of training firms using … | | | |
|---|---|---|---|---|---|
| | | Coefficients 1985 | Coefficients 1995 | Coefficients 1998 | Coefficients 2001 |
| Data 1985 | 0.247 | 0.247 | 0.231 | 0.268 | 0.258 |
| Data 1995 | 0.155 | 0.188 | 0.155 | 0.187 | 0.187 |
| Data 1998 | 0.178 | 0.179 | 0.149 | 0.178 | 0.177 |
| Data 2001 | 0.176 | 0.179 | 0.148 | 0.177 | 0.176 |
| | Actual change 1985–2001 | Explained difference 1985–2001 | | | |
| | –0.071 | –0.068 | –0.083 | –0.091 | –0.082 |

With the Oaxaca-Blinder decomposition discussed in section 3, we can try to shed more light on the explanatory power of the independent variables used in the analyses. As a starting point, we predict the share of training firms based on the different combinations of data and coefficients for the 4 points in time under consideration. For instance, the probit coefficients resulting from the cross-section estimation with 1985 data will be combined with the 2001 values of the independent variables to predict the share of training firms in 2001. This procedure provides an answer to the question: if firm behaviour (as expressed by probit coefficients) were the same in 2001 as in 1985, but the firm composition[24] of the economy changed, what would the share of training firms be?

The fixed-effects estimation in table 1 showed that firm-size coefficients changed compared with the probit estimations and the latter should not be interpreted causally. Is our decomposition biased since it is based on these cross-sectional probit estimates? Recall that what we are looking for is the best prediction of the share in 2001, assuming that firm behaviour did not change. As regards firm size, we know that from 1985 to 2001, many very small firms newly entered the market.[25] The question now is whether the firm-size coefficients in the 1985 probit provide a good representation of the new firms' training propensity in order to predict the effect of the increase in the number of small firms. If these new small firms behave as the small firms in 1985 did, the coefficients of 1985 provide a good prediction of the training propensity of these new firms.[26] If, however, new small firms are systematically different from existing small firms (with respect to

[24] The same holds for structural changes on the supply side.

[25] The trend towards a larger share of very small firms is well documented (FSO 2000). These firms mainly originate in the expanding service sector.

[26] Note that best prediction does not imply causality. The explanatory contribution of the firm-size variables should therefore not be interpreted causally. Instead, firm size serves as proxy for different unobserved firm characteristics.

their training behaviour), the prediction based on the 1985 coefficients is not valid. Then, the part explained by firm size might be too high in the decomposition.

The discussion in the previous paragraph highlights the main assumption of the decomposition analysis: we assume that the behaviour of the firms as represented by the cross-section coefficient estimates is constant over time. Given this assumption, we show to what extent the change in the share of training firms over time can be explained by changes in the independent variables.

Table 2 displays the actual share of training firms and predictions of this share based on different data combinations. The third column presents predictions based on coefficients from a probit with the 1985 data. The 1985 probit equation was used to predict the share of training firms in the year 1995 using the values for the independent variables in 1995. The share for the year 1998 was predicted using the values of the independent variables in 1998 etc. The predictions using the 1985 coefficients and the 2001 data (i.e. 0.219 and 0.179 respectively) correspond exactly with the computation of the first term in the second bracket of formula (5) (see section 3). The upper panel of the table presents predictions based on cross-section probits without supply-side variables, but with firm characteristics (firm size, industry, firm type) and region dummies as independent variables. The lower panel of the table presents the same predictions, but including supply-side factors (share of 16-year-olds and share of grammar school pupils) in the probits from which the coefficients are taken.

The predictions in the upper panel of table 2 are not particularly good. The predicted difference between the share of training firms in 1985 and 2001 estimated with the different cross-section coefficients lies between −2.6 and −3.3 percentage points, whereas the real difference amounts to −7.1 percentage points.

