
Working Paper No. 179

# Open Labor Markets and Firms' Substitution between Training Apprentices and Hiring Workers

Manuel Aepli and Andreas Kuhn

Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

UNIVERSITÄT BERN

Working Paper No. 179

# Open Labor Markets and Firms' Substitution between Training Apprentices and Hiring Workers

Manuel Aepli and Andreas Kuhn

March 2020



The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Program of the Swiss State Secretariat for Education, Research, and Innovation (SERI).

www.economics-of-education.ch

# Open Labor Markets and Firms' Substitution between Training Apprentices and Hiring Workers

Manuel Aepli, Swiss Federal Institute for Vocational Education and Training, and University of Bern*

Andreas Kuhn, Swiss Federal Institute for Vocational Education and Training, University of Bern, and IZA

March 2020

**Abstract**

In this paper, we study whether Swiss employers substitute between training apprentices and hiring cross-border workers. Because both training apprentices and hiring skilled workers are costly for firms, we hypothesize that (easier) access to cross-border workers will lead some employers to substitute away from training their own workers. We account for potential endogeneity issues by instrumenting a firm's share of cross-border workers using a firm's distance to the national border and therefore its possibility to fall back on cross-border workers to satisfy its labor demand. We find that both OLS and 2SLS estimates are negative across a wide range of alternative specifications, suggesting that firms substitute between training and hiring workers when the supply of skilled workers is higher. Our preferred 2SLS estimate implies that the increase in firms' share of cross-border workers within our observation period, from 1995 to 2008, led to about 3,500 fewer apprenticeship positions (equal to about 2% of the total number of apprentice positions).

*JEL classification: D22; J23; J61; M53*
*Keywords:* immigration; cross-border workers; firm behaviour; labor demand; substitution effects; apprenticeship training

*We thank Michael Gerfin, Samuel Mühlemann and Jürg Schweri for helpful comments and suggestions. We are grateful for participation in and helpful inputs from the Doctoral Course Program of the Swiss Leading House on Economics of Education, Firm Behaviour and Training Policies. We also profited from helpful comments of participants at the Brown Bag Seminar at the University of Bern, the CVER Conference 2019 at the London School of Economics, and the second BIBB Conference on the Economics of VET 2019 in Siegburg.
Contact: Manuel Aepli, Kirchlindachstrasse 79, CH-3052 Zollikofen, manuel.aepli@ehb.swiss; Andreas Kuhn, Kirchlindachstrasse 79, CH-3052 Zollikofen, andreas.kuhn@ehb.swiss.

# 1 Introduction

Integrating the labor markets of their member countries has been a key objective within the European Union. One of the main rationales of these efforts being that these policies are designed to spur gains in innovation and productivity and, consequently, growth (e.g. Hunt and Gauthier-Loiselle, 2010; Peri, 2012). However, economic theory also predicts that there will be distributional effects from immigration, at least in the short run. This implies that some groups will presumably suffer from immigration, even when its overall effect is positive on the host country (e.g. Bansak *et al.*, 2015; Borjas, 2014). Consistent with this, several recent empirical studies document that recent waves of immigration have been accompanied by a rise right-wing parties that make efforts to put restrictions on further immigration in various European countries (e.g. Brunner and Kuhn, 2018; Edo *et al.*, 2019; Halla *et al.*, 2017).

In this paper, we estimate – against this broader background – employers' short-run substitution between the training of resident apprentices and the hiring of cross-border workers (hereafter CBWs), i.e. non-resident immigrant workers who live in one of the neighboring countries but who work regularly in Switzerland. Prospective apprentices are not yet in the labor market, they are not organized or politically represented, and they therefore have low bargaining power. They may thus be especially at risk of being substituted with CBWs. Moreover, because training apprentices is costly for employers – average training costs over the full training period equal almost 100,000 Swiss francs, equivalent to about 1.28 times the annual median wage in 2016 – we hypothesize that access to CBWs will decrease the costs of hiring skilled workers and will therefore lead some firms, in the medium term, to substitute away from training their own workforce to hiring more CBWs (i.e. easier access to CBWs may make the labor market more competitive, thereby removing some of employers' incentives to train their own workforce; cf. Acemoglu and Pischke, 1999). Similar to Kerr *et al.* (2015) and others, we focus on employers' decisions using comprehensive firm-level data from the Swiss business census. We believe that the substitution of employers between training and hiring workers is a relevant topic from an educational-policy point of view as well, because apprentices are quantitatively of importance within the Swiss educational system (around two-thirds of young people attend some kind of apprenticeship after comulsory schooling) and because the Swiss apprenticeship system is often viewed as one of the key sources of the country's economic success.

