Four Essays on the Economics of Vocational Education and Training

Inaugural dissertation in fulfilment of the requirements for the degree of

Doctor rerum oeconomicarum

at the Faculty of Business, Economics and Social Sciences
of the University of Bern

Submitted by

BARBARA MUELLER

from Melchnau (BE), Switzerland

2016

The faculty accepted this work as dissertation on 25th February 2016 at the request of the two advisors Prof. Dr. Stefan C. Wolter (University of Bern) and Prof. Dr. Stefan Boes (University of Lucerne), without wishing to take a position on the view presented therein.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Swiss apprenticeship system is strongly market-based and constitutes the main educational pathway at the upper–secondary level of each youth cohort. Recent literature in the field of the economics of vocational education and training (VET) has shed much insight on the underlying economic rationale of firms to train apprentices. The aim of this thesis is to contribute to knowledge of and empirical evidence on the functioning of the Swiss apprenticeship system by studying the interaction with its other market participants, that is, the young people who decide to undertake apprenticeship training. After investigating the long-term dynamics of the Swiss apprenticeship market with a large administrative data set in Chapter 2, the focus shifts to the analysis of outcomes at the individual level of the youngsters who transition from compulsory schooling to apprenticeship training, and, later on, move to the labor market. For this purpose, we use the longitudinal data set *TREE (transition from education to employment)* that comprises PISA literacy test scores for the year 2000 of pupils aged 15 years along with their individual background characteristics and detailed information on their later educational and work pathways. Chapter 3 analyzes the role of easy-to-observe attributes of apprenticeship applicants versus hard-to-obtain information about their ability in firms' hiring decisions. Chapter 4 empirically evaluates the short- and long-term effectiveness of interim solution programs designed to enhance the chances of unsuccessful applicants in the apprenticeship market. Chapter 5 focuses on the labor market outcomes of VET graduates and analyzes the transferability of the human capital acquired from apprenticeships.

1.1 Background

Having an upper-secondary education diploma is becoming increasingly important for successful and enduring labor market integration in all industrialized countries (OECD, 2012a). There is ample international evidence on the pecuniary and non-pecuniary benefits of education for both the individual and the society as a whole (Card, 1999; Harmon et al., 2003; Lange and Topel, 2006; Oreopoulos and Salvanes, 2011; OECD, 2010, 2012b). However, countries differ markedly with respect to their institutional arrangements regarding skill formation at the upper-secondary level. Firm-based VET is of central importance to the educational system in several countries, such as Germany, Austria, Denmark, the Netherlands and Switzerland (OECD, 2009a,b; Wolter and Ryan, 2011). Mass apprenticeship systems are characterized by a legislative framework, regulated programs of learning with fixed duration, formal assessment, and recognized certification. These institutional features play an important role toward maintaining the sustainability of these systems (Acemoglu and Pischke, 2000; Ryan, 2000; Malcomson et al., 2003; Dustmann and Schoenberg, 2012; Steedman, 2012). In Switzerland, two-thirds of a cohort undertake a firm-based apprenticeship at the upper-secondary level, whereas only approximately 20 percent of a cohort choose college-bound high school (SKBF-CSRE, 2014). Nevertheless, because of the high degree of permeability between apprenticeship training and higher education, the tertiary graduation rate in Switzerland hovers around the OECD average (OECD, 2014). Rates of (lifetime) returns to Swiss apprenticeships training are found to be relatively high (Weber et al., 2001). Similar to other countries with predominant apprenticeship systems, youth unemployment rates are rather low and school-to-work transitions are more smooth in general (OECD, 2000; Ryan, 2001; ILO, 2014). The apparently successful performance in providing large numbers of young people with recognized qualifications demanded by the labor market has recently reinforced widespread policy interest in apprenticeships in many other countries (OECD, 2000; UK Parliament, 2009; Hoffman, 2011; Symonds et al., 2011; Steedman, 2012; President Obama's 2014 State of the Union Address, 2014).

The question of why German firms seem to considerably invest in apprenticeship training—providing workers with skills that should be transferable to other employers—has led to a strand of economic literature extending the classical Becker (1962) model with its prediction that firms will not pay for general on-the-job training in perfect markets. The newer training literature suggests that firms' training investment is motivated by several kinds of labor market frictions and information asymmetries that lower post-training mobility and compress the wage structure for higher skill levels, enabling firms to recoup training costs by accruing rents after training (Acemoglu and Pischke, 1998, 1999; Leuven, 2005). Cost–benefit surveys for Switzerland repeatedly show that training firms incur no net costs on average (e.g., Strupler and Wolter, 2012). This result is in line with the standard Becker (1962) model for general on-the-job training in competitive labor markets, where trainees bear the costs of general training through lower wages. While post-training benefits due to strong employment protection are the most decisive factor for firms offering training in Germany, the Swiss case illustrates that labor market frictions per se are not a necessary condition for the sustainability of an apprenticeship training system (Muehlemann et al., 2010). Importantly, it has also been shown that *non-training* firms in Switzerland would face substantial net costs if they were to take up apprenticeship training (Wolter et al., 2006; Muehlemann et al., 2010).

Whereas firms base their decision about whether or not to train, or whether to train or to hire from the external market, on cost–benefit considerations of the respective strategies (Stevens, 1994a; Blatter et al., 2015), policy makers and society hold high expectations on firms' readiness to train. This is because apprenticeship training constitutes the most important part of the upper-secondary education system in Switzerland. These different expectations on the apprenticeship system ideally—but not automatically—coincide with each other.

To this end, on the institutional level, the apprenticeship system is collectively governed by the Confederation, cantons and corresponding professional organizations (SBFI-SERI, 2015b). These three main partners jointly design and regulate approximately 230 occupational tracks in the so-called VET ordinances, aiming to ensure high quality standards and constant adaptation of specific VET programs to the current needs of the labor market. The ordinances cover the legally relevant aspects applicable to a given occupation: they define the occupational profile, training content, criteria to be met by qualified workers in the occupation, training duration, and qualification procedures. Depending on the occupational track, apprenticeships consist of firm-based on-the-job training (3-4 days a week) in combination with formal education in public vocational schools (1-2 days), which lasts for 3-4 years. These federally regulated training occupations cover all domains of the economy and a wide range of intellectual aspiration levels. At the lower end, apprenticeship training is also meant to attract youngsters who might not be capable of completing a full-time-schooling alternative.

On the other hand, besides the high regulation of occupational tracks, the apprenticeship system is fully market-based. The recent socio–political aim to enhance the overall upper-secondary graduation rate from 90 to 95 percent by 2015 (EVD/EDI/EDK, 2011) mainly depends on the capability and willingness of firms to successfully integrate vast cohorts of compulsory school leavers into apprenticeship training. Unlike in a system of general education, youngsters need to have a training contract with a firm. This requires early vocational career orientation of compulsory school graduates, matching supply and demand for apprenticeships in different vocational tracks, and matching prerequisites of apprenticeship applicants with the requirements of firms. In contrast to a general education system, which necessitates a period of on-the-job training after one takes up employment at a firm, apprenticeship training entails early specialization within an occupation. There is, however, no guarantee of employment as a skilled worker in the training firm after graduation. As the majority of Swiss firms provide training without bearing training costs—presumably without coupling the training decision with future skill needs—there is no incentive for the firm to retain a former apprentice to recoup investments.

This thesis aims to contribute to the empirical knowledge on the functioning of the Swiss apprenticeship system with respect to its longitudinal dynamics, its interaction with the behaviors and decisions of heterogeneous young people in the apprenticeship market, and its success at providing knowledge and skills that can be effectively put to use in the labor market after completing training. These aspects are discussed as seperate chapters in this thesis, as outlined below.

1.2 Outline of the thesis

Chapter 2 is devoted to an analysis of the long-term dynamics of training participation of Swiss firms. Since the mid-1990s, the apprenticeship systems in Germany and Switzerland have been mainly discussed in politics and the media under the light of too low a training participation rate of companies, calling for more full-time schooling alternatives or subsidies for training firms. Given that the share of training firms has declined in past decades (most markedly, from approximately 25% in 1985 to nearly 15% in the mid-1990s), a negative trend in firms' willingness to train has been suspected to fundamentally threaten the Swiss apprenticeship system. We use full population longitudinal data of the firm census from 1985 to 2008 to assess to what extent and how the decline in the share of training firms can be explained by a range of independent variables. Besides several firm characteristics that have been shown to be related with training costs (Schweri et al., 2003; Muehlemann et al., 2007), we include supply-side factors such as demographic development of the relevant age cohorts, which have been traditionally ignored in the empirical discussion. Pooled probit models, fixed-effects models and decomposition techniques for dichotomous outcomes (Fairlie, 2003) show that the variation in the share of training firms can be explained to a large extent. Overall, we do not find a *ceteris paribus* decline over time in firms' willingness to train apprentices. The main reasons for the observed aggregate decrease are increasing numbers of (new) very small firms, shifts in industry composition, a reduction in the number of young people, and an increasing share of young people opting for high school. Further, we find a minor, but nevertheless significant, impact of the business cycle on firms' training activity. While most of these factors do not point to increasing structural deficits in the functioning of the apprenticeship system, the (small) effect of industry shifts towards modern services points to the importance of maintaining regulated occupational tracks so that they are up-to date and flexible enough to adapt to new developments in the economy.

However, despite responses to supply-side variation in the apprenticeship market, the so-called "Lehrstellenbarometer" (apprenticeship barometer; OPET, 2001), a yearly cross-sectional survey of youngsters and firms as of 1997, consistently shows a non-negligible share of compulsory school leavers who do not find an apprenticeship position. Simultaneously, firms cite the unsuitability of candidates as the main reason for thousands of unfilled vacancies. While there are no adequate data to analyze the source of potential mismatch problems in the apprenticeship market, the next chapter analyzes the question on the manner in which apprenticeship applicants are sorted into apprenticeships.

Chapter 3 is devoted to an analysis of the selection decision of firms in the hiring process of apprentices. In the public discussion, stereotyping is claimed to play too dominant a role in firms' hiring decisions, such that applicants with unfavorable attributes (e.g., low parental socioeconomic status or migration background) are at a disadvantage to secure (good) apprenticeships, presumably irrespective of their true ability. Using the sample of compulsory school leavers from the TREE data, who were in the apprenticeship market the year after the PISA-test 2000, we analyze whether and to what extent firms successfully obtain and consider information about an applicant's difficult-to-observe abilities in the hiring process for apprenticeship posts. The PISA reading literacy competence test scores provide us with an ability proxy that is only observable by the researcher, not by recruiters of training firms. On the one hand, the literature predicts that when information about the abilities of job seekers is difficult to obtain, statistical discrimination by employers may be an efficient strategy in the hiring and wage-setting process (Phelps, 1972; Spence, 1973; Arrow, 1973; Aigner and Cain, 1977). On the other hand, this strategy might be costly. Due to specific institutional regulations, such as standardized content, fixed duration, and little scope to adjust prearranged wage profiles over the apprenticeship period, we hypothesize that firms' expectation error might be subject to an asymmetric risk: hiring someone whose ability level considerably lies below the expected level can lead to severe costs, while it is not apparent that firms would profit much in the reverse case.

Following the procedure in Farber and Gibbons (1996), we use the test score information cleaned from the part that is explainable by observables, representing the ability component that is hard-to-observe for outsiders. We then go a step further and explicitly differentiate between positive and negative deviations from the predicted ability level. The empirical results are as follows. First, we find that a deviance in the PISA test scores (from what one would predict based on easy-to-obtain observable characteristics) significantly influences the probability of succeeding in the transition to a firm-based apprenticeship but in a non-symmetric way. In line with our hypothesis, only those with a test result below their predicted result (the so-called "underachievers") have significantly lower chances of securing an apprenticeship. Second, as for the resulting allocation of successful applicants into different intellectually demanding vocational tracks, ordered probit estimations show rather symmetric effects; hard-to-get ability information is considered in a way that significantly increases allocative efficiency at both ends of the distribution. Taken together, the results implicate considerable pre-market employer learning. Further, results regarding longer-term outcomes suggest that additional revelation of ability occurs during the subsequent training period. Apprentices who are PISA overachievers are less likely to face problems such as dropping out, repeating a year of apprenticeship, changing the vocational track, or fail in the final exam. In contrast, PISA underachievers who, despite their lower-than-expected ability, successfully secure an apprenticeship are disproportionally more likely to be exposed to these problematic events. This provides an additional explanation for why firms seem to place more emphasis on detecting underachievers rather than overachievers in the course of the hiring process.

Chapter 4 evaluates the short- and long-term consequences of following so-called interim solution programs. During the last few decades, a variety of such non-certifying programs have been established to enhance the chances of unsuccessful applicants in the apprenticeship market. These programs bridge the one-year gap between compulsory schooling and upper-secondary education with additional schooling or practical training. The role and effectiveness of interim solution programs is controversial, and, as far as we know, has not yet been empirically investigated. This chapter analyzes how gap years spent in *interim solution programs* affect, first, the chances of entering certifying education in the subsequent year; second, the intellectual aspiration level of the certifying education taken up, and third, the chances of successfully completing upper-secondary education by age 21. Following an interim solution program is compared to both having a gap year after compulsory schooling *without educational activity* and securing *direct entry* into certifying education. Propensity score matching techniques with multiple treatments (Rosenbaum and Rubin, 1983; Lechner, 2001) and the unusually rich information in the TREE data are used to build adequate control groups.

The results show substantial program effects compared to having a no educational activity during the gap year. Participating in interim solution programs enhances subsequent chances to enter upper-secondary education by approximately 26 percentage points, decreases the probability to only enter certifying education at a low intellectual aspiration level (as compared to middle/high-level tracks) by approximately 25 percentage points, and increases the graduation probability at age 21 by approximately 30 percentage points. The estimated program effects for both participants and non-participants are similar. In turn, there is no evidence of positive program effects when program participants are compared to those who directly enter upper-secondary education. The intellectual aspiration level of the subsequent certifying track is not affected on average, and the probability of having no diploma or enrolment by age 21 is slightly lower (by 5 percentage points). However, there is evidence of some positive program effects for specific sub-groups: participants from low-level compulsory school tracks and those with low PISA literacy test scores are more likely to enter more demanding certifying tracks than otherwise identical peers with direct entry. Overall, estimated program effects are only small in either direction when comparing interim solutions to direct entries.

As a whole, the most important implication from a policy point of view are severe short- and long-term consequences of not following any kind of (transitory) educational activity directly after compulsory schooling. Such gap years are very strong predictors for a failed upper-secondary education career as a whole. This might go along with social costs that might best be countered at an early stage.
According to our results, the group at risk can broadly be characterized as follows. They show comparable initial school performance to the direct entry group (which is heterogeneous) and rather better performance than the interim solution group. However, their parental background is the least favorable on average, with respect to family structure, socioeconomic index, and number of books at home. Additionally, they exhibit a higher tendency towards school absenteeism during compulsory school. In bivariate (but not multivariate) comparisons, parental support in scholastic matters and (self-assessed) school effort are less favorable, too. Importantly, for virtually all individuals in gap years or interim solutions, we find comparable individuals in the direct entry group (the reverse does not hold, however), implying that the mix of their characteristics should not impede, per se, integration into certifying education or training.
Thus, preventing pupils from completely dropping out of the educational system by the end of compulsory school seems very important. Recent policy efforts to improve early detection and individual guidance (case management) during the end of compulsory schooling already go in this direction.

Chapter 5 analyzes the specificity and transferability of human capital acquired in apprenticeship training by analyzing inter-firm and occupational mobility and their (causal) effects on post-training wages one year after graduation. While the economic rationale for a comprehensive work-based apprenticeship system is to provide trainees with a set of clearly defined occupational skills, there is a lack of empirical evidence on how successful the Swiss apprenticeship system is in producing these sorts of skills. As shown in Stevens (1994b), one potential challenge of apprenticeship training is that, in the presence of monopsony power of firms, there is an inherent incentive for firms to distort the regulated training content towards firm-specific components, in an effort to reduce across-firm mobility of workers and extract a rent after training. The existing empirical literature on mobility after apprenticeships refers mainly to Germany and might not be generalizable to other countries. A recent German study found that pure firm changes and occupation-and-job changes causally result in average wage losses (Fitzenberger et al., 2015). By analyzing the outcomes in Switzerland, we can shed light on the outcomes of mobility from comprehensive apprenticeships schemes (like those in Germany) under more lightly regulated labor market conditions, similar to those that prevail, for example, in English-speaking countries.

The data–base for the analysis is composed of individuals surveyed in TREE who completed their apprenticeship training by 2005 and are observed in the labor market in the subsequent wave. To estimate the effects of firm- and occupation-specific components of human capital, we build three groups that identify those changing firms within occupations (firm movers) and those changing firms across occupations (occupation changers), as opposed to staying in the training firm and occupation (stayers). As for the occupational boundaries, we use a rather broad definition of the occupational field (the two-digit occupational level within the Swiss occupational nomenclature of jobs). Firm mobility within one year after completing an apprenticeship is shown to be high (49 percent of all apprenticeship graduates, including occupation changers), whereas mobility out of the learned occupational field is limited (7 percent of all graduates).

In ordinary least squares (OLS) wage regressions, we find no significant differences in wages of firm stayers and firm movers within their learned occupational field, even when we control for ability and match quality proxies. Occupation changers, however, earn almost 5 percent less. We then apply a treatment regression approach that considers the effect of the endogenously chosen multinomial-valued mobility decision on wages (Deb and Trivedi, 2006). The results still implicate no wage effect for firm movers who stay within the learned occupational field, but there is a negative wage differential of approximately 9 percent for occupation changers. Additional results show that wage cuts upon changing the occupation depend on the distance between occupations.

Overall, the results show high transferability of occupational skills across firms and do not provide evidence that regulated training contents are distorted by firms towards components that are too specific. This finding is in line with other empirical evidence for Switzerland (e.g., Muehlemann et al., 2010) implying that—in the lightly regulated Swiss labor market—firms cannot rely their training decisions and behaviors on the presence of labor market frictions.

Chapter 2

The training participation of firms between 1985 and 2008

2.1 Introduction

[1]Fluctuations in the apprenticeship market cause regular and intense discussions in Swiss politics and the media. It has been feared, for example, that the market-based apprenticeship system is increasingly failing to provide many young people with apprenticeship training.

The quantitative importance of problems in the apprenticeship market is controversial. The most important question from a policy view-point 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 position but cannot find one is unknown. The "Lehrstellenbarometer" (apprenticeship barometer) provides partial answers: it is an annual cross-sectional survey conducted since 1997 that gives an indication of how many young people are still looking for an apprenticeship position in April (1st survey) and in August (2nd survey) of the survey year. Of the young people looking for an apprenticeship position at the beginning of the year, around 15% each year had only received a place in a one-year preparatory course or the like, and around 5% each year had found no solution at all by August. Even though these shares show some variations between 1997 and 2014, the broad picture has remained rather constant (see for example OPET, 2001, and all the other yearly publications).

[1] This chapter is partly based on—and an extension of—the paper Schweri, Juerg and Barbara Mueller (2007): Why has the share of training firms declined in Switzerland, ZAF - Journal for Labour Market Research, 40(2/3), 149-167.

Although these figures seem quite low in comparison with other countries (e.g., for Germany, see Buchholz et al., 2012), they still indicate that a segment of compulsory school graduates cannot be integrated (directly) into the apprenticeship system. Public discussions are, therefore, intense and several political initiatives have been launched at the federal and cantonal level, many of them demanding state intervention in the apprenticeship market, such as firm subsidies, tax relief, or the expansion of full-time schooling. Similar discussions have been underway in Germany and Austria, too.

One important reason put forward in favor of policy interventions is that firms do not train enough apprentices, and that, in particular, their willingness to train has declined over time. This development has often been diagnosed from a simple descriptive indicator, namely the share of training firms in the economy, which has decreased from over 24% in 1985 to around 18% in more recent years. We focus on this argument and show that the causes of the decline in this indicator have to be considered before far-reaching conclusions can be drawn.

To this end, we analyze 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. As the share of training firms results from both supply-side and demand-side forces in the apprenticeship market, a ceteris paribus decreasing willingness of firms to train—supposedly due to increased costs for a specific firm in providing a specific sort of training—is not the only potential explanation for decreased shares of training firms in the economy.

Supply-side factors are important variables that have been largely ignored in political discussions as well as empirical research. In Switzerland, most young people start an apprenticeship directly after compulsory schooling. Therefore, the number of young people looking for apprenticeship places should be strongly influenced by demographic development of the relevant youth cohorts. This number should also be determined by the educational preferences of young people, for example, the observed shift towards more tertiary education. Changes in the supply of potential apprentices will affect the outcome in the apprenticeship market.

On the other hand, changes in the economic composition offer a potential mechanism for a decline in the aggregate demand for apprentices. The cost structure of training provision, and thus a firm's demand for apprentices, has been shown to be considerably related to firm characteristics, such as firm size, industry, and location (Schweri et al., 2003; Muehlemann et al., 2007). Even without changes in the cost structure of training provision for a specific firm-type, a change in the composition of firm characteristics in the economy could result in a change in overall training provision over time. For example, Sheldon (2005) pointed out that the ongoing shift towards the service sector might weaken the apprenticeship system because such industries gain importance where apprenticeship training is observed to be less widespread.

We use the full population data of the Swiss firm census from 1985 to 2008 to analyze whether, to what extent, and how the change in the aggregate training activity of Swiss firms can be explained. These data allow us to shift the analysis to the individual firm level, and at the same time, they provide us a comprehensive picture of economic composition over two decades. We exploit the information on whether a firm is training at a specific point in time, the set of firm characteristics, and matched information on the local supply-side situation and industry-specific GDP growth. By applying pooled probit estimations, fixed-effects methods, and the decomposition technique for dichotomous outcomes proposed by Fairlie (2003), we aim to answer whether and to what extent the change in the share of training firms can be attributed to shifts in average firm characteristics, to changes in the supply of potential apprentices, and further, to fluctuations in the business cycle. By doing so, we also test the hypothesis that there is an "unexplainable" decline in firms' propensity to train apprentices.

The remainder of this chapter is organized as follows. Section 2.2 introduces the Swiss apprenticeship system and Section 2.3 the relevant literature. The estimation strategy is presented in Section 2.4. Section 2.5 describes the data source and provides descriptive information. Section 2.6 contains the empirical results, and Section 2.7 concludes with a discussion of the findings.

2.2 The Swiss apprenticeship system: institutional background

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 program 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 high 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 (SKBF-CSRE, 2011). Vocational training in a dual-training program usually lasts 3-4 years. A few of the approximately 230 occupations permitted an apprenticeship period of just two years in the past (mostly in the retail sector). Firms report fairly low dropout 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 "Berufsmatura" (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's degree. The quality of the training provided in Switzerland, which combines school lessons (for 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 (see, for example, Bierhoff and Prais, 1997). 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 through 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).

The apprenticeship system is market-based: young people have no guarantee of receiving an apprenticeship position nor are firms obliged to train apprentices. The apprenticeship market can, therefore, be seen as a sub-market of the labor market. Although employers' organisations often issue salary recommendations for apprentices, their salaries are determined by the employing company and are not regulated by law or governed on the basis of collective agreements between trade union federations and employers' federations.[2] It is crucial for the justification of our estimation strategy that the apprenticeship market works as a market and is not dominated by the state's or associations' regulations (e.g., by means of collusion).[3] 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 firms were asked about the cost and benefit of their apprenticeship training programs.

Figure 2.1: Boxplot of monthly wages (average across apprenticeship years, CHF) in selected occupations; duration of apprenticeship periods in brackets. Data Source: Schweri et al. (2003)

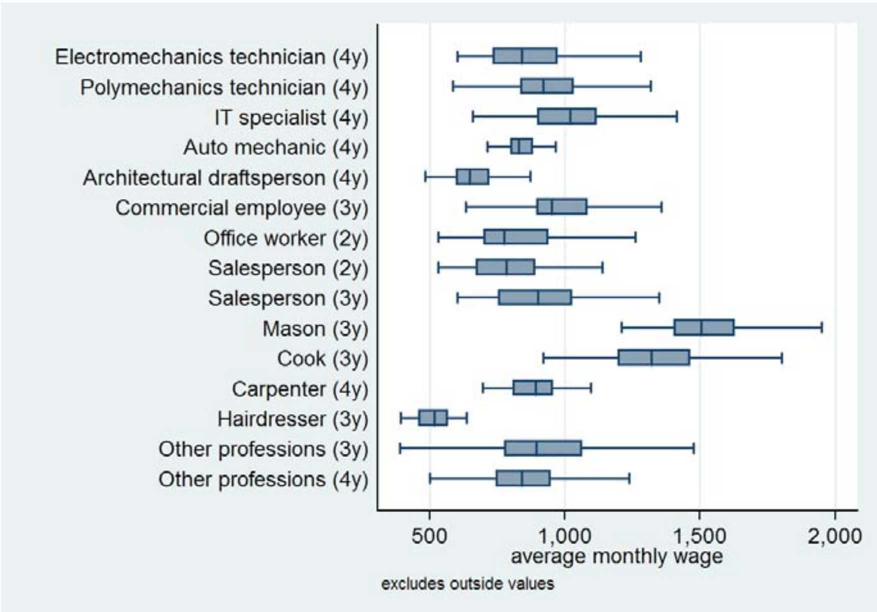


Figure 2.1 shows the wage variance between the most important occupations in terms of numbers of apprentices. Although wages do not vary within one firm for apprentices in the same occupation, figure 2.1 exhibits a high degree of wage variance within as well as between occupations, as can be expected in a free market where wages adapt to differing conditions and scarcities between occupations, industries, regions, and individual firms.

---

2 In a recent comparison of apprenticeship wages in Germany, Britain, and Switzerland based on evidence from both national statistics and on-site interviews with managers in the metalworking industry, Ryan et al. (2012) suggest that there exists some informal pay-coordination within employers organizations in Switzerland, which, however, rather depend on social norms.

3 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).

2.3 Related literature on the training decision of firms

The economic literature lists two main reasons for the provision of training by firms (Lindley, 1975): firms have a "production-oriented" motive and profit by substituting unskilled or skilled workers with (cheaper) apprentices, or firms train out of an "investment-oriented" motive to meet their future needs for skilled workers, recouping their investments after training. As apprenticeship training in Germany, Switzerland, Austria, and other European countries is regarded to provide general skills of potential use to many firms, the investment-oriented motive lies outside the classical prediction of Becker (1962) that firms never pay for general training in perfect markets. The reason is that firms cannot extract a rent from skilled workers when labor markets are competitive, so they cannot recoup their investments after training. Empirical evidence showing substantial firm investments in German apprenticeship training (see Harhoff and Kane, 1997; Acemoglu and Pischke, 1999) has led to a new strand of economic literature on firms' motivation to train apprentices. To explain firms' investments, the new training literature focuses on the role of market imperfections that reduce workers' mobility and create a wedge between productivity and wages at higher skill levels (Stevens, 1994b; Acemoglu and Pischke, 1998, 1999; Leuven, 2005). The training provided in frictional labor markets will, however, remain below the social optimum that is achieved in competitive labor markets where individuals pay for their training. A number of studies have attempted testing the premises of this new training literature (Acemoglu and Pischke, 1998; Clark, 2001, 2002; Beckmann, 2002; Euwals and Winkelmann, 2004; Dustmann and Schoenberg, 2009) by comparing wages between those workers who stayed in their training firm after completing the training (stayers) and those who left (movers). Depending on the exact assumptions about labor market frictions, different hypotheses about movers' and stayers' wages emerge. The mover–stayer literature is, however, far from conclusive (see Clark, 2002).

Recent cost–benefit surveys for Switzerland show that Swiss training firms bear no training costs on average: about two-thirds of firms provide training at a negative net cost by covering all costs related to training with the apprentices' productive work during the apprenticeship period (Schweri et al., 2003; Muehlemann et al., 2007; Strupler and Wolter, 2012). This result is in line with Becker's classical prediction that apprentices will pay for general training through a low training wage. 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 behavior 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 (Wolter et al., 2006; Muehlemann et al., 2010). Importantly, the decision to train has been shown to significantly depend on a firms' expected net cost or net gain from training (Muehlemann et al., 2007). Net costs and net gains are observed to vary between different firm characteristics, such as firm size and industry, and between training occupations and regions, but also within the same occupation and industry.

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, leading to considerable net costs on average (Beicht et al., 2004). A recent comparison of Swiss and German data shows that the difference in the cost–benefit structure can be explained by differences in relative wages of apprentices and skilled workers and the higher contribution to the production process in Switzerland (Dionisius et al., 2009). This differences are consistent with the strong differences in the regulation of the labor market in both countries (Muehlemann et al., 2010). The Swiss labor 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 labor market regulations; there are more than 50% of German apprentices (as opposed to 36% in Switzerland) still employed by their original training firm one year on. This is not to say that Swiss labor markets are entirely frictionless. Actually, one-third of Swiss apprentices create net costs for their training firms that have to be recouped somehow. Recent evidence shows, for example, that Swiss firms offer more training in less dense regional labor markets, where the poaching thread of other firms is lower and mobility costs higher (Muehlemann and Wolter, 2011), and in the presence of high external hiring costs (Blatter et al., 2012). There has been found evidence for the presence of moderate, but however significant, monopsony power over skilled as opposed to unskilled workers in the Swiss labor market (Muehlemann et al., 2013). However, monopsony power is shown to be even larger over trainees than over skilled workers, providing additional source of evidence why the training activity of the majority of Swiss firms is well described by a production-oriented training strategy.

Since the cost–benefit data are more recent, there is no empirical evidence available on whether costs to provide training have changed over the period of our interest. Yet, there are no studies, as far as we know, that analyze the longitudinal dynamic of the Swiss apprenticeship training system at the firm level.[4] Developments over time have, however, been analyzed for Austria by Stoeger and Winter-Ebmer (2001) for the period 1976 to 1998. They find a strong negative time trend in training activity that is only marginally explained by firm characteristics. The largest part of the decline remains unexplainable. They did not include supply-side factors, however, of which we expect a separate effect as explained in the Introduction.

Supply-side information has been included by Muehlemann and Wolter (2007) in a cross-sectional analysis of firms' training activity based on Swiss cost–benefit data. They find sizeable effects for both the number of young people and the share of high school enrolment on training provision. While higher numbers of applicants increase the probability for a successful match between potential training firms and candidates, a higher share of high school enrolment presumably attracts the most able school leavers, lowering firms' expectation on the average ability level of the remaining pool of potential applicants. Demography has also been shown to provide a relevant explanation for developments in the number of apprentices over time at the aggregate cantonal level between 1988 and 2004 (Muehlemann et al., 2009), with much higher effects than those found for the business cycle. Existing studies regarding the effect of business cycles on apprenticeship training mostly show a procyclical relationship (for Switzerland see Schweri and Mueller, 2008; Muehlemann et al., 2009; for Germany see Dietrich and Gerner, 2007; Troltsch and Walden, 2010). In economic downturns, when business is low and orders are insufficient, employers seem to be reluctant to hire additional workers, such as new apprentices. This stands in contrast to the often found countercyclical effects between business-cycle and training incidence of incumbent employees (see Brunello, 2009, for a review on the international literature on the relationship between the business cycle and training and between the business cycle and apprenticeship).

[4] With the exception of analyses with older versions of this data for the Federal Statistical Office, including data up to 2001,2005, and 2008, respectively (Mueller and Schweri, 2006; Schweri and Mueller, 2007, 2008; Mueller and Schweri, 2012a).

2.4 Estimation Strategy

As discussed in the previous sections, the Swiss apprenticeship market strongly relies on market forces. As a whole, the functioning of the system is rather consistent with the result in highly competitive markets, where it is assumed that apprentices' wages are adjusted such that the cost of training is allocated to the apprentice. The number of apprenticeship places in the economy is thus expected to be determined by supply and demand in the apprenticeship market. The share of training firms is a result of the market outcome. To test whether firms' propensity to train has declined ceteris paribus, we want to estimate firms' demand for apprentices. With the firm-level data we describe in section 2.5, we would therefore like to estimate a firm demand function:

$$A_{it}^d = \gamma_t + \gamma_w w_{it} + X_{it}^d \gamma_d + \mu_{it} \tag{2.1}$$

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}^d, w_{it} in equilibrium. The supply side is represented by

$$A_{it}^s = \delta_t + \delta_w w_{it} + X_{it}^s \delta_s + \upsilon_{it} \tag{2.2}$$

Since $A_{it}^d = A_{it}^s$ in equilibrium, w_{it} is endogenous in equation (2.1). The classical solution is to use supply shifters X_{it}^s as instruments for w_{it} in the demand equation (2.1). Our data set (see section 2.5) provides us with supply shifters, but not with wage data. We therefore cannot estimate the structural equation (2.1) and instead estimate the reduced form equation where training propensity A_{it} is a function of all the independent variables in the model.

$$A_{it} = \beta_t + X_{it}^s \beta_s + X_{it}^d \beta_d + \epsilon_{it} \tag{2.3}$$

One of 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 apprentices has been steadily declining over time. Assuming that no relevant time-variant variables have been omitted from the X matrices, this corresponds to the hypothesis that β_t is declining for higher t in a *ceteris paribus* consideration.

Supply shifters will be the share of 16-year-olds in a region and the share of high 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. Our hypothesis is that an increase in the number of young people or a decrease in the share of high 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 β_t. The hypotheses on the effect of the demand-side independent variables, i.e. on the β_d, are discussed in the next section where we present the data set.

Equation (2.3) 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 analyze the binary training decision:

$$I_{it} = \begin{cases} 1 & \text{if } A_{it} > 0 \\ 0 & \text{if } A_{it} = 0 \end{cases} \tag{2.4}$$

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 et al. (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 analyzed, 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) and Muehlemann and Wolter (2011).

In this chapter, we concentrate on I_{it} as a binary training decision variable. 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.[5]

After discussing the estimation results, we proceed with the assessment of the relative importance of the different variables for explaining the observed decrease in the share of training firms from 1985 to 2008. To this end, we use the decomposition idea introduced by Blinder (1973) and Oaxaca (1973) and apply it to an analysis of differences across time.[6] The change in average training activity $\bar{I}^{08} - \bar{I}^{85}$ is:[7]

$$\bar{I}^{08} - \bar{I}^{85} \quad = \quad \hat{\beta}^{08}\bar{X}^{08} - \hat{\beta}^{85}\bar{X}^{85} \tag{2.5}$$

After extending the right-hand side with $\hat{\beta}^{85}\bar{X}^{08} - \hat{\beta}^{85}\bar{X}^{08}$, equation (2.5) can be written as

$$\bar{I}^{08} - \bar{I}^{85} \quad = \quad (\bar{X}^{08} - \bar{X}^{85})\hat{\beta}^{85} + (\hat{\beta}^{08} - \hat{\beta}^{85})\bar{X}^{08} \tag{2.6}$$

The change $\bar{I}^{08} - \bar{I}^{85}$ can thus be decomposed into two parts: a part that is explained by changes over time in the distribution of independent variables (the first part of equation (2.6), also called the "endowment effect"), and another part that cannot be explained and is attributed to a change in the coefficients (the second part of equation (2.6), the so-called "unexplained part"). The unexplained part should only be interpreted with caution, as it not only captures behavioral differences but also reflects the portion of the gap that is due to differences in unobserved endowments. As in most other studies using this technique, we only interpret that part of the gap that can be explained by measurable characteristics X.

The Blinder–Oaxaca decomposition can also be applied to non-linear models with dichotomous dependent variables (Fairlie, 2003).[8] As we want to analyze the source of differences in the 0/1 training decision over time, this technique is more appropriate. The decomposition

[5] Estimation results for the share of apprentices among all employees in a firm have been published in Mueller and Schweri (2006); Schweri and Mueller (2008); Mueller and Schweri (2012a). The results do not alter our main conclusions.

[6] The typical application is to analyze the source of differences between two groups, e.g., wage differentials between males and females or between natives and migrants.

[7] In fact, the decomposition can be used for any combination of the points in time under scrutiny. We use the two extreme points in time (1985 and 2008) in the formulas to simplify the discussion.

[8] An early approach to decompose differences in proportions is also provided by Gomulka and Stern (1990), who additionally also apply it to an "over time" perspective.

can then be written as

$$\begin{aligned}\bar{I}^{08} - \bar{I}^{85} \quad = \quad & \left[\sum_{i=1}^{N^{08}} \frac{F(X^{08}\hat{\beta}^{85})}{N^{08}} - \sum_{i=1}^{N^{85}} \frac{F(X^{85}\hat{\beta}^{85})}{N^{85}}\right] + \\ & \left[\sum_{i=1}^{N^{08}} \frac{F(X^{08}\hat{\beta}^{08})}{N^{08}} - \sum_{i=1}^{N^{08}} \frac{F(X^{08}\hat{\beta}^{85})}{N^{08}}\right] \end{aligned} \tag{2.7}$$

The basic idea is the same as above. However, equations (2.5) and (2.6) are not valid for non-linear models since $\bar{I} = F(\bar{X}\hat{\beta})$ does not necessarily hold as in linear models. Equation (2.7), therefore, averages predicted training probabilities $F(.)$, where $F(.)$ is the cumulative distribution function from the standard normal distribution in the case of probit estimations. In order to evaluate the total contribution of the independent variables (endowments) to the gap in the share of training firms across time (the first part of equation (2.7)), we only need to estimate two sets of predicted probabilities by holding the $\hat{\beta}$ vector constant and then take the difference between the average values. However, one must decide which $\hat{\beta}$ vector to use in order to weight the first term of the decomposition. An equally valid decomposition can be done by either weighting the changes in X by, for example, the coefficients of $\hat{\beta^{85}}$ (as in equations (2.5) to (2.7)) or by the coefficients of $\hat{\beta^{08}}$. Therefore, it has become a popular alternative to weight the term by coefficients derived from a pooled estimation. The results presented in section 2.6 will stem from the coefficients estimated based on both single years and by pooling all waves of our data.[9]

The Fairlie decomposition technique also allows us to further decompose the explained part in order to identify the contribution of each independent variable to the total gap. The empirical importance of supply-side factors can thus be compared with that of the demand-side factors. The linear decomposition for a single variable can be easily seen in the first part of equation (2.6): the changes in the means of the independent variables are multiplied by the respective coefficients. The total explained difference thus results from summing up the changes caused by shifts in the distribution of the individual independent variables. In the case of binary outcomes, a detailed decomposition is less straightforward, because the independent contribution of one variable depends on the value of the other variables through the respective non-linear function. Furthermore, the equation has to be adapted for $N^{85} \neq N^{08}$.

[9] In the same vein as Fairlie (2003), we include year dummies in the probit estimation of the pooled version but do not use them to calculate the detailed decomposition described later.

Assuming an identical number of observations over time, the resulting change in the average predicted probabilities due to the independent contribution of X_1 to the gap can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\Big[F(X_{1i}^{08}\hat{\beta}_1^{85}+X_{2i}^{85}\hat{\beta}_2^{85})-F(X_{1i}^{85}\hat{\beta}_1^{85}+X_{2i}^{85}\hat{\beta}_2^{85})\Big] \tag{2.8}$$

The contribution of a single variable to the explained gap is thus evaluated by replacing the distribution of this variable with the distribution of the same variable in another year while keeping the distribution of the rest of the independent variables X_2 at their initial levels. A property of this procedure is that the sum of the single contributions will be equal to the total contribution of all variables to the gap.

The procedure described in Fairlie (2003) entails one-to-one matching of firms between the two years of interest. If the sample size varies between groups (years), the estimation involves drawing random samples from the larger group. Our reported results stem from estimations with 1000 random replications. Since the separate contributions of each variable may be sensitive to the ordering of variables, we choose to randomize the ordering of switching distributions, too.[10]

While cross-section analysis as described above may serve well to describe and predict firms' training propensities based on observed characteristics for a specific point in time, it is less suited to identify causal effects. Unobserved heterogeneity is a potential problem in our estimations. As discussed in section 2.5, only a limited number of independent variables are available. Many firm and regional labor market characteristics that might influence a firm's training decision are not observed in our data. If these unobserved variables, denoted henceforth by c_{it}, are part of ϵ_{it} in equation (2.3), and thus correlated with the observed independent variables, the estimated coefficients of the latter will be biased.[11]

Given a linear model and assuming that unobserved firm characteristics c are constant over time,

$$A_{it} = X_{it}\beta + c_i + \mu_{it} \tag{2.9}$$

[10] The STATA module to perform decomposition techniques as described in Fairlie (2003) has been provided by Jann (2006).

[11] 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.

we can get rid of the unobserved heterogeneity c_i in equation (2.9) by time-demeaning all variables:

$$(A_{it} - \bar{A}_i) \quad = \quad (X_{it} - \bar{X}_i)\beta + (\mu_{it} - \bar{\mu_i}) \tag{2.10}$$

Equation (2.10) describes the linear fixed-effects model that can be consistently estimated under the usual OLS assumptions.

Using a binary dependent variable I_{it} that indicates whether a firm trains, we estimate two different fixed-effects models. First, we perform a fixed-effects estimation based on a linear probability model (LPM). This allows us to compare results with an ordinary LPM based on pooled cross-sectional estimations. Second, we also estimate a conditional logit fixed-effects model for comparison. This method is generally preferable; however, it suffers from some drawbacks (Wooldridge, 2002); only observations that show a change in the dependent variable over time contribute to the maximum likelihood. Because many firms in our sample either always train or never train, we lose about 80 percent of all observations and probably end up with a sample that no longer represents the population of interest (then, we only analyze a sample of ex-post observed changes over firms that might have been especially sensitive to changes in some firm-characteristics or regional circumstances). Further, there is no satisfactory way to derive partial effects: the conditional maximum likelihood does not provide estimates of the individual fixed effects, which would be needed to compute marginal effects that are comparable to ordinary binary models.

2.5 Data and hypotheses

Our basic data set consists of firm census data of the Swiss Federal Statistical Office (FSO). The firm census was conducted out in 1985, 1991, 1995, 1998, 2001, 2005, and 2008. As of 2008, the firm census was replaced by register data (called STATENT), which do not allow direct comparison with the firm census data. Therefore, our observation period is 1985 to 2008. We exploit the information on whether a firm has currently employed one or more apprentices. The question about apprentices was not asked in 1991. A major advantage of the data is that the firm census encompasses the full population of Swiss firms in the industry and service sectors.[12] We only use the data of private marked-based firms for our analysis. Since we can trace firms that existed in more than one of the survey years, we can construct a panel data set that does not suffer from attrition due to non-response. The survey frequency of every 3-4 years is not a major problem for our purpose since apprenticeships last that long, too. The large gap between 1985 and 1995 is unfortunate, however, and additional surveys would have enhanced the analytical possibilities. The advantage of a population data set also comes at the price of a reduced set of variables[13], which, however, we enhance with additional statistical information from the FSO, as described further below.

Figure 2.2 presents descriptive information on the training activities of firms over the period 1985 to 2008.

Figure 2.2: Swiss firms' training activities (1985-2008)

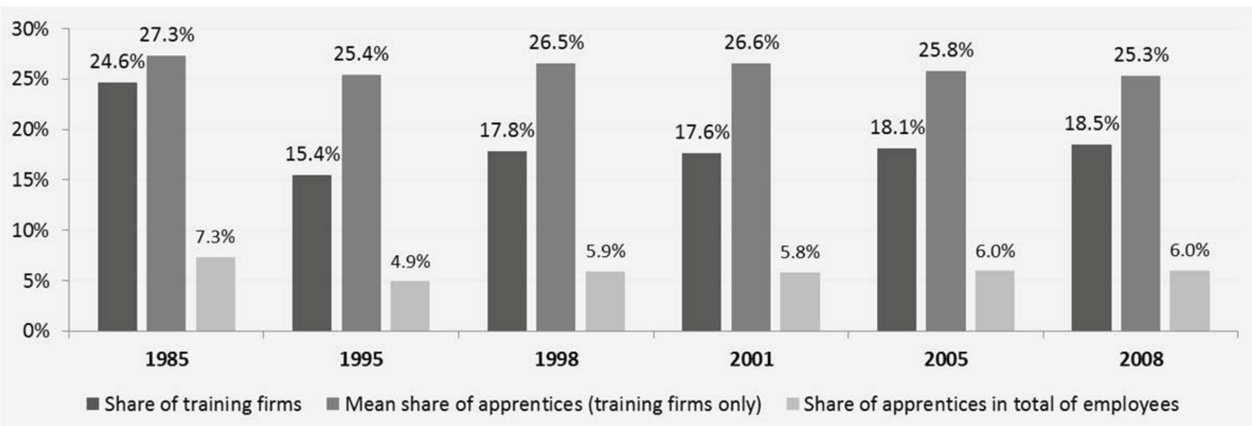


Figure 2.2 shows how the share of training firms in our sample has evolved over time

[12] The firm census for the primary sector differs with regards to survey years and is not used for our analysis.

[13] 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.

(dark bars). There was a sharp decline between 1985 and 1995, following which the share of training firms increased again. However, the indicator still shows a decline of 6.1 percentage points (24.6%-18.5%) between our first wave (1985) and our last wave (2008).

The mean share of apprentices in the training firms[14] (grey bars) amounts to 25% and has hardly changed over time. Therefore, the variation in the share of apprentices among the total number of employees in the economy (bright bars) mainly seems to be caused by the variation in the share of training firms. The indicator on whether a firm trains apprentices (1) or does not train apprentices (0) will therefore be the dependent variable of interest in our estimations.

The following additional information on firms are available from the firm census data and are included as independent variables: survey year, firm size, industry, firm type (independent firm, headquarters or branch), region (i.e., canton or greater region), and area type (rural or urban). Table 2.1 shows the univariate distribution of all variables in our final data.[15]

Three noticeable points with respect to the distribution of firms need to be highlighted. First, the total number of firms has markedly increased by 25% from 1985 to 2008, with a large part of this increase occurring between 1985 and 1995, naturally raising the denominator of the share of training firms as of 1995. In contrast, the number of training firms mirrors the pattern of training activity described in Figure 2.2: it was highest in 1985 and lowest in 1995. Second, the increase in the total number of firms was accompanied by an increasing share of small firms with less than two employees (this share increased from 33% in 1985 to 44% in 2008). Third, we observe a shift in the industry composition away from the industry and traditional service sectors towards modern (skill-intensive) services, the latter expanding from 24% of all firms in 1985 to 37% in 2008.