In the lower panel of table 2, the predicted differences are between –6.8 and –9.1 percentage points. So these predictions are much closer to the true value, which shows again that supply-side factors are important factors that should be included in longitudinal analyses. It is striking, however, to see that the coefficients of the 1985, 1998 and 2001 models provide good predictions for all years with the exception of 1995. Conversely, the 1995 coefficients generate clearly lower predictions for all years. The "outlier" 1995 may be due to the exclusion of business cycle effects, as mentioned earlier.[27]

For the next step, we calculate the contribution of each independent variable to the total predicted difference between 1985 and 2001 using 1985 coefficients.[28] Table 3 shows what contribution the different independent variables make to explaining the total predicted difference of −6.8 percentage points (see table 2).[29]

Table 3

**Non-linear decomposition of the difference in the shares of training firms 1985–2001**

| | |
|---|---|
| Number of observations | 269,978 |
| Difference 1985 to 2001 | –0.0711 |
| Total explained | –0.0678 |
| **Contributions of single variables:** | |
| Regional dummies | –0.0001 |
| Industry dummies | –0.0046** |
| Firm size | –0.0238** |
| Firm type | –0.0002 |
| Rural area | –0.0000 |
| Demography | –0.0219** |
| Share of grammar school pupils | –0.0172** |

*Notes: * and ** denote significance at the 1 percent and 1 per mill levels respectively.*

Changes in four variables contribute significantly to the explained change from 1985 to 2001: firm size, industry, demography and the share of grammar school pupils. The structure of the Swiss economy with regard to firm-size composition and, to a lesser extent, industry composition has thus changed in a way that has led to a decrease in the share of training firms, even though *ceteris paribus* the training propensity of individual firms had not changed.

[27] Another explanation might be measurement error: The number of apprentices that was reported by firms in 1995 is slightly lower than the number of apprentices according to other data of the Federal Statistical Office, while the figures are quite close for other years. The FSO does not have an explanation for this, however, and we could not find any obvious distortion for particular subgroups in the 1995 data.

[28] The predicted difference results from 0.247 − 0.179 = 0.068, see column three in table 2. The probit output for 1985 data can be seen in the appendix (table 6).

[29] Since the results of a non-linear decomposition for the contribution of each variable to the gap may be sensitive to the ordering of variables, we followed the procedure used by Fairlie (2003) and randomized the ordering of switching distributions by using a large number of simulations. The point estimates are the mean results of a process that is replicated 1000 times.

These demand-side factors explain nearly half of the change (42 %). Supply-side factors, namely demography and the share of grammar school pupils, explain the larger part (58 %). This result again supports our view that supply-side factors should not be ignored when analysing the developments on the apprenticeship market.

## 6 Discussion and conclusions

The share of firms training apprentices is an indicator that receives much public attention in Switzerland. The indicator shows a decline of 7.1 percentage points from 1985 to 2001, the first and (up to now) the most recent years for which data are available. This has often been interpreted as a sign of firms' reduced involvement in apprenticeship training. The variation over time in the share of training firms has been analysed in this paper with pooled probit models, fixed-effects models and a non-linear decomposition technique. The main findings are that the unexplained part of the difference between 1985 and 2001 reduces to about 0.5 to 2 percentage points in the different models when all available independent variables are controlled for. There is no clear trend in the unexplained part that would indicate a (negative or positive) trend in the individual firms' willingness to train apprentices.

Part of the decline between 1985 and 2001 can be explained by demand-side factors. Changes in the firm size and – to a lesser extent – the industry composition of the Swiss economy have led to a decrease in the share of training firms. Including supply-side factors, namely the share of sixteen-year-olds in the working-age population as well as the share of grammar school pupils, allows us to explain most of the variation. These supply-side factors provide more than half of the explanation for the observed decline according to a decomposition analysis.