We focus on the substitution with CBWs for both substantive and methodological reasons. First, and in contrast to many immigrants both working and living in Switzerland (resident immigrant workers, henceforth[1]), CBWs presumably have skills similar to those of natives and they have the advantage that they usually speak the same language as natives (e.g. Stöhr, 2015). We therefore expect that the substitutability will be higher between apprentices and CBWs than between apprentices and non-native workers in general – many of whom are low skilled and/or do not speak any of the country's official languages.[2] Second, CBWs are quantitatively relevant in the Swiss labor market: CBWs made up about 4.8% of the total workforce in the year 2008 (i.e. the year which marks the endpoint of our empirical analysis), representing about 18.2% of the total foreign workforce in that year. Moreover, in relative terms, the increase in CBWs has recently been much greater than in resident immigrant workers; between 1995 and 2008 (our analysis period), the total number of CBWs increased by about 46.1% (from 146,773 to 214,377); then, between 2008 and 2018, it increased by an additional 46.4% (to a total of 313,926 CBWs).[3] Third and finally, the focus on CBWs gives us the opportunity to implement a simple but powerful instrumental-variable approach based on the country's geography. Obviously, a firm's possibility to fall back on CBWs depends on its location relative to the national border. The closer to the border a firm is located, ceteris-paribus, the larger the pool of CBWs potentially available for this firm. Thus, we consider the shortest travel time from a firm's location (i.e. the municipality it is located in) to the country border as a valid instrument for the share of CBWs a firm employs, as we will argue in much more detail below.

Our empirical approach is very close to the one used by Dustmann *et al.* (2017), who exploit the introduction of a policy in 1991 which led to a sudden and large influx of Czech CBWs into the German labor market. Specifically, the policy allowed Czech workers to seek employment, but not residence, in border municipalities (see also Moritz, 2011). They find a moderate negative effect on native wages and a rather large negative effect on native employment – the

[1] This group also includes persons born in Switzerland with non-native parents, because citizenship is not automatically granted for persons born in the country.

[2] In general, immigrant inflows may have very different effects, depending on whether they are substitutes or complements with natives (e.g. Peri and Sparber, 2009). For Switzerland, Gerfin and Kaiser (2010) argue that natives and resident immigrant workers are imperfect substitutes.

[3] These figures are taken from the employment statistics of the Swiss Federal Statistical Office, available at: `https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/surveys/es.html`. Note that there are slight discrepancies between these figures and those derived from the data used in the analysis below (due to, for example, different sample definitions).

authors find that a one percentage point increase in the number of Czech CBWs decreased native employment by about 0.93% three years after the policy has been implemented. They also find that the employment effects are more pronounced for unskilled workers, who represent a closer substitute for inflowing Czech workers. Our empirical strategy is also close to that of Beerli *et al.* (2018), who use a difference-in-differences approach to study how the removal of all remaining immigration restrictions for workers from the European Union increased the number of CBWs to Switzerland – and how this in turn affected employment and wages in the regions close to the border. Overall, they find that the removal of the remaining restrictions led to an increase in overall native employment, arguing that the policy increased overall labor demand by firms.

Broadly consistent with this evidence, we find that firms which have easier access to hiring skilled workers partially substitute the training of resident apprentices with hiring additional cross-border workers. Our OLS and 2SLS estimates are consistent with each other and are robust across a wide range of robustness checks. Moreover, as expected, 2SLS estimates are larger (in absolute value) than the corresponding OLS estimates, and they are precisely enough estimated to rule out statistical equivalence between OLS and 2SLS estimates. Our preferred 2SLS estimate implies that an increase in the average share of CBWs across firms of 1.14 percentage points (corresponding to the observed increase in the share of CBWs between 1995 and 2008, i.e. our analysis period) leads to a corresponding decrease in the number of apprenticeship positions of about 2% (equal to about 3,500 apprenticeship positions in absolute terms); a significant effect in both statistical and substantive terms.

In contrast to our study, previous studies on the effects of immigrants on the Swiss labor market have mainly focused on the overall effect of immigration on either wages and/or employment among residents (i.e. natives and previous immigrants).[4] For example, Gerfin and Kaiser (2010), using the structural skill-cell approach, estimate the wage effects from immigration for the period 2002 to 2008. Their findings are mixed: While natives face no wage pressure on average, previously immigrants' wages are negatively affected. Across educational subgrops, highly-educated workers incur the largest wage losses. Favre *et al.* (2013) reach similar conclusions, applying a shift-share instrumental-variable approach and focusing on the period between

[4]To save space, we refrain from discussing the voluminous international empirical literature on the effects of immigration. Among many others, both Borjas (2014) and Bansak *et al.* (2015) contain such an overview.