The literature has derived the following hypotheses of how these firm characteristics affect firms' training activity: larger firms are known to have a higher training propensity (Neubaeumer and Bellmann, 1999; Franz et al., 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

[14] The share of apprentices in a firm is computed as the ratio of the number of apprentices to the total number of employees (including apprentices), where employees are measured by so-called full-time equivalents (FTEs).

[15] We do not present descriptive statistics for all regions (cantons) and detailed industry categories, but we do report summarized descriptive information at higher aggregated levels.

Table 2.1: Descriptive statistics: distribution of variables (1985-2008)

	1985	1995	1998	2001	2005	2008
Number of firms	275873	330620	338971	343577	333753	347060
Number of training firms	67891	50991	60332	60574	60447	64273
	%	%	%	%	%	%
Training firm: no	75.39	84.58	82.20	82.37	81.89	81.48
Training firm: yes	24.61	15.42	17.80	17.63	18.11	18.52
Firm size: <2 FTE	32.76	39.32	42.50	43.34	43.33	44.13
Firm size: 2 FTE	20.23	18.08	17.97	17.06	16.46	15.74
Firm size: 3-4 FTE	18.82	16.58	16.04	15.54	15.54	14.93
Firm size: 5-9 FTE	14.55	13.49	12.14	12.23	12.52	12.60
Firm size: 10-19 FTE	7.03	6.59	5.93	6.02	6.23	6.37
Firm size: 20-49 FTE	4.23	3.88	3.50	3.71	3.84	3.99
Firm size: 50-99 FTE	1.40	1.21	1.11	1.20	1.20	1.26
Firm size: 100-149 FTE	0.43	0.36	0.35	0.39	0.37	0.41
Firm size: 150-249 FTE	0.31	0.27	0.24	0.27	0.28	0.30
Firm size: 250-499 FTE	0.16	0.15	0.14	0.17	0.17	0.18
Firm size: 500-999 FTE	0.06	0.05	0.05	0.05	0.05	0.06
Firm size: >1000 FTE	0.02	0.02	0.02	0.02	0.01	0.02
Firm type: headquarter	4.96	4.84	3.69	3.23	2.96	2.91
Firm type: independent firm	81.20	81.46	84.98	85.87	86.29	86.12
Firm type: branch	13.84	13.69	11.33	10.90	10.75	10.97
Area type: urban	75.07	75.72	75.68	76.40	76.74	77.14
Area type: rural	24.93	24.28	24.32	23.60	23.26	22.86
Greater region: Lake Geneva	19.11	18.82	18.26	18.13	18.22	18.83
Greater region: Espace Mittelland	21.90	21.57	21.25	20.78	20.53	19.95
Greater region: North-western CH	12.21	12.42	13.05	13.01	12.84	12.86
Greater region: Zurich	17.79	18.35	18.05	18.49	18.40	18.37
Greater region: Eastern CH	15.08	14.47	14.86	14.66	14.67	14.28
Greater region: Central CH	8.65	9.22	9.54	9.95	10.24	10.54
Greater region: Ticino	5.25	5.15	5.00	4.97	5.09	5.17
Industry: traditional industry	12.11	10.39	9.78	9.53	8.97	8.63
Industry: modern industry	3.45	3.36	3.02	3.05	3.07	2.93
Industry: construction	10.38	10.65	10.63	10.54	10.76	10.93
Industry: traditional services	49.82	44.72	44.37	41.91	41.75	41.14
Industry: modern services	24.24	30.87	32.20	34.96	35.45	36.38
Demography	2.28	1.65	1.71	1.71	1.73	1.76
Share of high school students	18.42	25.57	25.66	25.37	26.20	26.65
GDP-growth	4.00	−0.02	3.09	0.97	2.50	2.07

Sample: second and third sector, all firms that are private and marked-based

training. According to table 2.A1 in the appendix, there is a strong bivariate relationship between firm size and training activity. In the year 2008, the share of training firms was 5% in the smallest firm size class (less than two full-time equivalents) and 89% in the highest firm size class (over 1000 full-time equivalents). Thus, we include dummies representing twelve different firm size classes in our analysis.

Industries differ by production technology and skill needs. The training strategies of firms will, therefore, differ between industries: in the construction industry, for example, apprentices can typically work productively early on. These firms may find it favorable to employ apprentices (instead of unskilled workers with higher wages) and thus should have a high training propensity. Service industries have a poorer tradition in apprenticeship training than crafts and manufacturing. We further distinguish between traditional and modern sectors for descriptive purposes. Modern services (skill-intensive) and modern manufacturing (highly technology intensive industries) might operate in a faster changing environment than traditional sectors, which might impede the training of apprentices over several years. According to the descriptive statistics in table 2.A1, the modern services sector records the lowest share of training firms. In the multivariate analysis, we include a full set of industry dummies representing 45 different industries via the 2-digit numeric code of the *General Classification of Economic Activities* in order to control for as much firm heterogeneity as possible.

With regard to firm type, single firms and headquarters probably have a wider array of activities, whereas specialized branches of a firm might show a smaller propensity to train. As for firm location, it is well known that the French- and Italian-speaking regions of Switzerland are home to more full-time vocational schools and fewer dual apprenticeship places. The educational system is regulated by cantons and might have an influence on the educational decisions of school leavers as well as on the training decisions of firms. In order to capture regional heterogeneity, we include the most detailed set of regional dummies in our analyses (cantons or greater regions, depending on the specification). We can also differentiate between urban and rural areas: in rural districts, the reputation effects of apprenticeships (with customers) might be more important, and poaching might be less of a problem. One might therefore expect the share of training firms to be higher in rural areas. Table 2.A1 in the appendix shows that the bivariate relationships between training activity and the described firm characteristics are in line with these expectations.

In order to include supply-side information as indicated in the previous sections, we match the following variables to the firm census data: the share of 16-year-olds among the working-age population in the canton of a firm's location and the share of high school students among the 16-year-olds in the canton of a firm's location. These data are derived from other statistics published by the Swiss FSO.[16]

We use cantonal level information since cantons form strong political entities in Switzerland. Cantons vary markedly in the composition of their population. Moreover, the authority to plan and implement educational policies resides with cantons and not with the federal authorities. Therefore, the share of 16-year-olds as well as the share of high school pupils among 16-year-olds is expected to vary markedly between cantons.

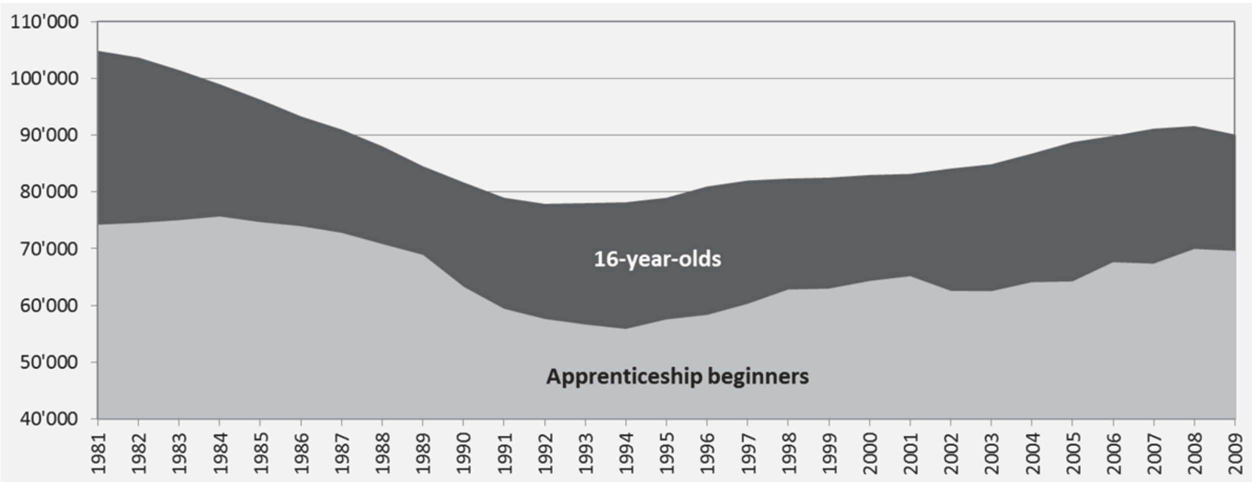
Figure 2.A1 in the appendix illustrates the heterogeneity between cantons in these vari-

[16] That is, data on the population structure (Statistik des jährlichen Bevölkerungsstandes (ESPOP)) and data on high school enrolment (Statistik der Lernenden (Schüler/innen und Studierende)).

ables and shows a clear bivariate cross-sectional relationship between the cantonal share of training firms and 16-year-olds as well as the cantonal share of training firms and high school enrolment.

As seen from figure 2.3, demography is also very likely to influence the variance in the outcomes of the apprenticeship market over time.

Figure 2.3: Number of 16-year-olds and apprenticeship beginners (1981-2009)



The number of 16-year-olds showed a significant downward trend from the early eighties to the mid-nineties and rose again afterwards. The number of people starting apprenticeships[17] seems to mirror this trend. The share of training firms might, therefore, be affected by the number of apprenticeship candidates available in the market. We include demographic information in the regressions in the form of the share of 16-year-olds among the working-age population of a canton. Averages over all firms for this variable over time can be found at the end of table 2.1.

The share of 16-year-olds opting for high school accounts for shifts in the preferences for general versus vocational education. An increase in the share of high school pupils, as observed between 1985 and 1995 (table 2.1), will reduce the number of candidates in the apprenticeship market and is, therefore, expected to lower the overall share of training firms. Not including this variable might bias the time dummies, which we want to reflect possible time trends in firm behavior in our estimations.[18]

[17] Data source: Swiss Federal Statistical Office, Statistik der Lernenden (Schüler/innen und Studierende.)

[18] Including the share of high 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 high school instead. This is unlikely, however: in the short run, high schools have a given infrastructure and stock of teachers. The number of high school places is therefore inelastic in the short run.

In order to capture fluctuations in the business cycle, we match the official GDP growth derived from the Swiss national accounts ("Volkswirtschaftliche Gesamtrechnung") at the most detailed industry level for which data are available.

Figure 2.4: Apprenticeship beginners and GDP growth (1981-2009)

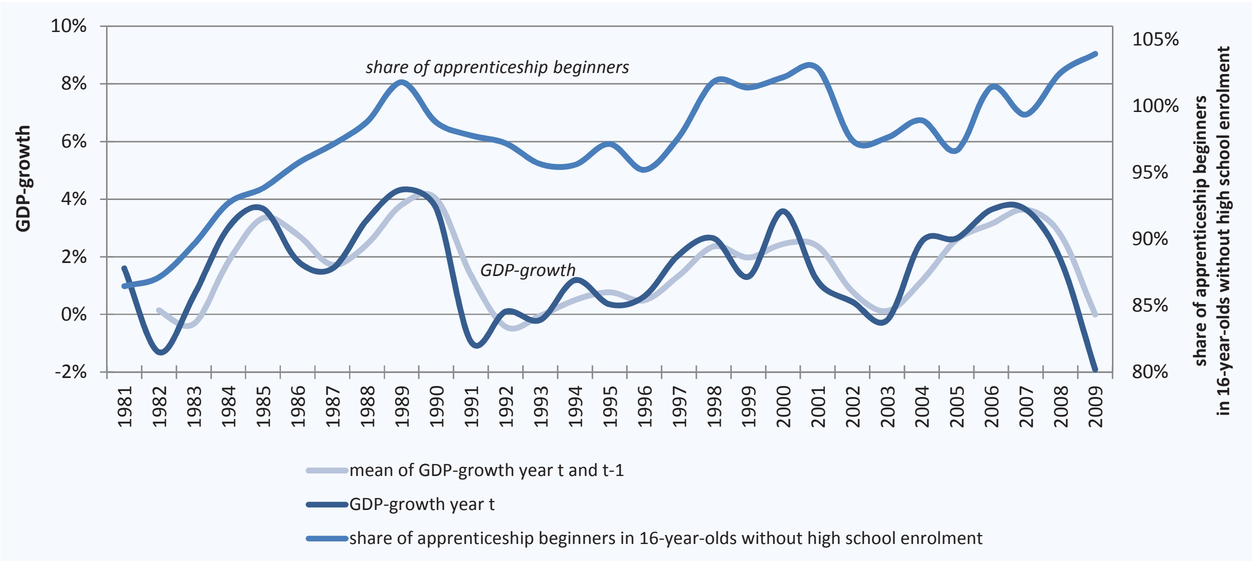


Figure 2.4 illustrates the development of the share of apprenticeship beginners (in the population of 16-year-ols without high school enrolment) and GDP growth between 1981 and 2009. It shows that the share of apprentices in potential apprenticeship applicants broadly follows the movement of GDP growth since the early nineties. At the beginning of the figure and the very end, however, the two curves are rather independent from each other. A tentative interpretation could be that the business cycle might be more important when the numbers of 16-year-olds increase (figure 2.3). However, this hypothesis will not be testable with the limited observation points of the firm census data.

For all the variables matched to the firm census data based on their location (demography and share of high school students) or industry (GDP growth), we match the mean across the year of the firm census and the two preceding years. This is our preferred definition, because the firm census measures the current number of apprentices, that is, apprenticeship beginners over three years, on average.

Empirical evidence supports this: both the number and the share of high school pupils in the relevant age cohort change only gradually and do not show cyclical fluctuations. For example, a study by the Federal Statistical Office which aims to predict the development of pupil numbers confirms that the number of high school pupils did not depend on business cycles in the past (FSO, 2004b). The same result has been found in Muehlemann et al. (2009).

2.6 Results

2.6.1 Changes in firms' training propensity between 1985 and 2008

In this section, we discuss the estimation results with a focus on the time-differences between observation periods.

Table 2.2 presents pooled cross-sectional 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 on firms' propensity to train, they can only partly explain the different training levels across time. Controlling for these factors, the average training propensity of firms was 3.0 percentage points higher in 1985 than in 2008 (compared with a difference of 6.1 percentage points in the descriptive statistics; see also figure 2.2).

Table 2.2: Estimation of firms' training propensity

	(1) **Probit**	(2) **Probit**	(3) **Probit**	(4) **LPM**	(5) **FE-LPM**	(6) **FE-LPM^{sl}**	(7) **FE-logit**
Firm size: 2 FTE	0.146***	0.146***	0.146***	0.093***	0.011***	0.054***	0.206***
Firm size: 3-4 FTE	0.264***	0.265***	0.265***	0.185***	0.023***	0.092***	0.350***
Firm size: 5-9 FTE	0.387***	0.387***	0.387***	0.275***	0.058***	0.183***	0.722***
Firm size: 10-19 FTE	0.501***	0.501***	0.501***	0.352***	0.117***	0.315***	1.291***
Firm size: 20-49 FTE	0.600***	0.600***	0.600***	0.441***	0.204***	0.477***	2.003***
Firm size: 50-99 FTE	0.688***	0.688***	0.689***	0.544***	0.314***	0.653***	2.812***
Firm size: 100-149 FTE	0.743***	0.744***	0.744***	0.626***	0.407***	0.792***	3.481***
Firm size: 150-249 FTE	0.767***	0.767***	0.767***	0.666***	0.464***	0.876***	3.897***
Firm size: 250-499 FTE	0.805***	0.805***	0.805***	0.739***	0.562***	1.036***	4.698***
Firm size: 500-999 FTE	0.822***	0.822***	0.822***	0.783***	0.638***	1.121***	5.220***
Firm size: >1000 FTE	0.844***	0.844***	0.844***	0.843***	0.697***	1.160***	5.548***
Firm type: headquarter	0.048***	0.048***	0.048***	0.062***	0.018***	0.042***	0.193***
Firm type: branch	−0.023***	−0.023***	−0.023***	−0.032***	−0.028***	−0.072***	−0.304***
Industry dummies (45)	Yes	Yes	No	Yes	Yes	Yes	Yes
Rural	0.012***	0.012***	0.012***	0.012***	−0.012***	−0.056***	−0.233***
Regional dummies (26)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year 1985	0.030***	0.005*	0.003	−0.002	−0.002	−0.003	−0.046
Year 1995	−0.040***	−0.038***	−0.035***	−0.035***	−0.046***	−0.187***	−0.775***
Year 1998	−0.008***	−0.007***	−0.007***	−0.008***	−0.016***	−0.062***	−0.256***
Year 2001	−0.008***	−0.007***	−0.005***	−0.005***	−0.014***	−0.060***	−0.241***
Year 2005	−0.004***	−0.004***	−0.004***	−0.004***	−0.011***	−0.049***	−0.190***
Demography		0.033***	0.033***	0.051***	0.046***	0.106***	0.460***
Share of high school students		−0.001***	−0.001***	−0.001***	−0.001***	−0.005***	−0.023***
GDP growth (coeff*10)			0.013***	0.016***	0.005***	0.004	0.014
(Pseudo) R2	0.2070	0.2070	0.2071	0.192	0.120	0.060	0.062
N	1969854	1969854	1969854	1969854	1969854	427878	427878

* p<0.05, ** p<0.01, *** p<0.001, based on (cluster-)robust standard errors.
Reference group: Firm size <=2 FTE, independent firm, urban district, traditional services, Year 2008.
FTE = full-time equivalents. Note: Probit results show marginal effects at the mean of the other explanatory variables, FE-Logit results show coefficients (no marginal effects).
^{sl} sample restricted to the sample of model (7).

The second column shows the estimation 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 16-year-olds in the adult population dropped from 2.27% to 1.65% (for all of Switzerland). Based on model 2, this decrease would, *ceteris paribus*, have caused a drop of 2.05 percentage points in the share of training firms. The share of high school pupils significantly affects the training propensity in the expected direction, too: the higher the (cantonal) share of compulsory school leavers opting to go to high school, the lower the firms' training activity. While most of the coefficients do not react strongly to the inclusion of these supply-side factors, the coefficient of the dummy variable for the year 1985 drops substantially from 3.0 to 0.5 percentage points. The difference in firms' training activity between 1985 and 2008 can, therefore, be largely explained by the independent variables. Although the inclusion of supply-side factors does not increase the goodness-of-fit of the model, it does crucially affect the conclusions that can be drawn concerning the development of firms' training propensity: the small unexplained difference does not suggest a major change in firms' ceteris paribus willingness to train.

Furthermore, another factor might be responsible for the remaining differences, namely the business cycle. We do not have firm level data related to business performance. We, however, include information on economic growth at the sectoral level in model 3 and find a positive effect on training propensity. The inclusion of this variable further mitigates the estimated differences across years. The small estimated difference of 0.3 percentage points between 1985 and 2008 is not statistically significant anymore; the observed drop of 6.1 percentage points can, therefore, be fully explained by including all variables in model 3. This estimation, however, cannot fully explain the substantial drop observed in 1995.[19]

Model 5 in table 2.2 presents the results of the fixed-effects estimation. Model 4 shows the underlying pooled linear probability model for comparison purposes. The coefficients of LPM model 4 are quite similar to the marginal effects of probit model 3.

The major change in fixed-effects model 5 compared to the former LPM and probit models occurs with firm size: notably, the dummies indicating smaller firm sizes show a massive decline in coefficient size. Therefore, the firm size coefficients of cross-sectional estimations seem to have absorbed considerable unobserved cross-sectional heterogeneity related to firm size; they should not be interpreted causally. According to the fixed-effects estimates, the growing or shrinking of firms induces much smaller changes in a firm's training propensity

[19] An 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.

than implied by cross-sectional estimations.

The effect of the demography variable is confirmed in model 5, and the coefficient is even larger than the corresponding value in the pooled probit models (though it is slightly smaller than that in LPM model 4). The change in demography between 1985 and 1995, which would have caused a drop of 2.0 percentage points in the share of training firms according to model 2 (as computed with the underlying data), now suggests a decrease of 2.85 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 explained by a combination of demand-side and supply-side factors; there is, for example, no significant difference left between 1985 and 2008. There is, in addition, no evidence of a downward trend in firms' ceteris paribus willingness to train since the training propensity rather slightly increased after 1995.

The discussed results are confirmed in the last estimation (model 7), where we estimate a fixed-effects conditional logit that takes into account the binary nature of the dependent variable within a panel design. As discussed in section 2.4, this method analyzes the sub-sample of firms that showed a change in the dependent variable across time, and there is no straightforward way to compute marginal effects. Therefore, using the same sub-sample, we additionally show a linear fixed-effects estimation (model 6) for comparison. As expected, the coefficients of most variables in model 6 are higher than those in model 5, since they are obtained by using only the sample of firms that showed a change in training activity at least once. The only qualitative difference between models 6 and 7 compared to the models using the full sample, concerns the insignificant coefficient on economic growth. As the coefficient itself seems to be only a little smaller in model 6 than in model 5, this result might be due to the smaller sample size together with the fact, that the causal effect is not large anyway; 1 percentage point higher GDP growth enhances training propensity only by 0.005 percentage points (model 5).

2.6.2 Decomposition analysis

So far, we have seen that, after controlling for all the information at hand, there is no significant *ceteris paribus* difference in the estimated training propensity of firms between the years 1985 and 2008 as measured by a year dummy. With the decomposition technique discussed in section 2.4, we can try to shed more light on the power of single independent variables in explaining and predicting the changes in training activity over time.

As a starting point, we predict the share of training firms based on the different combinations of data and coefficients for the 6 points in time under consideration. For instance, the probit coefficients resulting from the cross-sectional estimation with data for 1985 are combined with the values of the independent variables of the other years to predict the share of training firms over time, holding training behavior related to each firm attribute at a constant level. This procedure provides an answer to the following question: if firm behavior (as expressed by probit coefficients) were the same as in 1985, but the firm composition of the economy or important surrounding factors changed over time, what share of training firms would we predict to observe in the other years? In other words: given the main assumption of constant behavior for this decomposition technique, we show to what extent the change in the share of training firms over time can be explained by changes in the independent variables.

The decompositions are based on coefficients from the cross-sectional estimations. However, the fixed-effects estimation in table 2.2 showed that firm size coefficients change considerably compared with the cross-sectional estimations and that the latter should not be interpreted causally. Is our decomposition biased since it is based on cross-sectional probit estimates? Recall that we are attempting to make the best prediction, assuming that firm behavior did not change. The best prediction, however, 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.[20]

Table 2.3 displays the actual share of training firms and predictions of this share based on different data combinations. The upper panel of the table presents predictions based on cross-sectional probits without supply-side variables but including firm characteristics (firm size, industry, and firm type) and region dummies as independent variables. The lower panel of the table presents the same predictions but includes supply-side factors (share of 16-year-olds

[20] As regards firm size, we know that from 1985 to 2008, many very small firms newly entered the market. The trend towards a larger share of very small firms is well documented (FSO, 2004a); these firms mainly originate in the expanding service sector. 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 the same as the small firms in 1985 did, the coefficients of 1985 provide a good prediction of the training propensity of these new firms. If, however, new small firms are systematically different from existing small firms (with respect to their training behavior), the prediction based on the 1985 coefficients is not valid. Then, the part explained by firm size might be too high in the decomposition. For this reason and in the absence on reasonable theoretical or practical guidance, we do not imply that coefficients of any particular year should be preferred over the others.

and share of high school pupils) in the underlying probit estimations.

Panels A and B are identically organized. Column (1) presents predictions for all years based on coefficients from a probit estimation using data from 1985. For illustration purposes, the probit coefficients for the year 1985 and the values of the independent variables in 1995 are used to predict the share of training firms in the year 1995; the share for the year 1998 is predicted using the values of the independent variables in 1998, etc. The predictions using the coefficients from 1985 and the data from 2008 (i.e. 0.220 and 0.179, respectively) correspond exactly with the computation of the first term in the first bracket of equation (2.7) in section 2.4. The underlying estimations in column (8) differ from the pooled version in column (7) in two respects: cantons, instead of greater regions, are controlled for,[21] and GDP growth is included in the model (see table 2.A2 in the appendix).

Table 2.3: Predicting the share of training firms using data and cross-sectional probits for all periods with two different sets of independent variables

PANEL A: Only firm characteristics/regions used in estimations

Underlying data	*Actual share (observed)*	*Predicted share (out of sample prediction) using cross-section estimates of single years:*							
		(1) Coeff. 1985	*(2) Coeff. 1995*	*(3) Coeff. 1998*	*(4) Coeff. 2001*	*(5) Coeff. 2005*	*(6) Coeff. 2008*	*(7) Coeff. pooled*	*(8) Coeff. pooled*
Data 1985	0.246	0.246	0.168	0.208	0.207	0.213	0.219	0.209	0.208
Data 1995	0.154	0.230	0.154	0.189	0.189	0.194	0.199	0.191	0.191
Data 1998	0.178	0.220	0.145	0.178	0.178	0.182	0.186	0.180	0.180
Data 2001	0.176	0.219	0.144	0.176	0.176	0.180	0.184	0.178	0.178
Data 2005	0.181	0.220	0.145	0.177	0.177	0.181	0.186	0.179	0.179
Data 2008	0.185	0.220	0.145	0.177	0.177	0.181	0.185	0.179	0.179
	Actual change 1985-2008	*Explained change 1985-2008*							
	-0.061	-0.026	-0.023	-0.030	-0.030	-0.032	-0.033	-0.030	-0.029

PANEL B: All variables used in estimations

Underlying data	*Actual share (observed)*	*Predicted share (out of sample prediction) using cross-section estimates of single years:*							
		(1) Coeff. 1985	*(2) Coeff. 1995*	*(3) Coeff. 1998*	*(4) Coeff. 2001*	*(5) Coeff. 2005*	*(6) Coeff. 2008*	*(7) Coeff. pooled*	*(8) Coeff. pooled*
Data 1985	0.246	0.246	0.231	0.271	0.257	0.265	0.257	0.251	0.249
Data 1995	0.154	0.185	0.154	0.186	0.186	0.191	0.198	0.180	0.175
Data 1998	0.178	0.177	0.148	0.178	0.177	0.181	0.186	0.172	0.177
Data 2001	0.176	0.178	0.148	0.177	0.176	0.180	0.185	0.172	0.170
Data 2005	0.181	0.179	0.149	0.178	0.177	0.181	0.186	0.173	0.176
Data 2008	0.185	0.179	0.150	0.179	0.178	0.182	0.185	0.174	0.176
	Actual change 1985-2008	*Explained change 1985-2008*							
	-0.061	-0.067	-0.081	-0.091	-0.080	-0.083	-0.071	-0.078	-0.072

Note: Underlying probit estimation results of Panel B can be found in the appendix, table 2.A2.

The predictions in Panel A of table 2.3 are not particularly good. However, they demon-

[21] In underlying single cross-section estimations (1) - (6) we cannot control for cantons because supply-side factors are included on a cantonal level, too.

strate that a part of the decline in the share of training firms between 1985 and the subsequent periods may be attributed to a change in the structural composition of the economy. The predicted difference between the share of training firms in the years 1985 and 2008 ranges from 2.3 to 3.3 percentage points, depending on the underlying coefficients, whereas the real difference in the share of training firms amounts to 6.1 percentage points.

In Panel B, the predicted differences range from -6.7 to -9.1 percentage points. These predictions are closer to the true value, thus demonstrating once again that the inclusion of supply-side factors in longitudinal analyses is important. In the absence of year dummies in the underlying estimations, however, we even over-predict the decline in training firms with the variables at hand. Most of the models provide quite good predictions for all years, with the exception of 1995. None of the models predicts the substantial drop in the share of training firms in 1995 or the full extent of the increased training propensity since 1998.

Finally, we calculate the contribution of each independent variable to the total predicted difference between 1985 and 2008 using coefficients derived from estimations based on the data for 1985, the data for 2008, and the pooled data among all years. The pooled versions correspond to the pooled versions of table 2.3 with the exception of the inclusion of year dummies in the underlying regression, because this is methodologically more adequate in the detailed decomposition. Table 2.4 shows the contribution made by the different independent variables to the total predicted difference.

Table 2.4: Detailed decomposition of the difference in the shares of training firms (1985-2008)

	Decomposition based on ... (table/model)							
	coeff. 1985 (2.A2/1)		*coeff. 2008* (2.A2/2)		*coeff. pooled (all)* (2.A2/7)		*coeff. pooled (all)* (2.2/3)	
Actual change 1985-2008	−0.0609		−0.0609		−0.0609		−0.0609	
Explained change 1985-2008	−0.0672		−0.0714		−0.0763		−0.0585	
Contribution of variables:								
Regional dummies	−0.0001	0.1%	−0.0006***	0.9%	−0.0004***	0.6%	−0.0006***	1.1%
Industry dummies	−0.0036***	5.4%	−0.0120***	17.8%	−0.0090***	11.8%	−0.0081***	13.9%
Firm size	−0.0204***	30.4%	−0.0222***	33.0%	−0.0208***	27.3%	−0.0208***	35.6%
Firm type	−0.0004**	0.6%	−0.0012***	1.8%	−0.0009***	1.1%	−0.0008***	1.4%
Rural area	−0.0001**	0.1%	−0.0004***	0.6%	−0.0002***	0.3%	−0.0002***	0.3%
Demography	−0.0209***	31.0%	−0.0154***	22.9%	−0.0279***	36.5%	−0.0184***	31.4%
Share of high school pupils	−0.0217***	32.3%	−0.0196***	29.2%	−0.0171***	22.5%	−0.0065***	11.1%
GDP growth							−0.0031***	5.2%
Explained change 1985-2008	−0.0672	100.0%	−0.0714	100.0%	−0.0763	100.0%	−0.0585	100.0%
N	275873		347060		1969854		1969854	

* p<0.05, ** p<0.01, *** p<0.001
Note: Underlying probit estimation results can be found in table 2.A2 in the appendix and table 2.2.
The contributions of all categories within a group (region, industry, firm size, firm type) have been added up to a group total. Significances refer to joint significances.
The pooled versions are based on estimations with year dummies as additional control variables; this procedure has been suggested by Fairlie (2003), see also Jann (2006). The explained part thus deviates slightly (by 0.0017) from the one reported in table 2.3, Panel B. While the inclusion of year dummies is more accurate in table 2.4, it would result in perfect predictions in table 2.3 and thus not provide any information at all.

Changes in four variables, firm size, industry, demography, and the share of high school students, contribute substantially to the explained changes from 1985 to 2008. 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. These demand-side factors explain around one-third to over one-half of the change, depending on the underlying coefficients: while training behavior has been hardly unaffected with regard to firm size (contributions using the coefficients for 1985 and 2008 = 30.4% and 33.0%, respectively) the same does not hold for industries, as the coefficients for 1985 and 2008 lead to different contributions to the explained change (1985 coefficients: 5.4%; 2008-coefficients: 17.8%). Supply-side factors, namely demography and the share of high school pupils, explain the other part (contributions using the coefficients for 1985 and 2008 = 63.3% and 52.1%, respectively). The decomposition in the last column also includes GDP growth and shows that business cycle effects play only a minor role in explaining the difference between 1985 and 2008. Moreover, it shows a substantially lower contribution of the share of high school pupils. The inclusion of detailed regional control dummies in the underlying pooled regression (see model (3) in table 2.2) mitigates the effect of the cantonal high school share. The effect of this variable seems to be overestimated in the former models, where only greater regions were controlled for; the cantonal high school share seems to capture some additional between-cantonal variation in educational systems, and thus, the effect of this variable is presumably overestimated in the other longitudinal predictions.

Overall, the result continues to support our view that supply-side factors should not be ignored when analyzing trends in the apprenticeship market.

2.7 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 6.1 percentage points from 1985 to 2008, the oldest and (up to now) the most recent years, respectively, 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 was analyzed in this chapter with pooled probit models, fixed-effects models, and a non-linear decomposition technique. The main findings indicate that the unexplained part of the difference between 1985 and 2008 reduces to zero in the different models when all available independent variables are controlled for. There is no clear trend in the unexplained part to indicate a (negative or positive) trend in individual firms' willingness to train apprentices.

Part of the decline between 1985 and 2008 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 16-year-olds in the working-age population as well as the share of high school pupils, allow us to explain most of the variation. According to the results of the decomposition analysis, these supply-side factors provide approximately half the explanation for the observed decline. 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 favor 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 additional new very small-sized 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 that this factor, which is often mentioned in conjunction with global trends such as "the knowledge society", plays a major role. 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.[22]

[22] Actually, the Swiss Federal Office of Professional Education and Technology is currently modernizing the

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, per se, no reason to worry. The increasing share of young people choosing high school instead of vocational training is another reason for the decrease in the share of training firms. Future increase in the share of high school pupils will obviously impact the apprenticeship system. However, a change in firms' demand for human capital, that is, a possible substitution of apprentices with high school pupils, again, provides no reason for state intervention aimed at increasing the share of training firms. If, however, the state expands high schools without a change in firms' human capital needs, this might deprive the apprenticeship market of highly performing youngsters, thus reducing the training propensity of those firms that offer apprenticeship places to 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 cannot secure apprenticeships and does 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 position? 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), they seek positions in other occupations where no vacant places are available, or they did not acquire the necessary qualifications in compulsory school to qualify 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. Ideal employer–employee matched data, together with individual data on all compulsory school leavers and their educational pathways, would enhance analytical possibilities. In principle, the new data structure of the FSO, which refers to matched registered data based on personal identification numbers, is a step in this direction.

We do not claim to have analyzed all the challenges facing the apprenticeship market in this chapter, but we have focused on one indicator, the share of training firms, which has been prominently discussed in the Swiss media and politics. We have shown that the decline in the share of training firms compared to 1985, in itself, is not a sufficient argument for policy interventions. The decline can be largely explained by the emergence of new very

training regulations for all roughly 250 occupations, together with employer associations and cantons.

small firms and by supply-side factors such as demography. The supply-side factors have to be included in a longitudinal analysis to provide a complete and undistorted picture of the developments in the apprenticeship market over time. The aggregate picture shows that as of the mid-nineties, the supply of potential apprentices (without high-school enrolment) has increased again due to demographic reasons and the apprenticeship market has been able to absorb—with some fluctuations—even higher shares of these youngsters in recent times compared to 1995 (both in relative and absolute terms). The most important reasons for this picture might be the rather favorable cost–benefit ratio of training for the training firms, and presumably, also of some non-training firms around the break-even to train.

2.A Appendix

Table 2.A1: Descriptive statistics: share of training firms by firm characteristics (1985-2008)

	1985 %	1995 %	1998 %	2001 %	2005 %	2008 %
Firm size: <2 FTE	9.38	4.35	5.55	5.59	5.11	5.10
Firm size: 2 FTE	18.94	12.48	15.20	14.57	15.27	15.00
Firm size: 3-4 FTE	28.49	19.66	23.51	23.23	23.79	24.34
Firm size: 5-9 FTE	36.76	26.45	31.67	31.32	32.18	32.75
Firm size: 10-19 FTE	43.74	30.64	38.16	37.56	40.21	41.60
Firm size: 20-49 FTE	53.21	38.35	46.40	46.65	48.04	50.76
Firm size: 50-99 FTE	63.73	47.36	58.19	56.32	58.76	61.89
Firm size: 100-149 FTE	74.68	56.95	64.53	64.33	65.71	67.52
Firm size: 150-249 FTE	78.33	64.06	69.98	67.24	67.43	70.48
Firm size: 250-499 FTE	86.36	74.95	75.00	74.48	75.31	73.35
Firm size: 500-999 FTE	91.12	82.39	84.07	75.53	74.16	82.13
Firm size: >1000 FTE	93.94	88.68	73.08	82.69	82.61	89.29
Firm type: headquarter	44.61	30.92	38.48	40.11	41.36	41.84
Firm type: independent firm	23.29	14.40	16.38	16.19	16.56	16.67
Firm type: branch	25.18	16.02	21.73	22.31	24.17	26.82
Area type: urban	24.22	14.81	17.04	16.81	17.08	17.50
Area type: rural	25.78	17.35	20.15	20.28	21.51	21.95
Greater region: Lake Geneva	20.25	13.00	15.22	15.04	15.45	15.95
Greater region: Espace Mittelland	27.59	17.61	20.24	20.29	20.61	21.66
Greater region: North-western CH	26.09	15.17	16.91	17.51	18.09	18.25
Greater region: Zurich	20.95	12.42	14.93	14.65	15.17	15.72
Greater region Eastern CH	26.59	18.31	21.20	21.12	21.89	22.41
Greater region: Central CH	29.76	18.34	20.12	19.18	19.88	19.21
Greater region: Ticino	22.75	13.12	14.95	13.99	13.81	14.19
Industry: traditional industry	29.15	19.22	23.22	22.99	24.26	24.56
Industry: modern industry	30.83	17.98	22.99	23.05	24.01	24.31
Industry: construction	37.07	23.28	27.43	26.53	28.53	28.15
Industry: traditional services	21.47	14.36	17.16	17.72	18.08	18.81
Industry: modern services	22.58	12.69	13.36	12.90	12.92	13.40

Figure 2.A1: Share of training firms, demography and high school enrolment by canton

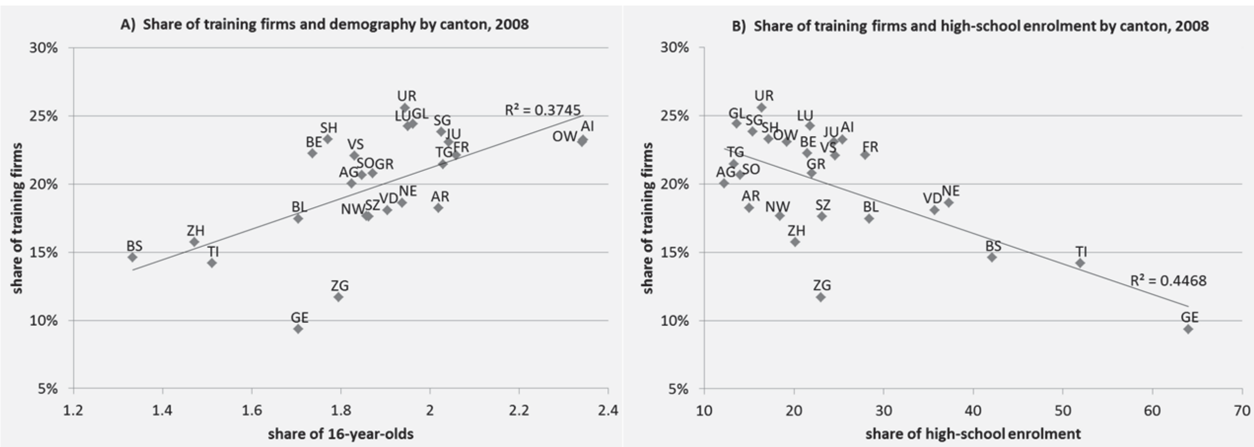


Table 2.A2: Estimation of firms' training propensity; cross-sectional estimations for single years and pooled

	(1) **1985**	(2) **1995**	(3) **1998**	(4) **2001**	(5) **2005**	(6) **2008**	(7) **pooled**	(8) **pooled**
2 FTE	0.145***	0.131***	0.146***	0.135***	0.157***	0.154***	0.145***	0.145***
3-4 FTE	0.274***	0.234***	0.263***	0.254***	0.275***	0.281***	0.264***	0.264***
5-9 FTE	0.399***	0.346***	0.392***	0.378***	0.397***	0.400***	0.386***	0.386***
10-19 FTE	0.506***	0.442***	0.504***	0.491***	0.523***	0.534***	0.50***	0.501***
20-49 FTE	0.599***	0.545***	0.600***	0.595***	0.612***	0.633***	0.599***	0.599***
50-99 FTE	0.668***	0.638***	0.701***	0.682***	0.702***	0.722***	0.687***	0.688***
100-149 FTE	0.727***	0.713***	0.746***	0.742***	0.751***	0.763***	0.743***	0.744***
150-249 FTE	0.740***	0.763***	0.778***	0.759***	0.766***	0.782***	0.766***	0.767***
250-499 FTE	0.773***	0.824***	0.804***	0.803***	0.809***	0.804***	0.804***	0.805***
500-999 FTE	0.785***	0.859***	0.840***	0.805***	0.797***	0.833***	0.822***	0.822***
>1000 FTE	0.792***	0.879***	0.829***	0.840***	0.845***	0.86***	0.844***	0.844***
Firm type: headquarter	0.052***	0.027***	0.051***	0.060***	0.061***	0.051***	0.046***	0.046***
Firm type: branch	−0.045***	−0.034***	−0.024***	−0.024***	−0.009***	−0.003	−0.023***	−0.023***
Industry dummies (45)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional dummies (7)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Regional dummies (26)								Yes
Demography	0.048***	0.059***	0.069***	0.048***	0.050***	0.027***	0.053***	0.055***
Share high school students	−0.004***	−0.002***	−0.002***	−0.002***	−0.002***	−0.002***	−0.002***	−0.000***
GDP-growth (coeff*10)								0.033***
(Pseudo) R2	0.1792	0.1862	0.2041	0.2056	0.2237	0.2333	0.2047	0.2061
N	275873	330620	338971	343577	333753	347060	1969854	1969854

* p<0.05, ** p<0.01, *** p<0.001, based on robust standard errors.
Reference group: Firm size <=2 FTE, independent firm, urban district. FTE = full-time equivalents.
Note: Probit results show marginal effects at the mean of the other explanatory variables.

Chapter 3

The role of hard-to-obtain information on ability for the transition into VET

3.1 Introduction

In this chapter,[1] we attempt to investigate the effect of easy-to-observe characteristics and hard-to-observe ability on the success of applicants in the transition from compulsory schooling to apprenticeship training.

Difficulties in identifying the true productive capacity of heterogeneous workers have played an important role in labor economics for a long time. The literature predicts that, in the absence of accurate information on ability, firms will base hiring or wage-setting decisions on easy-to-observe signals and thus screen and statistically discriminate on education level, gender, ethnicity or other readily available factors that are assumed to be correlated with the missing information (See Phelps, 1972; Spence, 1973; Arrow, 1973; Aigner and Cain, 1977). An indirect test of statistical discrimination is provided by the employer learning literature (Farber and Gibbons, 1996; Altonji and Pierret, 2001), wherein information on cognitive ability that is only observable to the researcher, e.g., Armed Forces Qualification Test scores (AFQT), is found to have increasing influence on wages as workers gain experience, indicating that workers' true productivity is gradually revealed over time to the labor market.

The transition into apprenticeship training can be seen as a sorting process of young people without labor market experience into firms and into a set of rather standardized jobs,

---

[1] This chapter is based on the idea and co-autorship of my supervisor Stefan C. Wolter and is a slightly adapted version of the published paper Mueller and Wolter (2014), *The role of hard-to-obtain information on ability for the school-to-work transition.* Empirical Economics 46(4), 1447-1471.

where bad matches are comparatively costly to adjust. Using unique data comprising PISA literacy test scores along with a set of information on individuals that is observable to the hiring firms, we analyze whether and to what extent hard-to-get ability information affects the hiring process and whether there is further revelation of hard-to-observe ability during the subsequent training period.

As an apprentice needs to be hired by a firm for the entire training period, apprenticeship training entails early integration into the labor market (at approximately age 16). Unsuccessful applicants mostly pursue non-certifying intermediate school programs that are designed to increase the prospects on the apprenticeship market the year after.

Employers' screening devices thus play an important role in the sorting process of young adults into vocational education. In the public discussion, stereotyping is claimed to play a (too) dominant role in firms hiring decisions, such that applicants with unfavorable attributes, as low parental socioeconomic status, migration background, low-level compulsory school track attendance and bad school marks, are at a disadvantage to get (good) apprenticeship places, presumably irrespective of their true ability. However, the latter is difficult to assess for firms because, in the absence of uniform school standards and external exams in Switzerland, compulsory schooling outcomes are not well comparable across schools and classes, which in turn reinforces the incentive for firms to take into account additional ability proxies—this may include, for example, family background characteristics—in order to enhance the accuracy of their ability believes.

Although employment decisions solely based on easy-to-observe factors as described are cheap to make, they may also be costly, especially in the setting of apprenticeship training. Economic rationale suggests a potentially high interest among firms in seeking hard-to-get information about ability before selecting apprenticeship applicants: in contrast with ordinary work contracts, apprenticeship contracts cannot be terminated easily[2], and wages for each apprenticeship year as well as training standards are fixed beforehand over the defined training period (3 or 4 years, depending on the training profession). Furthermore, the successful completion of the apprenticeship largely depends on academic performance at the vocational school: A severe mismatch between apprentice and the intellectual aspiration level of the training profession—especially the negative case where apprentices are overwhelmed—

[2] An ordinary working contract can be terminated by the employer without giving reasons, while an apprenticeship contract can be terminated only under certain specified conditions (see http://www.lehrvertrag.ch/).

potentially results in a drop out of training and, as a consequence, in sunk costs for the training firm.

To test whether a student's hard-to-observe ability is revealed and accounted for within the transition from schooling to market-based upper-secondary education or whether allocation into vocational tracks is solely based on easy-to-observe factors, we make use of the longitudinal data set TREE, that comprises PISA 2000 test scores of pupils at age 15 along with individual background characteristics and detailed information on their further educational and working pathways. The PISA reading literacy competence test provides us with an ability proxy that is only observable by the researcher, not by recruiters of training firms. We do not claim that this ability measure encompasses all ability dimensions that might be important for firms; however, reading literacy is expected to be an indispensable part of relevant competencies to successfully pass apprenticeship training (especially but not exclusively the school based part of it) and to successfully manage working life afterwards. Following the procedure in Farber and Gibbons (1996), we use the test score information in its orthogonalized form, thus already cleaned from the part that is explainable by observables, leaving the ability component that is hard-to-observe for outsiders. We then go one step further and analyze an—as far as we know—unaddressed topic in the existing literature by explicitly differentiating between so-called *overachievers* and *underachievers*. This enables us to test whether hard-to-observe ability is revealed and accounted for (if at all) symmetrically. Thereby, we can test separately whether hard-to-observe ability information is gathered and used in favor of those applicants who are—based on the PISA test score—better then they appear to be *(overachievers)* and whether ability is revealed to the disfavor of those students who create an overall outward impression that is better than their (PISA) performance *(underachievers)*.