These results are important for policy making since the observed decline in the share of training firms has been used repeatedly as an argument in favour of policy intervention. Do the factors we have found to explain the decline warrant new policy interventions and regulations? One major reason for the decline in the share of training firms is the increase in the share of very small firms in the economy. Since the increase is mainly due to new, additional very small firms, the increase does not crowd out apprenticeship places and does not constitute a serious challenge for the apprenticeship system. The changing industry structure has had a moderate effect on the change in the share of training firms. We do not find a major role of this factor which is often mentioned in conjunction with global trends such as "the knowledge society" and such like. Nevertheless, the increasing share of modern service firms might be a challenge for the apprenticeship system in the future since modern services traditionally have lower training propensities than other industries. It is important that the existing regulations, namely the training regulations that shape every occupation, remain up-to-date and flexible enough to adapt to new developments in the economy.[30]

Turning to supply-side factors, we found a decline in the share of training firms due to changes in demography. This proves the flexibility of the (market-based) system and is no reason for worries. The increasing share of young people choosing grammar school instead of vocational training is another reason for the decrease in the share of training firms, though the results were less unequivocal than for firm size, industries and demography. If the share of grammar school pupils increases in the future, this will obviously have an influence on the apprenticeship system. But a change in the firms' demand for human capital, i.e. a possible substitution of apprentices with grammar school pupils, again provides no reason for state intervention in order to increase the share of training firms. If, however, the state expands grammar schools without a change in firms' human capital needs, this might deprive the apprenticeship market of highly performing youngsters and thus reduce the training propensity of those firms that offer apprenticeship places for high performers. To sum up, the factors we have identified as causing the decline in the share of training firms do not, in our view, call for new state interventions.

All the same, there is a rather small, but not negligible group of young people that do not find apprenticeship places and do not follow post-compulsory education. We have shown that there is no negative time trend in firms' willingness to train apprentices. Why then do not all young people find an apprenticeship place? The most likely explanation is mismatch problems, since at the same time, thousands of apprenticeship places remain vacant every year. Either young people do not find vacancies close to their place of residence (regional mismatch); or they look for places in other occupations, where no vacant places are available; or they did not acquire the

[30] Actually, the Swiss Federal Office of Professional Education and Technology is currently modernising the training regulations for all roughly 250 occupations, together with employer associations and cantons.

necessary qualifications in compulsory school for the vacant places (skills mismatch).

Future empirical research should address this question and test whether mismatch phenomena can explain why some young people do not acquire a post-compulsory education. Further research topics include the impact of the business cycle on the apprenticeship market and the behaviour of firms in the public sector. Ideal employer-employee matched data would also permit the inclusion of individual characteristics of apprentices and apprenticeship candidates in the analyses, such that a system of demand and supply functions as outlined in section 3 could be estimated.

We do not claim to have analysed all of the challenges for the apprenticeship market in this paper, but have focused on one indicator, the share of training firms, which is prominently discussed in Swiss media and politics. We have shown that the decline in the share of training firms in itself is not a sufficient argument for policy interventions. The decline can largely be explained by the emergence of new, very small firms and by supply-side factors such as demography. The latter factors have to be included in a longitudinal analysis to provide a complete and undistorted picture of the developments on the apprenticeship market over time.

## References

*Acemoglu, D., & Pischke, J.-S.* (1998): Why do firms train? Theory and Evidence. Quarterly Journal of Economics, 113, 79–119.

*Acemoglu, D., & Pischke, J.-S.* (1999): Beyond Becker: Training in Imperfect Labour Markets. Economic Journal, 109, 112–142.

*Acemoglu, D., & Pischke, J.-S.* (2000): Certification of training and training outcomes. European Economic Review, 44, 917–927.

*Becker, G. S.* (1962): Investment in Human Capital: A Theoretical Analysis. Journal of Political Economy, 70 (5), 9–49.

*Beckmann, M.* (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, 48, 368–389.

*Beicht, U., Walden G., & Herget H.* (2004): Kosten und Nutzen der betrieblichen Berufsausbildung in Deutschland. Berichte zur beruflichen Bildung (Heft 264). Bielefeld: Bertelsmann.