2002 and 2010. Moreover, their analysis also covers CBWs. They find no overall negative effects of immigrants or CBWs on employment, neither for Swiss-born nor for previously immigrated workers. However, they find statistically significant effects for certain subgroups. A one percent increase in the workforce due to immigration leads to a drop in the employment rate of high skilled Swiss-born workers of 0.3 percentage points. Moreover, while immigration does not affect the employment rate of previously immigrated workers, CBW inflows do: a one percent increase in the workforce due to CBWs decreases the employment rate of previously immigrated workers by 0.2 percentage points. Similarly, Basten and Siegenthaler (2019) find that immigration had limited effects had limited negative effects on natives' employment as well as on their wages. Ruffner and Siegenthaler (2016) explicitly focus on the increase in the number of CBWs in a difference-in-difference approach and conclude that CBWs helped firms overcome their skill shortages and increased firms' productivity without crowding out native employment. The study by Beerli *et al.* (2018), already mentioned above, also finds that the overall effects of immigration were positive. Our study complements this literature by providing estimates on the substitution of apprentices with CBWs, a question which has not yet been adressed in the empirical literature so far. It also illustrates that there is nothing contradictory in finding both positive and negative effects of immigration at the same time and that immigration policy may therefore involve difficult policy trade-offs.

The remainder of this paper is organized as follows. Section 2 introduces Switzerland's immigration policy during the analysis period, describes the institutional background of the Swiss apprenticeship system, and formulates our main hypothesis. Section 3 presents the main data sources and the construction of the key variables, along with descriptive statistics of the key variables. Section 4 discusses our empirical strategy. Section 5 presents the main results, as well as several robustness checks. Section 6 concludes.

# 2 Background

## 2.1 Immigration to Switzerland

Ever since the end of World War II, the share of immigrants in the Swiss labor market has been, more or less steadily, increasing. Starting in the boom of the 1950s, Switzerland became the host country for a large number of immigrants.[5] Consequently, in the year 2016, the share of non-native residents in Switzerland reached nearly 25%. Initially, and until about the 1980s, immigrants tended to be unskilled laborers (e.g. construction workers), and a majority of them came from geographically close countries (most importantly, from Italy and Spain). Afterwards, the composition of immigrants changed in two notable ways. First, from about the 1990s onwards, immigration became more heterogeneous; for example, there was a significant influx of immigrants from former Yugoslavia during and following the Balkan wars. The other notable shift was towards more highly skilled workers, starting in the late 1990s to early 2000s, and which continues to this day.[6]

Figure 1 shows how the number of CBWs and of resident immigrant workers, respectively, changed over time (relative to the size of the total workforce). Both the number of CBWs and resident immigrant workers increased quite sharply in this period, but the share of CBWs increased relatively more than the share of resident immigrant workers.

Figure 1 around here

In 2008, at the end of our analysis period, most of the total of 217,011 CBWs lived in France (119,685), followed by Italy (47,904), Germany (42,753), and Austria (6,670). No other country is mentioned in the official statistic on CBWs for 2008.[7] Within Switzerland, 67% percent of all CBWs worked in only three cantons in 2008. 54,808 CBWs (25%) worked in the canton of Geneva, which is located in the far west of Switzerland and almost entirely surrounded by France. In the north of Switzerland, bordering both Germany and France, the two half-cantons of Basel hosted another 22% of the total of CBWs. Italian is the official language in the canton of Ticino in the south of Switzerland, which together with its proximity to Italy, attracted 44,941 CBWs in 2008 (20.7%).[8]

[5]For a long-term overview of Swiss immigration legislation, see Ruedin *et al.* (2015), for example.

[6]In 1999, Switzerland signed the bilateral "Agreement on the Free Movement of Persons" (AFMP) with the European Union, which Swiss voters approved in the national plebiscite in 2000. It took effect in the year 2002 and secured the same legal labor status for both immigrants and CBWs as for native employees (see Ruffner and Siegenthaler, 2016, for details). The new system further eased the possibility for foreign individuals to work within Switzerland, and boosted both the number of resident immigrant workers and of CBWs. We skip the details of this policy because our identification strategy does not rely on variation in immigration policy over time (as discussed below in section 4).