The remainder of this chapter is organized as follows: the next section provides an overview of the related literature. In section 3.3, we present the empirical strategy. Section 3.4 presents the data and in section 3.5 we show and discuss the empirical results. Finally, section 3.6 concludes with a summary and discussion of our findings.

3.2 The selection procedure of firms and the role of information on hard-to-observe ability: related literature and hypotheses

Firm-based apprenticeship training requires hiring by an employer willing to train the applicant. The employer and the parents of the apprentice sign a work and training contract that is binding for both parties till the end of the apprenticeship training but does not guarantee further employment once the apprenticeship training is over. The search and selection process for apprenticeship positions is comparable to an ordinary job search procedure: firms announce apprenticeship openings, potential apprentices apply for these openings, applicants undergo a selection process with interviews or even tests and, finally, receive the apprenticeship post or must continue searching. The latter may be accompanied by a process of adapting expectations about the apprenticeship track aspiration level one might be suited for: after unavailingly applying for high-prestige apprenticeships, applicants might adjust their aspirations downward and begin to apply for less demanding vocational tracks (occupations).

It has been of public concern for several years that finding an (good) apprenticeship post is—irrespective of their true potential—becoming more difficult for compulsory school leavers with unfavorable (easy-to-observe) characteristics.[3] Recent empirical evidence implies, too, that firms' selection of apprentices is strongly linked to individual background characteristics and former schooling outcomes, whereas the latter are regarded as to provide only limited ability information, too (Haeberlin et al., 2004; Hupka et al., 2006; Imdorf, 2006; Neuenschwander and Malti, 2009). It is therefore not clear whether the selection practices of firms brings about an allocation that is superior to the one that would be accessible by pure stereotyping on easy-to-observe applicants' characteristics that are correlated with ability.

If the costs to overcome imperfect ability information are sufficiently high, the concept of statistical discrimination predicts that firms base their hiring decisions on all easy-to-observe ability indicators that are assumed to be correlated with the missing information (Phelps, 1972; Arrow, 1973; Aigner and Cain, 1977). Firms are thus assumed to build expectations

[3] As a consequence thereof, worker's organizations (*Travailsuisse* and the *Association of Commercial Employees*) have published guidelines to sensitize firms for fair selection practices, namely not to place weight on applicant's family background and not to overweight former school types (http://www.zukunftstattherkunft.ch). There has also been launched an online platform that allows preselection of applicants based on anonymized application dossiers that only include information of objective relevance (http://www.weareready.ch).

about an applicants' ability by considering all observable characteristics that provide additional ability information due to different group specific (expected) mean abilities. Whereas this form of stereotyping ("first moment"' statistical discrimination) is discriminating on the individual level (there is unequal treatment of actually equal able individuals), it is not discriminating on average and not discriminating against certain groups: there is equal treatment of individuals with equal *expected* ability.[4]

The role of easy-to-observe and initially hard-to-observe ability has been investigated by the employer learning literature by exploiting the US NSLY79 data which contain an ability measure (the Armed Forced Qualification test-scores) that is only observable to researchers but hard-to-observe for employers. Farber and Gibbons (1996) find that the part of AFQ test score[5] information that was not predictable by observables at market entry (the residualized test score) becomes increasingly correlated with wages as market experience increases: employers learn. Altonji and Pierret (1997, 2001) simultaneously investigate employer learning and statistical discrimination and find that wages not only become more dependent on a worker's ability (AFQ test-scores) but at the same time also become *less* dependent on easily observable characteristics. These results suggest that firms initially form beliefs about the productivity of a worker using statistical discrimination, i.e., based on educational credentials and ethnicity; as true productivity is revealed over time, firms revise their beliefs accordingly.[6] Using the same data Lange (2007) finds that employers learn quickly; initial expectation errors decline on average by one-half within the first three years and thus restrict the importance for job market signaling of schooling decisions. The tradeoff between screening upon hiring and employer learning has further been found to be different for different educational levels (Arcidiacono et al., 2010): whereas ability is revealed to the labor market only gradually for high school graduates it is observed nearly perfectly from the very beginning of the careers of college graduates. Information that are typically included in resumes of college graduates and thus are easily observable to recruiters of firms (such as for example college major, grades or standardized test scores) are, however, not included in the analysis; they are shown to be strongly correlated with AFQ test-scores and thus expected to be a likely explanation for immediate ability revelation in the college market.

4 Unless the case where firms are risk averse and there is unequal variance in productivity across groups.

5 Farber and Gibbons (1996) only use those parts of the AFQ test that resemble very much the PISA test scores used in our data.

6 Evidence for employer learning has also been found in Great Britain (Galindo-Rueda, 2003) and, in the case of blue-collar workers, in Germany (Bauer and Haisken-DeNew, 2001).

In contrast to the described studies our analysis puts its focus more heavily on the time of the hiring process and therefore on the question whether new applicants on the labor market are only evaluated based on their easy-to-observe characteristics (including school marks and level of school track) or whether hard-to-observe ability is detected and used by recruiting firms as well. Due to special institutional circumstances described below we then extend our analysis to test whether hard-to-obtain information on ability affects the labor market entry in symmetric or asymmetric way. The reason for doing so is the hypothesis that firms' expectation error might be subject to an asymmetric risk, too: hiring someone whose ability level considerably lies below the expected level can lead to severe costs as described below, while it is not apparent that firms would profit much in the reverse case.

There are at least three arguments that explain why one expects Swiss training firms to invest considerably in learning about an applicant's ability prior to hiring him or her. Whereas the first argument is a general one, the second and third arguments demonstrate also the potential rationale for the asymmetry hypothesis.

First, there is a widespread disbelief among firms in the credibility of observable schooling outcomes at compulsory school level (Moser, 2004; Imdorf, 2009). It arises from the lack of uniform school standards, the lack of standardized tests in compulsory schools (grades are only comparable within classes or teachers), the lack of homogeneous curricula and the opaque tracking mechanism into different levels at around age 12.

Second, the sorting process of young school leavers into firms and occupations is not based on trial and error (job shopping), as, for example, in the US (see Topel and Ward, 1992). Training contracts have a fixed duration of 3 to 4 years—depending on the training occupation—and cannot be terminated as easily as ordinary working contracts. There is no scope for adjusting training content below a defined minimum level (both in terms of topics that must be covered and in terms of complexity that must be taught), nor is it possible to downwardly adjust training wages. Wages are fixed in advance for the entire period of the training contract.

Third, the earlier phase of apprenticeship training is typically related to net costs for firms; some apprenticeship tracks (mostly technical apprenticeships) are associated with considerable firm net investments even until the end of training (Muehlemann et al., 2007; Wolter and Ryan, 2011). Premature terminations of apprenticeship contracts are therefore costly, so are efforts and investments of firms to prevent drop outs e.g. by providing additional coaching in case of arising scholastic problems. Academic difficulties at the vocational school due to a bad match between trainee and the intellectual aspiration level of the vocational track are the primary reason for premature terminations of apprenticeships (Stalder and Schmied, 2006) and therefore cognitive skills as measured in PISA should be an important ability information for the training firms.

In line with these considerations it is observed that firms try to diminish the risk of running into problems by investing in learning about a trainee's ability before hiring. They perform screening on the basis of several instruments (see Muehlemann et al., 2007) such as application letters, school-reports, interviews and so-called trial days. Due to the mentioned distrust of firms towards the informational value of school grades and the (intransparent) sorting into school types at compulsory level, some firms are increasingly requiring test results of external screening tests as part of the application dossier. These standardized aptitude tests, sold and administered by specialized private firms and financed by the applicants and their parents, promise to assess the knowledge in school subjects and general cognitive skills in order to provide information on an applicant's ability to potentially pass the desired apprenticeship training (see Siegenthaler, 2011).

However, the observation that a part of the firms try to get superior ability information prior to hiring does not automatically imply, first, that they actually make use of this information in the hiring process, second, that they actually gain ability information that is relevant to predict the desired outcome such as apprenticeship success and third, that they actually gain ability information that is of additional value compared to easy-to-get information.[7]

[7] Siegenthaler (2011) for example analyzed the informational value for the case of the privately sold aptitude test "multicheck retail sale" and found that the test results do not improve firms' ability to predict apprenticeship success once easy-to-obtain information provided in application dossiers (former school grades and the level of compulsory schooling) is taken into account.

3.3 Empirical Strategy

Successful transition and apprenticeship as explained above are represented by three different dependent variables. A dummy variable indicates whether an applicant succeeds in seamlessly entering certifying firm-based apprenticeship training. A second variable indicates the intellectual standard of the vocational program the successful applicants follow.[8] The third variable reflects the occurrence of problems during training. The exemplary econometric model is thus:

$$Successful\ Transition_i^* = y_i^* = \alpha_i + X_i\beta + B_i^*\pi + \epsilon_i \tag{3.1}$$

$$Successful\ Transition_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

where the vector X stands for those student characteristics that are easily observable by the market and that might be correlated with ability and B_i^* represents the part of students' ability that is unobservable at first glance.

To test whether hard-to-observe ability influences the success[9] of transition and training, we need an ability measure, B_i^*, that is known to the econometrician and relevant for the transition but, at the same time, cannot be easily observed by recruiters of training firms. If B_i^* is taken into account by employers although hard-to-observe, then its estimated effect $\hat{\pi}$ should significantly differ from zero. If, however, decisions are only made based on easy-to-observe characteristics, the coefficient $\hat{\pi}$ should not be significant.

The variable that allows us to differentiate between observable and unobservable ability is given in our data by the PISA literacy test score. PISA test scores are observable to researchers but unknown to teachers, parents, employers, pupils, or any other person or institution. This enables us to create a variable that represents the unobservable part of

[8] The intellectual aspiration level of apprenticeship training is also relevant for the second transition, the one from the apprenticeship training into the labor market. Bertschy et al. (2009) have shown that this level affects the chances of seamless transition in a significant and causal way.

[9] Although we define an immediate transition from school to work as a successful transition, this does not mean that all of the unsuccessful applicants would have been better off if they had succeeded immediately in finding an apprenticeship. For some of the unsuccessful candidates the delay of transition might even improve the match between their ability, expectations and the demands of their future employers.

a student's ability, namely, the part that is orthogonal to employers' readily available information. Following the procedure of Farber and Gibbons (1996), we define B_i^* to be the residual from a regression of B_i on all observable students' characteristics, X, that might act as ability-predicting factors. We first regress test scores on the type of school track, school grades and individual background variables such as immigration background, gender, parental education, and other information that is known or assumed to be correlated with school performance and observable to outsiders because it is either included in school reports or is common information in letters of application. As employers might also look out for non-cognitive traits such as motivation, dependability or social behavior (that reflect other dimensions of ability but are potentially correlated with cognitive ability, too), we include variables on non-cognitive skills and behavioral information that should be strongly related to information that most employers will observe easily in the course of the hiring process.

The unobservable part of ability, B_i^*, can then be obtained by subtracting expected ability (predicted test scores) from observed test scores.

$$B_i^* = B_i - E^*(B_i|X_i) = B_i - X_i\hat{\gamma} \tag{3.3}$$

The corresponding OLS regression accounts for almost 40% of the variance in PISA test scores (see table 3.2 in section 3.4), showing that a substantial part of the Pisa scores can be determined by characteristics that are easily observable by employers.[10]

We then go one step further and split the residuals, B_i^*, along their sign into two parts: the positive and negative residual PISA score. We also create two dummy variables: The first dummy variable indicates whether a student belongs to the group of *underachievers*, that is, if actual ability, B_i, lies a considerable amount, t, below the ability level, $E^*(B_i|X_i)$, predicted by observable characteristics (B_i^* is negative and exceeds a certain threshold). The

[10] We have to concede the fact that employers have easy-to-obtain information on the candidates like e.g. health problems that might be correlated with the PISA results and that are not observable by the researchers. In this case, the additional information available to employers helps them to estimate the true ability more accurately then our own regressions would suggest. This carries the potential risk that the researcher would wrongfully qualify a candidate as an over- or underachiever, whereas the employer had accurately assessed the true ability based on his/her easy-to-obtain information. In order to minimize the risk of an unjustified classification of the individual, we use a large threshold for the creation of the dummy variables of 73 PISA points (= one proficiency level) around the predicted score. As it is unlikely that an easy-to-obtain information not available to the researchers would influence the predication to such an extend (keeping in mind that we already control for observables that are likely to be highly correlated with information that we might miss), we assume that this potential bias is negligible. As far as this assumption is testable, we do not find evidence for violations (p=0.784 in a RESET test).

second dummy variable indicates whether someone belongs to the group of overachievers, that is, if actual ability lies to a considerable amount, t, above the level predicted by observable characteristics (B_i^* is positive and exceeds a certain threshold). These dummy variables allow us to test whether hard-to-observe ability is revealed symmetrically, if at all, at both ends of the residual distribution.

$$Underachiever_i = \begin{cases} 1 & \text{if } B_i^* < 0 - t \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

$$Overachiever_i = \begin{cases} 1 & \text{if } B_i^* > 0 + t \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

We choose the threshold, t, to be a value that has some established implication: t equals the number of score points that lie within the same PISA proficiency level according to OECD (2001) and thus comprises a span of 73 test score points. Therefore, *under*- and *overachievers* are defined as those students whose realized PISA score differs by more than one proficiency level from what would have been expected according to their observable characteristics. Note that the term *"-achiever"* always refers to the PISA achievement relative to the expectations throughout this chapter.

In order to take into account the generated regressor character of predicted PISA test-scores and thus of the dummies for under- and overachievers, we calculated bootstrapped standard errors (1000 replications).

3.4 Data and Descriptive Statistics

The data set we use is the Swiss longitudinal data set *Transition from Education into Employment (TREE)* that followed the pupils that had been tested in the *Programme for International Student Assessment 2000 (PISA 2000)*.

For the scope of our analysis, we restrict the TREE sample to those compulsory school graduates who were in the apprenticeship market the year after the PISA test, distinguishing between the successful ones, namely, those who take up firm-based apprenticeship training after compulsory school, and the unsuccessful ones who find themselves in some kind of non-certifying interim solution or gap year despite having had applied at least once for an apprenticeship in a firm. This leaves us with 2097 pupils that are observed in *PISA 2000* and in the follow-up survey.

One core information for our analysis is derived from PISA reading literacy test results. The focus of *PISA 2000* was on testing reading literacy of 15-year-olds in 43 participating countries (OECD, 2001); mathematical and scientific literacy were investigated only for half of the students. PISA measures competencies in points with a mean of 500 points for all participating countries and a standard deviation of 100 points. PISA reading literacy is measured by a composite test score that summarizes the results from three reading literacy scales. The "retrieving information" scale reports on students' ability to locate information in a text. The "interpreting texts" scale report on the ability to construct meaning and draw inferences from written information. A "reflection and evaluation" scale reports on students' ability to relate text to their knowledge, ideas and experiences. The average reading literacy test score of Swiss pupils turned out to not significantly deviate from the OECD mean (OECD, 2001); however, there was a comparatively large overall variation in student performance with rather strong social selectivity found for Switzerland.

Additionally, experts have divided the scale into six different proficiency levels (very low, low, medium low, medium high, high, very high). For our analysis, we define *under*- and *overachievers* to deviate by more than one proficiency level (73 score points) from what one would predict (see section 3.3 and below for the prediction model). To give an impression on the difference between two adjacent proficiency levels: students proficient at level 3 (medium low) are capable of reading tasks of moderate complexity, such as locating multiple pieces of information, making links between different parts of a text, and relating it to familiar everyday knowledge. Students proficient at level 4 (medium high) are capable of difficult reading tasks, such as locating embedded information, construing meaning from nuances of language and critically evaluating a text (OECD, 2001).

As independent variables other than PISA reading literacy test score results we use information on those characteristics that are easily observable by employers. A description of these variables, along with PISA test scores, is provided in table 3.1 and already shows strong bivariate relationships between test-scores and easy-to-observe individual characteristics. For detailed variable definitions see table 3.A1 in the appendix 3.A.

Table 3.1: Descriptives—univariate and bivariate (with PISA Test Scores)

		*Distribution of PISA test scores**			
Independent variables	Share (%)	Mean*	Std. Dev*	Min*	Max*
PISA literacy test score		496.6	78.4	198.0	812.9
Achieves-as-expected	78.7	497.4	62.5	286.5	643.9
Underachiever	10.3	384.5	57.6	198.0	545.5
Overachiever	11.0	596.0	57.1	446.1	812.9
Male	52.0	492.8	78.4	198.0	737.5
Female	48.0	500.7	78.2	250.6	812.9
Swiss	81.1	508.5	73.6	198.0	812.9
Immigrant: second-generation	8.6	457.5	76.6	283.6	622.5
Immigrant: first-generation	10.3	435.8	77.9	250.6	634.6
¡ Age 16	71.4	504.3	77.0	198.0	812.9
¿= Age 16	28.6	477.4	78.7	255.0	704.3
Parental education: comp. school	32.0	474.7	82.2	198.0	738.7
Parental education: upper sec. II	38.3	510.3	73.2	257.7	812.9
Parental education: tertiary	29.7	502.6	75.7	268.1	737.5
Family structure: nuclear	78.2	499.5	78.1	198.0	738.7
Family structure: single	11.7	490.0	75.8	267.1	697.3
Family structure: mixed	6.5	493.9	73.9	257.7	812.9
Family structure: other	3.5	457.9	91.3	255.0	670.5
Track lower sec II: no selection	5.7	491.8	71.8	272.0	676.1
Track lower sec II: low	33.1	444.4	73.7	198.0	653.8
Track lower sec II: medium	44.7	521.7	67.0	294.9	812.9
Track lower sec II: high	16.5	535.0	61.4	357.3	738.7
	Mean	*Std. Dev.*	*Min*	*Max*	*corr PISA*
Mark in test language	4.74	0.61	1.00	6.00	0.22
Mark in mathematics	4.71	0.77	1.00	6.00	0.10
Mark in science	4.86	0.66	2.00	6.00	0.20
Instrumental motivation	2.87	0.73	1.00	4.00	0.02
Absenteeism	1.30	0.45	1.00	4.00	−0.16
Sense of belonging	3.31	0.52	1.00	4.00	0.08

The school track at the lower secondary level is represented by a variable with four categories: *high-level school track* that prepares students for high school, *intermediate-level school track*, *basic-level school track* (which has no requirements) and *school track with no selection*. The share of pupils in different tracks varies between cantons; we thus account for regional variations in educational systems by controlling for *cantons* in all estimations. Further information on the academic performance of a student is given by school marks in annual school reports. These reports are important components of applications for apprenticeships.

The PISA data provide us with information on *school marks in the regional (test) language*, *mathematics* and *sciences*.

The next set of variables we use reflects easy-to-observe information on pupils' individual and parental background as typically specified in the CV of apprenticeship applicants. There is information on *gender*, student's *age* (as some of the ninth graders are one year older due to repetitions of school years) and *migration status*. The latter is represented by two dummy variables, one for *second-generation immigrants* (born in Switzerland, but with both parents born outside Switzerland) and one for *first-generation immigrants* (born outside Switzerland). Further, there is information on *highest-achieved parental education* (no post-compulsory education, upper secondary level, tertiary level) and *family structure* (nuclear, single, mixed, other).

Finally, we also use a set of PISA 2000 variables (OECD, 2002) that should be good proxies for information on non-cognitive ability that most employers might observe easily in the hiring process. The index of *instrumental motivation* was derived from students' reports on how often they study to increase their job opportunities, ensure that their future will be financially secure, and enable them to get a good job. A four-point scale was used with response categories almost never, sometimes, often and almost always. The PISA index of *sense of belonging* was derived from students' reports on whether their school is a place where they feel like an outsider, make friends easily, feel like they belong, feel awkward and out of place, other students seem to like them, or feel lonely. The PISA index of *absenteeism* was derived from students' reports on how often they missed school, skipped classes and were late for school in the two last weeks.[11] We include these measures on non-cognitive ability and social behavior as z-standardized indices into our regressions (mean of zero and standard deviation of 1).

As discussed in section 3.3, we use PISA test scores to create a variable that represents the unobservable part of student's ability; that is, we want to filter out ability information that is not predictable by observable individual or group characteristics. The results of the corresponding OLS regression are presented in table 3.2 (model 4). A substantial part of the variation in PISA test scores can be determined by easy-to-observe characteristics

[11] Although we have a very rich set of background variables on the students, it is still possible that the employers can collect additional information that is not observable by the researchers. If this information would be correlated with the deviation from the predicted PISA scores, an omitted variable bias would occur. However, to the extent that these unobservables are correlated with the observable characteristics the bias is minimized (see also footnote 10).

(approximately 40%). The coefficients show significantly lower test scores for males, for older pupils (repetitions), for immigrants, for pupils with less educated parents, for pupils living in a family structure which is 'difficult to describe' (other), for pupils in lower levels of compulsory school tracks, for pupils with less favourable school marks in the test (regional) language, in mathematics and in sciences, and for pupils with higher absenteeism and lower sense of belonging.

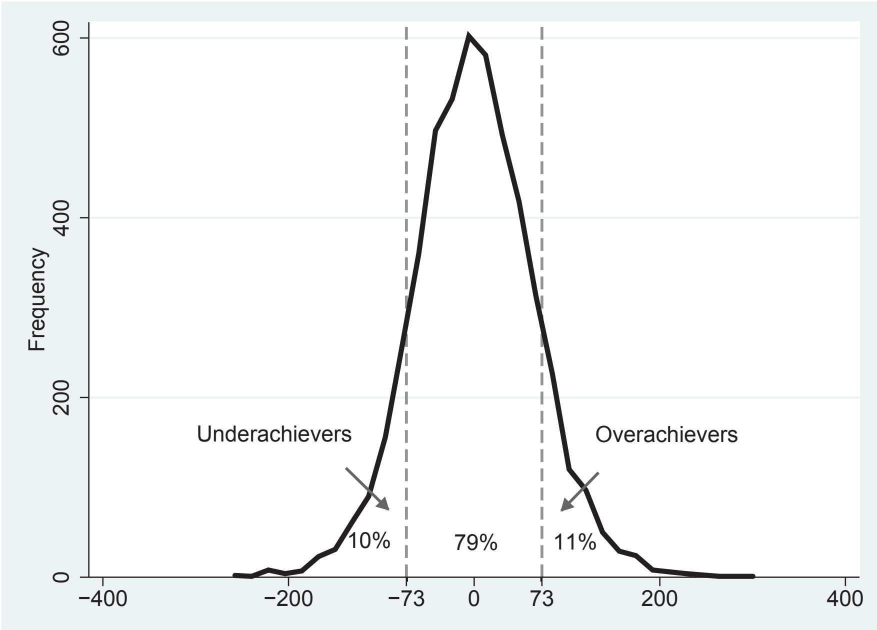
Table 3.2: Estimation results: OLS PISA literatcy test scores

	(1)	(2)	(3)	(4)
Female	6.607*		6.478*	6.921*
	(3.184)		(2.829)	(2.826)
Immigrant: second generation	−37.021**		−25.557**	−23.665**
	(6.057)		(5.085)	(5.063)
Immigrant: first generation	−56.198**		−33.927**	−33.093**
	(5.708)		(4.837)	(4.791)
Age 16	−23.756**		−18.439**	−17.655**
	(3.628)		(3.200)	(3.199)
Parental education: upper sec.	19.101**		10.677**	11.126**
	(4.077)		(3.478)	(3.476)
Parental education: tertiary	15.227**		5.477	6.111+
	(4.239)		(3.597)	(3.576)
Family structure: single	−7.592		−1.812	−0.966
	(4.803)		(4.199)	(4.167)
Family structure: mixed	−5.159		3.127	3.650
	(6.429)		(5.776)	(5.770)
Family structure: other	−37.120**		−20.128*	−18.052*
	(9.540)		(9.205)	(8.995)
School mark in test language		17.840**	14.764**	14.881**
		(2.674)	(2.670)	(2.675)
School mark in mathematics		4.705*	4.769*	4.431*
		(2.022)	(1.995)	(1.997)
School mark in sciences		12.429**	10.143**	9.885**
		(2.575)	(2.502)	(2.512)
Track lower sec II: high-level		124.446**	113.028**	112.258**
		(5.723)	(5.663)	(5.647)
Track lower sec II: medium-level		82.060**	73.878**	73.441**
		(3.761)	(3.651)	(3.624)
Track lower sec II: no selection		7.267	4.763	7.972
		(10.149)	(9.554)	(9.532)
Instrumental motivation				−0.224
				(1.418)
Absenteeism				−6.237**
				(1.548)
Sense of belonging				2.594+
				(1.413)
Constant	495.542**	269.332**	302.840**	304.532**
	(6.682)	(17.332)	(17.382)	(17.587)
N	2097	2097	2097	2097
R-squared	0.180	0.340	0.388	0.395

+ p<0.10, * p<0.05, ** p<0.01, robust standard errors in parentheses.
Reference group: male, Swiss parents, age¡16, highest parental education: compulsory school, nuclear family structure, low-level compulsory school, cantons (22) controlled for in all models.

Figure 3.1 shows the distribution of the residuals resulting from regressing test scores on students attributes. For 78% of the pupils, the regression model is able to predict PISA test scores within the range of one competence level (73 points). Approximately 11% of the observations at each end of the residual distribution are identified as *overachievers* (positive deviation larger than one proficiency level) or *underachievers* (negative deviation larger than one proficiency level). For all these observations, realized test scores of under- and overachievers lie outside the 95% confidence interval of the predicted value.

Figure 3.1: Distribution of unexplained PISA test scores (residuals)



Note: The unexplained part of PISA test scores is computed by subtracting predicted test scores (OLS regression on all observables) from realized test scores (see section 3.3 and Model 4 of table 3.2).
73 score points refer to one PISA literacy competence level according to OECD (2001).

For the dependent variables, shown in table 3.3, we will first analyze the indicator of whether somebody who has applied for a training position succeeds in *entering a certifying apprenticeship* directly after compulsory schooling or not. The share of those without an apprenticeship one year after the PISA test is 25%.

The second dependent variable provides information on the *intellectual aspiration level* of the vocational track for those who have started apprenticeship training. The aspiration levels for 101 different vocational tracks were rated on a scale ranging from 1 to 6 by an expert group of vocational advisers (for details and studies using this variable see Stalder, 2011).[12]

Approximately 45% of all apprentices in our sample follow an apprenticeship track of high intellectual aspiration level (5 or 6), e.g., toward a certificate as a commercial employee,

[12] We have imputed missing information (1.4%) about the aspiration level of less common tracks by regressing on training duration (years), amount of vocational schooling lectures (hours per year) and an interaction term between the two. These factors strongly explain the aspiration level (R-squared of 84%).

IT technician, electronic technician, draughtsman, chemist, or optician. Approximately 25% do an apprenticeship with a low intellectual aspiration level (1 or 2), such as hairdressers, gardeners, bakers, painters, salespeople, florists, cooks, carpenters, or cosmeticians. As shown in table 3.3, there is a relationship in the data between one's reading literacy, as measured by PISA, and the intellectual aspiration level of the vocational track someone follows.[13] Due to the ordinal character of the intellectual aspiration level we perform ordered probit estimation and present average partial effects for the highest aspiration level.

Table 3.3: Dependent variables—univariate and bivariate (with PISA Test Scores)

		*Distribution of PISA test scores**			
Outcome variables	Share (%)	Mean*	Std. Dev*	Min*	Max*
Educational status (N=2079)					
Non-certifying/no education (0)	24.7	480.5	85.2	255.0	812.9
Certifying apprenticeship training (1)	75.3	501.9	75.3	198.0	737.5
Training aspiration level (N=1578)					
very low (1)	13.8	457.9	74.5	198.0	640.9
low (2)	11.0	472.7	78.9	278.5	737.4
lower medium (3)	12.8	481.9	70.2	268.8	671.7
upper medium (4)	16.9	493.6	70.6	268.1	670.5
high (5)	9.6	528.7	65.6	327.0	668.7
very high (6)	35.9	531.6	65.6	300.8	737.5
Problems in training (N=1382)					
No problems (0)	80.7	511.7	73.7	198.0	737.5
Any problems (1)	19.3	482.4	71.7	298.3	670.5

To test whether hard-to-observe ability has further effects beyond the success of the school-to-work transition, we analyze an indicator that takes a value of 1 for evidence of problems during apprenticeship training, such as repetition of an apprenticeship year, change of training occupation, failure on the final exam or dropping out of training. The share of apprentices who had at least one of those critical events is 19% of those who are observable in the data for the standard duration of their training.

[13] PISA test-scores have not been used by the experts to assess the aspiration level.

3.5 Results

3.5.1 Probability of directly entering a certifying apprenticeship training

The first probit model in table 3.4 only includes characteristics that are easily observable by firms and might be used by these to form expectations of students' ability. The estimated average marginal effects show that many of the easy-to-observe characteristics play a statistically significant role. Applicants coming from a medium-level compulsory school track have e.g. a 9.1 percentage point higher probability of entering apprenticeship training than those from basic-level tracks. The probability of successfully applying to a firm is, however, highest for those from high-level compulsory schools (with a difference of 19.0 percentage points). The school mark in mathematics is important as well: having a better mark of one unit (for example a mark of 5.0 (good) instead of 4.0 (sufficient)) increases the probability of successfully applying for an apprenticeship by 4.8 percentage points. In contrast, marks in sciences and in the test language (regional language) seem to have no additional effect.

As for the background variables, females, immigrants (first- and second-generation), pupils with higher absenteeism and those living in single parent households or patchwork families are significantly less likely to be in certifying apprenticeships one year after PISA (all else being equal); having parents with upper secondary education (as opposed to compulsory school education) has a positive effect. On the other hand, having parents with tertiary education does not significantly increase the probability of having a smooth transition, although the coefficient goes in the expected direction. Apart from the gender variable, all the coefficients point in the same direction as the coefficients in the PISA test score regression in table 3.2 in section 3.4.

Model 2 additionally includes the PISA literacy test score. The test score is scaled (1=73 PISA points), so that the effect size can be interpreted as the effect of a variance of 73 PISA points (which is equal to one competence level). The coefficient shows no significant effect on successfully applying for apprenticeship training, implying that hard-to-observe ability would not be relevant for a successful transition. The fact that some of the observable characteristics still have a significant coefficient even when the PISA score is controlled for should not immediately be interpreted as a sign of discrimination of these school leavers as it is highly unlikely, that the PISA score incorporates all relevant information for the employers.

Table 3.4: Estimation results: Probability of directly entering apprenticeship training

Probit estimation: 1=direct entry into certifying apprenticeship, 0 otherwise							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PISA Literacy test score/73		0.014					
		(0.010)					
Predicted PISA score/73			0.120**	0.119**	0.118**	0.119**	
			(0.015)	(0.015)	(0.015)	(0.015)	
Residual PISA score/73				0.014			
				(0.011)			
Negative PISA residuals*(-1)/73					−0.041*		
					(0.020)		
Positive PISA residuals/73					−0.014		
					(0.021)		
Underachiever						−0.063*	−0.068*
						(0.031)	(0.031)
Overachiever						−0.000	−0.003
						(0.031)	(0.030)
Female	−0.167**	−0.169**					−0.168**
	(0.017)	(0.017)					(0.018)
Immigrant: second generation	−0.107**	−0.102**					−0.107**
	(0.032)	(0.032)					(0.032)
Immigrant: first generation	−0.098**	−0.090**					−0.098**
	(0.029)	(0.029)					(0.029)
Age 16	−0.017	−0.014					−0.018
	(0.019)	(0.020)					(0.019)
Parental education: upper sec.	0.050*	0.048*					0.049*
	(0.021)	(0.021)					(0.022)
Parental education: tertiary	0.020	0.019					0.019
	(0.022)	(0.022)					(0.023)
Family structure: single	−0.108**	−0.108**					−0.108**
	(0.025)	(0.025)					(0.025)
Family structure: mixed	−0.083*	−0.083**					−0.085*
	(0.032)	(0.032)					(0.033)
Family structure: other	−0.038	−0.034					−0.033
	(0.044)	(0.044)					(0.047)
School mark in test language	0.002	−0.001					0.002
	(0.016)	(0.017)					(0.016)
School mark in mathematics	0.048**	0.047**					0.050**
	(0.012)	(0.012)					(0.012)
School mark in sciences	0.007	0.005					0.007
	(0.015)	(0.015)					(0.014)
Track lower sec II: high-level	0.190**	0.168**					0.188**
	(0.034)	(0.039)					(0.035)
Track lower sec II: medium-level	0.091**	0.076**					0.090**
	(0.021)	(0.024)					(0.021)
Track lower sec II: no selection	−0.011	−0.014					−0.016
	(0.070)	(0.070)					(0.077)
Instrumental motivation	0.007	0.007					0.007
	(0.009)	(0.009)					(0.009)
Absenteeism	−0.020*	−0.019*					−0.019*
	(0.009)	(0.009)					(0.009)
Sense of belonging	0.012	0.011					0.012
	(0.009)	(0.009)					(0.009)
N	2097	2097	2097	2097	2097	2097	2097
Pseudo R-squared	0.164	0.165	0.088	0.089	0.090	0.090	0.166

+ p¡0.10, * p¡0.05, ** p¡0.01, Average marginal effects, bootstrapped standard errors in parentheses.
Reference group: Achieves-as-expected, male, Swiss parents, age¡16, parental education: comp. school, nuclear family structure, low-level compulsory school track, cantons (22) controlled for in all models.

Although we cannot exclude discrimination, the coefficients might also indicate that these variables are proxies for other important parts of the prospective apprentices that are taken into account in the hiring process.

Model 3 uses only the part of the hard-to-observe ability that is *predictable* by easy-to-observe characteristics of the applicants. The predicted PISA result thus replaces the easy-to-obtain information. According to the results, an additional predicted competence level would increase the chances to obtain an apprenticeship by 12.0 percentage points (with p=0.000).

Model 4 additionally includes the part of the PISA test-score that cannot be predicted by observables. This variable is not statistically significant and the point estimate is rather small (equivalent to the 1.4 percentage points in model 1). Hence given the difference between the predictable part of the PISA score and the hard-to-obtain part of it, one would deduce that the hard to obtain part of the PISA information plays no role in firms' decisions to recruit apprentices.

However, once we differentiate between a positive and a negative residual it becomes obvious that the non-existent effect of the residual is due to the asymmetric impact that the residual has on the probability of a successful transfer into apprenticeship training. If we distinguish between the negative and positive residual in the PISA scores (model 5) and create dummy variables for being either an *underachiever* or an *overachiever* in model 6 (along with predicted PISA) and model 7 (along with all easy-to-observe variables)[14], we see that only a negative deviation from the predicted PISA score seems to matter for employers. School leavers that score at least one competence level below the PISA score that one would have expected based on their easy-to-observe characteristics, have a probability for a successful transition that lies almost seven percentage points below the probability of school-leavers with identical observables. In contrast, school leavers with a substantively higher PISA score than one would have expected have the same probability for a successful transition as less able colleagues with the same observable characteristics.

Contrasting models 5, 6 and 7 additionally shows that using residuals (model 5) or dummy variables for large deviations from the predicted PISA results does not change the results,

[14] We have also estimated models that allow for the possibility that the effect of (positive/negative) residual PISA scores is different in magnitude depending on whether the difference is within the magnitude of one PISA competence level (small) or outside one competence level (the definition for under-/overachievers). Results show that small deviations are neither significantly different from large deviations nor significantly different from zero and thus more spurious.

neither does the inclusion of easy-to-observe variables instead of predicted PISA test scores (difference between model 6 and 7). The result that underachievers have significantly lower probabilities in seamlessly transition from school to apprenticeship therefore is robust across different model specifications.

One might argue that, during the hiring process, employers especially look out for non-cognitive skills or behavioral aspects of pupils, such as motivation, and that if PISA test scores are correlated with these other factors, our estimations might reflect the importance of these non-cognitive factors and not of cognitive abilities, which are already demonstrated through school marks or educational track. We cannot completely rule out such mechanisms, as they are difficult to test. However, including or excluding information that is contained in PISA like *instrumental motivation*, *sense of belonging* or *absenteeism* (as a behavioral measure at age 15) in all our estimations does not significantly affect our qualitative results.[15]

3.5.2 Intellectual aspiration level of the dual vocational track

This section describes the empirical results of ordered probit estimations for the question of how hard-to-observe ability components influence the intellectual aspiration level (from 1 to 6) of the apprenticeship track a school leaver successfully enters.

The first model in table 3.5 again only includes characteristics that are easily observable for the employer and might be used by outsiders to build expectations of students' ability. The estimated marginal effects show that many easy-to-observe characteristics that are correlated with PISA test-scores (table 3.2 in section 3.4) play again a significant role for the sorting into occupations (tracks) with different intellectual aspiration levels. Given the effect sizes, the school track followed in compulsory schooling is the essential criteria for the allocation into the different demanding occupational tracks, besides parental education, school marks and instrumental motivation.

Including the PISA test-score information in model 2 illustrates that most of the coefficients of those easy-to-observe variables decrease due to the correlation with PISA and that hard-to-observe ability measured by PISA has a highly significant separate effect on the aspiration level: a hard-to-observe difference of one proficiency level leads to a 7.5 percentage point higher probability of following a highest-level vocational pathway.

According to model 4, the unpredictable part of PISA has, however, a much smaller

[15] Estimations excluding non-cognitive variables are available upon request from the authors.

Table 3.5: Estimation results: Intellectual aspiration level of vocational track

Ordered probit estimation: aspiration level scaled from 1 (very low) to 6 (very high)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PISA Literacy test score/73		0.075∗∗					
		(0.011)					
Predicted PISA score/73			0.270∗∗	0.273∗∗	0.273∗∗	0.273∗∗	
			(0.015)	(0.015)	(0.015)	(0.015)	
Residual PISA score/73				0.078∗∗			
				(0.010)			
Negative PISA residuals*(-1)/73					−0.075∗∗		
					(0.021)		
Positive PISA residuals/73					0.081∗∗		
					(0.022)		
Underachiever						−0.110∗∗	−0.115∗∗
						(0.033)	(0.032)
Overachiever						0.124∗∗	0.113∗∗
						(0.030)	(0.030)
Female	0.013	0.003					0.010
	(0.020)	(0.020)					(0.020)
Immigrant: second generation	0.138∗∗	0.156∗∗					0.132∗∗
	(0.038)	(0.037)					(0.038)
Immigrant: first generation	−0.016	0.009					−0.022
	(0.036)	(0.036)					(0.036)
Age 16	−0.060∗∗	−0.041∗					−0.063∗∗
	(0.021)	(0.020)					(0.021)
Parental education: upper sec.	0.078∗∗	0.064∗∗					0.073∗∗
	(0.024)	(0.023)					(0.024)
Parental education: tertiary	0.055∗	0.047∗					0.050∗
	(0.024)	(0.024)					(0.024)
Family structure: single	0.027	0.029					0.028
	(0.030)	(0.030)					(0.031)
Family structure: mixed	−0.014	−0.022					−0.016
	(0.042)	(0.041)					(0.044)
Family structure: other	−0.020	−0.009					−0.029
	(0.044)	(0.042)					(0.046)
School mark in test language	0.049∗∗	0.035∗					0.048∗∗
	(0.017)	(0.017)					(0.017)
School mark in mathematics	0.058∗∗	0.053∗∗					0.060∗∗
	(0.013)	(0.013)					(0.014)
School mark in sciences	−0.025	−0.034∗					−0.023
	(0.015)	(0.015)					(0.015)
Track lower sec II: high-level	0.567∗∗	0.451∗∗					0.569∗∗
	(0.034)	(0.038)					(0.033)
Track lower sec II: medium-level	0.342∗∗	0.270∗∗					0.345∗∗
	(0.023)	(0.025)					(0.022)
Track lower sec II: no selection	−0.081	−0.081					−0.074
	(0.072)	(0.069)					(0.073)
Instrumental motivation	0.041∗∗	0.040∗∗					0.040∗∗
	(0.009)	(0.009)					(0.009)
Absenteeism	−0.001	0.006					−0.000
	(0.010)	(0.010)					(0.010)
Sense of belonging	0.012	0.010					0.013
	(0.009)	(0.009)					(0.009)
N	1578	1578	1578	1578	1578	1578	1578
Pseudo R-squared	0.089	0.098	0.068	0.077	0.077	0.075	0.095

+ p¡0.10, * p¡0.05, ** p¡0.01, Average marginal effects, bootstrapped standard errors in parentheses.
Reference group: Achieves-as-expected, male, Swiss parents, age¡16, parental education: comp. school, nuclear family structure, low-level compulsory school track, cantons (22) controlled for in all models.

effect than the predictable one, showing that the part of the effect of the observables that also explains differences in PISA scores is the most relevant and only to lesser extent the exact PISA scores.

The subsequent tests for asymmetric effects suggest that the part of ability that cannot be predicted by observables affects the outcome rather symmetrically: the effect of the residuals in both directions amounts to approximately 8 percentage points (model 5). Likewise, *PISA underachievers* (dummy-specification) are on average found in lower aspiration levels, and *PISA overachievers* are found in higher aspiration levels relative to otherwise identical school leavers whose ability is well predicted by easy-to-observe characteristics. Both coefficients of the under- and overachiever dummies are highly significant and approximately 11 percentage points in magnitude.

3.5.3 Failure and success in apprenticeship training

Table 3.6 shows the probit estimation results for having problems during training such as dropping out, repeating a year, changing training occupation or failing the external exam. The set of independent variables across models is the same as before, we additionally include the new information on the aspiration level of the vocational track.

The estimation results in table 3.6 imply the following: First, hard-to-observe ability has a significant effect: the higher the unpredictable PISA score at the time of application, the less likely are costly events during the apprenticeship period, such as dropping out, repeating a year, changing training occupation or failing the exam. A higher PISA score of one proficiency level (73 test-score points) decreases the probability of having problems by approximately 4.5 percentage points (models 2 and 4). The coefficients for large deviations presented by the under- and overachievers dummies show a rather symmetrical pattern for negative and positive deviations (model 6 and 7).

Second, in comparison with the previous analyses in section 3.5.1 and 3.5.2, (initially) hard-to-observe ability has gained in relative importance compared to the ability part that could be predicted by observables.[16] This finding goes in line with labor market theories that postulate that employers learn about ability of their employees as time goes by.

[16] The residual in model 4 has about half of the effect size of the predicted PISA score. In the previous regressions this relation was 1:8.5 (no significant effect of the overall residual) and 1:3.5.

Table 3.6: Estimation results: Problems in apprenticeship

Probit estimation: Problems in apprenticeship: 1=evidence of problems, 0=otherwise							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PISA Literacy test score/73		−0.043**					
		(0.012)					
Predicted PISA score/73			−0.088**	−0.097**	−0.097**	−0.094**	
			(0.020)	(0.020)	(0.020)	(0.019)	
Residual PISA score/73				−0.045**			
				(0.013)			
Negative PISA residuals*(-1)/73					0.051*		
					(0.024)		
Positive PISA residuals/73					−0.039		
					(0.025)		
Underachiever						0.095*	0.094*
						(0.038)	(0.039)
Overachiever						−0.073*	−0.066+
						(0.035)	(0.034)
Female	−0.022	−0.019					−0.021
	(0.022)	(0.022)					(0.022)
Immigrant: second generation	0.062+	0.048					0.061
	(0.037)	(0.037)					(0.037)
Immigrant: first generation	0.064+	0.051					0.067+
	(0.037)	(0.037)					(0.038)
Age 16	0.050*	0.040+					0.052*
	(0.024)	(0.024)					(0.024)
Parental education: upper sec.	−0.039	−0.032					−0.036
	(0.026)	(0.026)					(0.026)
Parental education: tertiary	−0.027	−0.026					−0.026
	(0.027)	(0.027)					(0.028)
Family structure: single	0.059+	0.059+					0.058+
	(0.032)	(0.032)					(0.033)
Family structure: mixed	0.082*	0.084*					0.082*
	(0.040)	(0.039)					(0.041)
Family structure: other	0.054	0.051					0.063
	(0.058)	(0.057)					(0.061)
School mark in test language	0.014	0.024					0.016
	(0.019)	(0.019)					(0.019)
School mark in mathematics	−0.012	−0.012					−0.015
	(0.015)	(0.015)					(0.015)
School mark in sciences	−0.029+	−0.024					−0.029+
	(0.017)	(0.017)					(0.018)
Track lower sec II: high-level	−0.079+	−0.030					−0.096*
	(0.044)	(0.046)					(0.045)
Track lower sec II: medium-level	−0.034	−0.004					−0.047
	(0.029)	(0.030)					(0.031)
Track lower sec II: no selection	0.068	0.068					0.066
	(0.073)	(0.073)					(0.075)
Instrumental motivation	0.008	0.006					0.006
	(0.010)	(0.010)					(0.010)
Absenteeism	0.016+	0.012					0.014
	(0.009)	(0.010)					(0.010)
Sense of belonging	−0.011	−0.011					−0.013
	(0.010)	(0.010)					(0.010)
Aspiration level (2)	0.056	0.067	0.053	0.063	0.063	0.063	0.068
	(0.043)	(0.043)	(0.045)	(0.045)	(0.045)	(0.044)	(0.045)
Aspiration level (3)	0.092*	0.097*	0.090*	0.094*	0.094*	0.092*	0.096*
	(0.040)	(0.039)	(0.040)	(0.040)	(0.040)	(0.039)	(0.040)
Aspiration level (4)	0.076*	0.090*	0.079+	0.092*	0.092*	0.088*	0.089*
	(0.038)	(0.038)	(0.041)	(0.041)	(0.041)	(0.040)	(0.040)
Aspiration level (5)	0.113*	0.135**	0.114*	0.135**	0.135**	0.124**	0.128**
	(0.044)	(0.044)	(0.046)	(0.046)	(0.046)	(0.044)	(0.044)
Aspiration level (6)	0.095**	0.118**	0.106**	0.128**	0.127**	0.123**	0.117**
	(0.036)	(0.036)	(0.037)	(0.037)	(0.037)	(0.036)	(0.036)
N	1382	1382	1382	1382	1382	1382	1382
Pseudo R-squared	0.114	0.122	0.097	0.106	0.097	0.108	0.123

+ p¡0.10, * p¡0.05, ** p¡0.01, Average marginal effects, bootstrapped standard errors in parentheses.
Reference group: Achieves-as-expected, male, Swiss parents, age¡16, parental education: comp. school, nuclear family structure, low-level compulsory school track, cantons (22) controlled for in all models.