*Bierhoff H., & Prais S. J.* (1997): From school to productive work. Britain and Switzerland compared. Cambridge: University Press 1997.

*Blinder, A. S.* (1973): Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources, 8 (4), 436–455.

*Clark, D.* (2002): Matching, Screening and Firm Investment in General Training: Theory and Evidence. Paper presented at the 10th International Conference on Panel Data, Berlin, July 5–6, 2002.

*Dustmann, Ch., & Schoenberg, U.* (2004): Training and Union Wages. IZA Discussion Papers 1435, Institute for the Study of Labor (IZA).

*Fairlie, R.* (2003): An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models, Yale University Discussion Paper No. 873, Economic Growth Center.

*Franz, W. & Soskice, D.* (1995): The German Apprenticeship System. In: F. Buttler, W. Franz, R. Schettkat, & D. Soskice (eds), Institutional Frameworks and Labor Market Performance. Comparative views on the U.S. and German economies (p. 208–234). London/New York: Routledge.

*Franz, W., & Zimmermann, V.* (2000): Die betriebliche Ausbildungsbereitschaft im technologischen und demographischen Wandel. ZEW Wirtschaftsanalysen (Band. 46), Baden-Baden: Nomos.

*FSO Federal Statistical Office* (2004): Schuelerinnen und Schueler der Sekundarstufe II: Entwicklungen und Perspektiven. Neuchâtel: Bundesamt fuer Statistik BFS (Bildungsmonitoring Schweiz).

*FSO Federal Statistical Office* (2000): Betriebszaehlung 1998: Schweiz und Kantone 1991/1995/1998. Neuchâtel: Bundesamt fuer Statistik BFS.

*FSO Federal Statistical Office/TREE* (2003): Wege in die nachobligatorische Ausbildung: Die ersten zwei Jahre nach Austritt aus der obligatorischen Schule – Zwischenergebnisse des Jugendlaengsschnitts TREE. Neuchâtel: Bundesamt fuer Statistik BFS (Bildungsmonitoring Schweiz).

*Gerfin, M.* (2004): Firm-Sponsored General Training in Frictional Labour Markets: An Empirical Analysis for Switzerland. IZA Discussion Papers 1077, Institute for the Study of Labor (IZA).

*Leuven, E.* (2005): The Economics of Private Sector Training: A Survey of the Literature. Journal of Economic Surveys, 19 (1), 91–111.

*Malcomson, J., Maw, J., & McCormick, B.* (2003): General training by firms, apprentice contracts, and public policy. European Economic Review, 47, 197–227.

*Muehlemann, S., Schweri, J., Winkelmann, R., & Wolter, S. C.* (2007): An Empirical Analysis of the Decision to

Train Apprentices. LABOUR: Review of Labour Economics and Industrial Relations, 21 (3), 419–441.

*Muehlemann, S. & Wolter, S. C.* (2007): Regional Effects on Employer-Provided Training: Evidence from Apprenticeship Training in Switzerland. ZAF Zeitschrift für ArbeitsmarktForschung, Heft 2+3/2007.

*Mueller, B., & Schweri, J.* (2006): Die Entwicklung der betrieblichen Ausbildungsbereitschaft. Eine Laengsschnittuntersuchung zur dualen Berufsbildung in der Schweiz, SIBP Schriftenreihe Nr. 31.

*Neubaeumer, R., & Bellmann, L.* (1999): Ausbildungsintensitaet und Ausbildungsbeteiligung von Betrieben: Theoretische Erklaerungen und empirische Ergebnisse auf der Basis des IAB-Betriebspanels 1997. In: D. Beer, B. Frick, R. Neubaeumer, & W. Sesselmeier (Hrsg), Die wirtschaftlichen Folgen von Aus- und Weiterbildung (p. 9–41). Muenchen/Mering: Rainer Hampp.

*Oaxaca, R.* (1973): Male–Female Wage Differentials in Urban Labor Markets. International Economic Review, 14 (3), 693–709.