[7]There is a small minority of Swiss CBWs, i.e. Swiss citizens who live in a neighboring country but work as CBWs in Switzerland. In the period 2008–2010, there were about 3.6% Swiss CBWs.

[8]All numbers in this paragraph are taken from the official statistic on cross-border commuters by the

There are three main reasons for this large influx of CBWs into the Swiss labor market. First, the removal of the (remaining) legal hurdles in the course of the implementation of the AFMP made it easy for Swiss firms to hire CBWs (and for CBWs to work in Switzerland, respectively). Second, there exist differentials in real wages (especially when also considering costs of living, which are considerably lower outside Switzerland) and employment opportunities.[9] Third, the country is landlocked and, due to its small size, the majority of its settlement regions are reachable by car within one hour from a neighboring country (this ignores the time from a CBW's home to the border, however). Finally, Switzerland is neither a linguistically homogeneous country nor are its national languages endemic to the country. In fact, the official language of all of the four countries that border Switzerland is also an official language of Switzerland.[10] Thus, language as a potential remaining barrier for labor market integration after removing legal regulations is basically irrelevant at the Swiss border. Due to this feature, we argue that CBWs' skill mix is relatively similar to that of the resident workforce (cf. appendix table B.1).[11]

## 2.2 The Swiss apprenticeship system

In Switzerland about two-thirds of a cohort enter some kind of apprenticeship training after compulsory schooling, which makes apprenticeship training the most important educational track at the upper secondary level (SERI, 2014). There are presently about 230 different learnable occupations to choose from, and regular apprenticeship durations vary between two and four years. In most cases, apprentices sign an apprenticeship contract with a training firm, which also pays them a wage (which is considerably lower than the wage of a trained worker in the same occupation). The firm is responsible for the practical part of the apprenticeship training. Apprentices attend vocational school for one or two days a week and spend the rest

---

Federal Statistical Office; available at: `https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/surveys/ccs.html`. Note that there are slight discrepancies between these figures and those derived from the data used in the analysis below (due to, for example, different sample definitions).

[9] In fact, there is suggestive evidence that similar CBWs get paid substantively less than native workers (SECO, 2019).

[10] Lichtenstein, a very small country located between Switzerland and Austria, is as good as fully integrated in the Swiss labor and other markets. For this reason, it is considered being a part of Switzerland in many statistics – including the official statistics on CBWs.

[11] Consequently, mean wages of CBWs (gross monthly median wage of 5,591 CHF) are also more similar to natives' wages (median wage of 5,872 CHF) than those of resident non-native workers (median wage of 4,961 CHF); all wages are computed from the Swiss Earnings Structure Survey of 2004.

of the week in their training firms (the much less frequent alternative being full-time vocational school).[12]

### Employers' motives to provide apprenticeship positions: The costs and benefits of apprenticeship training in Switzerland

The Swiss apprenticeship system relies heavily on the participation of (mostly) private firms, which voluntarily decide whether they want to train apprentices or not. From the employer's point-of-view, there is a key trade-off between the costs and the benefits of training apprentices (e.g. Muehlemann and Wolter, 2014; Wolter *et al.*, 2006). On the one hand, training firms bear quite substantial costs including, among other things, wages paid to the apprentices and wages for instructors providing the institutionally required on-the-job training for the apprentices. In 2009, these costs averaged almost 100,000 Swiss francs over the full training period, an equivalent to about 1.28 annual median wages. On the other hand, training firms may profit from the productive output by the apprentices during their time spent within the firm; in many cases, this output is not substantially different from that of other employees (at least at the end of the training period).

Consistent with this, empirical studies on the costs and benefits of apprenticeship training in Switzerland (Mühlemann *et al.*, 2007; Schweri *et al.*, 2003; Wolter and Strupler, 2012) show that the benefits exceed the costs during the apprenticeship period for about two-thirds of all Swiss training firms. Thus, the cumulated benefits from apprentices' productive work until the end of the training period equals or surpasses the cumulated costs of training for these firms. The main reason for this is that apprentices in many occupations become relatively productive early on in their apprenticeship, while their wages remain relatively low – compared to those of fully trained workers.