3.5.4 Further investigations—using an alternative ability measure

In this section we test the robustness of our results against an alternative ability measure provided by the data, namely the PISA literacy test score in mathematics[17].

Since mathematical literacy has been tested only incidentally in PISA 2000, the sample shrinks to 54% of our initial sample. The correlation coefficient between test scores in reading and test scores in mathematics is $r = 0.56$ in this subsample.

We replicate our analysis by again assuming that these test scores are, first, correlated with the ability information that employers would like to have[18], second, hard-to-observe to the labor market but known to the researcher, and third, comparable across individuals.

In spite of having substituted the dependent variable, the new PISA test score model in table 3.7 looks very similar to the one presented in table 3.2 on page 66. A comparison of the two PISA regressions brings about two notable findings: First, in contrast to the results for reading literacy, females have lower test scores in mathematics by nearly half a PISA proficiency level compared to men. Second, the performance gap between natives and immigrants persist (approximately half a PISA proficiency level) and is even slightly larger in mathematics than in reading.

As for the effect of hard-to-observe ability on transition and apprenticeship success, the main results in table 3.7 show to be nearly identical to those of the previous sections: when it comes to the hiring decision, only negative ability deviations from expected ability seem to matter and positive deviations are not considered by the market. Again, the hard-to-observe part of PISA scores affects the sorting into aspiration levels rather symmetrically. As for problems during apprenticeship we find, again, a significant effect of hard-to-observe ability. With regards to the last outcome and in contrast to the analysis using reading literacy and the whole sample, we do, however, not find evidence for a better performance of overachievers, but only higher difficulties for underachievers.

Overall, the main results are in line with our previous findings: First, firms seem to look out for (hard-to-observe) ability that shows to matter for the success of apprenticeship training. Second, we find evidence for our hypotheses that firms especially try to diminish the

[17] Empirical studies based on NSLY79 data (see section 3.2) often use fathers' education as an alternative ability measure that is assumed to be hard-to-observe. Throughout this chapter, however, we regard parental education as to be easy-to-observe, as parental education / occupation is standard information in application dossiers of apprenticeship applicants.

[18] Potential training firms often not only complain about insufficient language skills of applicants, but also about insufficient mathematical knowledge of compulsory school leavers.

Table 3.7: Estimation results: replication using PISA literacy test scores in mathematics

	OLS: PISA	Probit: Entry			Oprobit: Aspiration level			Probit: problems		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PISA Literacy test score/73		0.030*			0.069**			−0.038**		
		(0.013)			(0.012)			(0.015)		
Predicted PISA score/73			0.154**	0.154**		0.228**	0.228**		−0.081**	−0.079**
			(0.016)	(0.016)		(0.017)	(0.017)		(0.023)	(0.022)
Residual PISA score/73			0.030*			0.068**			−0.039*	
			(0.013)			(0.013)			(0.015)	
Negative Residuals*(-1)/73				−0.042*			−0.047*			0.066**
				(0.021)			(0.022)			(0.024)
Positive Residuals/73				0.014			0.096**			−0.000
				(0.026)			(0.025)			(0.031)
Female	−30.597**	−0.148**			0.020			−0.038		
	(4.065)	(0.023)			(0.026)			(0.028)		
Immigrant: second generation	−32.945**	−0.055			0.137**			0.040		
	(8.433)	(0.043)			(0.050)			(0.048)		
Immigrant: first generation	−35.985**	−0.045			−0.063			0.052		
	(7.588)	(0.038)			(0.046)			(0.048)		
Age 16	−15.486**	−0.017			−0.030			0.058+		
	(4.637)	(0.026)			(0.027)			(0.031)		
Parental education: upper sec.	1.074	0.049+			0.045			−0.018		
	(4.926)	(0.028)			(0.032)			(0.034)		
Parental education: tertiary	2.482	0.013			0.038			0.020		
	(5.638)	(0.031)			(0.033)			(0.037)		
Family structure: single	2.911	−0.095**			0.053			0.095*		
	(6.818)	(0.035)			(0.042)			(0.041)		
Family structure: mixed	13.956+	−0.092*			−0.003			0.141**		
	(7.986)	(0.045)			(0.057)			(0.053)		
Family structure: other	−10.752	0.001			−0.007			−0.037		
	(11.178)	(0.059)			(0.052)			(0.088)		
School mark in test language	1.666	0.001			0.055*			0.010		
	(4.514)	(0.022)			(0.022)			(0.025)		
School mark in mathematics	21.142**	0.060**			0.027			0.000		
	(3.480)	(0.018)			(0.019)			(0.020)		
School mark in sciences	13.511**	−0.009			−0.010			−0.010		
	(3.654)	(0.019)			(0.020)			(0.021)		
Track lower sec II: high-level	112.906**	0.161**			0.398**			−0.082		
	(9.338)	(0.052)			(0.050)			(0.059)		
Track lower sec II: medium-level	73.038**	0.064*			0.236**			−0.051		
	(5.194)	(0.031)			(0.033)			(0.040)		
Track lower sec II: no selection	9.229	0.020			−0.067			0.024		
	(14.886)	(0.092)			(0.094)			(0.089)		
N	1144	1144	1144	1144	875	875	875	764	764	764
(Pseudo) R-squared	0.389	0.168	0.133	0.133	0.087	0.065	0.066	0.139	0.103	0.105

+ p¡0.10, * p¡0.05, ** p¡0.01, Average marginal effects, bootstrapped standard errors in parentheses.
Reference group: Achieves-as-expected, male, Swiss parents, age¡16, parental education: comp. school, nuclear family structure, low-level compulsory school, cantons (22) controlled for in all models, intellectual aspiration level controlled for in models 8-10.

risk of negative ability surprises when hiring. Third, hard-to-observe ability plays a significant but minor role in the hiring decision but gains relative importance during apprenticeship training.

3.6 Summary and conclusion

The objective of this chapter was to analyze whether and to what extent employer learning about hard-to-observe ability takes place at the very beginning of a worker's career, namely, in the transition process from compulsory schooling to marked-based upper-secondary education in Switzerland. In light of the fact that apprenticeship contracts have standardized content and fixed duration and leave little scope to adjust prearranged wage profiles over the training period, our first aim was to analyze whether employers try to obtain more revealing information about an applicant's ability *before* hiring rather than only relying on readily available information. We test how deviation in the PISA 2000 test scores from what one would predict based on observable characteristics influences successful transition and training. As the institutional setting may particularly provide incentives for firms to gather and use hard-to-get information in order to avoid hiring applicants whose ability level considerably lies *below* the expected level (so called *underachievers*), we allow for asymmetric effects of hard-to-get ability information in the hiring process. Second, we analyze whether hard-to-observe ability is further revealed in the course of the apprenticeship period and becomes observable through training outcomes.

Our results suggest that hard-to-observe ability plays a significant role in transition as well as the training success, but not always in a symmetric manner. Regarding applicants' transition success, we find that only PISA underachievers are affected by pre-market employer learning. They are less likely to successfully apply for apprenticeships than their otherwise identical peers. Overachievers, in turn, do not seem to benefit from having more academic potential than one would expect. Therefore, costly-to-observe ability components are only revealed in the hiring process at the lower end of the residual distribution, indicating that firms use pre-market learning in particular to minimize the downward risk of a mismatch.

For the resulting allocation of successful applicants into different intellectually demanding vocational tracks, we find, however, rather symmetric effects; hard-to-get ability information is revealed in a way that significantly increases allocative efficiency at both ends of the distribution.

The results regarding long-term outcomes suggest, however, that there is still additional revelation of ability during the subsequent training period. Apprentices who are PISA overachievers are less likely to face problems, such as dropping out, repeating a year of apprenticeship, changing vocational track or final exam failure. In contrast, PISA underachievers who, despite their lower-than-expected ability, successfully find an apprenticeship are disproportionally more likely to be exposed to these problematic events. The fact that even underachievers who successfully find an apprenticeship show inferior outcomes during the apprenticeship period provides an additional explanation for why firms seem to place more emphasis on detecting under- rather than overachievers in the course of the hiring process.

We showed in this chapter that, in the case of costly and far-ranging hiring decisions, such as apprenticeship training contracts, information that cannot be observed easily is already used by employers at the initial stage of the hiring process and that applicants that differ from their apparently similar peers in regard of their ability are therefore treated differently.

However, due to the nature of our data, we can only observe the outcome of the hiring process and not the behavior of firms themselves. Given the observation that some easy-to-observe characteristics like nationality, school track or parental background have an impact on these outcomes even after controlling for observable cognitive and non-cognitive competencies, future research should address the question, whether the impact of these characteristics is due to discrimination or whether they stand for information that so far only firms observe but not the researchers.

Finally, our finding that negative and positive hard-to-observe ability surprises are revealed and accounted for in an asymmetric way—presumably depending on labor market or institutional arrangements—could be further tested within the setting of the standard employer learning literature.

3.A Appendix

Table 3.A1: Variable definition

Variable[a]	**Definition**
Certifying education	Dep. variable. Equals 1 if pupil enters a certifying upper-sec. education in the form of apprenticeship training directly after compulsory school; 0 otherwise (non-certifying interim solutions or no education).
Aspiration level	Dep. variable. Aspiration level of 101 vocational tracks (expert ratings) from 1 (low) to 6 (high).
Problems in training	Dep. variable. Equals 1 if apprentice exhibits repetition of apprenticeship year, failure in exam, change in training occupation or drop out.
PISA test score	Reading literacy test score from the PISA 2000 survey.
Underachiever	Equals 1 if PISA test score is considerably (1 competence level) lower than predicted by observables; 0 otherwise.
Overachiever	Equals 1 if PISA test score is considerably (1 competence level) higher than predicted by observables; 0 otherwise.
Female	Equals 1 if female; 0 if male.
Nationality	Dummies representing 3 categories: "Swiss" (born in Switzerland with at least one parent born in Switzerland), "second-generation immigrant" (pupil born in Switzerland but parents born outside Switzerland), "first-generation immigrant" (pupil and parents foreign born).
Age 16 at PISA survey	Equals 1 if pupil aged 16 at the time of PISA 2000; 0 if aged 15.
Parental education	Dummies representing 3 categories of highest parental education: compulsory school, upper-secondary education, tertiary education.
Family structure	Dummies representing 4 categories: nuclear, single, mixed and other, where the last category also covers missing information.
Mark in test language	Mark in test language (German, French, Italian, depending on linguistic region) in last school report. Metric scale: 1-6 (1=lowest, 6=highest).
Mark in mathematics	Mark in mathematics in last school report (1=lowest, 6=highest).
Mark in sciences	Mark in sciences (mean across biology, chemistry, physics, sciences) in last school report. Metric scale: 1-6 (1=lowest, 6=highest).
Level compulsory school	Dummies for the school track that was attended at the time of the PISA 2000 survey: low-level compulsory school (e.g. Realschule), medium-level compulsory school (e.g. Sekundarschule), high-level compulsory school (e.g. Pro-Gymnasium) and "no selection" (integrated track).
Regions (cantons)	Dummies for 22 Swiss cantons (= states).
Instrumental motivation	PISA index[b] derived from students' reports on how often they study to increase their job opportunities, to ensure that their future will be financially secure, and to enable them to get a good job (3 items).
Sense of belonging	PISA index[b] derived from students' reports on whether their school is a place where they feel like an outsider, make friends easily, feel like they belong, feel awkward/out of place, other students seem to like them, or feel lonely (6 items).
Absenteeism	PISA index[b] derived from students' reports on how often they missed school/classes and were late for school in the two last weeks (3 items).

[a] All variables except for the outcome variables are measured at the time of PISA 2000. [b] z-standardized indices used.

Chapter 4

The effect of gap years and interim solutions on educational outcomes

4.1 Introduction

Having an upper-secondary education diploma is becoming increasingly important for successful and enduring labor market integration in all industrialized countries (OECD, 2012a). There is ample international evidence on pecuniary and non-pecuniary benefits of education at the individual and social level (Card, 1999; Harmon et al., 2003; Lange and Topel, 2006; Oreopoulos and Salvanes, 2011; OECD, 2010, 2012b). Consequently, the sociopolitical objective to equip virtually all people of newer cohorts with formal post-compulsory education has been placed high on the Swiss political agenda in recent years. In the course of this process, the federal state and cantons have jointly defined the aim to enhance the upper-secondary graduation rate by age 25 from 90 to 95 percent by 2015 (EDK, 2006; EVD/EDI/EDK, 2011)[1]. One of the identified starting points is to work against the steadily increasing graduation age at upper-secondary level (EDK, 2006), mainly caused by decreasing rates of *direct* entries into certifying education and, inversely, by increasing rates of young people in non-certifying gap years after compulsory schooling. According to official statistics[2], about one quarter of young people do not follow certifying education in the year after compulsory school (age sixteen). Amongst them, a majority are observed to follow so-called "interim solution

[1] EVD: Federal Department of Economic Affairs, EDI: Federal Department of Home Affairs, EDK: Swiss Conference of Cantonal Ministers of Education.

[2] See the *key indicators on education* provided by the Swiss federal statistical office (FSO) at http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/17/blank/01.indicator.404301.4084.html, Verläufe und Übergänge - Übergang in die Sekundarstufe II.

programs". Similar developments are observed in Germany, where the rate of young people in the so-called transitory system has been even more pronounced (approximately 40% in 2004 compared to 15% in Switzerland, see Hupka-Brunner et al., 2011). These interim solution programs share some features with active labor market programs and are especially designed to bridge the gap between lower-secondary and (labor market-based) upper-secondary education, providing those who have neither found an apprenticeship training place nor admission to general education schooling with relevant skills in order to enhance their chances of successfully entering certifying education the subsequent year. The main objective of this chapter is to evaluate the short- and long-run impact of these programs.

Similar to other active programs, the prevalence and quantitative importance of interim solutions is sometimes suspected to have the undesirable side effect of an institutionalized reinforcement of their necessity by, for example, increasing the expectations of training firms or schools regarding educational prerequisites of applicants by facilitating the postponement of educational decisions and serious search efforts of compulsory school leavers for apprenticeship places, or by exculpating compulsory schools from adequately preparing pupils to directly enter certifying tracks. Nevertheless, given the dominant role of these programs in Switzerland, a pragmatic question that can be asked is whether the intended program objectives are met such that participants fare better than they would in an alternative state.

Empirical knowledge on the effects of different sorts of gap years is important because they potentially induce individual and social costs. First, lagged entries into certifying education are associated with individual opportunity costs, especially if gap years are attributed to redundant loops and waiting years that do not enhance—or even lower—the changes of acquiring a certifying education afterwards (e.g., by stigmatization). In addition, in the case of interim solution programs, there might be private direct costs (e.g., admission fees) and, moreover, public direct costs and opportunity costs arising from the public supply of most of these programs. On the other hand, there might be large individual and social gains arising from interim solution programs, namely if these programs are effective measures to, first, integrate individuals into upper-secondary education who otherwise would not have succeeded in taking up any further studies or training; second, increase individuals' chances to enter more challenging tracks than otherwise possible; or, third, increase the possibility of entering an educational track with a better match to someone's interests and thus with lower risks of dropping out of education or undergoing later career adjustments. So far, little is known about the causal effects of interim solutions on further educational involvement and success. Similarly, little is known about the effects of gap years without any educational activity.

This chapter analyzes how gap years spent in *interim solution programs* affect, first, the chances of entering certifying education in the subsequent year; second, the intellectual aspiration level of the certifying education taken up; and third, the success in acquiring upper-secondary education by age 21, compared to both *direct entries* into certifying education and to gap years *without educational activity* after compulsory schooling. As interim solution programs might be more useful to some individuals in some cases, we will place emphasis on detecting potential heterogeneity in program effects and on a detailed discussion of the (non-)selection into programs.

The lack of empirical evidence on the consequences of gap years and the effectiveness of interim solutions can be ascribed to the lack of longitudinal administrative data on educational pathways in Switzerland. We use the Swiss follow-up survey of the PISA 2000-cohort *TREE* (*transition from education to employment*), providing the only (large-scale) data on individual trajectories and later outcomes. Other studies based on these data suggest considerable discontinuities and lagged entries after compulsory schooling, but also considerable numbers of young people with gap years finally succeeding in taking up certifying education (Keller et al., 2010; Hupka-Brunner et al., 2011). Furthermore, Buchholz et al. (2012) find that the Swiss education system, even though sharing many features with the German system, is more effective in integrating a vast majority of young people into market-based upper-secondary education, especially those with only low cognitive competences who are regarded as not capable of undergoing training in Germany. Buchholz et al. (2012) suggest that the Swiss interim solution system might thus rather operate as a “jumping board”, whereas related programs in Germany more often lead into a “dead end”. The effect of different sorts of gap years has not been analyzed directly in these studies, however.

As the route someone follows after compulsory schooling is not randomly decided, an assessment of the causal effect of different kinds of gap years on further educational success requires speculation on what would have happened to a person had s/he taken up another decision. We address this fundamental evaluation problem by applying propensity score matching techniques with multiple treatments (Rosenbaum and Rubin, 1983; Imbens, 2000; Lechner, 2001) based on the information from the *TREE* data. The data allow us to observe a wide set of background information on compulsory school graduates, including information on PISA 2000 test scores, former school performance, expectations and behavior, parental background, and regional information. We justify our estimation technic by assuming that the treatment-control-group design based on this unusually rich set of variables allows us to properly estimate counterfactual outcomes.

The remainder of this chapter is structured as follows: more detailed information on the institutional setting and the specific research questions are discussed in Section 4.2. Section 4.3 describes the data and the sample. The evaluation framework and estimation strategy is presented in Section 4.4. Section 4.5 describes the treatment and control group construction, and Section 4.6 is dedicated to the results of the empirical analyses. Section 4.7 concludes the chapter.

4.2 Institutional setting and hypothesis

In Switzerland, compulsory schooling ends after nine years (approximately at age 15), and youngsters have to decide on their further educational pathway. While only a small share decides to follow a general education track, the vast majority opts for VET-based education (see SKBF-CSRE, 2011). There are about 250 different vocational tracks regulated by the federal state, with a high diversity of tracks regarding the intellectual aspiration level. As following dual apprenticeship education requires having a contract with a training firm, upper-secondary education is prone to market forces for a considerable part of a cohort. It requires early vocational career orientation of compulsory school graduates, a match between supply and demand for apprenticeship places in different vocational fields, and a match between prerequisites of applicants and requirements of firms.

Even though the graduation rate at the upper-secondary education level has been increasing over the last decade (according to federal statistics from 85.6% in 1990 to 93.7% in 2010)[3], the share of *direct* entries after compulsory school into certifying education has decreased from 84 percent (year 1990) to 75 percent (year 2000, year 2010), mainly due to a decrease in *immediate* transitions into vocational education and training (VET).[4,5]

At the same time, the share of compulsory school graduates following a non-certifying intermediate school year has increased from 9% to 14%. As non-school-based activities are not captured by official statistics, there is no information on the activity followed by the rest of the youngsters. They might either follow non-school-based interim solutions (language stays, au pair work, traineeships) or, alternatively, do something that is not related to training or education, such as jobbing, being unemployed, travelling or staying at home.

The role of (school or practical based) interim solution programs is to offer a means of bridging the gap between lower and upper-secondary education by providing those with transition problems with more general education, first practical working experience, career orientation support, labor market relevant values or better language skills, depending on

---

[3] See the *key indicators on education* provided by the Swiss federal statistical office (FSO) at http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/17/blank/01.indicator.405101.4015.html, Abschlussquote auf der Sekundarstufe II.

[4] See the *key indicators on education* provided by the Swiss federal statistical office (FSO) at http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/17/blank/01.indicator.404301.4084.html, Verläufe und Übergänge - Übergang in die Sekundarstufe II.

[5] Note that all these figures are based on cross-section data, as it has not been possible in official statistics to longitudinally observe educational pathways (or pathways out of the educational system) at the individual level until recently. As of the year 2012, this will be possible because of the introduction of an individual identification number.

the concrete program (see Egger, Dreher & Partner AG, 2007, for a detailed description on the system of interim solutions). Programs can be scholastic or practical oriented, or a combination of the two. Many programs are administered by cantonal vocational training departments and provided by either public or private institutions. There are substantial differences between cantons with respect to the degree of coordination of these programs and with respect to the quantitative importance of such programs. There is a wide heterogeneity of programs, too, all with respect to content, target population, admission costs, and providers.

Large parts of programs are solely school based (so-called "10th school years") and provide participants with additional general education. Amongst them, some programs put higher focus on enhancing linguistic competence (especially for participants with migration background) and some programs are targeted towards preparatory schooling for a certain occupational field, eventually combined with practical pre-trainings within school or in firms.

These programs are sometimes fee based and mostly require early application. To find admission, applicants have to produce evidence of considerable past search effort and current motivation towards educational activities. Applicants with the least severe scholastic, personal, or familiar deficits are most likely to be chosen as participants (Egger et al., 2007); the youngsters most at risk, namely those lacking school motivation or those with the most problematic deficit structure, are unlikely to be reached by those programs. Apart from privately organized and financed solutions, alternative options are provided by so-called "pre-apprenticeships" or other practical-oriented preparatory courses. The latter can be started even after the beginning date of the new school year. They still require, however, the participation of firms and a certain basic motivation on the part of the applicant.

For compulsory school graduates who, by the end of compulsory schooling, have not yet found any bridging solution for the following year, there are low-threshold programs (so-called "motivation semesters") that are funded by the unemployment insurance office and organized by job placement institutions on behalf of social security services of municipalities. Apart from unemployment registration, there are hardly any admission requirements to be fulfilled.

The systemic function of a variety of these non-certifying interim solutions can be summarized as follows (Meyer, 2003):

- *Compensatory function*: To smooth away scholastic, linguistic, motivational, behavioral, or other deficits to enable participants to achieve success in the post-compulsory educational market and in the world of work

- *Orientation function*: To support young people in deciding on their further educational pathway and in their choice of profession

- *Buffer function*: To provide compulsory school graduates with a sort of daily structure if they cannot enter post-compulsory education due to shortages in apprenticeship openings or due to minimum age requirements, as for example in the health sector ("institutionalized waiting room").

The importance of the last-mentioned counter-cyclical absorber function is bolstered by official statistics that show fluctuations in the share of young people in interim solutions that are diametrical to the business cycle and to entries into apprenticeship trainings (FSO, 2007a,b). However, even in times of economic downturns, there are firms claiming to find it difficult to find adequate apprentices due to insufficient academic endowment or lack of maturity among applicants (LINK, 2009), implying that some sort of qualification mismatch or mismatch between supply and demand for specific vocational tracks is likely to exist. The different functions of interim solutions are thus not easy to disentangle. In many cases, compulsory school graduates who have not found an apprenticeship place in their desired occupational field prefer to wait a year instead of taking up a different apprenticeship (LINK, 2009). For some challenging apprenticeship tracks (e.g., IT-technicians, commercial employees), many firms explicitly demand a 10th school year from applicants who have only followed a low-level compulsory school track. The compensation function of interim solutions is thus not limited to supporting weak performers in getting any (possibly low-level) apprenticeship place, but presumably plays also an important role in enabling high motivated and well performing pupils of low-level compulsory schools to enter an apprenticeship at the highest aspiration level. However, it might be generally difficult to differentiate between the compensation function (which assumes accumulation of new human capital) and the pure signaling potential of interim solutions.

Even though the effectiveness of interim solution programs in enabling subsequent entry into certifying education has not yet been investigated, there are some case studies, implying success rates of programs of around 70 percent (Egger et al., 2007).

The issue of whether successful subsequent entries into certifying education should directly be attributed to interim solution programs depends on the hypothetical entry rate that would have occurred in the absence of such programs. For example, if only the waiting queue function was of importance, one might argue that there are alternative activities that could, during the year of waiting for an apprenticeship or school place, keep young people "off the streets" as well, such as unskilled work, possibly leading to equal chances to start a certifying education afterwards.

In order to assess whether interim solution programs meet their intention to causally enhance subsequent chances, we first compare upper-secondary entry rates of *interim solution program* participants with entry rates of those *without educational activity* during the gap year, taking into account the fact that the two groups might substantially differ from each other with respect to individual characteristics. The latter seems to be particularly important, because it is straightforward to expect that those who choose interim solution programs (instead of no educational activity) are a positively selected group with regard to ability, motivation, parental support, educational expectations, and other factors that might lead to higher success rates even without participating in programs.

A similar argument can be made for the second outcome of interest, the intellectual aspiration level of the certifying track taken up after the gap year. If gap years are about waiting for better apprenticeship places (prolonged search), then one should not expect differences in chances between the two gap year groups, all else being equal. In case of effective programs, however, those from interim solution programs should fare batter. As for the comparison to the group with a *direct entry*, one would expect higher aspiration levels for the interim solution group as well.

The third outcome under investigation, the certification rate by age 21, helps us to explore the long-term effects of having gap years after compulsory schooling.

4.3 Data and sample

We make use of the following information in the TREE data: first, we restrict the sample to individuals in low- or medium-level compulsory school tracks in ninth grade (tracks with basic requirements, tracks with extended requirements, and tracks without selection). By doing so, we drop those in (pre-)gymnasial tracks. In principle, this group just continues gymnasial schooling at upper-secondary level and does not belong to the sample at risk of having involuntary gap years. This is not to say that observed gap years in the remaining sample always are of involuntary nature. As described in section 4.2, gap years may occur due to failed applications at schools and firms, prolonged search, the decision to enhance chances in firms or schools through more schooling, and, finally, a lack of motivation to follow post-compulsory education. The data do not allow us to distinguish between the reasons for gap years in a clear-cut manner. However, the data do provide us with a lot of background information that helps us to predict activities after compulsory school. Hence, we restrict the sample to cases with non-missing information in important predicting variables (individual characteristics) and, further, to cases with non-missing information in the first outcome variables.

The information on the activity after compulsory school stems from the first TREE wave (t1), the wave in the year after the PISA 2000 survey. It refers to a point in time approximately nine months after the end of compulsory schooling. There are 3,311 cases for which we have valid background information and information on activities after compulsory schooling.

We distinguish between three groups, also denoted as *treatment groups*: the group with *direct entry into certifying education*, the group that follows some kind of *interim solution program* as described in section 4.2, and the *no education* group that does not follow any educational activity during this year. In the following, the latter groups are sometimes also referred to as the two gap year groups. While the group with no educational activity is easy to define, there is a wide range of heterogeneous programs pooled under the term *interim solutions*. In our data, approximately two-thirds in this group follow an additional (non-certifying) school year or a so-called motivational semester, and approximately 15 percent follow a pre-apprenticeship (Vorlehre) or an internship. The rest follows some kind of preparatory classes (Vorkurs), language studies (also au pairs) or other activities related to schooling or training. Table 4.1 shows the distribution of the activity in the first TREE wave: 76.56 percent of our sample are in certifying education, 20.45 percent in an interim solution, and 2.99 percent without educational activity.

Table 4.1: Outcomes by treatment groups

Treatment (t1)	N	%	Outcome 1 Entry by t2	Outcome 2 Low aspiration level	Outcome 3 Diploma by t6
Direct entry	2535	76.56%	100.00%	19.76%	96.22%
Interim solution	677	20.45%	72.23%	25.98%	89.04%
No education	99	2.99%	48.48%	48.94%	55.77%
Total sample	3311	100.00%	776	3'067	2'254

Note: the shadowed categories are the groups used for comparisons

In order to evaluate the effect of uneven transitions on educational outcomes, we start to analyze the effect of the two kinds of gap years on successful transition in the subsequent year (TREE wave t2). The success rate of entering certifying education in the second year after compulsory schooling (outcome 1) amounts to 72.23 percent among those coming from *interim solutions* and to 48.48 percent among those *without educational activity*. Because the *direct entry* group was successful by definition at the time of wave 2, this group is not used for comparison.

As for the second outcome of interest (outcome 2), we analyze whether we find disadvantages of lagged entries with respect to the intellectual aspiration level of the chosen certifying education. According to table 4.1, the most successful group, that is, the group with the lowest share of individuals in low-level upper-secondary tracks, is found among the group with a *direct entry* (19.76%), followed by the group coming from *interim solutions* (25.98%) and by the group coming from a gap year *without educational activity* (48.94%). Note that intellectual aspiration level refers to the year of entry into certifying education and thus either stems from wave t1 (direct entry) or t2 (lagged entry), depending on the starting time. The aspiration levels for 101 different vocational tracks were rated by an expert group of vocational advisers (for details and studies using this variable see Stalder, 2011).[6] For individuals following a full-time schooling variant of apprenticeship training, we assume the aspiration level of the equivalent firm-based apprenticeship training occupation. Gymnasium and other general education tracks are assumed to be high level.

The last outcome of interest (outcome 3) is whether or not someone has acquired a diploma[7] at the upper-secondary level six years after compulsory schooling (TREE wave t6).

6 We have imputed missing information (1.4%) about the aspiration level of less common tracks by regressing on training duration (years), on the amount of vocational schooling lectures (hours per year) and on an interaction term between the two. These factors strongly explain the aspiration level (R-squared of 84%).

7 There is a small minority of individuals that has not yet finished initial education, but is however enrolled in upper-secondary education and is near graduation. We regard this minority as successful, even though there remains some residual risk that some of these individuals might fail in the end.

This is around age 21 for most of the individuals, depending on their age at the time of the PISA test in 2000. If an individual had not acquired a diploma until spring 2006 and is not enrolled in certifying education at this point in time, he or she is regarded as not having acquired upper-secondary education. According to table 4.1, the graduation rate is 96.22 percent among those with a *direct entry*, 89.04 among those having followed an *interim solution program*, and 55.77 percent among those *without educational activity* in the year after compulsory schooling. As it also comes out of table 4.1, there is attrition in our sample between wave t1 and t6 of around 32 percent. We assume that attrition is ignorable.

The empirical strategy in the remainder of this chapter is to estimate the outcome differences between the three treatment groups that can be causally attributed to the activity at the time of t1. In order to build valid control groups, we make use of the following information measured in the ninth grade of compulsory schooling: for the *sociodemographic and family background characteristics,* we use gender, age, immigration status, family structure, highest achieved parental education, parental socioeconomic status (HISEI), and an index for the number of books at home. Information on *school performance and ability* is represented by the compulsory school track level, school marks in both regional language and mathematics, and, finally, PISA test scores in both reading literacy and mathematics/sciences. Apart from scholastic ability measures, the data also provides us with information on *personal attitudes and motivational context variables*: we use indices on school absenteeism[8], effort and perseverance, family support in school matters, and expected socioeconomic status at the age of 30. To capture *regional heterogeneity*, we use detailed information on the place of residence of the respondents, that is, dummies for so-called “greater regions”, a variable indicating urbanity, dummies for municipality size, and a variable with information on the cantonal supply of interim solution programs according to administrative data from the Swiss Federal Office. Detailed information on the creation of all these variables is presented in table 4.A1 in the appendix. For descriptive statistics and the distribution of characteristics by treatment groups, we refer to table 4.2 and its discussion in section 4.5.

[8] We do not know, however, the exact reasons for absenteeism at compulsory school; it might be caused, for example, by motivational deficits, family problems, or health problems.

4.4 Empirical strategy

4.4.1 The evaluation problem and propensity score matching

To estimate the causal impact of gap years on later educational outcomes, it is necessary to know what the outcome would have been if individuals had not chosen gap years. Because a person cannot be a participant and a non-participant in a gap year at the same time, we can never observe the counterfactual, namely the outcome that would have resulted if an individual had made an alternative choice. This is the so-called fundamental evaluation problem (see Heckman et al., 1999). The potential-outcome approach to causality suggested by Roy (1951) and Rubin (1974) provides the standard analytical framework to address this issue.

Assuming for now that we only face two groups (e.g., gap years versus certifying education), the evaluation problem can be exposed and solved as follows: Let D be the binary indicator describing treatment status with $D = 1$ for participants and $D = 0$ for non-participants. Y_1 and Y_0 denote the potential outcomes for participants and non-participants in both hypothetical states.

As we only observe the outcome for each individual in one regime (Y_1 if $D = 1$ and Y_0 if $D = 0$), the observed outcomes can be written in terms of potential outcomes as

$$Y = DY_1 + (1 - D)Y_0 = Y_0 + D(Y_1 - Y_0) \tag{4.1}$$

The causal effect of treatment D is defined by the difference of the two potential outcomes $\Delta = Y_1 - Y_0$. However, Δ can never be known, because the counterfactual individual outcome (Y_1 if $D = 0$ and Y_0 if $D = 1$) is unobservable.

Nevertheless, under certain assumptions that will be discussed further below, important parameters can be estimated at the population level, that is the *average treatment effect on the treated (ATT)*, the *average treatment effect on the untreated (ATU)* and the *average treatment effect (ATE)*. They can be written as follows:

$$\Delta_{ATT} = E(Y_1 - Y_0|D = 1) \quad = E(Y_1|D = 1) - E(Y_0|D = 1) \tag{4.2}$$

$$\Delta_{ATU} = E(Y_1 - Y_0|D = 0) \quad = E(Y_1|D = 0) - E(Y_0|D = 0) \tag{4.3}$$

$$\Delta_{ATE} = E(Y_1 - Y_0) \quad = ATT \times P(D = 1) + ATU \times P(D = 0) \tag{4.4}$$

If treatment effects are heterogeneous, the mean impact of attending the program on those who attend the program (ATT) might differ from the mean impact of attending the program on those persons not attending the program (ATU) and from the mean impact of attending the program on a randomly selected person in the population (ATE).

The parameter of interest in evaluation studies is most often the ATT. While $E(Y_1|D=1)$ in (4.2) can be estimated by the sample mean of Y_1 in the subsample of participants, the counterfactual mean $E(Y_0|D=1)$ remains unknown. The true parameter Δ_{ATT} is thus only identified if we can substitute the unobserved expected counterfactual outcome of participants $E(Y_0|D=1)$ by the observed mean outcomes of non-participants $E(Y_0|D=0)$, that is if $E(Y_1|D=1) = E(Y_0|D=0)$.[9] In a natural experiment, this is ensured because the assignment of D is random. Therefore, the two groups should not systematically differ in pre-treatment characteristics and non-participants ($D=0$) can be used to estimate $E(Y_0|D=1)$. However, there is no random assignment in our case and participants are likely to differ in a systematic manner from non-participants (selection bias). We cannot estimate Δ_{ATT}, Δ_{ATE}, or Δ_{ATU} without imposing assumptions.

One possible *identifying assumption* to solve the identification problem[10] is the *conditional independence assumption (CIA)* proposed by Rubin (1977). The CIA states that, given a set of observable covariates X that are not affected by treatment D, potential outcomes are independent of treatment assignments:

$$Y_0, Y_1 \perp D|X \tag{4.5}$$

This way of identification is at the core of the matching estimators as well as of the OLS-regression. While the OLS-regression is based on parametric assumptions, the idea of matching is to pair each program participant with an observably similar non-participant and interpret the difference in their outcomes as the effect of the program.

The *CIA* assumption[11], also referred to as *unconfoundedness*, *ignorable treatment assign-*

9 This applies analogously for identifying Δ_{ATU}, where we need $E(Y_1|D=1) = E(Y_1|D=0)$ because $E(Y_1|D=0)$ is unobservable. To identify Δ_{ATE}, we need both.

10 Heckman et al. (1999) and Angrist and Krueger (1999) provide an overview on a wide set of available identification and estimation strategies.

11 In order to identify the Δ_{ATT} in equation (4.2), the weaker assumption that participation D is independent of the non-treatment outcome is sufficient, that is $Y^0 \perp D|X$. The possibility that self-selection depends on the treated outcome Y^1 does not have to be ruled out. Symmetrically, to identify Δ_{ATU} it is sufficient to assume $Y^1 \perp D|X$. The additional challenge when estimating the Δ_{ATE} in equation (4.4) is thus that both counterfactual outcomes $E(Y^0|D=1)$ and $E(Y^1|D=0)$ have to be constructed by assuming (4.5).

ment, or *selection on observables* assumption, therefore implies that the selection bias can be removed by conditioning on observable characteristics X. To justify the CIA assumption that underlies this procedure, it is crucial to identify and observe all variables that could be mutually correlated with assignment and potential outcomes. Naturally, the plausibility of this assumption heavily relies on the quality and richness of the data. As our data described in section 4.3 contains an unusual large and informative set of information, we assume that CIA holds in our case.

Under CIA, the difference and hence the source of the bias between the treated and the control group can be removed by constructing a comparison group that is as similar as possible to the treatment group based on statistical matching on covariates or by applying regression techniques. The latter, however, imposes functional form assumptions that might not be adequate. In turn, matching on all covariates of a high dimensional vector X becomes problematic, as it might be impossible to find identical control-group individuals with respect to all characteristics (the so-called curse of dimensionality).

A solution to the *dimensionality problem* is the use of the *propensity score (PS)* as a balancing score. The propensity score $P(X) = P(D = 1|X)$ is defined as the probability $0 < P(X) < 1$ of participating in a treatment given characteristics X. Rosenbaum and Rubin (1983) show that if potential outcomes are independent of treatment conditional on covariates X (that is if the CIA holds), they are also independent of treatment conditional on a balancing score $b(X)$.

The *CIA* based on the propensity score $P(X)$ is the key assumption of the propensity score matching technic and can be written as

$$Y_0, Y_1 \perp D|P(X) \tag{4.6}$$

A further requirement to perform propensity score matching is the *common support* or *overlap* condition, which can be expressed as

$$0 < P(D = 1|X) < 1 \tag{4.7}$$

The *common support* condition[12] states that persons with the same X values must have a positive probability of being both participant and non-participant (Heckman et al., 1999). If

[12] This condition reduces to $P(D = 1|X) < 1$ for Δ_{ATT} and to $0 < P(D = 1|X)$ for Δ_{ATT}.

there are X for which everyone gets the treatment, the matching procedure fails to construct the counterfactual outcome for these individuals. While both linear regression and matching methods rely on the CIA assumption, the common support assumption is specific only to matching methods; linear regression extrapolates empty cells by functional form assumptions.

To estimate the propensity score, any discrete choice model such as the logit or probit model can be used. When estimating the variance of the treatment effect, one should address the fact that the propensity score represents an estimated rather than a known value. A typical solution that will be followed in our analysis is the usage of bootstrapping methods.

4.4.2 Extension to the multiple treatment case

As described in section 4.3, the individuals in our study face three different alternatives after compulsory schooling, namely 1) direct entry into certifying education, 2) institutionalized interim solutions, and 3) unstructured gap years. Our aim is to estimate the causal effect of 2) versus 3) on successful entry into certifying education the year after the gap year, but also the causal effect of 2) or 3) versus 1) on avoiding low education level tracks and on later educational outcomes. In short, our individuals face a setting of multiple treatments[13].

According to Lechner (2001) and Imbens (2000), the major properties shown by Rubin (1977) and Rosenbaum and Rubin (1983) discussed above can be generalized to the multiple treatment case. In the notation of Lechner (1999), the two-state case can be extended as follows:

The outcomes of $(M+1)$ mutually exclusive treatments are denoted by $\{Y_0, Y_1, ..., Y_M\}$. It is assumed that each participant receives exactly one of the treatments. Therefore, for any participant, only one component of $\{Y_0, Y_1, ..., Y_M\}$ can be observed in the data, and the remaining M outcomes are counterfactuals. Participation in a particular treatment m is indicated by the variable $S \in \{0, 1, ...M\}$. The focus now lies on the pair-wise comparison of the effects of treatments m and l. The average treatment effects parameters are

$$\Delta^{ATE}_{m,l*} = E(Y_m - Y_l) \qquad = E(Y_m - E(Y_l) \tag{4.8}$$

$$\Delta^{ATE}_{m,l} = E(Y_m - Y_l|S = m, l) \qquad = E(Y_m|S = m, l) - E(Y_l|S = m, l) \tag{4.9}$$

$$\Delta^{ATT}_{m,l} = E(Y_m - Y_l|S = m) \qquad = E(Y_m|S = m) - E(Y_l|S = m) \tag{4.10}$$

[13] In line with the literature, all options in the choice set of individuals are called 'treatment' in the following.

where $\Delta^{ATE}_{m,l*}$ denotes the expected effect of treatment m relative to treatment l for a participant randomly drawn from the population and, similarly, $\Delta^{ATE}_{m,l}$ denotes the same effect for a participant randomly selected from the group of participants in either m or l. Both average treatment effects are symmetric in the sense that $\Delta^{ATE}_{m,l} = -\Delta^{ATE}_{l,m}$. Further, $\Delta^{ATT}_{m,l}$ is the expected effect for an individual randomly drawn from the population of participants in treatment m only. If participants in treatments m and l differ in a way that is related to the distribution of X, and if the treatment effects vary with X, then $\Delta^{ATT}_{m,l} \neq -\Delta^{ATT}_{l,m}$. Since $\Delta^{ATU}_{m,l} = \Delta^{ATT}_{l,m}$ it follows that $\Delta^{ATE}_{m,l}$ is a weighted combination of $\Delta^{ATT}_{m,l}$ and $\Delta^{ATT}_{l,m}$, where weights are given by the participation probabilities in the respective states m and l.

Imbens (2000) and Lechner (2001) show that all these parameters can be identified under the extended CIA assumption (and a generalization of the balancing score property), because it identifies $E(Y_l|S = m)$ for all combinations of l and m. The CIA for the case of multiple treatments is

$$Y_0, Y_1, ..., Y_M \perp S|P(X) \tag{4.11}$$

The common support condition further extends to the requirement that all individuals actually have the possibility of participating in all states such that all participants in a treatment have counterparts in the other groups.

Lechner (2001) also shows, that for the identification of $\Delta^{ATE}_{m,l}$, $\Delta^{ATT}_{m,l}$ and $\Delta^{ATT}_{l,m}$ the CIA is only necessary to hold in the subsample of participants in treatments m and l. This implies that only this subsample is necessary for the empirical analysis. Thus, when we are interested in comparing two programs for participants in one of those two, the existence of multiple treatments can in fact be ignored, since individuals who do not take part in either program are not needed for identification.

There are several alternative ways to estimate the propensity scores within this framework (Lechner, 2001, 2002). One of them is to use a multinomial probit or logit model including all alternatives, where $P_{l|ml}(X)$ can be derived from $P_l(X)/[P_l(X)+P_m(X)]$. As multinomial probit models are computational burdensome and multinomial logit models are based on the IIA-assumption, an alternative way is to estimate a series of $M(M-1)/2$ different binary choice models. Lechner (2002) finds only little difference in the relative performance of these estimation strategies. As we only consider a small amount of alternative activities (treatments) in this study, we will perform separate binary logit models to estimate our balancing scores. Our aim is to estimate $\Delta_{m,l}^{ATE}$ in equation (4.9) and $\Delta_{m,l}^{ATT}$ in equation (4.10) for all combinations of m and l.

As the multiple treatment case can be reduced to several estimations of binary treatments, we go back to our original notation in the following.

4.4.3 The matching procedure

a) Matching algorithms

Exact matching on the propensity score is difficult in practice, as there might not be many individuals in the comparison group who show exactly the same propensity score as individuals in the treatment group. Therefore, the objective is to match units that are sufficiently close to each other. A typical matching estimator for the ATT takes the following general form (see Smith and Todd, 2005)

$$ATT^M = \frac{1}{n_i} \sum_{i \in I_1 \cap S_P} \left(Y_{i1} - \sum_{j \in I_0} w_{ij} Y_{j0} \right) \tag{4.12}$$

where I_1 denotes the set of program participants, I_0 the set of non-participants, S_P the region of common support (see further below for ways of constructing this set), and n_1 the number of persons in the set $I_1 \cap S_P$. The match for each participant $i \in I_1 \cap S_P$ is constructed as a weighted average over the outcomes of non-participants, where the weights w_{ij} depend on the distance between the propensity scores $P_i(X)$ and $P_j(X)$.

For each individual i in the sample of participants, there has to be defined a neighborhood $C(P_i(X))$. Neighbors who are matched to i are non-participants $j \in I_0 | P_j(X) \in C(P_i(X))$. Alternative matching estimators differ in how the neighborhood is defined and in how the weights w_{ij} are constructed (see Heckman et al., 1997, 1998; Dehejia and Wahba, 2002; Heckman et al., 1998; Smith and Todd, 2005; Caliendo and Kopeinig, 2008). All matching estimators face a trade-off between bias and variance of the estimation; in small samples, the choice of matching algorithm can be important (Heckman et al., 1997). We therefore use and compare the results of (variants) of the following matching algorithms:

Nearest neighbor matching (NNM) is the traditional pairwise matching without replacement, where non-participants with the value of $P_j(X)$ that is closest to $P_i(X)$ are matched to each treated individual, that is $C(P_i(X)) = \min_j \|P_i(X) - P_j(X)\|, j \in I_0$. We use a variant of this algorithm that matches the five nearest neighbors (5NNM) with replacement within the region of common support. Each nearest neighbor receives equal weight in constructing the counterfactual mean outcome. By allowing for replacement, we increase the average quality of the matches, as we face rather small samples in some of the estimations. In turn, allowing for five instead of only one nearest neighbor may increase the bias due to poorer matches on average, but it also uses more information to construct the counterfactual for each participant and thus enhances the precision of the estimation.

If the closest neighbors are far away, nearest neighbor matching faces the risk of bad matches. This can be avoided by imposing a tolerance level on the maximum propensity score distance, the so-called *caliper*. In *caliper matching*, matches for individual i are then selected only if $C(P_i(X)) = \{P_j(X) | \|P_i(X) - P_j(X)\| < \epsilon\}, j \in I_0$, where ϵ stands for the maximum difference that we allow between the propensity scores. A variant of caliper matching is the *radius matching* suggested by Dehejia and Wahba (2002), which not only uses the nearest n neighbors within the defined caliper, but all possible matches within the specified radius. This algorithm allows an oversampling of good matches (if available) and disregards bad matches. A possible drawback is the arbitrary nature of the chosen caliper (Smith and Todd, 2005). We will show results for radius matching with caliper 0.02.