*OPET Federal Office for Professional Education and Technology* (2001): Lehrstellenbarometer 2001. Bern: Bundesamt fuer Berufsbildung und Technologie BBT.

*Ryan, P.* (1998): Is Apprenticeship Better? A Review of the Economic Evidence. Journal of Vocational Education and Training, 50(2), 289–325.

*Ryan, P.* (2001): The school-to-work transition: a cross-national perspective. Journal of Economic Literature, 39(1), 34–92.

*Schweri, J., Muehlemann, S., Pescio, Y., Walther, B., Wolter, S., & Zuercher, L.* (2003): Kosten und Nutzen der Lehrlingsausbildung aus der Sicht Schweizer Betriebe. Chur/Zuerich: Rueegger.

*Soskice, D.* (1994): Reconciling Markets and Institutions: The German Apprenticeship System. In: L. M. Lynch (eds.), Training and the Private Sector: International Comparison (p. 25–60). National Bureau of Economic Research, Comparative Labor Market Series. Chicago/London: University of Chicago Press.

*Stoeger, K., & Winter-Ebmer, R.* (2001): Lehrlingsausbildung in Oesterreich: Welche Betriebe bilden Lehrlinge aus? (Arbeitspapier Nr. 0110). Linz: Johannes Kepler Universitaet, Institut fuer Volkswirtschaftslehre.

*Weber, B.* (2004): Jugendarbeitslosigkeit in der Schweiz. Die Volkswirtschaft, 10, 43–46.

*Wolter, S. C., Muehlemann, S., & Schweri, J.* (2006): Why some firms train apprentices and many others do not. German Economic Review, 7 (3), 249–264.

*Wooldridge, J. M.* (2002): Econometric Analyses of Cross Section and Panel Data. Cambridge/London: MIT Press.

## Appendix

Table 4

### Frequencies for variables used in the analyses

(except canton dummies and full set of industry dummies)

| | 1985 | 1995 | 1998 | 2001 |
|---|---|---|---|---|
| **Firm size** | | | | |
| <2 FTE | 88,811 | 126,269 | 141,149 | 146,184 |
| 2 FTE | 54,953 | 58,309 | 59,750 | 57,639 |
| 3–4 FTE | 51,057 | 53,375 | 52,867 | 52,139 |
| 5–9 FTE | 39,161 | 42,994 | 39,431 | 40,509 |
| 10–19 FTE | 18,656 | 20,495 | 18,886 | 19,572 |
| 20–49 FTE | 11,128 | 11,716 | 10,734 | 11,708 |
| 50–99 FTE | 3,675 | 3,630 | 3,371 | 3,710 |
| 100–149 FTE | 1,121 | 1,052 | 1,080 | 1,224 |
| 150–249 FTE | 793 | 795 | 726 | 845 |
| 250–499 FTE | 409 | 428 | 431 | 533 |
| 500–999 FTE | 158 | 141 | 153 | 174 |
| >1000 FTE | 56 | 42 | 46 | 47 |
| **Firm type** | | | | |
| Independent firm | 220,611 | 263,269 | 281,828 | 289,307 |
| Corporate headquarters | 13,330 | 15,450 | 12,026 | 10,623 |
| Branch | 36,037 | 40,527 | 34,770 | 34,354 |
| **Industry groups** | | | | |
| Traditional manufacturing | 33,264 | 34,103 | 32,833 | 32,479 |
| Modern manufacturing | 9,558 | 11,310 | 10,480 | 10,698 |
| Construction | 28,611 | 35,147 | 35,996 | 36,175 |
| Traditional services | 126,094 | 132,418 | 134,853 | 128,680 |
| Modern services | 72,451 | 106,268 | 114,462 | 126,252 |
| **Greater regions (cantons)** | | | | |
| Lake Geneva Region (GE,VD,VS) | 51,657 | 59,992 | 59,754 | 60,320 |
| Espace Mittelland (BE, FR, JU, NE, SO) | 58,929 | 68,371 | 69,341 | 69,031 |
| North-western Switzerland (AG,BL,BS) | 32,933 | 39,769 | 43,047 | 43,624 |
| Zurich (ZH) | 48,117 | 58,796 | 59,545 | 62,091 |
| Eastern Switzerland (AI,AR,GL,GR,SG,SH,TG) | 40,710 | 46,227 | 48,930 | 49,128 |
| Central Switzerland (LU,NW,OW,SZ,UR,ZG) | 23,331 | 29,478 | 31,439 | 33,338 |
| Ticino (TI) | 14,301 | 16,613 | 16,568 | 16,752 |
| **District** | | | | |
| Urban | 190,091 | 225,869 | 231,822 | 238,391 |
| Rural | 79,887 | 93,377 | 96,802 | 95,893 |
| **Demography** (share of 16-year-olds among working age population) | 2.27 % | 1.65 % | 1.71 % | 1.71 % |
| **Share of grammar school pupils** | 18.8 % | 25.7 % | 25.7 % | 25.5 % |
| **Number of firms** | 269,978 | 319,246 | 328,624 | 334,284 |
| **Number of training firms** | 66,779 | 49,608 | 58,620 | 58,922 |
| **Number of apprentices** | 172,559 | 118,197 | 138,215 | 143,904 |