### The costs of hiring skilled workers

At the same time, empirical studies show that there are substantial costs associated with the hiring of skilled workers. In their study focusing specifically on Swiss employers, for example, Blatter *et al.* (2012) document that hiring costs for skilled workers, depending on firm size,

[12] See Wettstein *et al.* (2014) for additional background information on the Swiss VET system, and how it fits into the country's overall educational system. See also Wolter and Ryan (2011) for a more general discussion of apprenticeship systems.

equal about 10 to 17 weeks of wage payments on average. Thus an alternative, and possibly complementary, motivation for training apprentices – especially among those firms incurring net training costs – is the retention of fully trained apprentices by employers to satisfy their demand for skilled labor (e.g. Mohrenweiser and Backes-Gellner, 2010; Wolter and Strupler, 2012). This motive for training apprentices appears to be especially relevant for firms operating in tight labor markets (Mohrenweiser and Zwick, 2009; Muehlemann and Leiser, 2018).[13]

In line with this argumentation, Blatter *et al.* (2015) show that Swiss firms which are confronted with higher costs for hiring skilled workers from the external labor market tend to train more apprentices (and vice versa). They document relatively large effects of hiring costs on the number of training positions, finding that a one standard deviation increase in average hiring costs is associated with an increase in the number of apprentices by about half a standard deviation.

## 2.3 Substitution between training apprentices and hiring workers?

Taken together, there are two main motives for training apprentices that are potentially affected by the ease of access to CBWs. First, during their training, apprentices perform productive tasks that otherwise either unskilled or skilled workers perform. Because apprentices earn a much lower wage than (either skilled or unskilled) workers do, firms may exert a considerable cost advantage when using apprentices for productive work (e.g. Muehlemann and Wolter, 2014). If labor becomes more widely available and thus wages of skilled or unskilled workers decrease relative to apprentices' wages, however, this cost advantage of hiring apprentices shrinks.

Second, because both the training of apprentices and the hiring of skilled workers from the external labor market are very costly to employers, we expect that (easier) access to CBWs reduces the hiring costs for employers and may thus tip the balance in favor of hiring workers externally instead of training apprentices.

Overall, both arguments predict a partial substitution between training and hiring, and we therefore expect that easier access to CBWs will lead some employers to hire additional CBWs

[13]More recent studies have argued that other motives are relevant as well. For example, firms may use apprenticeship training to screen workers (Mohrenweiser *et al.*, 2017) or as a signaling device (Backes-Gellner and Tuor, 2010). More recently, Kuhn *et al.* (2019) have argued that local norms describing the role of the state also have an influence on firms' training decision to provide apprenticeship positions.

instead of training their own apprentices in the medium term.

# 3 Data

## 3.1 Firm-level data from the Swiss Business Census

The empirical analysis in this paper relies mainly on information drawn from the Swiss Business Census (*"Betriebszählung"*). The Business Census covers the population of all firms active in either the second or the third sector in Switzerland, and it includes information on a number of firm-level characteristics, such as the number of employees, industrial affiliation, legal status, geographic location, the number of apprentices, the number of CBWs, and the number of resident immigrant workers (i.e. workers without Swiss citizenship). We have to limit our analysis to the waves of 1995, 2005, and 2008 because only these three include information on the number of CBWs.[14] After 2008, the Business Census was replaced by the firm register dataset STATENT, which, however, does not include information on apprentices anymore and is thus unsuitable for our purpose. The available data thus restrict the analysis to the period between 1995 and 2008.

The only sample restriction that we additionally impose is that we exclude all firms with fewer than three employees (with a total of 492,278 observations across the three waves). These firms make up a significant fraction of all firms (about 43% of all firms in the three waves covered by our analysis), but they cover only a small fraction of all apprenticeship contracts (only about 1.5% of all apprenticeship contracts). Thus these observations are not really useful when studying firm-provided apprenticeship training, because most of them do not train any apprentices, and we therefore follow the practice of other studies on the subject and exclude these firms from our empirical analysis.[15]

Table 1 around here

As shown in table 1, our analysis covers a total of 645,137 observations at the firm×year-level, representing a total of 342,323 unique firm observations and 10,595,201 employee obser-

[14] However, we do use additional waves of the Business Census to approximate firms' age as described in more detail below.

[15] For example, in their study on the costs and benefits of apprenticeship training in Switzerland, Wolter and Strupler (2012) also focus on firms with at least three employees. Of course, however, we will show that our main results are robust with regard to this decision (see table 9).

vations (see also appendix table B.2). In the empirical analysis below, we use the absolute number of apprentices employed by a given firm as our main dependent variable and the share of CBWs as our main regressor (but we will show that these choices are innocuous; see section 5.2 below). Panel (a) of table 1 shows that there are, on average, 0.8 apprentices per firm (however, also note that less than one-third of the firms train any apprentices). It is also interesting to note that apprentices make up about 4.8% of the total workforce in our sample in 2008. CBWs represent only about 5.3% of the workforce, but about 20% of all non-native workers. Overall, the total number of apprentices in Switzerland grew by 37.3% from 1995 to 2008, while the number of CBWs increased by 50.5% during this time. The total workforce rose by 11.5% in the same period (cf. table B.2).