In contrast to the discussed estimators, the *kernel matching* algorithm does not only consider some of the observations of the comparison group; it uses a weighted average of *all* individuals of the comparison group to construct the counterfactual outcome. The weights depend on the distance between the treaded individual and each observation from the control group. Higher weight is placed on persons close in terms of X and lower weight on more distant individuals. The weights of the kernel matching estimator take the form

$$w_{ij} = \frac{G(\frac{P_j(X)-P_i(X)}{a_n})}{\sum_{k \in I_0} G(\frac{P_k(X)-P_i(X)}{a_n})}, \tag{4.13}$$

where $G(\cdot)$ is a kernel function and a_n is a bandwidth parameter[14]. Nonzero values of this weight implicitly define $C(P_i(X))$ for this version of matching. This algorithm achieves a higher precision because more information is used than just the nearest neighbors; as all comparison group observations are used to estimate the missing counterfactual outcome, a proper implementation of the common support condition is very important for bias reduction.

[14] We use an epanechnikov kernel with a bandwidth of 0.06, which is the default for kernel matching in the Stata application psmatch2.

b) Estimating treatment effects

In general, the propensity score matching estimator is simply the mean difference in outcomes of the treatment and control group over the common support, weighted by the weights given to the matched observations among the non-treated (to estimate the ATT) or among the treated (to estimate the ATU).

c) The common support condition

As discussed in section 4.4.1, the various treatment effects are only defined in the region of common support. Sufficient overlap is important to ensure that we compare comparable individuals. While for the estimation of the ATT it is sufficient to have potential matches in the control group, the estimation of the ATE additionally requires that the characteristics of the comparison group are also observed in the treatment group. If there are ranges without overlap and with a large number of observations off support, the estimated effects might only be viewed as relating to the sub-population within support. However, holes in the support (e.g., in the region close to 1) do not necessarily always constitute an asymptotic support (identification) problem, but might in practice also arise from a small sample size (Lechner, 2008). From a small sample argumentation, it is not beforehand clear whether it is better to allow for some mismatch or whether to adjust the support: Lechner (2008) notes that both may be misleading, ignoring the support problem or estimating treatment effects only within the support region. Lechner (2008) provides an approach to derive bounds for the true effect. There are different possibilities to check and impose the common support criterion (see Caliendo and Kopeinig, 2008). After a visual analysis of the density distribution of the propensity scores in the groups, the approach we follow is the *Minima and Maxima comparison*. We ignore all observations with propensity scores smaller than the minimum and larger than the maximum in the opposite group. This might be overrestrictive, if there are many observations very close to the bounds that are discarded. We face this situation in some of our estimations. For comparison to the basic version, we have also performed estimations that use a caliper criterion instead of the minima-maxima comparison around the bounds. Results did not change substantially by doing so (not presented).

Generally, it can be seen as an advantage of matching techniques that they explicitly highlight the common support problem, which would also be present in parametric estimations but overlooked due to functional form assumptions.

d) Testing the balancing between treatment and controls

The aim of propensity core matching is to construct a comparison group, that is, to balance the distribution of X in the treatment and control group by reweighting observations. Since we match observations on the basis of similar (but not necessarily identical) propensity scores, the quality of the matching procedure has to be checked by testing the balancing of the covariates after matching. There are several indicators applied in the literature to check the matching quality (see Caliendo and Kopeinig, 2008; Leuven and Sianesi, 2003).

An indicator called *standardized bias (SB)* has been suggested by Rosenbaum and Rubin (1985) and is defined for each variable X as "the difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups". The standardized bias before matching (SB_U) and after matching (SB_M) is computed as[15]

$$SB_U = \frac{100 \times \left(\bar{X}_{1U} - \bar{X}_{0U}\right)}{\sqrt{0.5 \times \left(V_{1U}(X) + V_{0U}(X)\right)}},\ SB_M = \frac{100 \times \left(\bar{X}_{1M} - \bar{X}_{0M}\right)}{\sqrt{0.5 \times \left(V_{1U}(X) + V_{0U}(X)\right)}} \tag{4.14}$$

There is no formal criteria in the literature for when a standardized bias is too large. Rosenbaum and Rubin (1985) suggest that a value of 20 is "large". In most empirical studies, a mean bias reduction below 3% or 5% after matching is seen as sufficient.

Another procedure is to use *t-test* for equality of covariate means in the treated and non-treated group, both before and after matching (Rosenbaum and Rubin, 1985). This check for significant differences is implemented by a regression of each variable on the treatment indicator. After matching, no significant differences should be found.

A similar approach (Leuven and Sianesi, 2003; Sianesi, 2004) is to estimate the propensity score for the treatment decision on all the variables before and after matching and then to compare both *Pseudo-R2* and p-values of the likelihood-ratio test for *joint significance* of all the regressors. After matching, the Pseudo-R2 should be very low and the regressors should not be significant for the treatment decision anymore.

Due to the dependence of some balancing tests on sample size (t-test, joint significance) and thus the higher probability to pass a test in small samples, we use all the described balancing tests in combination, but mostly lean on evaluating the standardized bias.

[15] Holding the variability constant, the difference between SB_U and SB_M represents the bias reduction, see http://www.stata.com/statalist/archive/2012-03/msg01111.html.

4.5 Estimation: The matching procedure

This section presents the matching procedure. We pairwise compare and match individuals of three groups to each other: the group without educational activity after compulsory schooling (denoted as *N.E.* for *no education*), the group in interim solutions (denoted as *I.S.* for *interim solution*), and the group with direct entry into certifying education (denoted as *D.E.* for *direct entry*). In subsection 4.5.1, we investigate the distribution of pre-treatment characteristics across groups. Subsection 4.5.2 presents the propensity score estimation. Subsection 4.5.3 discusses the balancing of covariates and common support issues after matching.

4.5.1 Distribution of characteristics before matching

Before performing the match, it is important to evaluate how the groups of interest initially differ from each other. Table 4.2 shows variable means, pairwise t-test results for equality of means across groups, and standardized biases for all variables before matching. Table 4.2 reveals considerable biases and thus unequal distributions for most of the variables across groups. Even though t-test results are not very reliable (small sample), we note that, in our case, they always point to a significant mean difference when the bias is most severe (>20).

The comparison of the *interim solution* group and the *direct entry* group reveals that individuals in interim solutions are (significantly) more often female, first generation immigrants, and older than the standard age (which could be due to school year repetitions in the past) than those with a direct entry. Their family background is less favorable, all in terms of parental education, family structure, socioeconomic status and the number of books at home. Further, they exhibit a less successful educational history: they more often come from low-level compulsory school tracks, they have lower school marks in the official regional language and mathematics, and they show inferior performance in the PISA tests in reading and mathematics/sciences. Individuals in the interim solution group on average express lower expectations with regard to their own socioeconomic status at age 30. There is also a higher share of individuals with no clear idea about their occupational aim compared to the group with a direct entry. Further, there are pronounced regional differences across the two groups that might reflect both institutional and regional effects. For example, the cantonal supply of interim solutions is higher among the group in interim solutions than among the group with direct entry (and also higher than in the group without educational activity).

Table 4.2: Distribution of characteristics between the groups (before matching)

Variables (at PISA 2000)	N.E. (No education)	I.S. (Interim Solution)	D.E. (Direct entry)	N.E.(1) vs. I.S.(0) comparison		I.S.(1) vs. D.E.(0) comparison		N.E.(1) vs. D.E.(0) comparison	
	mean	mean	mean	t-test	%bias	t-test	%bias	t-test	%bias
Female	0.495	0.722	0.496	−4.64***	-47.8	10.69***	47.7	−0.02	-0.2
Swiss	0.616	0.765	0.785	−3.20***	-32.6	−1.11	-4.8	−3.98***	-37.4
2nd generation immigrant	0.121	0.089	0.104	1.04	10.6	−1.19	-5.3	0.54	5.4
1st generation immigrant	0.263	0.146	0.111	2.96 **	29.1	2.53*	10.6	4.63***	39.6
>= Age 16	0.354	0.297	0.213	1.14	12.1	4.64***	19.4	3.34***	31.6
Parental education: compulsory	0.485	0.371	0.253	2.18*	23.1	6.12***	25.6	5.18***	49.4
Parental education: secondary II	0.242	0.322	0.376	−1.60	-17.7	−2.59 **	-11.3	−2.70 **	-29.1
Parental education: tertiary	0.263	0.273	0.340	−0.22	-2.4	−3.32***	-14.6	−1.61	-17.0
Parental educ: missing	0.010	0.034	0.031	−1.28	-16.3	0.42	1.8	−1.18	-14.6
Nuclear family structure	0.556	0.752	0.795	−4.13***	-42.0	−2.45*	-10.4	−5.74***	-52.8
Socioeconomic status (Index)	41.143	44.691	47.833	−2.28*	-24.8	−4.71***	-20.8	−4.18***	-44.9
Number of books at home (Index)	3.768	4.375	4.505	−3.91***	-41.1	−2.00*	-8.8	−4.76***	-48.5
Level compulsory school: medium	0.455	0.439	0.549	0.30	3.2	−5.13***	-22.2	−1.85+	-18.9
Level compulsory school: low	0.455	0.487	0.350	−0.61	-6.6	6.58***	28.1	2.13*	21.3
Level compulsory school: no select.	0.091	0.074	0.101	0.60	6.2	−2.11*	-9.5	−0.31	-3.3
Mark test language: insufficient	0.061	0.028	0.031	1.71+	15.8	−0.42	-1.8	1.63	14.1
Mark test language: sufficient	0.384	0.285	0.339	2.01*	21.0	−2.65 **	-11.6	0.93	9.3
Mark test language: good	0.465	0.628	0.596	−3.12 **	-33.1	1.50	6.5	−2.61 **	-26.5
Mark test language: missing	0.091	0.059	0.034	1.22	12.1	3.00 **	12.0	2.99 **	23.6
Mark mathematics: insufficient	0.121	0.087	0.064	1.10	11.1	2.16*	9.0	2.28*	20.0
Mark mathematics: sufficient	0.354	0.353	0.351	0.01	0.1	0.07	0.3	0.04	0.4
Mark mathematics: good	0.444	0.502	0.548	−1.07	-11.6	−2.12*	-9.2	−2.03*	-20.8
Mark mathematics: missing	0.081	0.058	0.037	0.90	9.1	2.38*	9.7	2.21*	18.6
Reading literacy PISA: 0	0.111	0.062	0.027	1.81+	17.5	4.49***	17.1	4.84***	33.6
Reading literacy PISA: 1	0.242	0.145	0.104	2.50*	24.8	3.00 **	12.4	4.36***	37.2
Reading literacy PISA: 2	0.283	0.287	0.243	−0.08	-0.8	2.34*	10.0	0.91	9.1
Reading literacy PISA: 3	0.253	0.335	0.373	−1.64	-18.2	−1.82+	-7.9	−2.44*	-26.2
Reading literacy PISA: 4	0.081	0.149	0.212	−1.83+	-21.5	−3.66***	-16.4	−3.17 **	-37.8
Reading literacy PISA: 5	0.030	0.022	0.041	0.50	5.1	−2.35*	-11.0	−0.55	-6.0
Math/science literacy: Q1	0.434	0.355	0.238	1.54	16.3	6.16***	25.7	4.47***	42.4
Math/science literacy: Q2	0.273	0.269	0.258	0.08	0.9	0.59	2.6	0.34	3.4
Math/science literacy: Q3	0.131	0.165	0.227	−0.86	-9.6	−3.47***	-15.5	−2.24*	-25.1
Math/science literacy: Q4	0.040	0.083	0.162	−1.47	-17.6	−5.21***	-24.3	−3.26***	-41.0
Math/science literacy: missing	0.121	0.129	0.116	−0.20	-2.2	0.90	3.8	0.16	1.6
Absenteeism (Index)	1.579	1.337	1.338	4.43***	43.9	−0.06	-0.3	4.46***	42.6
Family support (Index)	−0.176	0.080	0.054	−2.29*	-25.1	0.62	2.6	−2.37*	-23.7
Effort and perseverance (Index)	2.590	2.714	2.721	−1.87+	-20.3	−0.25	-1.1	−1.98*	-20.9
Exp. socioec. status age30: miss	0.232	0.273	0.225	−0.86	-9.4	2.64 **	11.2	0.17	1.8
Exp. socioec. status age30 (Index)	−0.445	−0.366	−0.172	−1.02	-10.9	−5.53***	-25.0	−3.20***	-34.9
Occupational aim: nursing	0.020	0.086	0.021	−2.28*	-29.5	8.37***	29.4	−0.02	-0.2
Greater region: Region lémanique	0.293	0.189	0.179	2.41*	24.4	0.60	2.6	2.87 **	27.0
Greater region: Espace Mittelland	0.182	0.290	0.195	−2.24*	-25.5	5.32***	22.1	−0.33	-3.4
Greater region: Nordwestschweiz	0.111	0.099	0.057	0.37	3.9	3.94***	15.8	2.25*	19.6
Greater region: Zurich	0.091	0.106	0.078	−0.47	-5.2	2.35*	9.8	0.46	4.6
Greater region: Ostschweiz	0.232	0.211	0.190	0.48	5.1	1.23	5.3	1.05	10.3
Greater region: Zentralschweiz	0.020	0.056	0.046	−1.51	-18.8	1.08	4.5	−1.22	-14.5
Greater region: Ticino	0.071	0.049	0.254	0.92	9.2	−11.90***	-59.9	−4.17***	-51.3
Municipality type: Rural	0.303	0.375	0.359	−1.39	-15.2	0.78	3.4	−1.14	-11.9
Location: village (less 3000)	0.071	0.179	0.130	−2.71 **	-33.1	3.23***	13.5	−1.74+	-19.8
Location: small town (3000-15000)	0.576	0.521	0.589	1.01	10.9	−3.18***	-13.7	−0.27	-2.7
Location: town (15000-100000)	0.192	0.211	0.203	−0.44	-4.8	0.48	2.1	−0.26	-2.7
Location: city (100000-1000000)	0.162	0.089	0.078	2.29*	22.1	0.93	4.0	3.01 **	26.0
Cantonal share interim solutions	15.196	17.082	12.677	−1.92+	-22.1	11.67***	49.2	2.89 **	31.0
N	99	677	2535	776		3212		2634	
Mean of absolute bias					16.9		14.4		22.0

+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001

Comparing the group *no education* to the group *direct entry*, we find differences that are similar to the ones discussed above, but even more pronounced for most variables. The overall picture therefore implies that individuals with direct entry exhibit the most favorable characteristics, and individuals without educational activity the least favorable ones. One notable exception to this simplification is the gender variable: Females are only overrepresented in interim solutions. This might be (at least partly) related to the occupational aim of "nursing", where we find the same picture.[16] It is important to remember, however, that the differences under discussion are raw differences and not effects from multivariate regressions.

Comparing the groups *interim solution* and *no education* directly, it shows that individuals with no educational activity not only have less favorable family background and less successful educational measures (especially language skills), but also differ with respect to personal attitudes related to motivational factors. They have—according to their self-assessment one year before—a significantly higher tendency towards school absenteeism, lower levels of effort/perseverance and a lower level of perceived family support in school matters. There are no such differences between those in interim solutions and those in certifying educations.

Overall, the distribution of the variables before matching is more biased between the groups *no education* and *interim solution* than between *interim solution* and *direct entry.* The largest overall bias is found between the groups *no education* and *direct entry.*

4.5.2 Estimation of the propensity score

In order to match observations of different groups on the basis of their similarity, we need to estimate the propensity score.[17] We use all the pre-treatment characteristics described in section 4.5.1 for the basic specification of the treatment decision. The average marginal effects resulting from pairwise probit and multinomial logit estimations are presented in table 4.3. Note that we use the pairwise binary variant for our final propensity score estimation.

The set of variables has been chosen on the basis of specification search (reducing the absolute bias in the balancing of covariates) and on the basis of theory and former empirical results on what could simultaneously affect the treatment decision and the outcomes. As it

[16] In the interim solution group, there is a higher share of individuals who want to pursue a career in the field of nursing. At the time of PISA 2000 (until the year 2003), the vocational training towards nursing or similar occupations had still required a minimum age of 18, which institutionally forced youngsters with career aspirations in the health sector (mostly women) into interim solutions.

[17] We have also tried alternative balancing scores such as the linear index or the logit of the propensity score. The quality of the match sometimes improved, but mainly due to more observations off support.

is not the goal of a propensity model to maximize the prediction of the treatment decision, but rather to reduce the bias in the estimated treatment effect, it is important to include variables that are assumed to be related to the outcome, even though their relation to the treatment decision might only be weak.[18]

For some of the variables in table 4.3 the estimation results lack statistical significance in all models. First, migration background has no effect on the treatment decision after accounting for other characteristics. This might be due to its correlation with other variables in the model, namely family background characteristics, educational history, or language skills. As the point estimate is not exactly zero and a potential effect on outcomes cannot be ruled out, we do not skip this information out of the model. The same holds for the indices on family support and effort/perseverance, for school marks in the test language (which might be correlated with the PISA reading literacy test results, amongst others) and, finally, for PISA test results in mathematics/sciences (which might be correlated with school marks in mathematics, amongst others). The lack of significance is, in some of the cases, likely to be related to the rather small sample size; skipping those variables could therefore introduce potential bias in the treatment effects to be estimated. The following multivariate results in table 4.3 are particularly notable.

The decision between *no education* and *interim solution* seems not to be driven by scholastic ability. There is no significant effect of school marks or PISA test scores. Having followed a low-level compulsory school track even increases the probability of interim solutions. However, the probability to have no educational activity is significantly related to poorer family background, especially to non-nuclear family structure, lower socioeconomic index and less books at home. Further, a higher tendency towards school absenteeism also increases the probability of not following any activity related to education. The results for the decision between *no education* and *direct entry* look quite similar. However, there is no gender effect, no significant effect of the number of books at home, but a significant effect of low parental education.

In turn, the decision between *interim solution* and *direct entry* is significantly affected by inferior compulsory schooling outcomes (school marks in mathematics, low-level compulsory school track, lower levels of PISA reading literacy test scores) as well as poorer family back-

[18] In turn, the inclusion of covariates that only affect the treatment decision does not improve the subsequent estimation of the treatment effect, but rather only increases the variance of the estimation or exacerbates the support problem (for a discussion and references see Caliendo and Kopeinig, 2008).

ground (low parental education, non-nuclear family structure). Interestingly, the number of books at home positively affects the probability to choose interim solutions. While career aspirations itself do not have a significant effect, there is weak evidence that higher uncertainty levels regarding career choice increases the probability to follow interim solutions. Additionally, sizeable effects on the choice of interim solutions are found for local characteristics and for the supply of respective programs.

Table 4.3: Treatment decision and propensity score estimation

	(1)	(2)	(3)	(4)	(5)	(6)
	Probit	Probit	Probit		Mlogit	
Variables (at PISA 2000)	N.E.(1) vs. I.S.(0)	I.S.(1) vs. D.E.(0)	N.E.(1) vs. D.E.(0)	N.E.	I.S.	D.E.
Female	−0.091***	0.130***	0.003	−0.005	0.133***	−0.128***
2nd generation immigrant	0.008	−0.020	0.000	0.000	−0.015	0.016
1st generation immigrant	0.004	0.024	0.008	0.006	0.023	−0.029
>= Age 16	0.008	0.011	0.009	0.006	0.009	−0.015
Parental education: compulsory	0.036	0.046**	0.022*	0.015*	0.038*	−0.054**
Parental education: tertiary	0.044	0.006	0.015	0.011	0.001	−0.012
Parental educ: missing	−0.073	0.018	−0.016	−0.018	0.013	0.005
Nuclear family structure	−0.087***	−0.035*	−0.039***	−0.030***	−0.028+	0.058***
Socioeconomic status (Index)	−0.002+	−0.001	−0.001+	−0.000+	0.000	0.001
Number of books at home (Index)	−0.022*	0.016**	−0.003	−0.003	0.015**	−0.012*
Level compulsory school: low	−0.055*	0.074***	−0.007	−0.008	0.076***	−0.068***
Level compulsory school: no selec.	−0.031	−0.056*	−0.006	−0.004	−0.052+	0.056*
Mark test language: insufficient	−0.046	−0.011	−0.007	−0.007	−0.010	0.017
Mark test language: good	−0.031	−0.018	−0.009	−0.006	−0.016	0.022
Mark test language: missing	0.111	0.088	0.058	0.044	0.076	−0.120
Mark mathematics: insufficient	0.025	0.043	0.004	0.005	0.034	−0.039
Mark mathematics: good	0.008	−0.051**	−0.007	−0.003	−0.047**	0.051**
Mark mathematics: missing	−0.084	−0.108	−0.048	−0.037	−0.100	0.137
Reading literacy PISA: 0	0.018	0.081*	0.019	0.014	0.068+	−0.083*
Reading literacy PISA: 1	0.026	0.018	0.015	0.012	0.014	−0.025
Reading literacy PISA: 2	0.013	0.015	0.003	0.004	0.009	−0.013
Reading literacy PISA: 4	0.000	−0.025	−0.007	−0.005	−0.021	0.026
Reading literacy PISA: 5	0.071	−0.035	0.014	0.020	−0.042	0.023
Math/science literacy: Q1	0.037	0.023	0.011	0.009	0.019	−0.028
Math/science literacy: Q2	0.030	0.020	0.013	0.011	0.016	−0.028
Math/science literacy: Q4	−0.025	−0.017	−0.020	−0.016	−0.015	0.031
Math/science literacy: missing	−0.005	0.024	0.005	0.004	0.019	−0.023
Absenteeism (Index)	0.046*	0.007	0.016*	0.012*	0.001	−0.013
Family support (Index)	−0.006	0.005	−0.001	−0.001	0.005	−0.004
Effort and perseverance (Index)	−0.009	−0.007	−0.004	−0.003	−0.005	0.008
Exp. socioec. status age30: miss	−0.027	0.030+	0.001	0.001	0.028+	−0.030+
Exp. socioec. status age30 (Index)	−0.021	−0.005	−0.008	−0.007	−0.004	0.010
Occupational aim: nursing	−0.130*	0.146***	0.000	−0.015	0.136***	−0.121***
Greater region: Espace Mittelland	0.003	0.049+	0.014	0.007	0.046+	−0.053+
Greater region: Nordwestschweiz	−0.038	0.131***	0.017	0.003	0.124***	−0.127***
Greater region: Zurich	−0.047	0.055*	0.001	−0.005	0.056*	−0.051+
Greater region: Ostschweiz	−0.019	0.127***	0.009	0.001	0.124***	−0.125***
Greater region: Zentralschweiz	−0.102	0.126***	−0.023	−0.028	0.131***	−0.104**
Greater region: Ticino	−0.051	−0.130***	−0.054**	−0.043**	−0.128***	0.171***
Municipality type: Rural	−0.001	−0.045**	−0.012	−0.008	−0.040*	0.047**
Location: small town (3000-15000)	0.076+	−0.036+	0.019	0.017	−0.039*	0.022
Location: town (15000-100000)	0.037	−0.033	0.005	0.005	−0.035	0.030
Location: city (100000-1000000)	0.103+	−0.127***	0.020	0.020	−0.126***	0.106**
Cantonal share interim solutions	−0.005*	0.007***	0.000	−0.001	0.007***	−0.006***
N	776	3212	2634		3311	
Pseudo-R2	17.9	15.6	17.4		15.4	

+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001; Average marginal effects

Reference categories are: male, Swiss, age 15 at the time of PISA2000, parental education: upper-secondary, other family structure, level compulsory school: medium, mark in test language: sufficient, mark in mathematics: sufficient, reading literacy PISA: level 3, math/science literacy: Q3 (third quartile), occupational aim: other, greater region: region lémanique, municipality type: urban, location: very small town.

4.5.3 Assessing matching quality and common support

Table 4.4 summarizes key measures of overlap and balancing after the pairwise matching of treatment and control group observations. The *treated* group has been chosen to be the smaller one in each pairwise comparison. However, this choice does not affect the estimated treatment effects in the next section, because reversing the groups only reverses the sign and the ATT and ATU (see section 4.4).

Table 4.4: Summary of balancing and common support

	'No education' vs. 'Interim solution'		'Interim solution' vs. 'Direct entry'		'No education' vs. 'Direct entry'	
Balancing	Before match	After match	Before match	After match	Before match	After match
Mean |bias|	16.9	4.4	14.4	1.9	22.0	4.0
Median |bias|	15.5	2.9	10.5	1.5	20.4	3.6
Pseudo-R2	17.9	1.4	15.6	0.4	17.4	2.5
N of sign. mean diff. (p<0.1)		0		0		0
Common Support	Treated (1)	Untreated (0)	Treated (1)	Untreated (0)	Treated (1)	Untreated (0)
Group	No education	Interim sol.	Interim sol.	Direct entry	No education	Direct entry
N	99	677	677	2535	99	2525
Range PS-Score	[0.032; 0.712]	[0.001; 0.719]	[0.020; 0.885]	[0.002; 0.736]	[0.004; 0.486]	[0.000; 0.443]
Mean PS-Score	0.262	0.108	0.337	0.177	0.116	0.035
Std. Dev PS-Score	0.179	0.110	0.179	0.140	0.109	0.048
% Off support	0.0	27.8	2.1	7.9	2.0	20.2

Note: based on Kernel matching on the initial sample and imposing the minimun-maximum criteria, 44 variables

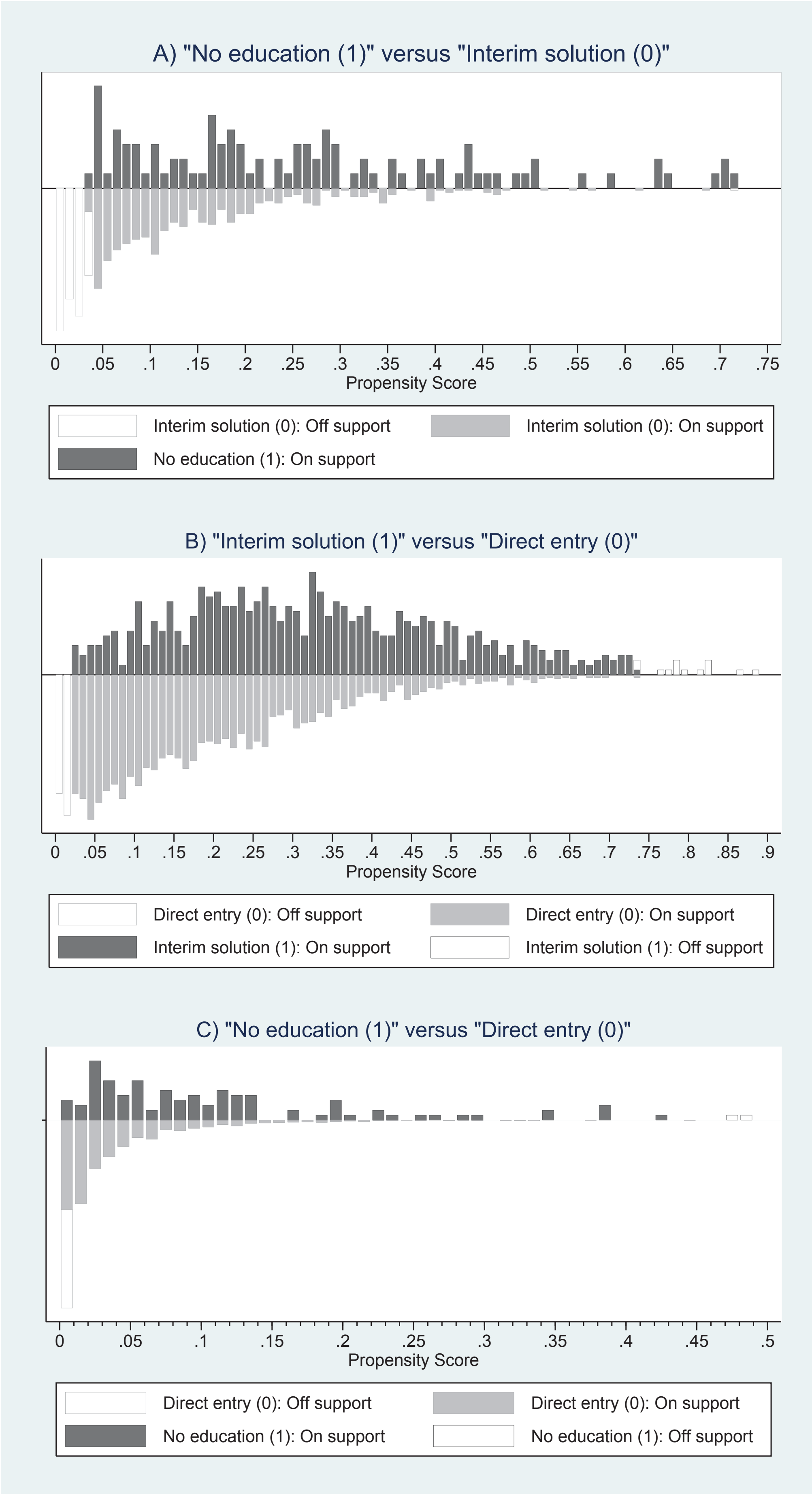
The estimated propensity scores based on the probit models in table 4.3 succeed in balancing the covariates in the matched samples quite well. The overall mean absolute standardized bias is reduced to 4.4% or lower in all combinations (4.4, 1.9, and 4.0).[19] The distribution of the bias among all variables is shown in figure 4.A1 in the appendix. It lies below 5% for most of the 44 variables and exceeds 10% in only a few cases. According to t-tests, there are no significant mean differences across groups after matching. Overall, matching quality seems satisfactory. The topic of common support, however, merits further investigation.

According to table 4.4, there are a considerable amount of cases without support when the *no education* group is part of the comparison (as control group). Further insight is provided by the graphical comparison of the density distributions of propensity scores in figure 4.1.

As for the comparison of *no education* (treated) versus *interim solution* (untreated), figure 4.1 shows that observations in the *no education* group can be successfully matched with controls of the *interim solution* group. The support is, however, very thin at the upper

[19] The psmatch2 package in stata only calculates the standardized bias for the treated group (command pstest). This is the relevant bias if we are only interested in the ATT. By reversing the treatment variables, a calculation of the bias for the untreated shows mean standardized biases of 3.4, 3.2 and 7.0, respectively. The latter is rather high, as it is difficult to find good matches for all *direct entry* individuals among the small (and very different) group of *no education* individuals. Applying an alternative matching algorithm (radius matching instead of kernel matching) leads to a smaller mean bias of 5.7

Figure 4.1: Distribution of the propensity score



end of the distribution. Some of these individuals would be dropped when applying a caliper criterion as described in section 4.4.3.

In turn, there is a lack of overlap at the lower tail of the distribution: there are many individuals in interim solutions (27.8%) with extremely low probability of being in the opposite group (lower than 0.001). According to the minimum-maximum-criteria, there are no comparable observations in the treatment group and the ATU cannot be estimated in this range, strictly speaking. It is notable, however, that more than half of the observations to the left of the treatment-group minimum are still placed within a propensity score distance (caliper) of 0.02. In order to reduce the common support problem, one could apply a more parsimonious specification of the propensity score model. However, excluding potentially important confounders can be very costly in terms of increased bias of estimated treatment effects and the plausibility of the CIA. We tried several specifications with a smaller variable set. It was, however, difficult to find a variant with acceptable values of remaining standardized biases.

As for the decision between *interim solution* (treated) and *direct entry* (untreated), figure 4.1 shows an overlap over a wide range of the propensity score distribution. According to table 4.4, the share of individuals that are off support is relatively low in both groups (2.1% in the treated group and 7.9% in the untreated group).

The last picture in figure 4.1 shows the propensity score distribution for the decision between *no education* (treated) and *direct entry* (untreated). The situation is somehow similar to the first picture. There is very low density for propensity scores above 0.15. Nevertheless, only 2% of the treated are "off support" according to the minimum-maximum criterion. At the lower end of the distribution, there are about 20% of untreated individuals without overlap in the treatment group. The propensity score difference to the first potential control is only 0.004, however. In this case, imposing the minimum-maximum criteria disregards potential matching of individuals that might in fact be very similar.

Overall, in all pairwise comparisons, we do not face severe common support problems when estimating the ATT. However, in estimations with *no education* as the control group, we cannot estimate the counterfactual outcome for those "non-treated individuals" (those in interim solutions or direct entries) who are extremely unlikely to be in the *no education* group. For interpretation and policy conclusions, the estimated ATU/ATE should thus not be interpreted as indicative for this sub-population.

4.6 Results

4.6.1 Short-term effects: Successful entry after the gap year

a) Baseline average treatment effects

After applying the propensity score matching techniques as described in section 4.4, we first compare the probability of the two "gap year" groups *no education* and *interim solution* to successfully master the entry into certifying education in the subsequent year.

Table 4.5: Effect on having had a successful entry by t2 (after the gap year)

Comparison: *No education* (1) versus *Interim solution* (0); Unmatched sample: T=99; NT=677											
Method	Quality	%Off Support		UMD		ATT		ATU		ATE	
	%bias	T	NT	Δ	S.E.	Δ	S.E.	Δ	S.E.	Δ	S.E.
1) Kernel c.s.	4.44	0.0	27.8	-0.237***	(0.049)	-0.283***	(0.062)	-0.261***	(0.069)	-0.264***	(0.063)
2) Radius c.s.	3.68	0.0	27.8	-0.237***	(0.049)	-0.263***	(0.069)	-0.274***	(0.076)	-0.273***	(0.069)
3) N(5) c.s.	3.94	0.0	27.8	-0.237***	(0.049)	-0.263***	(0.065)	-0.270***	(0.074)	-0.269***	(0.067)
4) OLS				-0.237***	(0.049)					-0.271***	(0.051)
5) OLS c.s.				-0.237***	(0.049)					-0.262***	(0.049)
6) OLS c.s.+w.				-0.237***	(0.049)	-0.274***	(0.061)	-0.256***	(0.072)	-0.259***	(0.066)

+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001
Bootstrapped standard errors in parenthesis (100 replications).
c.s.=min-max common support condition, w.=weights derived from kernel matching.

The results in table 4.5 provide evidence that students with no educational activity after compulsory schooling (treatment group) are by far less likely to subsequently enter certifying education. The probability difference of being successful amounts to approximately twenty-six to twenty-eight percentage points; the different matching techniques thus all produce quantitative similar effects. [20] The ATT and ATU are very similar to each other, too, not pointing to notable heterogeneity in the treatment effects between the treated and the untreated group. This topic, however, will be explored in more detail further below.

The lines 4 to 6 present the results for OLS estimations. Line 4 shows the estimated OLS coefficient of the treatment dummy after controlling for all variables also used in the matching procedure.[21] The effect amounts to -27.1 percentage points. Line 5 shows the same OLS coefficient after imposing the common support condition derived from the matching

[20] For comparison, we have also performed matching estimations without imposing the common support condition and by applying a caliper criterion of 0.02 percentage points in the radius and nearest neighbor matching estimation. The ATT is not affected, as the common support problem is negligible small for the treatment group. The estimated ATU (and thus the ATE) is reduced to approximately 22.5 percentage points in these estimations (results available on request). Because more control group individuals who are less comparable are incorporated into the analysis, these results may be less trustworthy. They do not change the main picture, though.

[21] Instead of estimating linear probability models, it might have been more straightforward to estimate non-linear probability models, such as logit or probit estimations. However, we prefer to show OLS estimations for reasons of comparison.

estimations. The estimated effect is slightly reduced to -26.2 percentage points. Therefore, results hardly change when we reduce the OLS sample to those individuals who have counterparts with similar characteristics in the other group. In the last line, we additionally use the weights estimated in kernel matching (line 1) to reweight the control group (ATT) and the treatment group (ATU) in the OLS estimation. This follows the idea of Ho et al. (2007) to use matching methods for preprocessing the data such that the treated group is as similar as possible to the control group and then applying standard parametric methods on these data. Overall, we find that our matching estimations discussed above provide results very similar to the results derived from OLS-regression on the raw and on preprocessed data.

In all estimations, we find large advantages for the *interim solution* over the *no education* group to successfully master the transition after the gap year. One could have expected that the raw difference in the share of successful transitions overstates the advantages of interim solutions because of *pre-treatment heterogeneity* across the two groups. However, a comparison between the unmatched difference (UMD) and the estimated differences after matching (ATT, ATU, ATE) shows that the selection bias even tends to go in the opposite direction: the treatment effect is slightly larger in absolute terms than the unmatched difference. Hence, the results do not support the view that better outcomes for the *interim solution* group compared to the *no education* group are simply due to more favorable pre-treatment characteristics.

b) Assessing treatment effect heterogeneity over the propensity score

Our results also help us to explore if there is *treatment-effect heterogeneity*. Under the CIA, treatment-effect heterogeneity is present whenever the ATT and the ATU substantially differ from each other. A comparison of the treatment effects for the two populations under investigation generally helps to answer how good target programs are. If those who join the treatment are those who profit most from treatment (positive selection with respect to program effects), a program can be regarded as well target and efficient. In our case, table 4.5 does not show notable differences in the ATT and ATU. Depending on the matching algorithm, one or the other is slightly larger and they do never statistically differ from each other. Therefore, it might be most safe to conclude that those who have chosen not to follow an institutionalized interim solution might have equally profited from these programs than those who have actually joined interim solutions (and vice versa). Note that these are average effects over all individuals.

A more detailed insight into selection procedures and treatment effect heterogeneity can be provided by analyzing treatment effects over the range of participation probabilities, as discussed in Lechner (2002); Brand and Xie (2010); Xie et al. (2012).

Figure 4.2: Treatment effect heterogeneity by ps-score

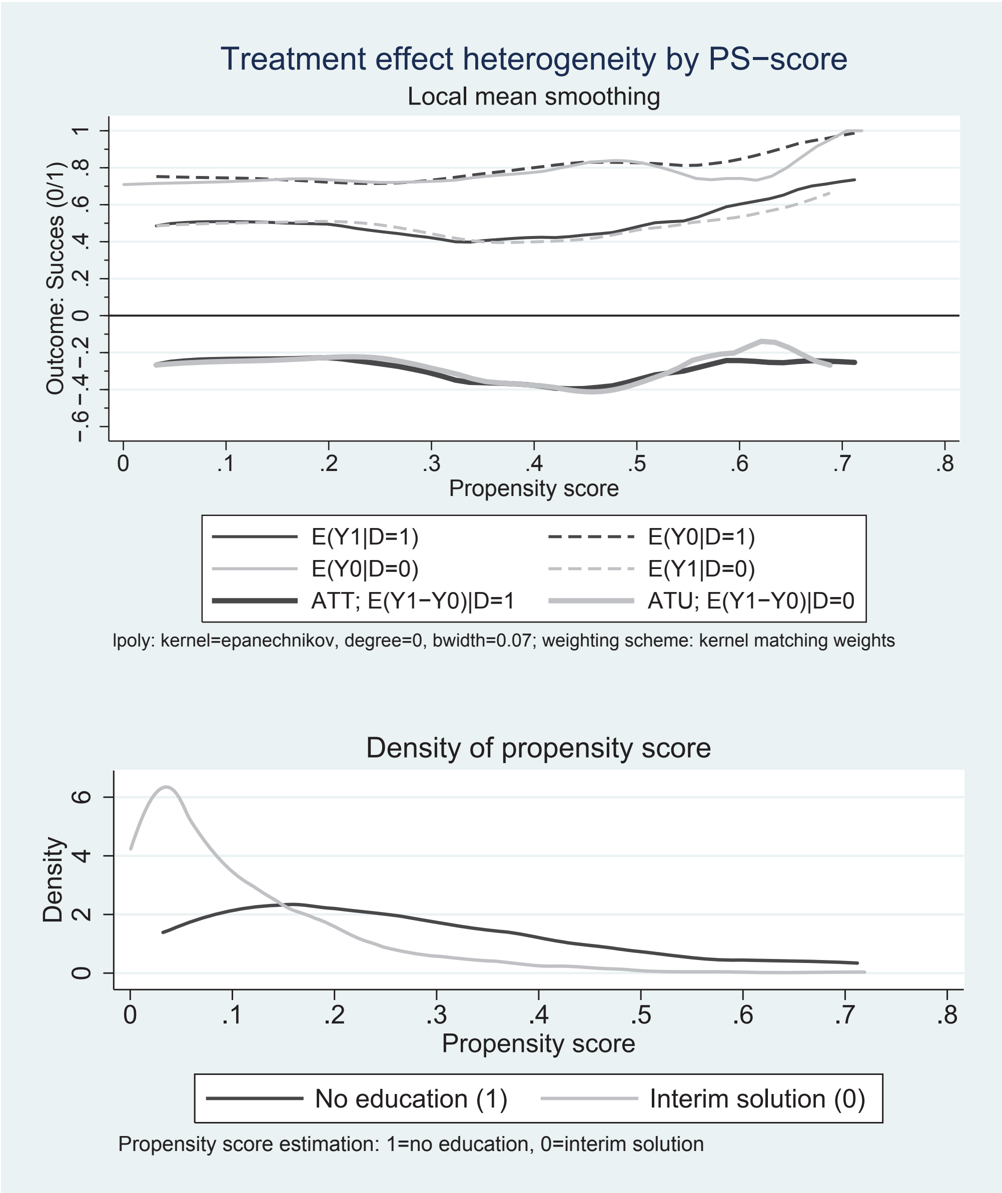


Figure 4.2 graphically shows the application of kernel weighted local-mean smoothing on observed outcomes, counterfactual outcomes, and resulting treatment effects for both the treatment and comparison group over the whole range of propensity scores.[22]

[22] To compute the counterfactual outcome for the treated and the ATT, the untreated group is weighted by the resulting weights of propensity score matching using the kernel matching algorithm. Accordingly, the group of the treated is reweighted to compute the counterfactual outcome for the untreated and the ATU.

First, figure 4.2 shows lower expected transition rates for the *no education* state (D=1) compared to interim solutions (D=0) over the whole range of propensity scores. The treatment effects (ATT, ATU) are rather constant at approximately 20 percentage points in absolute terms until a propensity score of 0.25, the range wherein a large mass of individuals can be found. The treatment effect then increases in absolute terms up to 40 percentage points until a propensity score of approximately 0.45 is achieved; that is, the negative effect of having a no-education gap year becomes larger with increasing propensity of having a no-education gap year. According to the density distribution of propensity scores, there are still many *no education* individuals observed within this range. These individuals would profit most from participating in interim solutions, as their expected counterfactual success rate increases (black dashed line) and their observed success rate decreases (solid black line).

This can be seen as a sign that the sorting into interim solutions might not be very efficient. However, the estimated treatment effects at even higher propensity scores of no education are again similar to those at the lower end of the propensity score distribution. However, as there are only a few individuals with high propensity scores and as there is only thin overlap between treated and controls in this region, the results from the propensity score range over 0.45 are potentially less reliable.

c) Assessing treatment effect heterogeneity by sub-groups

Next, we look at potential treatment effect heterogeneity between sub-groups of interest. We perform a disaggregated analysis by splitting the sample along the dimensions gender, migration background (yes versus no), compulsory school track (low versus medium), and results of PISA-2000 reading literacy tests (low versus high), one after another. That is, we perform exact matching on one dimension at a time. Results are presented in table 4.6.

Naturally, the sample size for certain sub-groups becomes very small. It gets more difficult to find good matches for each treatment and control observation within the same group. Overlap and matching quality are considerably reduced in some of the estimations. Further, due to high variances in the small samples, otherwise substantial treatment effects lack significance. Overall, effects should be interpreted with caution. However, as shown in the last column of table 4.6, OLS regressions using all available cases within a sub-group (no common support condition) come to very similar results for most of the sample-splits. The OLS results, however, show higher significances due to lower standard errors of the estimates.

Table 4.6: Treatment effects by sub-groups

Group	$PS_{D=1}$	% Bias	% Off supp. T	% Off supp. NT	N used T	N used NT	UMD	Matching results ATT	Matching results ATU	Matching results ATE	OLS
Female	0.199	2.9	6.1	36.0	46	313	-0.283***	-0.287**	-0.258*	-0.261**	-0.282***
							(0.070)	(0.094)	(0.100)	(0.091)	(0.071)
Male	0.389	6.5	10.0	20.7	45	149	-0.243***	-0.236*	-0.285*	-0.273*	-0.246***
							(0.067)	(0.118)	(0.140)	(0.115)	(0.073)
Swiss	0.239	5.1	1.6	23.7	60	395	-0.176**	-0.206*	-0.189*	-0.191*	-0.217***
							(0.061)	(0.088)	(0.092)	(0.087)	(0.063)
Migrated	0.480	7.9	21.1	55.1	30	71	-0.315***	-0.475***	-0.443*	-0.452**	-0.466***
							(0.085)	(0.138)	(0.184)	(0.149)	(0.097)
CS: low	0.346	8.1	11.1	30.0	48	266	-0.227***	-0.161	-0.255*	-0.241*	-0.221**
							(0.068)	(0.111)	(0.107)	(0.096)	(0.073)
CS: med	0.369	5.5	17.8	34.7	37	194	-0.253***	-0.285*	-0.190	-0.206	-0.247***
							(0.070)	(0.144)	(0.135)	(0.127)	(0.076)
PISA: low	0.324	4.4	7.9	21.0	58	264	-0.222***	-0.281**	-0.206+	-0.220*	-0.236***
							(0.065)	(0.091)	(0.111)	(0.098)	(0.068)
PISA: high	0.283	8.6	8.3	30.0	33	240	-0.233**	-0.318*	-0.208	-0.221	-0.276***
							(0.076)	(0.125)	(0.177)	(0.152)	(0.081)

+ $p<0.10$ * $p<0.05$, ** $p<0.01$, *** $p<0.001$
Bootstrapped standard errors of matching results in parenthesis (100 replications).
Matching algorithm: kernel matching imposing common support. OLS: no common support condition.
$PS_{D=1}$: the mean propensity score of the treated ('no education'). %Bias: standardized absolute bias after matching.
Note: The sample 'Migrated' consists of both first and second generation immigrants.
The sample 'CS: low' consists of individuals from compulsory schools (CS) with both low-level and no selection.
The sample 'PISA: low' consists of individuals with PISA-literacy-levels up to 480 score points (medium low).