Table 5

**Relative frequencies of firms and apprentices across different independent variables**

| | Share of firms | | | | Share of training firms | | | | Share of apprentices among all employees | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1985 | 1995 | 1998 | 2001 | 1985 | 1995 | 1998 | 2001 | 1985 | 1995 | 1998 | 2001 |
| Total | 100 | 100 | 100 | 100 | 24.7 | 15.5 | 17.8 | 17.6 | 7.4 | 5.0 | 6.0 | 5.9 |
| **Firm size** | | | | | | | | | | | | |
| <2 FTE | 32.9 | 39.6 | 43.0 | 43.7 | 9.4 | 4.3 | 5.5 | 5.5 | 9.8 | 4.8 | 6.3 | 6.5 |
| 2 FTE | 20.4 | 18.3 | 18.2 | 17.2 | 19.2 | 12.5 | 15.5 | 14.8 | 11.0 | 7.0 | 8.9 | 8.5 |
| 3–4 FTE | 18.9 | 16.7 | 16.1 | 15.6 | 28.8 | 20.1 | 23.9 | 23.6 | 11.7 | 7.9 | 9.4 | 9.2 |
| 5–9 FTE | 14.5 | 13.5 | 12.0 | 12.1 | 37.1 | 27.1 | 32.3 | 31.9 | 10.5 | 7.2 | 8.7 | 8.6 |
| 10–19 FTE | 6.9 | 6.4 | 5.8 | 5.9 | 44.4 | 31.5 | 39.0 | 38.2 | 8.4 | 5.6 | 6.9 | 6.8 |
| 20–49 FTE | 4.1 | 3.7 | 3.3 | 3.5 | 54.1 | 39.9 | 47.8 | 47.7 | 6.5 | 4.3 | 5.2 | 5.3 |
| 50–99 FTE | 1.4 | 1.1 | 1.0 | 1.1 | 64.4 | 49.1 | 59.6 | 56.8 | 5.1 | 3.6 | 4.0 | 4.2 |
| 100–149 FTE | 0.4 | 0.3 | 0.3 | 0.4 | 75.1 | 57.9 | 65.1 | 64.5 | 4.6 | 3.4 | 3.8 | 3.4 |
| 150–249 FTE | 0.3 | 0.3 | 0.2 | 0.3 | 78.4 | 64.9 | 71.1 | 67.6 | 4.3 | 3.2 | 3.4 | 3.6 |
| 250–499 FTE | 0.2 | 0.1 | 0.1 | 0.2 | 86.1 | 75.5 | 74.5 | 73.7 | 4.4 | 3.3 | 3.6 | 3.6 |
| 500–999 FTE | 0.1 | 0.0 | 0.1 | 0.1 | 91.8 | 80.9 | 83.0 | 74.7 | 4.6 | 3.8 | 4.2 | 3.3 |
| >1000 FTE | 0.0 | 0.0 | 0.0 | 0.0 | 92.9 | 88.1 | 69.6 | 80.9 | 5.4 | 3.3 | 3.4 | 3.8 |
| **Firm type** | | | | | | | | | | | | |
| Independent firm | 81.7 | 82.5 | 85.8 | 86.6 | 23.4 | 14.5 | 16.4 | 16.1 | 8.2 | 5.4 | 6.5 | 6.3 |
| Corporate headquarters | 4.9 | 4.8 | 3.7 | 3.2 | 44.9 | 31.2 | 38.8 | 40.3 | 6.2 | 4.6 | 5.4 | 5.3 |