Furthermore, we use the Business Census to construct several firm-level controls (descriptives for most of these are shown in panel (b) of table 1). The controls include firms' size (i.e. their total number of employees) and firm size squared, 18 industry dummies, a dummy taking the value one if firm $i$ is a private firm and zero if it is a public firm, and a dummy indicating foreign ownership. Moreover, we add firms' approximated age[16] and firms' share of resident immigrant workers to our firm specific controls. The intuition for including these two last control variables is set out in the appendix A dealing with the validity of the instrument.

## 3.2 Additional regional-level controls

Panels (c) and (d) of table 1 show descriptives for our additional location-specific controls. We also derive the log number of firms within the same labor market region, the log number of firms in the same industry within the same labor market, as well as the share of employees in the third sector within the respective labor market region from the Business Census. We employ some additional variables from the Federal Statistical Office to control for regional variation in the composition of the residents, e.g. the logarithm of municipalities' inhabitants and their population density. Furthermore, we calculate the median income for every labor market region from the Swiss Labor Force Survey 2014 (*"Schweizerische Arbeitskräfteerhebung"*) to control for regional income differentials. Finally, we include 26 cantonal and three language dummies.

[16]Concretely, we include a categorial variable $S_i$ for the number of sample appearance of firm $i$, e.g. for firm $i$ we do not observe in the waves 1991 and 1995 but in the waves 1998 to 2008, $S_i = 3$ in 2005 and $S_i = 4$ in 2008. Note that waves 1998 and 2001 are not part of our sample because they lack information on CBWs; however, we observe whether a firm existed in these waves.

Note that the regional variables are measured at either the municipal level or at the level of 106 labor-market regions (cf. last column of table 1).

Moreover, leaning on the analysis by Kuhn *et al.* (2019), we will also control for the social norm favoring the private over state provision of public goods, thereby influencing firms' investment in apprenticeship training. Controlling for this norm could be important in our context because there is a pronounced regional pattern of the strength of the norm. We follow Kuhn *et al.* (2019) and use the municipality results from two national-level votes on the role of the state within VET policy to measure this norm.[17]

While the principal aim of this paper is to analyze the effect of CBWs on firms' demand for apprentices, note that we observe concluded apprenticeship contracts – the number of which is likely affected by supply-side factors as well. For this reason we also include a firm's distance to the nearest high school, VET school, and full-time VET school as exogenous supply shifters in (most of) our estimates.

### 3.3 Distance to the national border

A final variable of interest is a firm's distance to the national border, which we will use as instrument for the share of CBWs within a firm (this idea is discussed in section 4 below in detail). To construct this variable, we use the matrix of travelling times put together by the Federal Office of Spatial Development, which contains the average travel time by car in minutes between any two Swiss municipalities (it also contains average travelling time when moving from one place to another within the same municipality). We set the minimum distance to the national border for every firm equal to the minimum travelling time necessary to drive by car from the municipality within which a given firm is located to one of the 138 different municipalities with a navigable border crossing on its territory. See figure 2 for a graphical representation of this variable.

Figure 2 around here

On average, firms included in our analysis lie 36.7 minutes by car away from the border. There is large variation in the travelling time to the border: The minimum travel time is only

[17] We use the mean share of supporting votes from two popular initiatives that demanded a stronger role of the state in the VET system (i.e. smaller values on this variable denote a stronger norm towards the private provision of training). See Kuhn *et al.* (2019) for additional details.

3.5 minutes (for the municipality of *Le Grand-Saconnex* in the canton of Geneva), while the maximum is 106.3 minutes (for the municipality of *Innertkirchen* in the canton of Bern; cf. figure 2), respectively. In the analysis below we use the natural logarithm of this variable, which ranges from 1.26 to 4.67 (see panel (e) of table 1 for the corresponding descriptives).