Shortly summarizing the results in table 4.6, there is not much substantial treatment effect heterogeneity between most of the sample-splits. The most striking difference in estimated treatment effects is found for the sample split by migration background. The average treatment effect amounts to approximately 20 percentage points for Swiss pupils (born in Switzerland with at least one parent born in Switzerland) and to approximately 40 percentage points for pupils having a migration background. Therefore, the latter would profit twice as much from following interim solution programs than Swiss pupils. However, migrated pupils have much higher propensity to choose no education gap years than Swiss pupils do (0.480 versus 0.239 among the no education group, column 2)[23]. They are located in the same propensity score range in figure 4.2 for which we have already detected large treatment effects of interim solution programs but rather high probabilities to choose no education.

[23] Even though the group of individuals with migration background is overrepresented in high propensity score ranges of no education, migration background does not directly affect the decision for no education gap years, but rather operates through characteristics that are related to migration background (see section 4.5.2).

4.6.2 Short-term effects: Aspiration level of educational track

a) Baseline average treatment effects

This section analyzes whether the different treatment groups enter upper-secondary education at different intellectual aspiration levels, conditional on having had a successful transition into certifying education by the second year after compulsory schooling.

As it is our hypothesis that interim solutions might be chosen, amongst others, to lower the probability of entering a track at low aspiration level, the dependent variable of table 4.7 is coded accordingly: 1 denotes low aspiration levels, and 0 middle to high aspiration levels of certifying education tracks.

Table 4.7: Effect on having entered at low (versus middle/high) aspiration level by t2

Method	Quality %bias	%Off Support T	NT	UMD Δ	S.E.	ATT Δ	S.E.	ATU Δ	S.E.	ATE Δ	S.E.
A) Comparison: *No education* (1) versus *Interim solution* (0); Unmatched sample: T=47; NT=485											
1) Kernel c.s.	3.85	10.6	25.8	0.230***	(0.068)	0.207*	(0.093)	0.250**	(0.089)	0.245**	(0.081)
2) Radius c.s.	4.94	12.8	26.8	0.230***	(0.068)	0.197+	(0.107)	0.318**	(0.111)	0.305**	(0.100)
3) N(5) c.s.	5.86	10.6	25.8	0.230***	(0.068)	0.214*	(0.102)	0.318***	(0.098)	0.307***	(0.089)
4) OLS				0.230***	(0.068)					0.219**	(0.068)
5) OLS c.s.				0.230***	(0.068)					0.229**	(0.072)
6) OLS c.s.+w.				0.230***	(0.068)	0.226*	(0.091)	0.225*	(0.092)	0.225*	(0.089)
B) Comparison: *Interim solution* (1) versus *Direct entry* (0); Unmatched sample: T=485; NT=2535											
1) Kernel c.s.	1.72	0.6	10.5	0.062**	(0.020)	-0.029	(0.023)	-0.018	(0.023)	-0.020	(0.021)
2) Radius c.s.	1.77	0.6	10.5	0.062**	(0.020)	-0.031	(0.023)	-0.022	(0.022)	-0.023	(0.021)
3) N(5) c.s.	2.40	0.6	10.5	0.062**	(0.020)	-0.021	(0.029)	-0.026	(0.026)	-0.025	(0.024)
4) OLS				0.062**	(0.020)					-0.023	(0.019)
5) OLS c.s.				0.062**	(0.020)					-0.028	(0.019)
6) OLS c.s.+w.				0.062**	(0.020)	-0.033	(0.023)	-0.022	(0.024)	-0.024	(0.023)
C) Comparison: *No education* (1) versus *Direct entry* (0); Unmatched sample: T=47; NT=2535											
1) Kernel c.s.	6.42	2.1	22.6	0.292***	(0.059)	0.191*	(0.087)	0.241**	(0.082)	0.240**	(0.081)
2) Radius c.s.	4.00	4.3	22.8	0.292***	(0.059)	0.154+	(0.085)	0.275**	(0.096)	0.272**	(0.094)
3) N(5) c.s.	5.00	2.1	22.6	0.292***	(0.059)	0.152	(0.107)	0.297*	(0.121)	0.293*	(0.118)
4) OLS				0.292***	(0.059)					0.180***	(0.053)
5) OLS c.s.				0.292***	(0.059)					0.165**	(0.057)
6) OLS c.s.+w.				0.292***	(0.059)	0.160**	(0.059)	0.198**	(0.070)	0.197**	(0.069)

+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001
Bootstrapped standard errors in parenthesis (100 replications).
c.s.=min-max common support condition, w.=weights derived from kernel matching.

Our first comparison (panel A) shows that those with a *no education* gap year have a higher probability of approximately 21 percentage points of entering at low aspiration level than those coming from an *interim solution* (ATT). The effect is slightly smaller than the unmatched difference of 23 percentage points. The reversed analysis shows that the causal gain of interim solutions is estimated to be somewhat larger for the group that effectively followed an interim solution (ATU). It amounts from 25 to 32 percentage points, depending on the matching algorithm. The pattern of ATU>ATT is evidence for positive sorting: individuals who profit the most from interim solutions with respect to low aspiration track avoidance are also more likely to choose them. There are similar patterns and effects revealed in panel C, where those with a *no education* gap year are compared to those with a *direct entry*.

Panel B compares *interim solution* to *direct entries*. The estimation results show that the unmatched difference of 6.2 percentage points in disfavor of the interim solution group disappears once treatment and control groups are matched on the basis of their observables. There is a switch in the sign from positive to negative and the point estimate is close to zero and not significant. Therefore, there is no evidence that those coming from interim solutions are harmed by not having a direct entry (which would have resulted in a positive sign). However, there is no evidence that they profit either.

These results are only partly in line with our expectation that interim solutions should lower the probability of following educational tracks at low-aspiration level. While this is the case in comparison to gap years without educational activity, it is not the case in comparison to direct entries into certifying education.

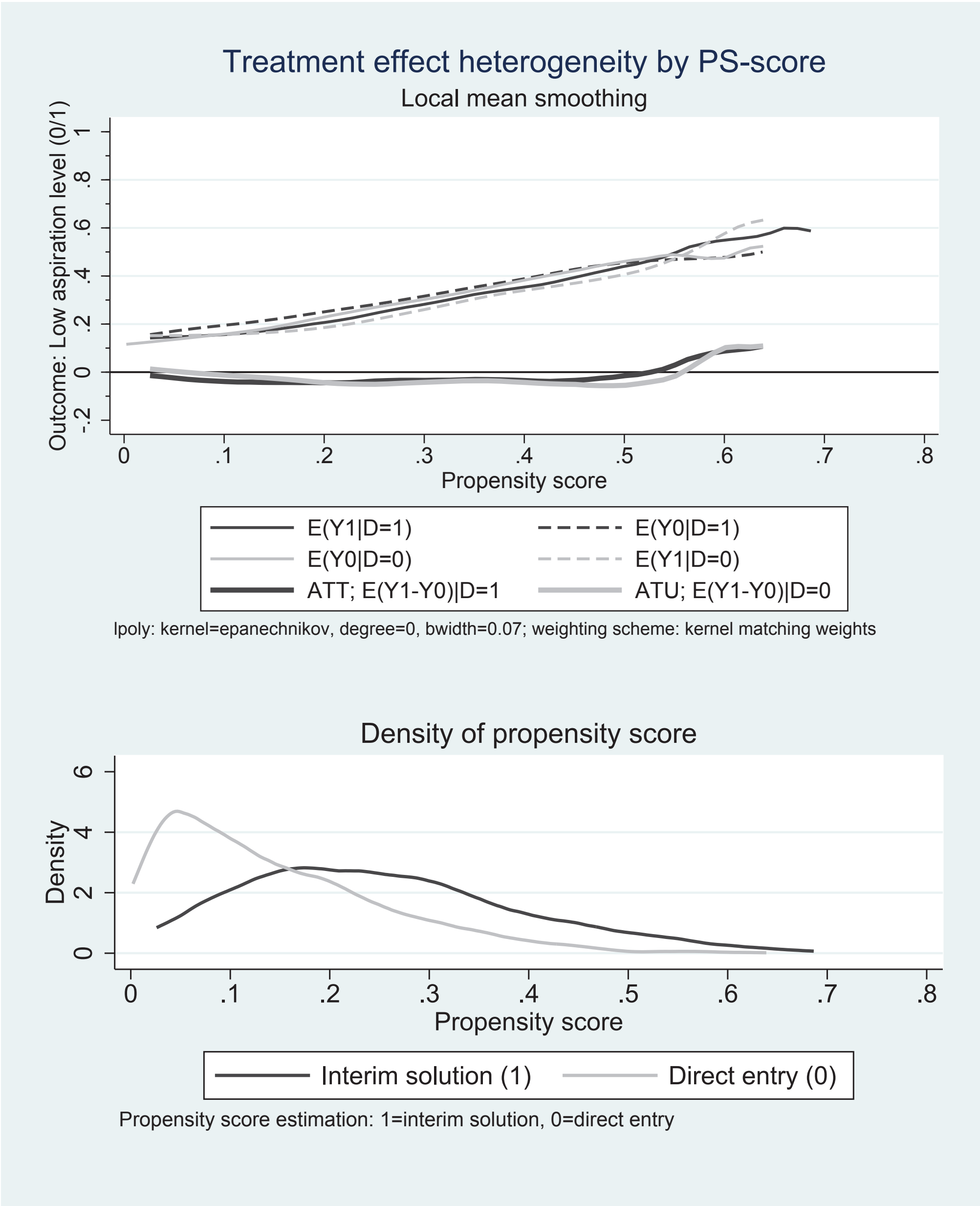
In the remainder of this section, we focus on the latter comparison and inspect whether there are specific propensity score ranges or sub-groups of compulsory school graduates for which—as an alternative to direct entries—interim solution programs causally help to avoid entering at low- instead of high-aspiration-level tracks.

b) Assessing treatment effect heterogeneity over the propensity score

First, figure 4.3 graphically shows that the same characteristics that increase the probability to choose an interim solution instead of a direct entry also enhance the probability to enter a low aspiration educational track instead of a more challenging one (increasing lines in the upper part of the figure). Over large parts of the propensity score range, the counterfactual outcome of the interim solution group (dashed black line) lies slightly above and runs parallel to the observed outcome (solid black line). The treatment effects are rather constant over the propensity score range, except for the lower and upper end of the propensity score distribution. At the lower end, the effect is even closer to zero. There is a switch in the sign at the upper end, where there are very little observations and estimations that are potentially less reliable and less relevant.

Overall, this analysis does not reveal substantial treatment effect heterogeneity in those propensity score ranges where the largest mass of individuals is found.

Figure 4.3: Treatment effect heterogeneity by ps-score



c) Assessing treatment effect heterogeneity by sub-groups

Finally, we again split the sample in order to investigate potential treatment effect heterogeneity by sub-groups. Results are presented in table 4.8. Most of the estimated treatment effects are still not significantly different from zero.

Table 4.8: Treatment effects by sub-groups: interim solution (1) vs direct entry (0)

Group	$PS_{D=1}$	% Bias	% Off supp. T	% Off supp. NT	N used T	N used NT	UMD	Matching results ATT	Matching results ATU	Matching results ATE	OLS
Female	0.325	2.6	0.3	10.7	336	1257	0.104***	−0.048	0.012	−0.002	−0.025
							(0.024)	(0.030)	(0.028)	(0.025)	(0.023)
Male	0.169	2.1	0.7	10.6	148	1143	−0.001	−0.053	−0.067*	−0.065*	−0.051
							(0.036)	(0.041)	(0.032)	(0.031)	(0.033)
Swiss	0.257	2.0	0.3	4.4	375	1903	0.075***	−0.023	−0.008	−0.011	−0.018
							(0.023)	(0.028)	(0.025)	(0.024)	(0.021)
Mig (2nd)	0.303	5.5	7.7	42.8	36	151	0.002	−0.131	−0.133	−0.133	−0.137*
							(0.062)	(0.120)	(0.088)	(0.081)	(0.058)
Mig (1st)	0.399	5.3	5.7	33.8	66	186	0.008	0.060	0.035	0.041	0.025
							(0.058)	(0.079)	(0.091)	(0.070)	(0.059)
CS: low	0.354	2.8	1.3	7.4	221	822	−0.009	−0.104**	−0.089*	−0.092**	−0.088**
							(0.036)	(0.040)	(0.038)	(0.035)	(0.034)
CS: med	0.231	1.8	1.3	10.2	224	1250	0.055*	0.007	−0.006	−0.004	0.006
							(0.023)	(0.029)	(0.031)	(0.029)	(0.023)
CS: no sel	0.428	9.0	29.4	50.6	24	124	0.191**	0.072	0.031	0.038	0.060
							(0.060)	(0.146)	(0.118)	(0.114)	(0.065)
PISA: low	0.304	2.6	0.4	6.6	225	884	0.027	−0.051	−0.081*	−0.075*	−0.045
							(0.035)	(0.043)	(0.038)	(0.036)	(0.033)
PISA: high	0.247	2.1	0.0	13.4	259	1376	0.055*	0.004	0.014	0.012	0.009
							(0.022)	(0.027)	(0.027)	(0.025)	(0.022)

+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001
Bootstrapped standard errors of matching results in parenthesis (100 replications).
Matching algorithm: kernel matching imposing common support. OLS: no common support condition.
$PS_{D=1}$: the mean propensity score of the treated ('interim solution'). %Bias: standardized absolute bias after matching.
'Mig (2nd)': second generation migrants; 'Mig (1st)': first generation migrants.
'CS: low': compulsory schooling (CS) at low-level; 'CS: med': compulsory schooling (CS) at medium-level,
'CS: no sel': compulsory schooling (CS) without selection (no tracking);
'PISA low' consists of individuals with 'PISA-literacy-levels up to 480 score points (medium low).

In terms of significance, the most reliable substantial treatment effect is found for the group of pupils coming from low-level compulsory school tracks. For this group, following interim solution programs instead of directly entering certifying education significantly lowers the probability of entering at a low aspiration level by about 10 percentage points. In contrast, pupils from compulsory school tracks with extended requirements (medium-level) do not profit.

For some other groups, we do find a significant ATU and ATE, but not a significant ATT: among those who directly entered certifying education, the group of males (ATU: -6.5 percentage points) and the group with low PISA-test-scores (ATU: -8.1 percentage points) would have significantly profited from interim solutions with respect to entering a higher aspiration level track.

The sample split by migration status reveals that the quantitatively most pronounced effect among all sub-samples is found for second generation migrants (-13 percentage points).

This effect is, however, not statistically significant due to large standard errors of the estimation. Notably, the OLS coefficient points to a quantitatively similar and significant effect of 13.7 percentage points. Therefore, one would probably find significant effects for second generation migrants in a larger sample by applying matching techniques.

Overall, compared to direct entries, interim solutions do not seem to lower the probability of pursuing an educational track at low aspiration level, at least not on average. There is evidence, however, that certain sub-groups with less favorable characteristics profit all the same, particularly graduates from low-level compulsory school tracks, but presumably also pupils with low PISA-scores and migration background, for which the effects are pronounced as well, but lack statistical significance in some of the estimations.

4.6.3 Long-term effects

We now turn to the longer-term influence of the activity directly after compulsory schooling and analyze the effect of having gap years or following interim solutions on educational outcomes at age 21. Table 4.9 shows the results for the outcome of *not having a diploma or no educational enrolment at age 21.*

Table 4.9: Effect on having no diploma or no enrolment at age 21

A) Comparison: *No education* (1) versus *Interim solution* (0); Unmatched sample: T=52; NT=456											
Method	Quality	%Off Support		UMD		ATT		ATU		ATE	
	%bias	T	NT	Δ	S.E.	Δ	S.E.	Δ	S.E.	Δ	S.E.
1) Kernel c.s.	5.08	5.8	22.1	0.333***	(0.049)	0.331***	(0.084)	0.299**	(0.112)	0.303**	(0.099)
2) Radius c.s.	6.55	5.8	22.6	0.333***	(0.049)	0.343***	(0.091)	0.281*	(0.121)	0.288**	(0.107)
3) N(5) c.s.	5.55	5.8	21.7	0.333***	(0.049)	0.306***	(0.087)	0.338***	(0.105)	0.334***	(0.093)
7) OLS				0.333***	(0.049)					0.329***	(0.052)
8) OLS c.s.				0.333***	(0.049)					0.314***	(0.056)
9) OLS c.s.+w.				0.333***	(0.049)	0.308***	(0.078)	0.279***	(0.082)	0.283***	(0.078)
B) Comparison: *Interim solution* (1) versus *Direct entry* (0); Unmatched sample: T=456; NT=1746											
1) Kernel c.s.	2.27	2.0	7.7	0.072***	(0.012)	0.050*	(0.020)	0.043*	(0.020)	0.044**	(0.017)
2) Radius c.s.	2.30	2.0	7.7	0.072***	(0.012)	0.050*	(0.020)	0.042+	(0.023)	0.044*	(0.020)
3) N(5) c.s.	2.53	2.0	7.7	0.072***	(0.012)	0.054*	(0.021)	0.041	(0.027)	0.044+	(0.023)
7) OLS				0.072***	(0.012)					0.051***	(0.012)
8) OLS c.s.				0.072***	(0.012)					0.047***	(0.013)
9) OLS c.s.+w.				0.072***	(0.012)	0.052***	(0.016)	0.041*	(0.019)	0.044**	(0.017)
C) Comparison: *No education* (1) versus *Direct entry* (0); Unmatched sample: T=52; NT=1746											
1) Kernel c.s.	4.63	3.8	43.5	0.405***	(0.029)	0.367***	(0.077)	0.317***	(0.077)	0.320***	(0.075)
2) Radius c.s.	3.62	7.7	44.0	0.405***	(0.029)	0.333***	(0.082)	0.316***	(0.095)	0.316***	(0.092)
3) N(5) c.s.	5.46	3.8	43.5	0.405***	(0.029)	0.356***	(0.078)	0.285**	(0.106)	0.288**	(0.102)
7) OLS				0.405***	(0.029)					0.362***	(0.029)
8) OLS c.s.				0.405***	(0.029)					0.354***	(0.036)
9) OLS c.s.+w.				0.405***	(0.029)	0.349***	(0.067)	0.301***	(0.068)	0.304***	(0.067)

+ p<0.10 * p<0.05, ** p<0.01, *** p<0.001
Bootstrapped standard errors in parenthesis (100 replications).
c.s.=min-max common support condition, w.=weights derived from kernel matching.

The unmatched difference between the treatment group *no education* and the two other groups (panel A and C) points to a much higher probability of not having a diploma or enrolment at age 21 for the group without educational activity directly after compulsory schooling. The UMD amounts to 33.3 (or 40.5) percentage points compared to those with an interim solution (or direct entry, respectively.) After matching, the difference to the group *interim solution* remains more or less at the same level (ATT: 33.1 percentage points), the difference to the group *direct entry* decreases only slightly (ATT 36.7 percentage points). Therefore, pre-treatment heterogeneity plays only a small role in explaining the differences.

Comparing the group *interim solution* to the group *direct entry*, we find that the unmatched difference of 7.2 percentage points decreases to a ATT of 5 percentage points and an ATU of 4.3 percentage points after matching. Therefore, there remains a comparatively small (but significant) higher risk for those with interim solutions to end up having no educational credential or enrolment six years after compulsory schooling.

4.7 Conclusion

This chapter analyzes the causal effect of non-certifying interim solution programs on subsequent educational outcomes by applying propensity score matching techniques. Outcomes of program participants are compared to those of matched non-participants without educational activity in the year after compulsory schooling and to those of matched non-participants who directly enter upper-secondary education.

We find substantial program effects when comparing participants with those having no educational activity during the gap year. Taking part in interim solution programs enhances subsequent chances to enter upper-secondary education by approximately 26 percentage points, decreases the probability to only enter at low intellectual aspiration level (as compared to middle/high-level tracks) by approximately 25 percentage points, and increases the graduation probability at age 21 by approximately 30 percentage points. Program effects for both participants and non-participants are similar, indicating that those without educational activity in the gap year would profit equally from following interim solution programs as participants do on average. Further, program participation rather uniformly enhances chances to enter upper-secondary education afterwards, with exceptionally large effects, however, for the sub-group of non-native youngsters (about 40 percentage points).

In turn, there is no evidence for positive program effects when we compare the outcomes of program participants to those of the other control group, namely of matched non-participants who directly enter upper-secondary education. The intellectual aspiration level of the subsequent certifying track is not affected, on average, and the probability of having no diploma or enrolment by age 21 even slightly decreases (by 5 percentage points) upon program participation. However, there is evidence for some positive program effects for specific sub-groups: those participants from low-level compulsory school tracks and those with low PISA literacy test scores are more likely to subsequently enter more demanding certifying tracks than their otherwise identical peers with direct entries.

Overall, estimated program effects are only small in either direction when comparing interim solutions to direct entries; the outcomes of the two groups do not differ by much. In contrast, effects are always positive and substantial when comparing interim solutions (or direct entries) to gap years without educational activity. Therefore, not being in education the year after compulsory schooling, in neither a certifying nor a non-certifying track, is a very strong and causal predictor for a failed upper-secondary education career. Only a small majority of the early dropouts are reintegrated at a later point in time in the educational system. Therefore, in order to enhance the upper-secondary graduation rate, one of the most important challenges might be to prevent pupils from completely dropping out of the educational system by the end of compulsory school. The issue of whether someone directly enters certifying education or via a non-certifying interim solution program does not appear to be of major relevance.

From a policy point of view, not only are the treatment effects insightful, but so are the findings on the selection behavior of pupils in different kinds of states after compulsory school. The different groups can broadly be characterized as follows: Compared to the group with direct entry into certifying education, interim solution participants have—on average—lower compulsory school performance (low-level tracks; low PISA-reading literacy; inferior marks in mathematics) and a less favorable family background (lower parental education; non-nuclear family structure). However, they have more books at home. Further, they are slightly more uncertain about their occupational aim, live in cantons with a higher supply of interim solutions (which does not prove causality, however), and are more often female. In turn, the group not following any educational activity in the year after compulsory school shows comparable initial school performance to the direct entry group and even better performance than the interim solution group. However, their parental background is the least favorable on average, all with respect to family structure, socioeconomic index, and number of books at home. Additionally, they exhibit higher tendencies towards school absenteeism during compulsory school. In bivariate (but not multivariate) comparisons, parental support in scholastic matters and (self-assessed) school effort and perseverance are less favorable, too.

Overall, the results regarding the selection behavior at the transition into upper-secondary education indicate that scholastic barriers do not seem to be the main sources for delayed entries. For virtually all individuals in interim solutions, we find similar individuals in the direct entry group (the reverse does not hold, however), implying that the mix of their characteristics should not impede per se the direct integration into certifying education. The same holds for the group in gap years without educational activity; they virtually all have counterparts in the two other groups. Therefore, there is no evidence that youngsters not being in certifying educational pathways should generally be incapable or unready to undergo some kind of certifying education.

In the light of our results, many policy efforts of recent years might be effective measures to work against delayed entries. The measures are, first, to earlier detect pupils at risk and accompany them through their earlier post-compulsory schooling years ("case management"), and second, to create new vocational tracks with lower scholastic entry-barriers (so-called "Attestlehren") but with institutionalized permeability towards more demanding tracks at a later stage. While the former should particularly decrease the number of pupils without any educational activity, the latter should also reduce the necessity to follow interim solutions as the only way to enhance the chances of getting into more demanding tracks afterwards. Naturally, the willingness of firms to provide such tracks is of crucial importance.

A shortcoming of the present analysis is that it is restricted to the compulsory school graduation cohort of the year 2000. Transitions to upper-secondary education might take place under different economic and institutional settings nowadays. Therefore, our results remain silent on the effects of recent reforms mentioned above. Although official population data do not suggest increasing (or otherwise changing) shares of direct entries into certifying education (see section 4.2), some mechanisms behind aggregate numbers might have changed all the same, potentially leading to different selection behavior or different treatment effect patterns.

Some important questions from a policy point of view remain open. First, whereas heterogeneity of effects between sub-groups was addressed to some extend (with very small group sizes, however), the potential heterogeneity of different programs was not investigated due to the lack of information in the data. Second, the results do not allow for quantifying the human capital effects (compensating for deficits) versus signalling effects of interim solution programs. Third, nothing can be said about how many postponed entries are due to postponed occupational decisions versus failures in the application process on the apprenticeship market. There remains the open issue on how the existence of such a transitory feeder system itself affects firms' willingness to train those coming directly from compulsory schooling or, generally, how the supply of such programs affects the demand for them (the latter is one of the open questions also formulated in a recent policy report by the SBFI-SERI (2015a)).

Future research should therefore investigate the same topic with newer data sources—ideally data sources that are larger but not less informative and contain several cohorts of compulsory school leavers, permitting the exploitation of information that is varying over regions as well as years.

4.A Appendix

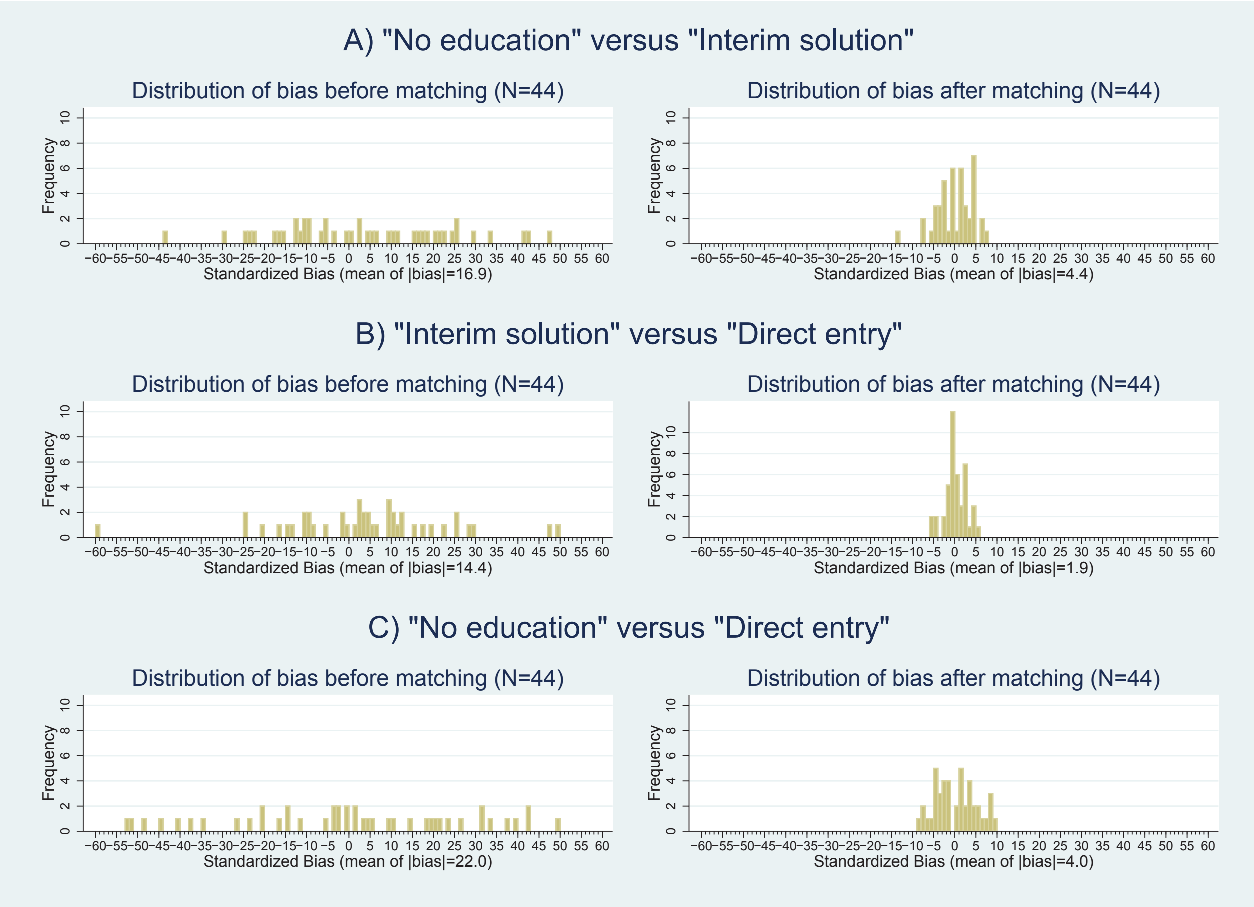
Table 4.A1: Variable definition

Variable$^{a)}$	**Definition**
Treatment status	3 categories representing the activity after compulsory schooling (TREE wave t1). *Direct entry*: the respondent follows a certifying education at upper-secondary level. *Interim solution*: the respondent follows a program that has some preparatory schooling or practical character. *No education*: the respondent is not involved in any educational activity.
Outcome 1: "Successful entry at t2"	Indicator variable that equals 1 if there is a successful transition into certifying education after a lagged entry (TREE wave t2); 0 otherwise.
Outcome 2: "Aspiration level of entry"	Indicator variable that equals 1 if the intellectual aspiration level of the certifying education is low (1 or 2) and equals 0 if the aspiration level is medium or high (3 to 6). The aspiration levels for 101 different vocational tracks were rated on a scale ranging from 1 to 6 by an expert group of vocational advisers (for details and studies using this variable see Stalder (2011)). See also section 4.3 and footnote 6 on how this scale has been extended to other educational tracks. The intellectual aspiration level refers to the entry year into certifying education and thus stems from wave t1 (direct entry) or t2 (lagged entry).
Outcome 3: "Graduation by t6"	Indicator variable that equals 1 if the respondent has acquired a diploma at upper-secondary level by the time six years after compulsory schooling (t6) or, alternatively, is enrolled in certifying education at the time of the survey t6; 0 otherwise.
Female	Equals 1 if female; 0 if male.
Nationality	Dummies representing 3 categories: Swiss: respondent born in Switzerland with at least one parent born in Switzerland; second-generation immigrant: respondent born in Switzerland but parents born outside Switzerland; first-generation immigrant: respondent and parents are foreign born.
Age 16 at PISA survey	Equals 1 if pupil aged 16 at the time of PISA 2000; 0 if aged 15.
Parental education	Dummies representing 4 categories of highest parental education: compulsory school, upper-secondary education, tertiary education, missing information.
Family structure	Equals 1 if the family structure is nuclear; 0 otherwise (single, mixed, other, missing information).
Socioeconomic status	The international socioeconomic index (ISEI) ranges from 16 to 90 and is constructed on the basis of ISCO-88-coded professions of the parents. We use the highest value among the parents.
Number of books	Index for the number of books at home as stated in the PISA-2000 survey.
Level of compulsory school	Dummies for the school track that was attended at the time of the PISA 2000 survey: low-level compulsory school (e.g. Realschule), medium-level compulsory school (e.g. Sekundarschule), and "no selection" (integrated track, mixed). Respondents of high-level compulsory school tracks (e.g. Pro-Gymnasium) are dropped from the analysis.
School mark in test language	4 categories for the school mark in test language (German, French, Italian, depending on the linguistic region) in the last compulsory school report: insufficient (<4), sufficient (from 4 to <5), good (from 5 to 6), missing.

School mark in mathematics	4 categories for the school mark in mathematics in last compulsory school report: insufficient (<4), sufficient (from 4 to <5), good (from 5 to 6), missing.
Reading literacy PISA	6 categories of reading literacy from the PISA 2000 survey. Test scores are summarized in the following levels: 0 "very low" [0;334.75], 1 "low" [334.75; 407.67], 2 "medium low" [407.67; 480.18], 3 "medium high" [480.18; 552.89], 4 "high" [552.89; 625.61], 5 "very high" [>625.61].
Math/sciences literacy PISA	5 categories of literacy in mathematics/sciences from the PISA 2000 survey. Mathematical and scientific literacy were each investigated only for approximately half of the students in PISA 2000. We pooled the resulting test scores in one variable and built 4 categories representing quartiles and one category representing missing values (11.9%).
Absenteeism	PISA index derived from students' reports on how often they missed school, skipped classes and were late for school in the two last weeks (3 items).
Family support	PISA index of family educational support derived from students' reports on how frequently their mother, father, or brothers and sisters worked with the student on what is regarded nationally as schoolwork. Students responded to each statement on a five-point scale with the following categories: never or hardly ever, a few times a year, about once a month, several times a month and several times a week.
Effort & perseverance	PISA index derived from the frequency with which students used the following strategies when studying: I work as hard as possible; I keep working even if the material is difficult; I try to do my best to acquire the knowledge and skills taught; and, I put forth my best effort.
Expected status age 30	Index (standardized) representing the expected socioeconomic status at age 30, constructed on the basis of the ISCO-88-coded expected job at age 30. We additionally include a dummy variable for missing information (23.5%). This variable is likely to capture, up to a certain degree, uncertainty about career choice.
Occupational aim "nursing"	Dummy variable that equals 1 if the expected job at age 30 is in the field of nursing. At the time of PISA 2000, the vocational training towards nursing or similar occupations had still required a minimum age of 18, which institutionally forced youngsters with career aspirations in the health sector (mostly women) into interim solutions.
Greater regions	7 categories of greater regions (official grouping of cantons) based on the location of compulsory school.
Municipality type "rural"	Dummy variable that equals 1 if school location is urban, 0 if rural.
School location	5 categories of municipality sizes (number of inhabitants) where the compulsory school is located: village (less 3000), small town (3000-15000), town (15000-100000), town (15000-100000), city (100000-1000000).
Cantonal share interim solutions	Cantonal share of pupils in interim solution programs in the year 2000 according to the administrative data base of the Federal Statistical Office (Bildungsstatistik).

[a] *All variables except for the outcome variables are measured at the time of PISA 2000.*

Figure 4.A1: Distribution of the standardized bias before and after matching



Chapter 5

Human capital specificity of VET—Evidence from mobility after graduation

5.1 Introduction

In this chapter,[1] we study the specificity and transferability of human capital by analysing inter-firm and occupational mobility of Swiss apprentices shortly after the conclusion of training. Countries with comprehensive work-based apprenticeship programs lend themselves to the study of occupation-specific human capital because they are structured along a multitude of well-defined occupations and corresponding educational tracks. Work-based apprenticeships are of central importance to the educational system in many countries, such as Austria, Denmark, Germany, the Netherlands and Switzerland (Wolter and Ryan, 2011). In the United Kingdom, policy initiatives have been proposed to reinvigorate the apprenticeship system UK Parliament, 2009: the Apprenticeships, Skills, Children and Learning Act of 2009), while calls for more vocational education tracks are a subject of public debate in the US (see Hoffman, 2011; Symonds et al., 2011; or President Obama's 2014 State of the Union Address[2]).

[1] This chapter is based on Mueller and Schweri (2015), *How specific is apprenticeship training? Evidence from inter-firm and occupational mobility after graduation,* Oxford Economic Papers 67(4), 1057-1077; and its earlier version: Mueller and Schweri (2012b), *The returns to occupation-specific human capital—Evidence from mobility after training*, Economics of Education Working Paper Series 0081, University of Zurich, Institute for Strategy and Business Economics (ISU).

[2] https://www.whitehouse.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address

The main economic rationale for a comprehensive work-based apprenticeship system is to provide trainees with a set of clearly defined and nationally tested occupational skills that are transferable to other firms after graduation. Occupational skills promise specialization gains due to more specialization than would be achieved in a purely general education system. Apprentices acquire occupation-specific human capital that enables them to immediately begin work as skilled workers after training; in a general education system, they would need to go through a period of on-the-job training when taking up employment at a firm.

However, this specialization may also be detrimental if the allocative efficiency of apprenticeship systems is inferior to that of general education systems. Specific human capital may impede workers from making efficiency-enhancing firm or occupation changes and lead to wage losses for those who want to or have to leave the training firm or the learned occupation.[3] This barrier becomes particularly important when job prospects on the labor market deteriorate because of the business cycle or changes in the skills needed in the economy, due to technological change or macroeconomic reallocation (Bassanini et al., 2007; Wasmer, 2006). Transferability might become even more important in the future, as some studies have indicated that occupational mobility has generally been increasing in recent decades (Kambourov and Manovskii, 2008; Lalé, 2012; Parrado et al., 2007).

The potential for gains from specialization as well as for allocative inefficiencies associated with apprenticeship systems depend on the transferability of the human capital acquired in apprenticeships. Firm-based apprenticeships have been suspected of conveying an overly specific, narrow set of skills.[4] To empirically assess the transferability of this human capital, we study the incidence of inter-firm and occupational mobility[5] of Swiss apprentices shortly after their training and the effect of these types of mobility on wages.

The literature on mobility after apprenticeships refers mainly to Germany and comes to heterogeneous results on the importance of inter-firm and occupational mobility for wages (see section 5.2). However, the evidence reported in the literature may not be generalizable outside of Germany where there are labor market institutions that limit the ability to observe the effects of mobility on wages (Muehlemann et al., 2010). Labor market institutions that hinder post-training mobility (such as work councils, industry-wide collective agreements,

[3] Furthermore, Lamo et al. (2011) compare the economic transitions in Poland and Estonia and conclude that overly specific training may increase unemployment.

[4] Heckman (1994) described the German apprenticeship program as 'very narrow technical training' with a 'rigid curriculum' that contributes 'to diminished options in later life' (p. 108).

[5] The terms 'occupational change' and 'occupational mobility' are used interchangeably in our analysis.

strong employment protection) are weaker or non-existent in Switzerland. By looking at outcomes in Switzerland, we can shed light on the outcomes of mobility from comprehensive apprenticeships schemes (like those in Germany) under more lightly regulated labor market conditions, similar to those that prevail, e.g., in English-speaking countries.[6]

We investigate the inter-firm and occupational mobility of a sample of apprentices who have just completed several years of training in firms (combined with vocational school) in a particular occupation. After training, apprentices must decide whether to (i) continue working for their training firm as a skilled worker, (ii) change firms within the learned occupation, or (iii) change firms and move out of the learned occupation.[7] Analysing the causal wage effect of trainees' mobility decisions by addressing the endogeneity of mobility allows us to assess the transferability of the trainees' newly acquired human capital to other firms and occupations.

We use a longitudinal data set that is based on the Swiss cohort of the *Programme for International Student Assessment (PISA) 2000* and matches employer and employee data. We exploit the employment information of workers one year after apprenticeship graduation, along with information on their training period. One advantage of this data is that all trainees are at the same stage of their labor market career; mobility immediately after training is not influenced by years of (additional) labor market experience.[8] Some confounding factors associated with years on the labor market include job-shopping, multiple changes and internal promotions. By avoiding these factors, the wage effects of mobility after training provide a 'purer' measure of the transferability of human capital acquired in training than analyses that compare learned and current occupations for employees with many years of labor market experience. Furthermore, the dataset contains open text information on the learned and current occupations (in addition to occupation codes) that we use to ensure accurate coding of occupation change. The wealth of background variables available in PISA allows us to control for important dimensions of individual heterogeneity such as socio-economic background, ability and the quality of worker-firm and worker-occupation matches. Finally,

[6] The Swiss labor market is one of the least regulated in Europe. Unlike in Germany, employment protection is low (OECD, 2004; Venn, 2009) and inter-firm mobility after apprenticeship is relatively high (Wolter and Ryan, 2011). Also see OECD (2009a) for differences in the institutional setup of the VET system in Switzerland and Germany.

[7] We restrict attention to these three alternatives. Occupation change always implies a firm change since we observe virtually no cases of occupation changes within the training firm after training.

[8] Firm or occupation movers may require more time than stayers before their wages reflect their individual performance potential. Firms would then pay wages equal to the expected value of movers' productivity, which will still allow us to identify the productivity differential between movers and stayers.

we address further sources of endogeneity of inter-firm and occupational mobility by means of the multinomial treatment regression model of Deb and Trivedi (2006).

The remainder of this chapter is structured as follows. The institutional background and related literature are discussed in section 5.2. Section 5.3 presents the data. The empirical strategy and identification issues are discussed in section 5.4. Section 5.5 is dedicated to the results of the empirical analyses. Section 5.6 presents conclusions.

5.2 Institutional Background and Related Literature

Cantonal authorities have to approve apprenticeship contracts to ensure that training firms fulfil the legal requirements (such as employing trained supervisors). While participation in the apprenticeship market is voluntary for firms and youngsters, federal laws regulate the occupations in which apprenticeships can be undertaken and provide various instruments for quality control. National training ordinances define every occupation's title, duration, educational objectives, curricula (including the number of lessons in vocational school) and procedures for the final exams. These regulations and the external certification reduce asymmetric information between the actors about training content and workers' ability (Acemoglu and Pischke, 2000; Malcomson et al., 2003) and ensure that firms provide a certain amount and quality of training that is transferable to other firms after graduation.

Successful graduates of an apprenticeship are awarded federally recognized diplomas and the respective titles that identify them as skilled workers in their occupation. The apprenticeship curricula provide for a mix of general and specialized occupational skills: for three or four years (in few cases, two years), apprentices work for three to four days a week in their training firm and attend vocational school for one to two days a week. In firms, they acquire all types of skills, learning about the firm's products and production technology, occupational tasks and general skills such as work values (accuracy, etc.). In vocational schools, apprentices attend general education classes and occupation-specific lessons. Industry training courses organized by employer associations complement the education in schools and firms by training all apprentices in a set of occupation-related skills defined in the training ordinance. Final exams consist of oral, written, and practical parts that test general and occupational knowledge and skills as defined in the training ordinance.

From a human capital perspective, apprentices acquire a mix of general and occupational skills that are transferable to other firms. Firms are free to provide additional firm-specific skills. Stevens (1994b) demonstrates that monopsony power due to imperfect competition can lead training firms to overinvest in the firm-specific element of training to reduce turnover and to capture the benefits of training at the expense of the worker. If the costs and benefits of firm-specific human capital are shared between training firm and apprentice (Becker, 1962; Hashimoto, 1981; Oosterbeek and Leuven, 2001), the workers' returns on firm-specific human capital should be observable in wages. Changing the employer after training should then lead, *ceteris paribus*, to a wage loss.

The concept of firm-specific human capital introduced by Becker (1962) extends to the idea of occupation-specificity of human capital (Shaw, 1987; Zangelidis, 2008; Kambourov and Manovskii, 2009; Sullivan, 2010): Occupation-specific human capital is transferable to other jobs within the same occupation but cannot be used in jobs outside that occupation. It can thus be classified as 'transferable' according to the definition by Stevens (1994b), in which 'it is of some value to at least one firm in addition to the training firm' (p. 540), but not perfectly general.[9] An exogenous change away from the learned occupation should entail a wage loss, ceteris paribus, because acquired occupation-specific skills cannot be put to use anymore. The causal wage difference between those changing occupation and those staying in an occupation can be interpreted as a measure of the transferability of training and, accordingly, as a measure of the occupation-specificity of the human capital that the training confers.

The empirical literature on mobility after apprenticeships refers mainly to Germany and comes to heterogeneous results on wage effects of inter-firm and occupational mobility. With respect to inter-firm mobility, von Wachter and Bender (2006) found causal evidence of initial wage losses for graduates leaving middle- and large-sized training firms at the time of graduation. In addition, they showed that initial sorting, adverse selection and endogenous job mobility bias ordinary least squares (OLS) regression results such that short-run wage losses are underestimated on average. Acemoglu and Pischke (1998) and Bougheas and Georgellis (2004) also found negative effects of leaving the training firm. Harhoff and Kane (1997) and Werwatz (1996) found some evidence for positive wage effects of leaving the training firm. Dustmann et al. (1997) found no significant mover-stayer wage differential, nor did Euwals and Winkelmann (2004), once they considered movers staying in the same firm size class. Winkelmann (1996) analyzed mobility patterns and found that apprentices' human capital was not less portable than that from other educations.

The effects reported for switching out of the learned occupation are similarly heterogeneous. Gathmann and Schoenberg (2010) found that task-specific human capital accounts for a part of the wage growth observed for medium-skilled workers. Fitzenberger and Spitz (2004) found positive effects of occupational changes. Werwatz (2002) found wage losses only for those occupational movers who ended up in unskilled jobs, showing that apprentices' hu-

[9] At first sight, such transferable human capital does not offer a lever for firms to retain workers as firm-specific human capital does. Yet, if firms have monopsony power within the boundaries of the occupation, occupation-specific skills could also be used to retain workers (Smits, 2007).

man capital is largely general. Clark and Fahr (2002) came to a similar conclusion; they estimated a 'worst-case scenario' where only one third of the human capital of exogenously displaced workers can be transferred beyond 1-digit occupations. Goeggel and Zwick (2012) looked at firm and occupation changes in the period immediately after graduation and reported heterogeneous wage effects; on average, they found positive effects for firm changes and negative effects for occupation changes.

We contribute to this literature by analysing firm and occupation changes in one estimation model. Many of the mentioned studies addressed the endogenous nature of mobility, but none of them analysed employer and occupation changes simultaneously.[10] This is necessary to disentangle the effects of firm- and occupation-specific human capital on wages in the early careers of apprenticeship graduates, because occupational change typically goes hand in hand with employer change. A new study by Fitzenberger et al. (2015), however, uses an estimation approach similar to ours for Germany. They use variation in regional labor market characteristics to analyse wage effects caused by both mobility across firms and mobility across occupations after graduation from apprenticeship. They find that pure firm changes and occupation-and-job changes result in average wage losses.

[10] For example, von Wachter and Bender (2006) and Werwatz (1996) dropped occupational changers when analysing the causal wage effect of employer changes. Clark (2000) and Clark and Fahr (2002) focused on displaced workers when analysing occupational changes. Goeggel and Zwick (2012) estimated separate models for employer change and occupational change; it remains unclear whether the correlation between both is taken into account. Other studies ignored the possibility that the wage effect of firm (occupation) changes might be partly driven by a loss of occupation-specific (firm-specific) human capital. An exception is Longhi and Brynin (2010), who studied inter-firm and occupational mobility in Britain and Germany, but did not address the endogeneity of mobility with respect to wages.

5.3 Data and Descriptives

The data base for our analysis is the Swiss TREE data. We use individuals who completed their apprenticeship by 2005 and use data including wave 2006 to identify their subsequent labor market outcome one year after graduation. Focusing on this period allows us to include the vast majority of the individuals with apprenticeships because these programs typically start immediately after compulsory school and end after two to four years.[11] We observe 1,618 individuals with a transition from apprenticeship to work or to another activity within a year after graduation. Of the graduates, 72 percent took up work, 15 percent were enrolled in further education, 4 percent were serving in the military, 3 percent were temporarily out of the labor force because they were travelling or engaged in language studies abroad, and 6 percent were unemployed.