| Branch | 13.4 | 12.7 | 10.6 | 10.3 | 25.7 | 16.6 | 22.4 | 23.3 | 6.3 | 4.2 | 5.0 | 4.8 |
| **Industry groups** | | | | | | | | | | | | |
| Traditional manufacturing | 12.3 | 10.7 | 10.0 | 9.7 | 29.0 | 19.2 | 23.1 | 22.9 | 5.6 | 3.9 | 5.1 | 5.2 |
| Modern manufacturing | 3.5 | 3.5 | 3.2 | 3.2 | 30.9 | 18.0 | 22.8 | 23.0 | 6.2 | 4.4 | 4.6 | 4.7 |
| Construction | 10.6 | 11.0 | 11.0 | 10.8 | 37.1 | 23.3 | 27.4 | 26.5 | 8.9 | 6.8 | 9.1 | 8.9 |
| Traditional services | 46.7 | 41.5 | 41.0 | 38.5 | 22.2 | 14.9 | 17.8 | 18.6 | 8.7 | 5.7 | 7.3 | 7.5 |
| Modern services | 26.8 | 33.3 | 34.8 | 37.8 | 21.6 | 12.3 | 12.8 | 12.3 | 6.9 | 4.2 | 4.4 | 4.1 |
| **Greater regions (cantons)** | | | | | | | | | | | | |
| Lake Geneva Region (GE,VD,VS) | 19.1 | 18.8 | 18.2 | 18.0 | 20.3 | 13.1 | 15.3 | 15.0 | 5.8 | 4.0 | 4.8 | 4.5 |
| Espace Mittelland (BE,FR,JU,NE,SO) | 21.8 | 21.4 | 21.1 | 20.7 | 27.8 | 17.8 | 20.3 | 20.3 | 8.4 | 5.8 | 6.8 | 6.5 |
| North-western Switzerland (AG,BL,BS) | 12.2 | 12.5 | 13.1 | 13.1 | 26.2 | 15.2 | 16.9 | 17.4 | 7.3 | 4.8 | 5.6 | 5.7 |
| Zurich (ZH) | 17.8 | 18.4 | 18.1 | 18.6 | 21.1 | 12.5 | 14.9 | 14.6 | 6.4 | 4.2 | 5.0 | 4.9 |
| Eastern Switzerland (AI,AR,GL,GR,SG,SH,TG) | 15.1 | 14.5 | 14.9 | 14.7 | 26.8 | 18.5 | 21.3 | 21.2 | 8.8 | 6.2 | 7.7 | 7.8 |
| Central Switzerland (LU,NW,OW,SZ,UR,ZG) | 8.6 | 9.2 | 9.6 | 10.0 | 30.0 | 18.5 | 20.2 | 19.3 | 9.5 | 6.2 | 7.4 | 7.1 |
| Ticino (TI) | 5.3 | 5.2 | 5.0 | 5.0 | 22.9 | 13.2 | 15.0 | 14.0 | 6.0 | 3.6 | 4.2 | 3.8 |
| **District** | | | | | | | | | | | | |
| Urban area | 70.4 | 70.8 | 70.5 | 71.3 | 24.2 | 14.8 | 16.9 | 16.6 | 7.1 | 4.8 | 5.6 | 5.5 |
| Rural area | 29.6 | 29.3 | 29.5 | 28.7 | 26.0 | 17.4 | 20.1 | 20.2 | 8.3 | 5.8 | 7.2 | 7.2 |