# 4 Empirical strategy

## 4.1 OLS estimates

To estimate the effect of CBWs on firms' training incidence, we start with the following basic regression equation:

$$A_i = \alpha + \beta CBW_i + \gamma x_i + \delta z_{j[i]} + \psi_{t[i]} + \epsilon_i \tag{1}$$

The dependent variable $A_i$ corresponds to the absolute number of apprentices employed by firm $i$, and the regressor of main interest is the share of CBWs working in firm $i$, denoted by $CBW_i$ in equation (1).[18] Most specifications will control for several firm-level characteristics described in section 3.1, denoted by $x_i$. The characteristics of region $j$ (see section 3.2), where firm $i$ is located in, enter as $z_{j[i]}$ on the right-hand side of equation (1). Finally, we also include a full set of census-year fixed effects, denoted by $\psi_{t[i]}$.

Whatever the exact set of controls, $\beta$ is the parameter of main interest, because it quantifies the partial association between a firm's share of CBWs and its number of apprentices. Throughout the analysis, we present robust standard errors that are clustered at the municipality level to account for potential correlation in the error terms of firms located within the same municipality (e.g. Cameron and Miller, 2015).

## 4.2 Instrumental-variable estimates

However, even after controlling for a large set of firm- and regional-level controls, OLS estimates of the parameter $\beta$ from equation (1) may suffer from endogeneity bias. Assume, for example,

[18] More specifically, the share of CBWs is the ratio of the number of CBWs to the total number of employees (including apprentices). Because apprentices are different from fully-trained workers, and because they implicitly appear on both sides of equation (1), we also checked that our results are robust to an alternative construction of this variable, not counting apprentices as part of a firm's workforce (see section 5.2 below).

that firms hire additional CBWs to satisfy an increasing demand for their products during an economic upswing. For the same reason, firms coincidently hire more apprentices performing productive tasks. In this case, OLS estimates would reveal a positive correlation between firms' share of CBWs and the number of apprentices they employ, though no causal interpretation applies.

To account for potential endogeneity issues of this kind, we propose an IV approach exploiting the fact that commuting distances impose considerable costs on CBWs, and distance to the border therefore constrains firms' hiring of CBWs.[19] Consequently, we instrument the share of CBWs employed by firm $i$ with the minimum distance to the national border of municipality $j$ firm $i$ is located, measured by the log travel time by car, $\ln(D_{j[i]})$.[20] Thus, we argue that variation in municipalities' minimum car-driving time to the national border induces quasi-experimental variation in firms' possibility to fall back on CBWs. The following two equations capture the essence of this approach:

$$A_i = \alpha + \beta \widehat{CBW}_i + \gamma x_i + \delta z_{j[i]} + \psi_{t[i]} + \epsilon_i, \tag{2a}$$

where $A_i$, $x_i$, $z_{j[i]}$, and $\psi_{t[i]}$ are the same as in equation (1) discussed above, but where $\widehat{CBW}_i$ comes from the following first-stage regression:

$$CBW_i = \pi_0 + \pi_1 \ln(D_{j[i]}) + \pi_2 x_i + \pi_3 z_{j[i]} + \psi_{t[i]} + \upsilon_i. \tag{2b}$$

As usual in such a setup, we use equation (2b) to empirically test the strength of the instrument (first-stage estimates, and associated test statistics, are discussed in section 5.1). The instrument should also meet the additional requirements of being as good as randomly assigned (possibly conditional on some set of controls) and not having a direct effect on the outcome (e.g. Angrist and Pischke, 2008; Wooldridge, 2010). We discuss these two additional requirements, which cannot directly be tested empirically, in appendix A and we present some robustness checks related to the concern of endogenous locational decisions of employers in section 5.2.

[19]More specifically, commuting costs consist of both direct transportation costs and the opportunity costs arising during the commuting time.

[20]We chose this parameterization of the instrument because it yields the highest fit in the first-stage regression. We also checked that other parameterizations yield similar results (see table 5.2 below).

Figure 3 around here

Figure 3 shows the spatial distribution of the share of CBWs at the district level. Supporting our argument from above, it is immediately evident that the density of CBWs decreases strongly with a region's distance to the national border. As a consequence, variation among the 148 districts is large and ranges from the district of *Gersau* in central Switzerland (located in the canton of Lucerne) with no CBWs at all to *Mendrisio* in the canton of Ticino with a share of CBWs of 34 percent.

# 5 Results

## 5.1 Main results

**OLS estimates**

Table 2 shows OLS estimates of a series of regressions that relate a firm's number of apprentices to its share of CBWs, according to the setup of equation (1).