We include working individuals[12] with non-missing values in the wage and mobility variables. Our final sample of employed individuals include 878 observations for the wage regressions. Due to the limited sample size, we do not split the sample between men and women. Female labor participation is similar to male participation at this age for apprenticeship graduates.[13] We control for occupation dummies to account for occupational segregation by gender in the labor market.[14]

The mobility behavior after apprenticeship is our primary interest; firm change is defined as working in a different enterprise than the training enterprise.[15] Occupation change is defined as a change away from the apprenticeship 2-digit occupation after the apprenticeship period, based on occupation codes used to classify occupation by the Swiss Federal Statistical

[11] Comparing our TREE sample with official numbers on apprenticeship graduates, graduates of four-year apprenticeships are slightly underrepresented.

[12] Using all VET graduates, the paper of Mueller and Schweri (2009) puts its focus on the determinants of various activities after apprenticeship graduation and includes a comparison of the mobility behavior of dual VET graduates compared to graduates of full-time vocational schools.

[13] In our data, female labor market participation of apprenticeship graduates is even higher (77% versus 66%). While both genders show very similar rates of unemployment (6%) and periods of being temporarily out of the labor force because they were travelling or engaged in language studies abroad (3%), only male participants are in military service (9%). The proportion of graduates who attend further certifying education (e.g., a university of applied sciences) is slightly higher for male graduates as well (16% versus 14%).

[14] Note that industry changes occur almost exclusively together with occupation changes in our data. Because all apprenticeship regulations refer to clearly defined occupations and not to industries, occupations are the relevant dimension in our context.

[15] Enterprises can consist of several establishments in the definition of the Federal Statistical Office. We use information from the firm census of the FSO to identify changes on enterprise level. This is our preferred definition because information on workers' abilities is likely to be available to the enterprise, not only to the establishment; see Euwals and Winkelmann (2004) for a short discussion.

Office (FSO).[16] There are thirty-nine 2-digit occupation categories.[17] Thus, we adopt a rather broad definition of an occupation to exclude changes between occupations entailing a very similar set of skills.[18] As a sensitivity check, we also provide results for 5-, 3- and 1-digit changes. As only three individuals showed an occupation change without a firm change (these were dropped from the analysis), we defined three mutually exclusive mobility categories. These are represented by the two dummy variables, *firm movers* who remain within the learned occupation and *occupation changers* who switched both firm and occupation, and by the reference category, *stayers.*

Table 5.1 describes the mobility behavior of employed apprenticeship graduates. Roughly one half of the graduates continue to work in their training firms, 42% change firms but not occupation, and 7% change firm and move out of the 2-digit occupation in which they were trained. The proportion of occupation changes is rather low one year after completing apprenticeship training. If we use the 5-digit definition, we find 4.5 percentage points more occupation changes than with the 2-digit definition.

Table 5.1: Mobility patterns after training (status one year after graduation)

	Digit-level of occupational change			
Category	5-digit	3-digit	2-digit	1-digit
Job in training firm ('Stayer')	51.03%	51.03%	51.03%	51.03%
Firm change within occupation ('Firm mover')	37.24%	41.23%	41.80%	42.60%
Firm change across occupation ('Occupational change')	11.73%	7.74%	7.18%	6.38%
Total	100.00%	100.00%	100. 00%	100.00%

[16] Apprenticeship occupations were pre-coded according to the official classification *BIS (Bildungsstatistisches Informationssystem)* of the Federal Statistical Office, which is designed to categorize apprenticeship occupations. However, the occupations of the jobs that were accepted after the apprenticeships are coded according to the Swiss Occupation Classification *SBN (Schweizerische Berufsnomenklatur 2000)*; this classification system is typically used for labor market analysis in Switzerland. To compare the learned and current occupations, we convert the vocational training occupations (i.e., *BIS* codes) into the Swiss Occupation Classification codes using the official 'thesaurus' developed by the Federal Statistical Office, which assigns about 19'000 job titles (open text) to the numeric classification scheme. While we could convert these codes into international classification codes, such as ISCO, we prefer to use the Swiss classification system, which is better suited to discriminating among the different Swiss apprenticeship occupations.

[17] Examples of these thirty-nine occupation categories include: '24: occupations in metalworking and engine construction (Berufe der Metallverarbeitung und des Maschinenbaus),' '26: occupations in wood processing and paper manufacture/paper conversion (Berufe der Holzverarbeitung sowie der Papierherstellung und -verarbeitung),' '28: occupations in the chemical and plastics industry (Berufe der Chemie und Kunststoffverfahren),' '61: occupations in the hotel and restaurant industry and domestic economy (Berufe des Gastgewerbes und Hauswirtschaftsberufe)' and '62: occupations in cleaning and hygiene (Berufe der Reinigung, Hygiene und Koerperpflege).'

[18] Task-based (Gathmann and Schoenberg, 2010) or skill-weights approaches (Lazear, 2009; Geel and Backes-Gellner, 2011) allow one to analyse the distance between occupations directly. The more classical approach that distinguishes among official occupations remains relevant because all state regulations (on training ordinances, curricula and diplomas) are based on these.

In studies using the German Qualification and Career Survey (GQCS), insight on the transferability of the human capital can be gained by using apprentices' own assessments about the proportion of skills acquired during apprenticeship that are applicable at their current job. Dustmann and Schoenberg (2007) use this information and report results for the degree of specificity of apprenticeship training in Germany (firm-specific: 4.5%, occupation-specific: 34.3%). Replicating their calculations using TREE respondents' assessements of the usefulness of what they learned in training (in firms as well as in school) to their current work, table 5.2 shows virtually identical results for Switzerland: Firm movers report values that are, on average, 2.4 percent below the values reported by firm stayers ('firm-specific component'), whereas occupation changers report values that are 34.4 percent below the average value of firm movers within the learned occupation ('occupation-specific component').

Table 5.2: Usefulness of the skills acquired in training for current job

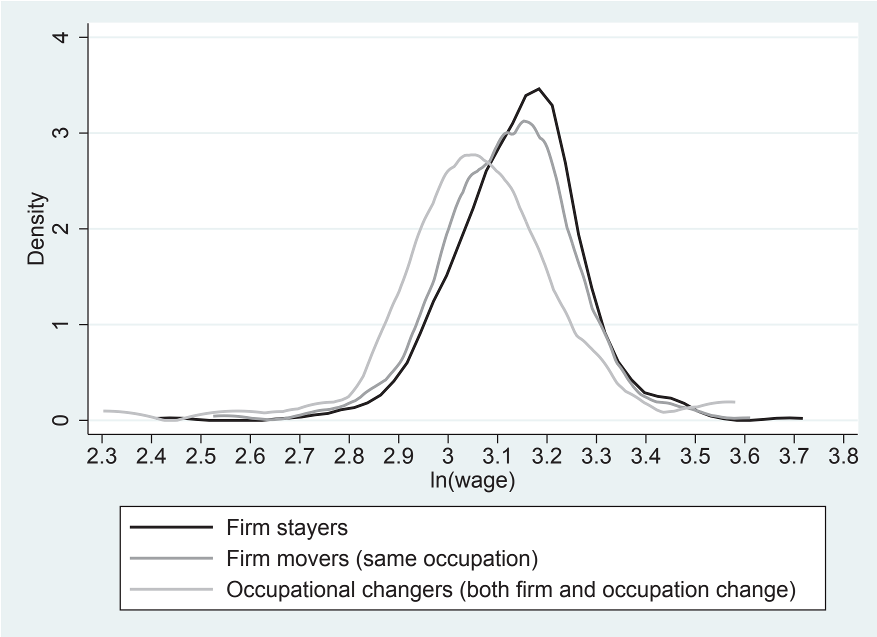
	Skills training firm		Skills voc. school		Mean of both sources	
	mean	*std.dev.*	*mean*	*std.dev.*	*mean*	*std.dev.*
Firm stayer	3.60	(0.62)	3.02	(0.78)	3.31	(0.58)
Firm mover (same occupation)	3.35	(0.76)	3.11	(0.81)	3.23	(0.64)
Occupational change	2.12	(0.95)	2.12	(0.94)	2.12	(0.84)

Note: The usefulness of the skills for the current job were rated by the respondents on a scale ranging from 1, 'not useful,' to 4, 'very useful.'

However, an analysis of (self-assessed) skill applicability might be distorted by similar endogeneity issues as discussed in the next session in the case of wages. In the remainder of the analysis, we restrict our attention to wages as our outcome of interest.

As the dependent variable, we use the logarithm of hourly wages in the current job. Descriptively, the wage distributions in figure 5.1 show that average wages are highest for stayers, followed by firm movers and occupation changers. Occupation changers earn the least on average: they earn 21.74 CHF per hour, compared with 22.86 CHF for firm movers and 23.31 CHF for stayers (see table 5.A2 in the appendix). Furthermore, the wage distributions are broader for movers than for stayers and even broader for occupation changers. This is also the pattern that we would expect for the (unobservable) distribution of wage offers because the offers come from very different firms in the case of occupation changers.

Figure 5.1: Distribution of the logarithm of hourly wages by mobility status one year after graduation



In our estimations, we always include variables on personal characteristics (gender, immigrant), parental education, language region, the apprenticeship training program (occupation and duration of apprenticeship, size of training firm, professional baccalaureate obtained), information on the period between end of apprenticeship and time of interview, size of the current firm and the unemployment rate in the occupation[19]. Further variables such as ability and match quality proxies as well as variables used for identification purposes are discussed in the next section. Table 5.A1 in the appendix contains variable definitions, and table 5.A2 presents descriptive statistics for all variables.

[19] Brunner and Kuhn (2014), and others cited there, have shown that unemployment at the time of labor market entry has a substantial effect on individuals' wages.

5.4 Estimation Strategy

5.4.1 Operationalization of the hypotheses

To estimate the effects of firm-specific and occupation-specific components of human capital, we analyse the wage effects of changing firm within occupation (firm movers) and changing firm across occupation (occupation changers), as opposed to staying in the training firm and the occupation (stayers). These wage differentials can, in principle, be estimated by an OLS log wage ($\ln w_i$) regression for person i:

$$\ln w_i \quad = \quad \beta_0 + \beta^{fm} fm_i + \beta^{oc} oc_i + x_i \beta^x + u_i \tag{5.1}$$

Our main hypotheses pertain to the wage effect of mobility.[20] Due to human capital losses, we expect the coefficients $\hat{\beta}^{fm}$ and $\hat{\beta}^{oc}$ for the firm mover (fm) and occupation changer (oc) dummy variables to be negative, with a larger wage loss for occupation changers than for firm movers. x is a vector of covariates, and u is the error term.

5.4.2 Potential sources of bias in OLS estimation

The OLS dummy coefficients may be biased due to four reasons: (i) heterogeneous occupation changes, (ii) optimizing behavior in job search, (iii) heterogeneity in worker-firm and worker-occupation matches and (iv) heterogeneity in worker ability. We discuss the likely direction of the biases introduced in OLS estimation, as well as our strategy to address these endogeneity issues.

First, the transferability of human capital from one occupation to another may not be constant. For instance, an electrician might still make some use of his technical skills as a car mechanic, but not as an office clerk. Thus, we would like the OLS coefficient of the occupational change dummy in equation (5.1) to identify the mean loss of human capital between two different occupations due to occupation-specific human capital. It does not identify a mean effect if occupation changers are more likely to choose a new occupation that

[20] Goeggel and Zwick (2012) used the 'wage mark-up' between the last apprenticeship wage and the first skilled wage as dependent variable, arguing that higher quality training firms pay higher apprenticeship wages. However, human capital theory predicts that trainees accept lower training wages in return for receiving general human capital. Lower wages might thus indicate a higher training intensity. Muehlemann et al. (2013) indeed find that Swiss firms whose staff provides more training hours to trainees pay lower apprenticeship wages. The wage mark-up is then not suited to measure human capital transferability. Instead, we include apprenticeship wage as control in our robustness checks.

allows a high transfer of human capital from their learned occupation. In this case, occupation changers transfer more of their skills than would be the case if changes were purely random. The OLS coefficient for occupation change will be biased upwards and underestimate the wage penalty for changing occupation.

Second, apprentices have to search for jobs as their apprenticeship contracts end. In search theory (Mortensen, 1986; Rogerson et al., 2005), the focus lies on the endogenous nature of inter-firm and occupational mobility with respect to wages. Search theory assumes that workers trade off the gains from accepting a job that offers a given wage with the expected gains from continuing to search and waiting for a higher wage offer. Apprentices will search for a post-training job and compare wage offers from the training firm, from other firms within the same training occupation and from firms in other occupations altogether. Occupational or inter-firm changes are realized when the wage offer in another firm or occupation exceeds the asking wage of the graduating apprentices. Because voluntary changers tend to benefit from mobility, their wage loss due to specific human capital will be underestimated.

Third, during the period of training, apprentices experience the quality of their match with the training firm and with the occupation. Matching theory (Jovanovic, 1979) highlights the idea that job mobility enhances efficiency when it improves the quality of worker-firm or worker-occupation matches. Apprentices might thus voluntarily change firm or occupation after training to dissolve a bad match. It is unclear whether these apprentices will have a higher or lower average match quality after their change than stayers, and whether match quality introduces any bias. Stevens (2003) convincingly argues that stayers will have below-average matches that they do not resolve because they would lose firm-specific returns. In the presence of returns to specific human capital, match quality differences might thus introduce an upward bias in the OLS coefficients for the firm mover and occupation changer dummies. We would underestimate the wage loss due to specific human capital.

Fourth, the literature suggests that mobility in the labor market depends on the unobserved ability of individuals. If abler apprentices receive better offers from outside firms, selection into mobility could be positive. However, the predominant case discussed in the literature is adverse selection. Gibbons and Katz (1991) analysed adverse selection in job changes when the ability of individuals is not observable by outside employers. In the model developed and tested by Acemoglu and Pischke (1998), apprenticeships serve as screening devices for firms to learn about the most able apprentices and retain them after graduation, while not retaining those that fail to attain a certain ability threshold. This leads to involuntary mobility of adversely selected graduates. In such a case, OLS overestimates the wage penalty for changers, as firm and occupation changers are less able on average.

The first three of the four sources of endogeneity discussed are presumed to induce an upward bias in the OLS regression estimates of our mobility dummy variables, which results in underestimation of the wage loss that would occur when an apprentice moves randomly (in a thought experiment) from one occupation or firm to another. If the size of a (fourth) possible bias due to ability sorting is small compared to the size of the other biases discussed, then OLS regression provides a lower bound for the causal wage penalty resulting from firm and occupation changes and hence for the returns to specific human capital.

5.4.3 Estimation Setup

Our first estimation model for equation (5.1) is OLS regression, augmented with observable proxy variables to account for some of the possible sources of bias. The second type of models are endogenous treatment regressions that account for the endogeneity problems.

According to the third and fourth sources of endogeneity discussed in the preceding section, the fm and oc dummy variables depend on individuals' unobserved abilities and match qualities in a sample with endogenously determined firm and occupation changes. With respect to ability, we control for individuals' grades in the final examination at the end of apprenticeship training, reading literacy test scores at age 15 (before training) from the international PISA 2000 survey[21], PISA mathematical self-concept, self-efficacy in work tasks during apprenticeship, extrinsic motivation, vocational baccalaureate, and previous school type on lower-secondary level. This proxy solution assumes that the unobserved, true ability and the firm mover and occupation change dummy variables are uncorrelated, conditional on the mentioned proxy variables. With respect to the quality of the trainee-firm and trainee-occupation matches, we control for the apprentices' statement whether they want to stay with the training firm after training, whether they like to perform the typical tasks of their apprenticeship occupations, for their assessments of the perspectives in the occupation and of the quality of their supervisor, all measured during training at least one year before graduation.

If the proxies for ability and match qualities are good proxies that eliminate or greatly

[21] Contrary to reading performance, PISA 2000 tested mathematics performance only for a subgroup of pupils.

reduce ability bias, the remaining two sources of endogeneity will bias the OLS coefficients of mobility upwards and underestimate the wage loss due to specific human capital. Hence, OLS provides a lower bound on the wage differentials. We treat the OLS coefficients as baseline results and address endogenous mobility by means of an endogenous treatment model (Heckman, 1978; Vella, 1998; Vella and Verbeek, 1999).

In our case, the endogenous regressor is multinomial instead of binary (on multiple treatments, see Lee, 1983; Dubin and McFadden, 1984; Dahl, 2002). It is a variable with three categories (i.e., two dummy variables) for firm stayers, firm movers and occupation changers. The first equation of the model consists of a discrete choice model for mobility decisions, and the second equation is a wage regression.

$$y_{ij}^* = \gamma_j^0 + x_i\gamma_j^x + z_i\gamma_j^z + v_{ij} \qquad j = st,fm,oc \tag{5.2}$$

$$\ln w_i = \beta_0 + \beta^{fm} fm_i + \beta^{oc} oc_i + x_i\beta^x + l_i\lambda + u_i \tag{5.3}$$

Equation (5.2) describes the choice of apprentices among three mobility alternatives based on the individual latent 'utilities' y_{ij}^*. The observed, optimal choice enters equation (5.3) as dummy variables fm and oc. x_i represents controls included in both equations, z_i represents variables that influence the mobility decision but do not influence wages. l_i represents the selection effect and will be derived from the choice equation, as explained below.

We use two different approaches to estimate equations (5.2) and (5.3). Both model the choice equation (5.2) as a multinomial logit, and the log wage equation (5.3) as a linear regression. Deb and Trivedi (2006) propose a simulated maximum likelihood method to estimate the joint distribution of endogenous treatment and outcome using a latent factor structure. Their model includes two lambda coefficients (for three choice categories) in the outcome equation that account for the selection effect.

We compare the results of the Deb and Trivedi (DT) method with those of a standard two-step control function (CF) approach. We derive Mills ratios from a first-stage multinomial logit, applying a generalized version of the Dubin and McFadden (1984) approach. Bourguignon et al. (2007) showed that this version ('DMF1' in their paper) performs well in Monte Carlo simulations of selection bias correction models and is their preferred option for small samples. The DMF1 approach involves one control function per choice alternative in the first stage, i.e., l_i includes three Mills ratios in wage regression (5.3). We assume that the treatment effect operates only through the intercept such that the coefficient vector β^x is the same for all groups (as in DT).[22] The two-step method has the advantage that it does not rely on simulation. Two-step methods are, however, inefficient and may result in inflated standard errors.

In both methods discussed, nonlinear functional forms identify the parameters even if the set of variables in both equations is identical. For more robust identification, we include variables (z_i) in the mobility equation (5.2) that are excluded from wage regression (5.3). If these variables are correlated with mobility but uncorrelated with wages (after controlling for x_i), the model identifies the causal effect of mobility on wages.

Ideally, these variables reflect exogenous variation in the demand for apprenticeship graduates at firm and occupation level. In the absence of direct measures in the data, we use matched information from other data sources to create measures at the level of the local labor market[23] (and industry) for average inter-firm and occupational mobility in the workforce and for employment growth within local industry. These measures for local labor market thickness or tightness will influence mobility behavior at the individual level. We use three such variables, for which we assume no correlation with realized wages. The first variable is the average regional quit rate of apprentices in the relevant industry. We calculated regional, industry-specific quit rates from two Swiss surveys on costs and benefits of firms from apprenticeship training that contain information on the percentage of apprentices staying with their training firm after training.[24] The second variable is the regional, occupation-specific share of workers up to age 25 that do not work in their learned occupation according to the Swiss population census 2000. The third variable is due to Neal (1995) who used level of employment and employment growth in industries as variables affecting switching behavior. As employment levels are not significantly related to mobility in our sample, we include only the employment growth variable. We expect that in industries with higher growth, appren-

[22] Due to the limited sample size, we are not able to allow for differences in the β^x coefficients across mobility groups as would be possible in a full switching regression.

[23] Local labor markets were defined as follows: individual regions encompass all municipalities that can be reached within 30 minutes by car from the apprentice's residential municipality. The computation is based on a distance matrix provided by the Federal Statistical Office (FSO). According to the Federal population census, the average commuting time in Switzerland was 20.1 minutes in 2000 and increased by less than 2 minutes since 1970 (Scherer et al., 2010). At the time of graduation, apprentices typically live at their parents' homes and are not very mobile. A distance of 30 minutes by car seems therefore a realistic radius.

[24] We thank Samuel Muehlemann and Stefan Wolter for the permission to use the cost-benefit data 2000 and 2004 (see Schweri et al., 2003, and Muehlemann et al., 2007). We use the pooled data set containing 4729 training firms. To account for possible outliers in quit rates, we use a dummy indicating a quit rate above the sample average.

tices are more likely to receive a job offer by their training firm and are more likely to stay. Vice versa, they are forced to leave the firm in the case of a non-growing industry, because the firm might not want to grow by hiring the former apprentice as a skilled worker.[25] We calculated employment growth from the firm census data. We matched all regional variables to the apprentices based upon their place of residence and occupation / industry during apprenticeship training. The relevant individual regions encompass all municipalities reachable within thirty minutes by car from the apprentice's residential municipality. We have thus created a specific region for every community in the data (see table 5.A1 in the appendix for more details).[26]

[25] Some authors use information on firm closures or sharp declines in firm size to analyse the effects of mobility (Fitzenberger and Spitz, 2004; Dustmann and Meghir, 2005). We do not find a significant effect of firm size reductions on mobility and there is no suitable data available on the former.

[26] Switzerland is composed of about 2700 communities. Apprentices in the sample lived in 390 different communities.

5.5 Results

5.5.1 Main findings

The first OLS wage regression in table 5.3 (second column) does not yet include ability or match quality proxies. The results show no significant difference in wages between stayers and firm movers, but wages for occupation changers are lower by 5.2% compared to stayers, and the difference is statistically significant.

We add ability and match quality proxy variables in the OLS wage regression in the third column. The coefficients of the mobility dummies remain very similar, with occupation changers earning 4.5% less than firm stayers. Looking at the new covariates, we see that the PISA literacy test score at age 15 is significant: an increase by one standard deviation on the international PISA scale, i.e. 100 score points, increases wages by 1.4%. We do not find a significant interaction effect between mobility and test score (results not shown). Individuals with higher mathematical self-concept, higher self-efficacy in work tasks and higher extrinsic motivation also earn significantly more, which shows that these ability and personality variables measured during training or before are relevant for later labor market outcomes. Other significant effects are in line with results well known from the literature: women earn less, employees in larger firms earn more (see Barth and Dale-Olsen, 2011), and there are wage differences across language regions and occupations.[27]

The multinomial logit results in the last three columns of table 5.3 allow us to assess the selection process into firm and occupational mobility. We present average marginal effects.

Many ability variables have no significant influence on mobility; yet, as the PISA reading test score increases by one standard deviation, the probability of occupation change reduces by 2.8 percentage points. Trainees that earn a vocational baccalaureate during their apprenticeship period are more likely to stay with their training firm. Most matching variables show highly significant and substantial effects: A positive assessment of the career prospects in the learned occupation and a high perceived quality of training both reduce the probability of changing occupation. Wanting to stay in the training firm strongly increases the probability of working in the training firm one year after training.

[27] Controlling for 13 industries using dummy variables does not change the results.

Table 5.3: Wage estimations: OLS, Deb-Trivedi (DT) and control function (CF) estimates; first stage mlogit

	ln(wage) OLS 1	ln(wage) OLS 2	ln(wage) DT model	ln(wage) CF estim.	MLOGIT Average marginal effects		
					Firm Mover	Occ.change	Stayer
Firm mover	−0.001	0.001	0.001	0.022			
	(0.010)	(0.011)	(0.021)	(0.036)			
Occupational change	−0.052***	−0.045**	−0.094***	−0.130**			
	(0.018)	(0.018)	(0.022)	(0.059)			
PISA test score reading literacy		0.014**	0.013**	0.012	0.017	−0.028**	0.011
		(0.006)	(0.006)	(0.008)	−0.021	−0.012	−0.02
Mathematics self-concept		0.011**	0.011**	0.012**	0.002	0.011	−0.012
		(0.005)	(0.005)	(0.006)	(0.018)	(0.011)	(0.017)
High GPA in apprent. training		0.006	0.005	0.003	0.049	−0.014	−0.036
		(0.011)	(0.011)	(0.011)	(0.037)	(0.024)	(0.035)
Self-efficacy in work tasks		0.025**	0.026***	0.026**	−0.022	0.026	−0.004
		(0.010)	(0.010)	(0.011)	(0.034)	(0.020)	(0.032)
Extrinsic motivation		0.019**	0.018**	0.018	−0.026	−0.034*	0.060*
		(0.010)	(0.009)	(0.012)	(0.032)	(0.018)	(0.031)
Vocational baccalaureate		−0.018	−0.018	−0.017	−0.064*	−0.023	0.087**
		(0.011)	(0.011)	(0.012)	(0.038)	(0.025)	(0.036)
High track in lower-sec. school		0.015	0.017	0.018	−0.082**	0.050**	0.032
		(0.011)	(0.011)	(0.011)	(0.037)	(0.022)	(0.036)
Likes tasks of training occupation		−0.006	−0.007	−0.010	0.027	−0.017	−0.010
		(0.008)	(0.008)	(0.008)	(0.026)	(0.014)	(0.025)
Wants to work in training firm		0.005	0.005	0.011	−0.204***	0.006	0.198***
		(0.010)	(0.011)	(0.013)	(0.030)	(0.018)	(0.029)
Perspectives in occupation		0.001	−0.001	−0.002	0.024	−0.032***	0.008
		(0.006)	(0.006)	(0.006)	(0.020)	(0.012)	(0.019)
Perceived quality of supervisor		0.010	0.008	0.006	0.000	−0.049***	0.048*
		(0.009)	(0.009)	(0.009)	(0.030)	(0.017)	(0.028)
Female	−0.041***	−0.033***	−0.033***	−0.035***	0.007	−0.006	−0.001
	(0.011)	(0.012)	(0.011)	(0.012)	(0.040)	(0.023)	(0.038)
Immigrant: second-generation	0.025	0.027	0.026	0.025	0.049	−0.019	−0.029
	(0.018)	(0.018)	(0.017)	(0.023)	(0.060)	(0.033)	(0.058)
Immigrant: first-generation	−0.006	−0.008	−0.01	−0.009	−0.075	−0.019	0.093*
	(0.016)	(0.016)	(0.016)	(0.019)	(0.057)	(0.033)	(0.054)
French-speaking part of CH	0.018	0.029*	0.029*	0.026	−0.042	0.028	0.014
	(0.012)	(0.017)	(0.016)	(0.019)	(0.059)	(0.035)	(0.056)
Italian-speaking part of CH	−0.099***	−0.084***	−0.080***	−0.078***	−0.081	0.089**	−0.008
	(0.020)	(0.020)	(0.020)	(0.024)	(0.070)	(0.038)	(0.068)
Parental educ.: upper-secondary	−0.003	−0.005	−0.004	−0.004	−0.015	0.015	0.000
	(0.010)	(0.010)	(0.010)	(0.011)	(0.035)	(0.020)	(0.033)
Parental educ.: tertiary	−0.003	−0.005	−0.008	−0.011	0.035	−0.055**	0.019
	(0.011)	(0.011)	(0.011)	(0.012)	(0.040)	(0.026)	(0.038)
Size of training firm: 11-100	0.012	0.010	0.011	0.012	−0.079**	0.007	0.072**
	(0.010)	(0.010)	(0.010)	(0.011)	(0.033)	(0.019)	(0.032)
Size of training firm: 100+	0.014	0.013	0.014	0.016	−0.086*	0.008	0.078*
	(0.014)	(0.014)	(0.014)	(0.014)	(0.049)	(0.028)	(0.047)
Size of current firm: 11-100	0.018*	0.020*	0.021**	0.021**	−0.050	0.013	0.037
	(0.010)	(0.010)	(0.010)	(0.011)	(0.034)	(0.021)	(0.033)
Size of current firm: 100+	0.058***	0.061***	0.063***	0.065***	−0.137***	0.023	0.114**
	(0.014)	(0.014)	(0.014)	(0.016)	(0.047)	(0.026)	(0.045)
Unemployment rate	−0.006	−0.008*	−0.007	−0.007	−0.004	0.008	−0.004
	(0.005)	(0.005)	(0.004)	(0.006)	(0.015)	(0.008)	(0.015)
Training occupation (duration):							
Agriculture and Forestry (3)	−0.125***	−0.119***	−0.112***	−0.110***	−0.051	0.066*	−0.015
	(0.022)	(0.024)	(0.023)	(0.032)	(0.081)	(0.040)	(0.080)
Production, Manufacturing (3)	−0.036**	−0.018	−0.016	−0.018	−0.043	0.026	0.017
	(0.018)	(0.019)	(0.019)	(0.021)	(0.065)	(0.035)	(0.061)
Production, Manufacturing (4)	−0.018	−0.011	−0.011	−0.013	0.002	−0.051	0.049
	(0.020)	(0.021)	(0.021)	(0.021)	(0.076)	(0.044)	(0.071)
Technicians, IT (4)	0.005	0.010	0.004	0.003	−0.019	−0.153**	0.172***
	(0.017)	(0.018)	(0.017)	(0.018)	(0.074)	(0.064)	(0.065)
Construction (3)	0.005	0.029	0.029	0.030	−0.133	0.003	0.130
	(0.022)	(0.024)	(0.023)	(0.025)	(0.088)	(0.046)	(0.083)
Construction (4)	−0.002	0.012	0.009	0.009	−0.123	−0.056	0.178**
	(0.026)	(0.026)	(0.026)	(0.034)	(0.099)	(0.056)	(0.090)
Retail and wholesale, transport (2)	−0.157***	−0.139***	−0.139***	−0.138***	−0.162**	−0.010	0.172**
	(0.019)	(0.020)	(0.020)	(0.022)	(0.072)	(0.042)	(0.068)

— *continued on next page …* —

— ... continued from previous page —

	ln(wage) OLS 1	ln(wage) OLS 2	ln(wage) DT model	ln(wage) CF estim.	MLOGIT Average marginal effects		
					Firm Mover	Occ.change	Stayer
Retail and wholesale, transport (3)	−0.043**	−0.034*	−0.033*	−0.029*	−0.199***	0.007	0.193***
	(0.018)	(0.019)	(0.018)	(0.016)	(0.066)	(0.039)	(0.060)
Catering and restaurant (2)	−0.097**	−0.077*	−0.077**	−0.074**	−0.118	0.001	0.117
	(0.039)	(0.040)	(0.039)	(0.035)	(0.129)	(0.071)	(0.128)
Catering and restaurant (3)	−0.091***	−0.074***	−0.074***	−0.076***	0.024	0.028	−0.051
	(0.018)	(0.020)	(0.019)	(0.023)	(0.069)	(0.038)	(0.068)
Office worker (2)	−0.066*	−0.068*	−0.058	−0.045	−0.337**	0.148**	0.189
	(0.040)	(0.041)	(0.040)	(0.048)	(0.165)	(0.060)	(0.148)
Health and welfare (3,4)	−0.021	−0.014	−0.013	−0.015	−0.023	0.014	0.009
	(0.018)	(0.019)	(0.018)	(0.018)	(0.066)	(0.047)	(0.061)
Tenure in month at current job	0.002	0.002*	0.002	0.003	−0.008	0.000	0.009*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)	(0.003)	(0.005)
Time between graduation & job	0.002	0.003	0.003	0.004	0.007	0.006	−0.013
	(0.003)	(0.003)	(0.003)	(0.003)	(0.010)	(0.005)	(0.010)
Spells between graduation & job:							
Any unemployment spell	−0.020	−0.020	−0.017	−0.020	0.263***	0.051**	−0.314***
	(0.015)	(0.015)	(0.015)	(0.020)	(0.054)	(0.023)	(0.057)
Any military service spell	−0.008	−0.017	−0.013	−0.012	0.028	0.064	−0.091
	(0.025)	(0.025)	(0.024)	(0.029)	(0.083)	(0.039)	(0.082)
Any travelling / at home spell	0.010	0.004	0.004	−0.000	0.169***	0.023	−0.192***
	(0.016)	(0.016)	(0.016)	(0.018)	(0.053)	(0.027)	(0.053)
Any educational activity spell	−0.005	−0.001	−0.000	0.002	−0.022	0.030	−0.008
	(0.019)	(0.019)	(0.019)	(0.026)	(0.067)	(0.030)	(0.069)
Any other job spell	0.013	0.010	0.008	0.005	0.142***	0.006	−0.147***
	(0.016)	(0.016)	(0.016)	(0.019)	(0.053)	(0.025)	(0.054)
Any other spell (missing info)	−0.012	−0.009	−0.009	−0.008	−0.022	0.005	0.017
	(0.017)	(0.017)	(0.017)	(0.017)	(0.057)	(0.036)	(0.051)
High regional quit rate					0.124***	−0.017	−0.107***
					(0.031)	(0.019)	(0.030)
Regional rate of occupation change					0.225	0.298*	−0.523
					(0.349)	(0.180)	(0.342)
Regional employment growth					−0.272	−0.188	0.461**
					(0.210)	(0.137)	(0.198)
Lambda / Mills 'Firm mover'			−0.002	−0.004			
			(0.023)	(0.017)			
Lambda / Mills 'Occ. change'			0.062***	0.047**			
			(0.016)	(0.020)			
Mills 'Firm stayer'				0.011			
				(0.013)			
Constant	3.163***	2.986***	3.001***	3.006***			
	(0.025)	(0.056)	(0.055)	(0.073)			
N	878	878	878	878	878		
R-sq / Pseudo R-sq	0.282	0.311		0.318	0.334		

* p<0.10, ** p<0.05, *** p<0.01
CF (control function) estimates: bootstrapped standard errors (2000 replications) to account for generated regressors.
Reference group: firm stayer; no professional baccalaureate; lower track in lower-secondary school; male; Swiss; German speaking part of Switzerland; highest parental education: compulsory school; firm size: 1-10 fulltime-equivalents; training occupation: commercial employee (3); no intervening spell between graduation and current job.

Overall, we find only limited evidence for adverse selection on ability in post-training mobility. A possible explanation is that positive and adverse selection exist at the same time, such that the mean effect of ability on mobility is rather small. Occupation changers have a somewhat lower PISA score and are less likely to earn a vocational baccalaureate, but the experience during training as represented by the match quality variables is more influential for mobility decisions than ability. The match quality variables, however, do not significantly influence wages in the OLS regression. Accordingly, ability and match quality variables are important to understand mobility decisions and wage setting, but exert little influence on the wage differentials between stayers, firm movers and occupation changers.

If our ability and match quality variables are good proxies, the remaining bias in OLS (from heterogeneous effects and endogenously chosen mobility) should be an upward bias. Thus, the OLS coefficient of 4.5% wage loss for occupation changers provides a lower bound (in absolute terms) for the short-term wage loss due to occupation-specific human capital acquired during apprenticeship training.

The last columns of table 5.3 show mlogit results for the mobility equation that acts as first stage for the endogenous treatment models. The three variables excluded from the (second stage) wage equation are jointly significant ($p < 0.0001$) in a Wald test. The regional quit rate and regional employment growth affect individuals' probability to stay in the firm significantly; the regional rate of occupation change increases individuals' probability to change occupation.[28]

The Deb and Trivedi (2006) endogenous treatment model accounts for the endogeneity of job and occupation changes. For most covariates, the 'DT' results in column 4 of table 5.3 are very similar to the OLS results. Yet, the wage loss of an occupation changer relative to a stayer increases from 4.5% to 9.4%. This result is in line with our prediction that—controlled for ability and match quality—OLS underestimates the wage loss due to switches out of the learned occupation. The lambda coefficient for occupation change is significantly positive, indicating that selection into mobility is not random, conditional on the observed x_i-variables.

The point estimate for firm movers is almost equal to zero and insignificant. Two interpretations are possible: Either there is no substantial firm-specific skills component on average, or the training firm acquires all the returns on firm-specific training investments. Irrespective of the interpretation, there is no evidence that firms distort training contents towards firm-specific skills in an effort to retain workers. Of course, training firms might offer contracts with modest post-training wages and back-loaded compensation that reward tenure as a strategy to retain apprentices (Stevens, 2004). Firm-specific human capital would thus matter for lifetime earnings. We would need data on apprentices with many years of post-training labor market experience to analyse such behavior by training firms.

[28] Concerning the variable 'regional rate of occupation change', its marginal effect is significant at the 10% level for occupation change against the two other alternatives (see table 5.3); however, its coefficient in the first stage DT mixed mlogit is significant at the 5% level between the alternatives staying and changing occupation (see table 5.A3 in the appendix).

5.5.2 Robustness Checks

First, we compare the results of the Deb and Trivedi approach with the two-step control function approach described in section 5.4.3. The fifth column in table 5.3 ('CF') shows that occupation change is again associated with a higher wage loss than in the OLS models, whereas firm stayers and movers show no significant wage difference. The point estimate of occupation change is larger than estimated by the DT-Method, however, the qualitative results remain the same. The two-step procedure comes at the cost of markedly higher (bootstrapped) standard errors leading to an imprecise estimation of the mobility effects.

Second, we redefine the occupation change dummy variable using 5-digit, 3-digit and 1-digit definitions in place of the 2-digit definition of an 'occupation' that we deem most adequate (see section 5.3). These different definitions of occupation change imply different delimitations between firm- and occupation-specific human capital in our setting (see table 5.1). Wage losses are smallest (6.3% in DT estimation, see table 5.4) when occupational changes mean small (5-digit) changes, which is what we would expect from human capital theory.

Table 5.4: Overview of key results obtained by varying the definition of occupational change

	(1) 5-digit	(2) 3-digit	(3) 2-digit	(4) 1-digit
Results of wage regression OLS				
Firm mover	−0.001	−0.000	0.001	−0.000
	(0.011)	(0.011)	(0.011)	(0.010)
Occupational change	−0.018	−0.038**	−0.045**	−0.044**
	(0.015)	(0.018)	(0.018)	(0.019)
Results of Deb-Trivedi treatment regression				
Firm mover	−0.006	−0.001	0.001	−0.000
	(0.021)	(0.021)	(0.021)	(0.020)
Occupational change	−0.063***	−0.083***	−0.094***	−0.093***
	(0.023)	(0.023)	(0.022)	(0.022)
Lambda 'firm mover'	0.004	0.000	−0.002	−0.001
	(0.023)	(0.023)	(0.023)	(0.023)
Lambda 'occ. change'	0.056**	0.058***	0.062***	0.062***
	(0.023)	(0.019)	(0.016)	(0.016)

* p<0.10, ** p<0.05, *** p<0.01
Controlled for the full set of covariates as in table 5.3
The definition in (3) is the one used throughout our analysis.
Note: The 4-digit-level coincides with the 3-digit-level nearly everywhere in the nomenclature.

Third, we change the calculation of the regional variables excluded from the wage equation (z-variables). If we define a 45-minute travel-time radius from the apprentices' places of residence instead of 30 minutes, the results (not shown) are very similar to those in table 5.3. The occupation change dummy coefficient is virtually identical with -.093.

Fourth, we estimate the DT model using only two, one or none of the z-variables in the mobility equation, thus relying on non-linear functional form for identification. The results are very similar; in particular the occupation change dummy remains significant and between -9.0% and -9.5%.

Finally, if we include the last training wage as an additional control for unobserved heterogeneity in training (Euwals and Winkelmann, 2004), it is weakly significant in the DT wage regression with positive sign. The wage effects of mobility remain virtually the same.

5.6 Conclusions

Firm-based apprenticeships have been suspected of transferring an overly specific, narrow set of skills. The literature available on human capital transferability for Germany comes to heterogeneous results and is probably not generalizable to other countries; however, several studies have reported high transferability of skills learned in apprenticeship.

We show that in Switzerland with its lightly regulated labor market, firm mobility within one year after completing an apprenticeship is high (49 percent of all apprenticeship graduates, including occupation changers), whereas occupational mobility is limited (7 percent of all graduates). In OLS wage regressions, we find no significant differences in wages of firm stayers and firm movers within their learned occupation defined at the two-digit occupation level, even when we control for ability and match quality proxies. Firm movers with occupation change, however, earn almost 5 percent less. Applying a treatment regression approach to account for endogenous mobility decisions, we still find no wage effect for firm movers who stay within the occupation, but a negative wage differential of approximately 9 percent for occupation changers.

We conclude that within the occupational field of an apprenticeship, the human capital acquired is widely transferable to other firms after graduation. Transferability is more limited for changes beyond the occupational field: apprenticeship graduates that undertake such changes earn lower wages, with the magnitude of the wage cut depending on the distance between occupations. Adverse selection and the match quality with the training firm and the training occupation hardly affect these findings: while the match quality matters for mobility decisions in our estimations, it does not affect post-mobility wages. Conversely, we find that several ability variables have an effect on wages (in the usual direction); however, for the mobility decision, only PISA-test scores show some adverse selection into occupation change.

While firm-specific human capital does not seem to play a significant role for Swiss apprentices' early labor market wages, on average, occupation-specific human capital is an important component of apprenticeship training and accounts for a portion of the return to training.

Overall, Swiss apprenticeship curricula and quality assurances as defined by legal regulations seem to be successful in producing transferable skills and knowledge such that apprenticeships are not restricted to a very narrow set of skills. There is no evidence that training firms distort the regulated training content towards firm-specific components in an effort to reduce across-firm mobility of workers. High across-firm transferability of acquired skills is also in line with the findings of cost-benefit studies for Switzerland, which show that Swiss apprentices bear the full costs of training on average. Still, the provision of widely transferable human capital should be stressed in ongoing curricula reforms. Our findings on the importance of occupation-specific human capital highlight the need for continuing vocational training to continuously update or retrain skilled workers according to changing labor market needs. Otherwise, apprenticeships might reduce long-term flexibility compared to a general education system that does not require a choice of an occupation at the age of 16. At the same time, we may also interpret the importance of occupation-specific human capital as necessary condition for gains that accrue from specialization, both at the individual level and for the economy as a whole. Relatively high rates of (lifetime) returns to apprenticeships (Weber et al., 2001), compared to individuals that do not follow any post-compulsory education, support the view that the apprenticeship system is successful in supplying young adults with the skills demanded by Swiss firms. More empirical studies are required to analyse the relative long-term costs and benefits of education systems that rely on vocational education with early specialization.

5.A Appendix

Table 5.A1: Variable definition

Variable	Definition
Ln(hourly wage) in current job	Natural logarithm of hourly gross wage in the current job (one year after graduation).
Mobility	Two dummy variables. 'Firm mover' equals 1 if individual does not work in the training firm (but in the learned occupation) one year after graduation. 'Occupation change' equals 1 if individual works in occupation (and firm) different from the apprenticeship occupation (firm) one year after graduation. Reference category: Firm stayer.
PISA test score for reading literacy	Reading literacy test score from the PISA 2000 survey (OECD, 2002); standardized with the international mean (500) and standard deviation (100) for the estimations.
Mathematics self-concept (PISA)	Mathematics self-concept measures whether one thinks positively about one's mathematics abilities. Measured in PISA 2000 survey, continuous scale (min: -1.62, max: 1.74). See OECD (2002) for details. Plus dummy for missing information. Questions: I get good marks in mathematics / Mathematics is one of my best subjects / I have always done well in mathematics.
High GPA in apprenticeship training	Dummy equals 1 when average grade in the final apprenticeship examination is above 5 (on a scale of 4=sufficient to 6=excellent); 0 otherwise. Plus dummy for missing information.
Self-efficacy in work tasks	Self-efficacy in work tasks reported by individuals in second last year of apprenticeship training. Scale 1 (min) to 4 (max). Plus dummy for missing information. Questions: I can always manage to solve difficult problems if I try hard enough / I am confident that I could deal efficiently with unexpected events / Thanks to my resourcefulness, I know how to handle unforeseen situations / I can usually handle whatever comes my way (1=not at all true; 4=exactly true).
Extrinsic motivation	Extrinsic motivation reported by individuals in second last year of apprenticeship training, scale 1 (min) to 4 (max). Plus dummy for missing information. Questions: To earn a lot of money, a good wage / To have a secure job position / To have a position with career opportunities / To have a job which is recognized and respected by others (1=totally subordinate; 4=very important).
Vocational baccalaureate	Dummy equals 1 if apprentice obtained a vocational baccalaureate (additional schooling) while serving his apprenticeship; 0 otherwise.
High track lower-secondary school	Dummy equals 1 if apprentice attended a lower-secondary school track with extended requirements (i.e., before starting apprenticeship); 0 otherwise.
Likes tasks of training occupation	Variable indicating how much the apprentice likes to perform the pertinent tasks of the apprenticeship occupation. Reported during training, ranging from 1 (don't like them at all) to 4 (like them very much).
Wants to work in training firm	Dummy indicating whether the apprentice wants to work in his/her training firm after graduation. Reported in second last year of apprenticeship training, 1 equals yes (exactly true [4] or moderately true [3])
Perceived perspectives in learned occ.	Apprentices' agreement to the statement that he/she will be able to make a living working in the occupation which he/she is trained in. Reported during training, ranging from 1 (not at all true) to 4 (exactly true).
Perceived quality of supervisor	Apprentices' rating of the pedagogical and professional competences of his/her supervisor in the training firm; index based on eight questions, reported during training, from 1 (not true at all) to 4 (totally true). Questions: I am very pleased with my vocational instructor / My vocational instructor explains my tasks well / Usually my vocational instructor tells me if I did a good job or not / If I ask a question, my vocational instructor takes time to explain / My vocational instructor praises me for the things I do well / I have a good relationship with my vocational instructor / My vocational instructor knows his/her area of expertise very well.
Female	Equals 1 if female; 0 if male.
Nationality	Dummies representing 3 categories: 'Swiss' (trainee born in Switzerland with at least one parent born in Switzerland), 'second-generation immigrant' (trainee born in Switzerland but parents born outside Switzerland), 'first-generation immigrant' (trainee and parents foreign-born).
Language region	Dummies representing Swiss language regions (German, French, Italian).
Parental education	Dummies representing 3 categories of highest parental education: compulsory school, upper-secondary education, tertiary education. Plus dummy for missing information.
Size of training firm	Dummies for 3 firm size categories (firm = establishment).
Size of current firm	Dummies for 3 firm size categories (firm = establishment) and a dummy for missing information.
Unemployment rate	The mean unemployment rate in the corresponding language region in the industry the apprentice was trained in. Source: Swiss Labour Force Survey, pooled annual data for 1996 to 2006.