Table 6

## Estimation of firms' training propensity

(dependent variable: training firm = 1, non-training firm = 0); cross-sectional probit estimations for individual years

| | 1985 firms | | 1995 firms | | 1998 firms | | 2001 firms | |
|---|---|---|---|---|---|---|---|---|
| | Coeffi-cients | Marginal effects | Coeffi-cients | Marginal effects | Coeffi-cients | Marginal effects | Coeffi-cients | Marginal effects |
| **Firm size (Reference = <2 FTE)** | | | | | | | | |
| 2 FTE | 0.472** | 0.148** | 0.582** | 0.133** | 0.593** | 0.150** | 0.561** | 0.140** |
| 3–4 FTE | 0.835** | 0.278** | 0.942** | 0.242** | 0.955** | 0.268** | 0.931** | 0.261** |
| 5–9 FTE | 1.154** | 0.403** | 1.254** | 0.358** | 1.291** | 0.400** | 1.258** | 0.386** |
| 10–19 FTE | 1.415** | 0.510** | 1.456** | 0.456** | 1.540** | 0.512** | 1.502** | 0.497** |
| 20–49 FTE | 1.694** | 0.602** | 1.722** | 0.565** | 1.806** | 0.612** | 1.783** | 0.603** |
| 50–99 FTE | 1.959** | 0.669** | 1.973** | 0.657** | 2.130** | 0.711** | 2.045** | 0.688** |
| 100–149 FTE | 2.288** | 0.726** | 2.209** | 0.725** | 2.301** | 0.750** | 2.271** | 0.744** |
| 150–249 FTE | 2.377** | 0.737** | 2.421** | 0.773** | 2.512** | 0.788** | 2.380** | 0.766** |
| 250–499 FTE | 2.747** | 0.770** | 2.733** | 0.825** | 2.619** | 0.804** | 2.594** | 0.801** |
| 500–999 FTE | 2.979** | 0.781** | 2.923** | 0.847** | 2.919** | 0.836** | 2.608** | 0.803** |
| >1000 FTE | 3.038** | 0.783** | 3.256** | 0.871** | 2.692** | 0.813** | 2.906** | 0.835** |
| **Firm type (Reference group = independent firm)** | | | | | | | | |
| Corporate headquarters | 0.186** | 0.056** | 0.145** | 0.029** | 0.235** | 0.055** | 0.269** | 0.064** |
| Branch | –0.150** | –0.040** | –0.196** | –0.033** | -0.093** | -0.018** | -0.085** | -0.017** |
| **Industry dummies (45)** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| **Rural district (Reference group = urban)** | 0.019* | 0.005* | 0.040** | 0.008** | 0.056** | 0.012** | 0.068** | 0.014** |
| **Demography** | 0.162** | 0.046** | 0.304** | 0.056** | 0.304** | 0.063** | 0.233** | 0.048** |
| **Share of grammar school pupils** | -0.012** | -0.003** | -0.010** | -0.002** | -0.008** | -0.002** | -0.009** | -0.002** |
| **Regional dummies (7)** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Pseudo R2 | 0.177 | | 0.189 | | 0.204 | | 0.204 | |
| N | 269,978 | | 319,216 | | 328,597 | | 334,284 | |

*Notes: Marginal effects are computed at the mean of the other explanatory variables.*
** and ** denote significance at the 1 percent and 1 per mill levels respectively, and are based on robust standard errors.*
*FTE= full-time equivalents. The equivalent of a full-time worker: for example, two people who each work half time correspond to one FTE.*