Tenure in month at the current job	Tenure in month at the current job, measured from job start (after graduation) until the (individual) interview date
Time between graduation and job	Number of month between apprenticeship graduation and job start, ranging from 0 to 12. Intervening spells are observed for 48% of the sample (25% of stayers, 60% of firm movers, 78% of occupational changers).
Any unemployment spell	Dummy equals 1 if the apprentice had any unemployment spell between graduation and current job
Any military service spell	Dummy equals 1 if the apprentice had any spell of serving military service between graduation and current job
Any travelling / at home spell	Dummy equals 1 if the apprentice had any spell of travelling or staying (longer) at home between graduation and current job
Any educational activity spell	Dummy equals 1 if the apprentice had any spell with educational activity (mostly language studies abroad) between graduation and current job
Any other job spell	Dummy equals 1 if the apprentice had any (short) job between graduation and current job
Any other spell (with missing info)	Dummy equals 1 if the apprentice had any spell with missing information on the activity between graduation and current job
Training occupation and duration	Dummies representing apprenticeship tracks and their duration. Training occupations are coded according to the nomenclature of occupations (Schweizerische Berufsnomenklatur SBN) at the 1-digit level and split according to the official duration of the corresponding apprenticeship track (2, 3 or 4 years).
High regional quit rate	Dummy indicating that the share of apprentices who leave their training firm after graduation is above average in this region and industry. Source: Cost-Benefit surveys, pooled data from 2000 and 2004.
Regional rate of occupation change	Regional[a)], occupation-specific share of workers up to age 25 that do not work in their learned occupation according to the Swiss population census 2000
Regional employment growth	The change in regional[a)] employment level in the corresponding industry (13 sections) between 2001 and 2005, i.e., between the time the apprentices started their training and the time they completed their training. Source: Swiss Firm Census.
	[a)] *Definition of regions: individual regions encompass all municipalities that can be reached within 20 minutes by car from the apprentice's residential municipality. The computation is based on a distance matrix provided by the Federal Statistical Office (FSO).*

Table 5.A2: Descriptive statistics (mean or %; std. dev.)

	Overall	Stayer	Mover	Occ.change
Total	878	484	331	63
Total percent	100.00	100.00	100.00	100.00
Hourly wage in current job (CHF)	23.01	23.31	22.86	21.74
	(3.17)	(3.04)	(3.10)	(4.07)
PISA test score for reading literacy	505.58	505.69	509.62	481.30
(here: unstandardized)	(75.39)	(77.88)	(70.08)	(83.58)
Mathematics self-concept (PISA)	0.11	0.18	0.03	0.12
	(0.85)	(0.86)	(0.84)	(0.87)
High GPA in apprenticeship training	21.87	20.76	23.98	17.47
Self-efficacy in work tasks	3.04	3.06	3.02	3.03
	(0.44)	(0.44)	(0.45)	(0.40)
Extrinsic motivation	3.31	3.35	3.29	3.22
	(0.45)	(0.45)	(0.45)	(0.46)
Vocational baccalaureate	23.01	28.35	17.71	15.87
High track lower-secondary school	63.90	66.74	61.04	60.32
Likes tasks of training occupation	3.43	3.49	3.41	3.11
	(0.64)	(0.59)	(0.65)	(0.82)
Wants to work in training firm	0.64	0.80	0.47	0.48
Does not want to work in training firm	0.36	0.20	0.53	0.52
Perceived perspectives in learned	3.16	3.18	3.19	2.83
occupation	(0.89)	(0.79)	(0.78)	(0.87)
Perceived quality of supervisor	3.28	3.34	3.24	3.10
	(0.50)	(0.47)	(0.51)	(0.57)
Female	58.54	55.36	63.49	52.38
Swiss	85.65	84.82	87.19	82.54
Immigrant: second-generation	6.26	5.58	6.54	9.52
Immigrant: first-generation	8.09	9.60	6.27	7.94
German-speaking part of Switzerland	69.13	72.77	65.94	61.90
French-speaking part of Switzerland	23.58	20.31	27.52	23.81
Italian-speaking part of Switzerland	7.29	6.92	6.54	14.29
Parental education: compulsory schooling	30.98	30.58	30.79	34.92
Parental education: upper-sec. education	39.29	39.06	38.15	47.62
Parental education: tertiary education	26.42	27.23	27.25	15.87
Parental education: no information	3.30	3.13	3.81	1.59
Regional unemployment rate per industry	3.49	3.04	3.54	3.82
	(1.46)	(1.38)	(1.44)	(1.97)
Tenure in month at the current job	6.95	8.02	5.83	5.82
	(3.61)	(3.31)	(3.47)	(4.18)
Time between graduation and job	2.17	0.92	3.36	4.19
	(3.04)	(2.12)	(3.24)	(3.52)
Training occupation (duration in years):				
Agriculture and forestry (3)	4.90	2.68	6.27	12.70
Production, manufacturing (3)	7.06	5.58	7.90	12.70
Production, manufacturing (4)	5.81	5.58	6.27	4.76
Technicians, IT (4)	9.34	12.28	7.08	1.59
Construction (3)	4.44	5.80	2.72	4.76
Construction (4)	3.19	4.24	1.91	3.17
Retail and wholesale, transport (2)	6.15	7.59	4.63	4.76
Retail and wholesale, transport (3)	6.95	8.93	4.63	6.35
Catering and restaurant (2)	1.25	0.67	1.91	1.59
Catering and restaurant (3)	8.31	4.24	12.81	11.11
Office worker (2)	1.14	1.34	0.54	3.17
Commercial employee (3)	33.37	34.60	32.43	30.16
Health and welfare (3, 4)	8.09	6.47	10.90	3.17
Size of training firm: 1-10	40.66	33.93	49.05	39.68
Size of training firm: 11-100	43.39	46.65	38.96	46.03
Size of training firm: 100+	15.95	19.42	11.99	14.29
Size of current firm: 1-10	38.72	33.04	46.32	34.92
Size of current firm: 11-100	40.32	43.30	37.33	36.51
Size of current firm: 100+	18.11	23.21	11.44	20.63
Size of current firm: Missing	2.85	0.45	4.90	7.94
Any unemployment spell	14.46	2.68	25.34	34.92
Any military service spell	3.99	2.46	4.90	9.52
Any travelling / at home spell	9.45	5.36	13.90	12.70
Any educational activity spell	7.97	3.35	11.99	17.46
Any other job spell	20.39	7.81	33.51	33.33
Any other spell (with missing information)	7.40	8.48	6.27	6.35
High regional quit rate	53.19	43.30	65.94	49.21
Regional rate of occupation change	21.96	22.17	21.56	22.79
	(5.83)	(5.57)	(5.99)	(6.58)
Regional employment growth in industry	−0.02	−0.02	−0.02	−0.03
	(0.08)	(0.08)	(0.07)	(0.07)

Table 5.A3: Wage estimations: OLS, Deb-Trivedi (DT); first stage: DT mixed mlogit

	ln(wage) OLS 2	ln(wage) DT model	MMLOGIT coefficients (base category=stayer)	
			Firm Mover	Occ.change
Firm mover	0.001	0.001		
	(0.011)	(0.021)		
Occupational change	−0.045**	−0.094***		
	(0.018)	(0.022)		
PISA test score reading literacy	0.014**	0.013**	0.015	−0.624**
	(0.006)	(0.006)	(0.166)	(0.276)
High GPA in apprent. Training	0.006	0.005	0.343	−0.076
	(0.011)	(0.011)	(0.285)	(0.542)
Mathematics self-concept	0.011**	0.011**	0.059	0.186
	(0.005)	(0.005)	(0.140)	(0.256)
Self-efficacy in work tasks	0.025**	0.026***	−0.061	0.566
	(0.010)	(0.010)	(0.265)	(0.470)
Extrinsic motivation	0.019**	0.018**	−0.364	−1.068**
	(0.010)	(0.009)	(0.256)	(0.426)
Professional baccalaureate	−0.018	−0.018	−0.668**	−0.975*
	(0.011)	(0.011)	(0.294)	(0.568)
High track in lower-sec. school	0.015	0.017	−0.425	0.769
	(0.011)	(0.011)	(0.292)	(0.511)
Likes tasks of training occupation	−0.006	−0.007	0.176	−0.284
	(0.008)	(0.008)	(0.207)	(0.339)
Wants to work in training firm	0.005	0.005	−1.687***	−0.967**
	(0.010)	(0.011)	(0.258)	(0.439)
percieved perspectives in training occupation	0.001	−0.001	0.032	−0.675**
	(0.006)	(0.006)	(0.153)	(0.269)
perceived quality of supervisor	0.010	0.008	−0.226	−1.200***
	(0.009)	(0.009)	(0.239)	(0.386)
Female	−0.033***	−0.033***	0.026	−0.054
	(0.012)	(0.011)	(0.312)	(0.544)
Immigrant: second-generation	0.027	0.026	0.286	−0.263
	(0.018)	(0.017)	(0.474)	(0.764)
Immigrant: first-generation	−0.008	−0.010	−0.708	−0.909
	(0.016)	(0.016)	(0.443)	(0.784)
French-speaking part of CH	0.029*	0.029*	−0.218	0.213
	(0.017)	(0.016)	(0.461)	(0.817)
Italian-speaking part of CH	−0.084***	−0.080***	−0.245	1.718*
	(0.020)	(0.020)	(0.556)	(0.923)
Parental educ.: upper-secondary	−0.005	−0.004	−0.046	0.328
	(0.010)	(0.010)	(0.272)	(0.459)
Parental educ.: tertiary	−0.005	−0.008	0.031	−1.133*
	(0.011)	(0.011)	(0.304)	(0.588)
Size of training firm: 11-100	0.010	0.011	−0.637**	−0.211
	(0.010)	(0.010)	(0.261)	(0.454)
Size of training firm: 100+	0.013	0.014	−0.679*	−0.215
	(0.014)	(0.014)	(0.385)	(0.641)
Size of current firm: 11-100	0.020*	0.021**	−0.362	0.133
	(0.010)	(0.010)	(0.266)	(0.494)
Size of current: 100+	0.061***	0.063***	−1.021***	−0.023
	(0.014)	(0.014)	(0.374)	(0.599)
Unemployment rate	−0.008*	−0.007	0.009	0.209
	(0.005)	(0.004)	(0.122)	(0.183)
Training occupation (duration):				
Agriculture and Forestry (3)	−0.119***	−0.112***	−0.119	1.378
	(0.024)	(0.023)	(0.655)	(0.941)
Production, Manufacturing (3)	−0.018	−0.016	−0.190	0.303
	(0.019)	(0.019)	(0.507)	(0.817)
Production, Manufacturing (4)	−0.011	−0.011	−0.208	−1.350
	(0.021)	(0.021)	(0.589)	(1.041)
Technicians, IT (4)	0.010	0.004	−0.903*	−3.973***
	(0.018)	(0.017)	(0.519)	(1.374)
Construction (3)	0.029	0.029	−1.100	−0.561
	(0.024)	(0.023)	(0.688)	(1.025)
Construction (4)	0.012	0.009	−1.379*	−2.087*
	(0.026)	(0.026)	(0.753)	(1.169)
Retail and wholesale, transport (2)	−0.139***	−0.139***	−1.355**	−1.476
	(0.020)	(0.020)	(0.558)	(0.990)
Retail and wholesale, transport (3)	−0.034*	−0.033*	−1.627***	−1.036

— continued on next page ... —

— ... *continued from previous page* —

	ln(wage) OLS 2	ln(wage) DT model	MMLOGIT coefficients (base category=stayer)	
			Firm Mover	Occ.change
	(0.019)	(0.018)	(0.510)	(0.901)
Catering and restaurant (2)	−0.077*	−0.077**	−0.930	−0.985
	(0.040)	(0.039)	(1.064)	(1.802)
Catering and restaurant (3)	−0.074***	−0.074***	0.364	0.585
	(0.020)	(0.019)	(0.552)	(0.917)
Office worker (2)	−0.068*	−0.058	−2.337*	1.819
	(0.041)	(0.040)	(1.267)	(1.324)
Health and welfare (3,4)	−0.014	−0.013	−0.100	0.287
	(0.019)	(0.018)	(0.491)	(1.025)
Tenure in month at current job	0.002*	0.002	−0.072*	−0.040
	(0.001)	(0.001)	(0.041)	(0.061)
Time between training & job (month)	0.003	0.003	0.088	0.206*
	(0.003)	(0.003)	(0.082)	(0.116)
Spells between graduation & job:				
Any unemployment spell	−0.020	−0.017	2.400***	2.453***
	(0.015)	(0.015)	(0.468)	(0.623)
Any military service spell	−0.017	−0.013	0.515	1.717*
	(0.025)	(0.024)	(0.671)	(0.975)
Any travelling / at home spell	0.004	0.004	1.603***	1.514**
	(0.016)	(0.016)	(0.442)	(0.661)
Any educational activity spell	−0.001	−0.000	−0.033	0.580
	(0.019)	(0.019)	(0.551)	(0.768)
Any job spell	0.010	0.008	1.211***	0.713
	(0.016)	(0.016)	(0.440)	(0.634)
Any other spell (with missing info)	−0.009	−0.009	−0.150	−0.246
	(0.017)	(0.017)	(0.424)	(0.847)
High regional quit rate			0.943***	0.118
			(0.251)	(0.430)
Regional rate of occupation change			3.276	8.603**
			(2.794)	(4.289)
Regional employment growth			−3.117*	−6.241**
			(1.618)	(3.074)
Lambda / Mills 'Firm mover'		−0.002		
		(0.023)		
Lambda / Mills 'Occupation change'		0.062***		
		(0.016)		
Constant	2.986***	3.001***	2.060	2.632
	(0.056)	(0.055)	(1.632)	(2.547)
N	878	878	878	

* p<0.10, ** p<0.05, *** p<0.01
Reference group: firm stayer; no professional baccalaureate; lower track in lower-secondary school; male; Swiss; German speaking part of Switzerland; highest parental education: compulsory school; firm size: 1-10 fulltime-equivalents; training occupation: commercial employee (3); no intervening spell between graduation and current job.

Bibliography

Acemoglu, Daron and Joern-Steffen Pischke (1998). Why do firms train? Theory and evidence. *Quarterly Journal of Economics 113*(1), 79–119.

Acemoglu, Daron and Jorn-Steffen Pischke (1999). Beyond Becker: Training in Imperfect Labour Markets. *Economic Journal 109*(453), F112–42.

Acemoglu, Daron and Jorn-Steffen Pischke (2000). Certification of training and training outcomes. *European Economic Review 44*(4-6), 917–927.

Aigner, Dennis J. and Glen G. Cain (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review 30*(2), 175–187.

Altonji, Joseph G. and Charles Pierret (1997). Employer Learning and the Signaling Value of Education. In I.Ohashi and T.Tachibanak (Eds.), *Industrial Relations, Incentives and Employment*, pp. 159–195. London: Macmillan Prese Lt.

Altonji, Joseph G. and Charles R. Pierret (2001). Employer Learning and Statistical Discrimination. *Quarterly Journal of Economics 116*, 313–350.

Angrist, Joshua D. and Alan B. Krueger (1999). Empirical strategies in labor economics. In O.Ashenfelter and D.Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 23, pp. 1277–1366. Elsevier.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010). Beyond Signaling and Human Capital: Education and the Revelation of Ability. *American Economic Journal: Applied Economics 2*(4), 76–104.

Arrow, Kenneth J. (1973). The theory of discrimination. In O.Ashenfelter and A.Rees (Eds.), *Labour Markets*. Princeton, NJ.: Princeton University Press.

Barth, Erling and Harald Dale-Olsen (2011). Employer size or skill group size effect on wages? *Industrial and Labor Relations Review 64*(2), 341–355.

Bassanini, Andrea, Alison L. Booth, Giorgio Brunello, Maria De Paola, and Edwin Leuven (2007). Workplace Training in Europe. In G.Brunello, P.Garibaldi, and E.Wasmer (Eds.), *Education and Training in Europe*, Chapter 8-13. Oxford: Oxford University Press.

Bauer, Thomas K. and John P. Haisken-DeNew (2001). Employer learning and the returns to schooling. *Labour Economics 8*(2), 161–180.

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

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 48*(3/4), 368–389.

Beicht, Ursula, Günter Walden, and Hermann Herget (2004). *Kosten und Nutzen der betrieblichen Berufsausbildung in Detuschland*. Berichte zur beruflichen Bildung (Heft 264). Bielefeld: Bertelsmann.

Bertschy, Kathrin, M. Alejandra Cattaneo, and Stefan C. Wolter (2009). PISA and the Transition into the Labour Market. *LABOUR 23*(1), 111–137.

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

Blatter, Marc, Samuel Mühlemann, Samuel Schenker, and Stefan C. Wolter (2012). Hiring Costs of Skilled Workers and the Supply of Firm-Provided Training. IZA Discussion Papers 6344, Institute for the Study of Labor (IZA).

Blatter, Marc, Samuel Muehlemann, Samuel Schenker, and Stefan C. Wolter (2015). Hiring costs for skilled workers and the supply of firm-provided training. *Oxford Economic Papers*.

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

Bougheas, Spiros and Yannis Georgellis (2004). Early Career Mobility and Earnings Profiles of German Apprentices: Theory and Empirical Evidence. *LABOUR 18*(2), 233–263.

Bourguignon, François, Martin Fournier, and Marc Gurgand (2007). Selection Bias Corrections Based On The Multinomial Logit Model: Monte Carlo Comparisons. *Journal of Economic Surveys 21*(1), 174–205.

Brand, Jennie E. and Yu Xie (2010). Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education. *American Sociological Review 75*, 273–302.

Brunello, Giorgio (2009). The effect of economic downturns on apprenticeships and initial workplace training: a review of the evidence. *Empirical Research in Vocational Education and Training 1*(2), 145–171.

Brunner, Beatrice and Andreas Kuhn (2014). The impact of labor market entry conditions on initial job assignment and wages. *Journal of Population Economics 27*(3), 705–738.

Buchholz, Sandra, Christian Imdorf, Sandra Hupka-Brunner, and Hans-Peter Blossfeld (2012). Sind leistungsschwache Jugendliche tatsächlich nicht ausbildungsfähig? Eine Längsschnittanalyse zur beruflichen Qualifizierung von Jugendlichen mit geringen kognitiven Kompetenzen im Nachbarland Schweiz. *Kölner Zeitschrift für Soziologie und Sozialpsychologie 64*(4), 701–727.

Caliendo, Marco and Sabine Kopeinig (2008). Some Practical Guidance For The Implementation Of Propensity Score Matching. *Journal of Economic Surveys 22*(1), 31–72.

Card, David (1999). The causal effect of education on earnings. In O.Ashenfelter and D.Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.

Clark, Damon (2000). How Transferable is German Apprenticeship Training? Paper presented at the iza seminar (14.11.2000), Nuffield College, Oxford and Centre for Economic Performance (CEP).

Clark, Damon (2001). Why Do German Firms Subsidize Apprenticeship Training? Tests of the Asymmetric Information and Mobility Cost Explanations. *Vierteljahreshefte zur Wirtschaftsforschung 70*, 102–106.

Clark, Damon (2002). Matching, Screening and Firm Investment in General Training: Theory and Evidence. 10th International Conference on Panel Data, Berlin, July 5-6, 2002 A2-4, International Conferences on Panel Data.

Clark, Damon and René Fahr (2002). The Promise of Workplace Training for Non-College Bound Youth: Theory and Evidence from German Apprenticeship. CEP Discussion Papers dp0518, Centre for Economic Performance, LSE.

Dahl, Gordon B. (2002). Mobility and the Return to Education: Testing a Roy Model with Multiple Markets. *Econometrica 70*(6), 2367–2420.

Deb, Partha and Pravin K. Trivedi (2006). Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: Application to health care utilization. *Econometrics Journal 9*(2), 307–331.

Dehejia, Rajeev H. and Sadek Wahba (2002). Propensity Score-Matching Methods For Nonexperimental Causal Studies. *The Review of Economics and Statistics 84*(1), 151–161.

Dietrich, Hans and Hans-Dieter Gerner (2007). The determinants of apprenticeship training with particular reference tobusiness expectations. *Zeitschrift für ArbeitsmarktForschung - Journal for Labour Market Research 40*(2/3), 221–233.

Dionisius, Regina, Samuel Muehlemann, Harald Pfeifer, Günter Walden, Felix Wenzelmann, and Stefan C. Wolter (2009). Costs and Benefits of Apprenticeship Training. A Comparison of Germany and Switzerland. *Applied Economics Quarterly (formerly: Konjunkturpolitik) 55*(1), 7–37.

Dubin, Jeffrey A. and Daniel L. McFadden (1984). An Econometric Analysis of Residential Electric Appliance Holdings and Consumption. *Econometrica 52*(2), 345–62.

Dustmann, Christian, Rob Euwals, and Arthur van Soest (1997). Wages in the first job after apprenticeship: movers versus stayers. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 30*(3), 671–674.

Dustmann, Christian and Costas Meghir (2005). Wages, Experience and Seniority. *Review of Economic Studies 72*(1), 77–108.

Dustmann, Christian and Uta Schoenberg (2007). Apprenticeship Training and Commitment to Training Provision. *Economics of Education Working Paper Series, University of Zurich, Institute for Strategy and Business Economics (ISU) No 0032*.

Dustmann, Christian and Uta Schoenberg (2009). Training and Union Wages. *The Review of Economics and Statistics 91*(2), 363–376.

Dustmann, Christian and Uta Schoenberg (2012). What Makes Firm-Based Vocational Training Schemes Successful? The Role of Commitment. *American Economic Journal: Applied Economics 4*(2), 36–61.

EDK (2006). Leitlinien zur Optimierung der Nahtstelle obligatorische Schule - Sekundarstufe II.

Egger, Dreher, and Partner AG (2007). Vertiefungsstudie Bildungsangebote im Übergang von der obligatorischen Schule in die Berufsbildung. Erstellt im Auftrag des BBT. Technical report, Egger, Dreher and Partner AG.

Euwals, Rob and Rainer Winkelmann (2004). Training Intensity and First Labor Market Outcomes of Apprenticeship Graduates. *International Journal of Manpower 25*(5), 447–462.

EVD/EDI/EDK (2011). Chancen optimal Nutzen. Erklaerung 2011 zu den gemeinsamen bildungspolitischen Zielen für den Bildungsraum Schweiz. Eidgenoessisches Volkswirtschaftsdepartement (EVD), Eidgenoessisches Departement des Innern (EDI), Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK). www.bildungsmonitoring.ch/sites/default/files/20110530/Erklaerung.pdf.

Fairlie, Robert W. (2003). An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models. Working Papers 873, Economic Growth Center, Yale University.

Farber, Henry S. and Robert Gibbons (1996). Learning and Wage Dynamics. *The Quarterly Journal of Economics 111*(4), 1007–1047.

Fitzenberger, Bernd, Stefanie Licklederer, and Hanna Zwiener (2015). Mobility across firms and occupations among graduates from apprenticeship. *Labour Economics 34*, 138–151. European Association of Labour Economists 26th Annual Conference.

Fitzenberger, Bernd and Alexandra Spitz (2004). Die Anatomie des Berufswechsels: eine empirische Bestandsaufnahme auf Basis der BIBB/IAB-Daten 1998/1999. Working Papers of the Research Group Heterogenous Labor 04-02, Research Group Heterogeneous Labor, University of Konstanz/ZEW Mannheim.

Franz, Wolfgang, Viktor Steiner, and Volker Zimmermann (2000). *Die betriebliche Ausbildungsbereitschaft im technologischen und demographischen Wandel*. ZEW Wirtschaftsanalysen (Band 46). Baden-Baden: Nomos.

FSO (2004a). *Betriebszaehlung 1998: Schweiz und Kantone 1991/1995/1998*. Neuchâtel: Federal Statistical Office.

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

FSO (2007a). *Bildungsmosaik Schweiz. Bildungsindikatoren 2007*. Neuchâtel: Bundesamt fuer Statistik BFS.

FSO (2007b). *Bildungsperspektiven. Szenarien 2007-2016 für die Sekundarstufe II*. Neuchâtel: Bundesamt fuer Statistik BFS.

Galindo-Rueda, Fernando (2003). Employer learning and schooling-related statistical discrimination in Britain. *IZA Discussion Paper no. 778, Institute for the Study of Labor, Bonn*.

Gathmann, Christina and Uta Schoenberg (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics 28*(1), 1–49.

Geel, Regula and Uschi Backes-Gellner (2011). Occupational mobility within and between skill clusters: an empirical analysis based on the skill-weights approach. *Empirical Research in Vocational Education and Training 3*(1), 21–38.

Gibbons, Robert and Lawrence F. Katz (1991). Layoffs and Lemons. *Journal of Labor Economics 9*(4), 351–380.

Goeggel, Kathrin and Thomas Zwick (2012). Heterogeneous Wage Effects of Apprenticeship Training. *Scandinavian Journal of Economics 114*(3), 756–779.

Haeberlin, Urs, Christian Imdorf, and Winfried Kronig (2004). *Chancenungleichheit bei der Lehrstellensuche: Der Einfluss von Schule, Herkunft und Geschlecht. NFP43 Synthesis 7.* Bern / Aarau: NFP43 / SKBF.

Harhoff, Dietmar and Thomas J. Kane (1997). Is the German apprenticeship system a panacea for the U.S. labor market? *Journal of Population Economics 10*(2), 171–196.

Harmon, Colm, Hessel Oosterbeek, and Ian Walker (2003). The Returns to Education: Microeconomics. *Journal of Economic Surveys 17*(2), 115–156.

Hashimoto, Masanori (1981). Firm-Specific Human Capital as a Shared Investment. *American Economic Review 71*(3), 475–82.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica 66*(5), 1017–1098.

Heckman, James J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica 46*(4), 931–959.

Heckman, James J. (1994). Is job training oversold? *The Public Interest Spring*, 91–115.

Heckman, James J, Hidehiko Ichimura, and Petra Todd (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies 65*(2), 261–94.

Heckman, James J, Hidehiko Ichimura, and Petra E Todd (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies 64*(4), 605–54.

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith (1999). The economics and econometrics of active labor market programs. In O.Ashenfelter and D.Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 31, pp. 1865–2097. Elsevier.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis 15*, 199–236.

Hoffman, Nancy (2011). *Schooling in the Workplace: How Six of the World's Best Vocational Education Systems Prepare Young People for Jobs and Life.* Cambridge, MA: Harvard Education Press.

Hupka, Sandra, Stefan Sacchi, and Barbara E. Stalder (2006). *Herkunft oder Leistung? Analyse des Eintritts in eine zertifizierende nachobligatorische Ausbildung anhand der Daten des Jugendlaengsschnitts TREE. Working Paper.* Basel: TREE.

Hupka-Brunner, Sandra, Thomas Meyer, Barbara E. Stalder, and Anita Keller (2011). PISA-Kompetenzen und Übergangswege: Ergebnisse aus der Schweizer TREE-Studie. In E.M.Krekel and T.Lex (Eds.), *Neue Jugend? Neue Ausbildung? Beiträge aus der Jugend- und Bildungsforschung*, pp. 173–188. Bonn: BIBB.

ILO, International Labour Organisation (2014). Global Employment Trends for Youth 2013: A Generation at Risk. Working papers, eSocialSciences.

Imbens, Guido W. (2000). The Role of the Propensity Score in Estimating Dose-Response Functions. *Biometrika 87*(3), 706–710.

Imdorf, Christian (2006). The selection of trainees in small and medium-sized enterprises. Integration and exclusion of immigrant youth at the transitional stage between school and vocational training in Switzerland. *Working paper (english version).* http://www.coreched.ch/publikationen/ISA_Durban06_Imdorf.pdf.

Imdorf, Christian (2009). Die betriebliche Verwertung von Schulzeugnissen bei der Ausbildungsstellenvergabe. *Empirische Paedagogik 23*(4), 392–409.

Jann, Ben (2006). Fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials. Available from http://ideas.repec.org/c/boc/bocode/s456727.html. Technical report.

Jovanovic, Boyan (1979). Job Matching and the Theory of Turnover. *Journal of Political Economy 87*(5), 972–90.

Kambourov, Gueorgui and Iourii Manovskii (2008). Rising Occupational And Industry Mobility In The United States: 1968-97. *International Economic Review 49*(1), 41–79.

Kambourov, Gueorgui and Iourii Manovskii (2009). Occupational Specificity Of Human Capital. *International Economic Review 50*(1), 63–115.

Keller, Anita, Sandra Hupka-Brunner, and Thomas Meyer (2010). *Nachobligatorische Ausbildungsverläufe in der Schweiz: Die ersten sieben Jahre. Ergebnisübersicht des Jugendlängsschnitts TREE, Update 2010*. Basel: TREE.

Lalé, Etienne (2012). Trends in Occupational Mobility in France: 1982-2009. *Labour Economics 19*(3), 373–387.

Lamo, Ana, Julián Messina, and Etienne Wasmer (2011). Are specific skills an obstacle to labor market adjustment? *Labour Economics 18*(2), 240–256.

Lange, Fabian (2007). The Speed of Employer Learning. *Journal of Labor Economics 25*(1), 1–35.

Lange, Fabian and Robert Topel (2006). *The Social Value of Education and Human Capital*, Volume 1 of *Handbook of the Economics of Education*, Chapter 8, pp. 459–509. Elsevier.

Lazear, Edward P. (2009). Firm-Specific Human Capital: A Skill-Weights Approach. *Journal of Political Economy 117*(5), 914–940.

Lechner, Michael (1999). Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption. IZA Discussion Papers 91, Institute for the Study of Labor (IZA).

Lechner, Michael (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In M.Lechner and F.Pfeiffer (Eds.), *Econometric Evaluation of Labour Market Policies*, pp. 43–58. Heidelberg: Physica.

Lechner, Michael (2002). Program Heterogeneity And Propensity Score Matching: An Application To The Evaluation Of Active Labor Market Policies. *The Review of Economics and Statistics 84*(2), 205–220.

Lechner, Michael (2008). A Note on the Common Support Problem in Applied Evaluation Studies. *Annales d'Economie et de Statistique* (91-92), 217–235.

Lee, Lung-Fei (1983). Generalized econometric models with selectivity. *Econometrica 51*(2), 507–512.

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

Leuven, Edwin and Barbara Sianesi (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components, Boston College Department of Economics.

LINK (2009). *Detaillierter Ergebnisbericht: Lehrstellenbarometer August 2009*. Bern: Bundesamt für Berufsbildung und Technologie BBT.

Longhi, Simonetta and Malcolm Brynin (2010). Occupational change in Britain and Germany. *Labour Economics 17*(4), 655–666.

Malcomson, James M., James W. Maw, and Barry McCormick (2003). General training by firms, apprentice contracts, and public policy. *European Economic Review 47*(2), 197–227.

Meyer, Thomas (2003). Zwischenlösung - Notlösung? In BFS/TREE (Ed.), *Wege in die nachobligatorische Ausbildung. Die ersten zwei Jahre nach Austritt aus der obligatorischen Schule. Zwischenergebnisse des Jugendlängsschnitts TREE*, pp. 101–109. Neuchâtel: Bundesamt fuer Statistik BFS.

Mortensen, Dale (1986). Job search and labor market analysis. In O.Ashenfelter and R.Layard (Eds.), *Handbook of Labor Economics*, Volume 2. North-Holland.

Moser, Urs (2004). *Jugendliche zwischen Schule und Berufsbildung. Eine Evaluation bei Schweizer Grossunternehmen unter Berücksichtigung des internationalen Schulleistungsvergleich PISA*. Bern: h.e.p. Verlag.

Muehlemann, Samuel, Harald Pfeifer, Guenter Walden, Felix Wenzelmann, and Stefan C. Wolter (2010). The Financing of Apprenticeship Training in the Light of Labor Market Regulations. *Labour Economics 17*(5), 751–874.

Muehlemann, Samuel, Paul Ryan, and Stefan C. Wolter (2013). Monopsony Power, Pay Structure, and Training. *Industrial and Labor Relations Review 66*(5), 1097–1114.

Muehlemann, Samuel, Juerg Schweri, Rainer Winkelmann, and Stefan C. Wolter (2007). An Empirical Analysis of the Decision to Train Apprentices. *LABOUR: Review of Labour Economics and Industrial Relations 21*(3), 419–441.

Muehlemann, Samuel and Stefan C. Wolter (2007). Regional effects on employer-provided training: Evidence from apprenticeship training in Switzerland. *Zeitschrift für ArbeitsmarktForschung - Journal for Labour Market Research 40*(2/3), 135–147.

Muehlemann, Samuel and Stefan C. Wolter (2011). Firm-sponsored training and poaching externalities in regional labor markets. *Regional Science and Urban Economics 41*(6), 560–570.

Muehlemann, Samuel, Stefan C. Wolter, Marc Fuhrer, and Adrian Wueest (2007). *Lehrlingsausbildung - oekonomisch betrachtet.* Zuerich/Chur: Rueegger Verlag.

Muehlemann, Samuel, Stefan C. Wolter, and Adrian Wueest (2009). Apprenticeship Training and the Business Cycle. *Empirical Research in Vocational Education and Training 1*(2), 173–186.

Mueller, Barbara and Juerg Schweri (2006). *Die Entwicklung der betrieblichen Ausbildungsbereitschaft. Eine Längsschnittuntersuchung zur dualen Berufsbildung in der Schweiz.* SIBP Schriftenreihe No 31. Zollikofen.

Mueller, Barbara and Juerg Schweri (2009). Berufswechsel beim Übergang von der Lehre in den Arbeitsmarkt. *Schweizerische Zeitschrift für Bildungswissenschaften 31*(2), 199–227.

Mueller, Barbara and Juerg Schweri (2012a). *Die Betriebe in der Berufsbildung: Entwicklungen 1985 bis 2008.* Neuchâtel: Federal Statistical Office FSO.

Mueller, Barbara and Juerg Schweri (2012b). The returns to occupation-specific human capital - Evidence from mobility after training. Economics of Education Working Paper Series 0081, University of Zurich, Institute for Strategy and Business Economics (ISU).

Mueller, Barbara and Juerg Schweri (2015). How specific is apprenticeship training? Evidence from inter-firm and occupational mobility after graduation. *Oxford Economic Papers 67*(4), 1057–1077.

Mueller, Barbara and Stefan Wolter (2014). The role of hard-to-obtain information on ability for the school-to-work transition. *Empirical Economics 46*(4), 1447–1471.

Neal, Derek (1995). Industry-Specific Human Capital: Evidence from Displaced Workers. *Journal of Labor Economics 13*(4), 653–77.

Neubaeumer, Renate and Lutz Bellmann (1999). *Die wirtschaftlichen Folgen von Aus- und Weiterbildung*, Chapter Ausbildungsintensität und Ausbildungsbeteiligung von Betrieben: theoretische Erklärungen und empirische Ergebnisse auf der Basis des IAB-Betriebspanels 1997, pp. 9–41. München: Hampp.

Neuenschwander, Markus P. and Tina Malti (2009). Selektionsprozesse beim Uebergang in die Sekundarstufe I und II. *Zeitschrift fuer Erziehungswissenschaft 12*(2), 216–232.

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

OECD (2000). *Form Initial Education to Working Life: Making Transitions Work*. Paris: OECD.

OECD (2001). *Knowledge and Skills for Life. First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

OECD (2002). *PISA 2000 Technical Report*. Paris: OECD.

OECD (2004). *Employment Outlook*. Paris: OECD.

OECD (2009a). *Learning for Jobs. OECD Policy Review of Vocational Education and Training. Initial Report*. Paris: OECD.

OECD (2009b). *Learning for Jobs. OECD Reviews of Vocational Education and Training. Switzerland*. Paris: OECD.

OECD (2010). *Improving Health and Social Cohesion through Education, Educational Research and Innovation*. OECD Publishing.

OECD (2012a). How Well Are Countries Educating Young People to the Level Needed for a Job and a Living Wage? Education Indicators in Focus 7, OECD Publishing.

OECD (2012b). What are the social benefits of education? In *Education at a Glance 2012: Highlights*. OECD Publishing.

OECD (2014). *Education at a glance 2014: OECD Indicators*. Paris: OECD.

Oosterbeek, Hessel and Edwin Leuven (2001). Firm-Specific Human Capital as a Shared Investment: Comment. *American Economic Review 91*(1), 342–347.

OPET (2001). Lehrstellenbarometer 2001. Technical report, Federal Office for Professional Education and Technology - Bundesamt fuer Berufsbildung und Technologie BBT, Bern.

Oreopoulos, Philip and Kjell G. Salvanes (2011). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives 25*(1), 159–84.

Parrado, Eric, Asena Caner, and Edward N. Wolff (2007). Occupational and industrial mobility in the United States. *Labour Economics 14*(3), 435–455.

Phelps, Edmund S. (1972). The Statistical Theory of Racism and Sexism. *The American Economic Review 62*(4), 659–661.

Rogerson, Richard, Robert Shimer, and Randall Wright (2005). Search-Theoretic Models of the Labor Market: A Survey. *Journal of Economic Literature 43*(4), 959–988.

Rosenbaum, Paul R. and Donald B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*(1), 41–55.

Rosenbaum, Paul R. and Donald B. Rubin (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistican 39*(1), 33–38.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers 3*(2), 135–146.

Rubin, Donald B. (1974). Estimating Causal Effects to Treatments in Randomised and Nonrandomised Studies. *Journal of Educational Psychology 66*, 688–701.

Rubin, Donald B. (1977). Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics 2*(1), 1–26.

Ryan, Paul (2000). The Institutional Requirements Of Apprenticeship: Evidence From Smaller EU Countries. *International Journal of Training and Development 4*(1), 42–65.

Ryan, Paul (2001). The School-to-Work Transition: A Cross-National Perspective. *Journal of Economic Literature 39*(1), 34–92.

Ryan, Paul, Uschi Backes-Gellner, Silvia Teuber, and Karin Wagner (2012). Apprentice pay in Britain, Germany and Switzerland: institutions, market forces, market power. Economics of Education Working Paper Series 0075, University of Zurich, Institute for Strategy and Business Economics (ISU).

SBFI-SERI (2015a). Bestandsaufnahme der Zwischenlösungen an der Nahtstelle I. Technical report, State Secretariat for Education, Research and Innovation (SERI).

SBFI-SERI (2015b). Vocational and professional education and training in Switzerland. Facts and figures 2015. Technical report, State Secretariat for Education, Research and Innovation (SERI).

Scherer, Roland, Kristina Zumbusch, Katja Schwanke, and Manfred Walser (2010). Die raumwirtschaftliche Bedeutung des Pendelns in der Schweiz. Kurzgutachten im Auftrag des BAV zum aktuellen Stand der Forschung. St. Gallen: Institut fuer oeffentliche Dienstleistungen und Tourismus.

Schweri, Juerg, Samuel Muehlemann, Yasmina Pescio, Belinda Walther, Sefan C. Wolter, and Lukas Zürcher (2003). *Kosten und Nutzen der Lehrlingsausbildung aus der Sicht Schweizer Betriebe*. Chur/Zürich: Rüegger.

Schweri, Juerg and Barbara Mueller (2007). Why has the share of training firms declined in Switzerland? *Zeitschrift fuer ArbeitsmarktForschung ZAF - Journal for Labour Market Research 40*(2/3), 149–167.

Schweri, Juerg and Barbara Mueller (2008). *Die Ausbildungsbereitschaft der Betriebe: Entwicklungen 1995 bis 2005. Analysen zur Betriebszählung 2005*. Neuchâtel: Bundesamt für Statistik BFS.

Shaw, Kathryn L. (1987). Occupation Change, Employer Change, and the Transferability of Skills. *Southern Economic Journal 53*(3), 702–719.

Sheldon, George (2005). *Die Berufstrukturelle Wandel der Beschäftigung in der Schweiz 1970-2000: Ausmass, Ursachen und Folgen*. Neuchâtel: Federal Statistical Office.

Sianesi, Barbara (2004). An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *The Review of Economics and Statistics 86*(1), 133–155.

Siegenthaler, Michael (2011). Can a standardised aptitude test predict the training success of apprentices? Evidence from a case study in Switzerland. *Empirical Research in Vocational Education and Training 3*(2), 105–128.

SKBF-CSRE (2011). *The Swiss Education Report 2010*. Aarau: Swiss Coordination Centre for Research in Education.

SKBF-CSRE (2014). *The Swiss Education Report 2014*. Aarau: Swiss Coordination Centre for Research in Education.

Smith, Jeffrey A. and Petra E. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(1-2), 305–353.

Smits, Wendy (2007). Industry-specific or generic skills? Conflicting interests of firms and workers. *Labour Economics 14*(3), 653–663.

Spence, Michael (1973). Job Market Signalling. *Quarterly Journal of Economics 87*(3), 355–374.

Stalder, Barbara E. (2011). *Das intellektuelle Anforderungsniveau beruflicher Grundbildungen in der Schweiz. Ratings der Jahre 1999-2005*. Basel: Institut fuer Soziologie der Universitaet Basel/TREE.

Stalder, Barbara E. and Evi Schmied (2006). *Lehrvertragsaufloesungen, ihre Ursachen und Konsequenzen. Ergebnisse aus dem Projekt LEVA*. Bern: Bildungsplanung und Evaluation der Erziehungsdirektion des Kantons Bern.

Steedman, Hilary (2012). *Overview of apprenticeship systems and issues : ILO contribution to the G20 Task Force on Employment*. Genf.

Stevens, Margaret (1994a). An Investment Model for the Supply of Training by Employers. *Economic Journal 104*(424), 556–70.

Stevens, Margaret (1994b). A Theoretical Model of On-the-Job Training with Imperfect Competition. *Oxford Economic Papers 46*(4), 537–62.

Stevens, Margaret (2003). Earnings Functions, Specific Human Capital, and Job Matching: Tenure Bias Is Negative. *Journal of Labor Economics 21*(4), 783–806.

Stevens, Margaret (2004). Wage-Tenure Contracts in a Frictional Labour Market: Firms' Strategies for Recruitment and Retention. *Review of Economic Studies 71*(2), 535–551.

Stoeger, Kaus and Rudolf Winter-Ebmer (2001). Lehrlingsausbildung in Oesterreich: Welche Betriebe bilden Lehrlinge aus? Working Paper 110, Johannes Kepler Universitaet, Institut fuer Volkswirtschaftslehre, Linz.

Strupler, Mirjam and Stefan C Wolter (2012). *Die duale Lehre eine Erfolgsgeschichte - auch für Betriebe. Ergebnisse der dritten Kosten-Nutzen-Erhebung der Lehrlingsausbildung aus der Sicht der Betriebe*. Glarus/Chur: Rueegger Verlag.

Sullivan, Paul (2010). Empirical evidence on occupation and industry specific human capital. *Labour Economics 17*(3), 567–580.

Symonds, William C., Robert Schwartz, and Ronald F. Ferguson (2011). *Pathways to prosperity: Meeting the challenge of preparing young Americans for the 21st century. Pathways to Prosperity Project*. Cambridge, MA: Harvard University Graduate School of Education.

Topel, Robert H. and Michael P. Ward (1992). Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics 107*(2), 439–479.

Troltsch, Klaus and Guenter Walden (2010). Beschaeftigungsentwicklung und Dynamik des betrieblichen Ausbildungsangebotes. Eine Analyse fuer den Zeitraum 1999 bis 2008. *ZAF - Zeitschrift fuer ArbeitsmarktForschung 43*, 107–124.

UK Parliament, corp creator (2009). *Apprenticeships, Skills, Children and Learning Act 2009: Elizabeth II. 2009. Chapter 22*.

Vella, Francis (1998). Estimating Models with Sample Selection Bias: A Survey. *Journal of Human Resources 33*(1), 127–169.

Vella, Francis and Marno Verbeek (1999). Estimating and Interpreting Models with Endogenous Treatment Effects. *Journal of Business & Economic Statistics 17*(4), 473–78.

Venn, Danielle (2009). Legislation, collective bargaining and enforcement: Updating the OECD employment protection indicators, www.oecd.org/els/workingpapers.

von Wachter, Till and Stefan Bender (2006). In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers. *American Economic Review 96*(5), 1679–1705.

Wasmer, Etienne (2006). General versus Specific Skills in Labor Markets with Search Frictions and Firing Costs. *American Economic Review 96*(3), 811–831.

Weber, Bernhard A., Aniela Wirz, and Stefan C. Wolter (2001). Switzerland. In C.Harmon, I.Walker, and N.Westergaard-Nielsen (Eds.), *Education and Earnings in Europe*. Cheltenham: Edward Elgar.

Werwatz, Axel (1996). How firm-specific is German apprenticeship training? Sonderforschungsbereich 373 Discussion Paper No. 1996-12, Humboldt Universitaet Berlin.

Werwatz, Axel (2002). Occupational Mobility after Apprenticeship - How Effective is the German Apprenticeship System? *Applied Economics Quarterly (Konjunkturpolitik) 48*(3-4), 279–303.

Winkelmann, Rainer (1996). Employment prospects and skill acquisition of apprenticeship-trained workers in Germany. *Industrial and Labor Relations Review 49*(4), 658–672.

Wolter, Stefan C., Samuel Mühlemann, and Jürg Schweri (2006). Why Some Firms Train Apprentices and Many Others Do Not. *German Economic Review 7*, 249–264.

Wolter, Stefan C. and Paul Ryan (2011). Apprenticeship. In E.A.Hanushek, S.Machin, and L.Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3, pp. 521–576. Elsevier.

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

Xie, Yu, Jennie E. Brand, and Ben Jann (2012). Estimating Heterogeneous Treatment Effects with Observational Data. *Sociological methodology 42*, 314–347.

Zangelidis, Alexandros (2008). Occupational And Industry Specificity Of Human Capital In The British Labour Market. *Scottish Journal of Political Economy 55*(4), 420–443.

Selbständigkeitserklaerung

Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemaess Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 26. September 2015

Barbara Mueller