Table 2 around here

First, the raw association between the number of apprentices and the share of CBWs, shown in column (1) of table 2, is negative but not statistically significant ($\widehat{\beta} = -0.220$). The next column adds census-year dummies, leaving the relevant estimate virtually unchanged ($\widehat{\beta} = -0.250$). Column (3) further adds firm-level controls, which yield a negative and statistically significant estimate ($\widehat{\beta} = -1.039$). Thus, compared to the previous two columns, this specification points to a possible substitution of employers between hiring CBWs and training apprentices. The change in the estimated coefficient also suggests that those firms that are, ceteris-paribus, more likely to hire apprentices are also more likely to hire CBWs. Further adding regional controls, including the full set of cantonal dummies, yields a comparable, yet somewhat smaller estimate of $\widehat{\beta} = -0.710$, suggesting that (some of) the regional-level characteristics are correlated with both the incidence of training and the likelihood of hiring CBWs (more precisely, the change in $\widehat{\beta}$ suggests that regional features that make firms in these locations more likely to train apprentices are negatively correlated with the share of CBWs). Finally, we further add the three distance-to-schools variables in the fifth, which again yields an estimate very close to that from the previous column. The estimate of $\widehat{\beta} = -0.709$ from the fifth column implies that a 10 percentage point increase in the share of CBWs is associated with 0.071 fewer apprentices per employer on average (we provide a more detailed discussion of the economic size of the estimated effects further below).

**2SLS estimates**

We next turn to the instrumental variable approach introduced in section 4.2, using a firm's distance to the national border as an instrument for its share of CBWs.[21]

Table 3 around here

Table 3 presents the resulting 2SLS estimates and also starts with a simple specification that includes no control variables at all. Instrumenting a firm's share of CBWs with its minimum distance to the border yields a negative and statistically significant estimate ($\widehat{\beta} = -1.778$). Moreover, in comparison with the corresponding OLS estimate, the 2SLS estimate is considerably larger (the increase in the estimate also makes up for the parallel increase in the standard error relative to the OLS estimate).[22]

We include census-year fixed effects in column (2) to account for changes in Switzerland's immigration policy during our sample period. This, however, hardly changes the relevant estimate (mirroring the pattern from table 2 above), which equals $\widehat{\beta} = -1.790$ in this specification.[23] In column (3), we further add our full set of firm-specific controls. The resulting estimate of $\widehat{\beta} = -1.958$ is somewhat smaller, but still close to that from the previous column, and it remains highly statistically significant. Column (4) of table 3 adds our set of regional controls. This yields a somewhat smaller (i.e. less negative) estimate of $\widehat{\beta} = -1.319$, which remains highly significant however.

Finally, the specification shown in column (5) also includes a firm $i$'s minimum distance to three schooling facilities (dual VET school, full school VET institutions, and high school).

[21] Appendix tables B.3 and B.4 show the corresponding first-stage and reduced-form estimates, respectively. Both effects are statistically significant and, more importantly, sizeable and rather precisely estimates. Also note that both are relatively stable across the different specifications.

[22] Consistent with this estimate, appendix figure B.1 shows that firms located in regions closer to the national border tend to train fewer apprentices on average than firms in regions further away from the border.

[23] We also estimated a model that included the interaction terms between the census-year dummies and the share of CBWs. The resulting 2SLS estimates on the interaction terms are positive, but statistically not significantly different from zero.

This specification, including the full set of controls, yields an estimate of $\widehat{\beta} = -1.445$ which remains highly statistically significant. This is our preferred estimate of the effect of CBWs on firms' provision of apprenticeship positions.[24] Compared to the corresponding OLS estimate from column (5) of table 2, the 2SLS estimate is about twice as large (i.e. more negative). Moreover, note that both OLS and 2SLS estimates are quite precisely estimated (note that the OLS and 2SLS estimates have similar t-values).

**Quantifying the economic size of the estimated effect**

Taken together, the estimates shown in tables 2 and 3 support our hypothesis that firms tend to substitute between training resident apprentices and hiring CBWs when there is easier access to CBWs. Also consistent with prior expectations, we find that 2SLS estimates are larger (i.e. more negative) than the corresponding OLS estimates. Before providing additional robustness checks, let us discuss the economic size of the estimates.

Some back-of-the-envelope calculations may illustrate the economic size of the estimated effect. Our preferred 2SLS estimate (i.e. that from column (5) of table 3) of $\widehat{\beta} = -1.445$ would imply that a ten percentage point increase in the mean share of CBWs decreases the number of apprentices trained by employers by 0.1446 on average. However, most firms presumably employ much smaller shares of CBWs (see table 1 again). Thus, as a more realistic benchmark, we may use the observed change in the share of CBWs between 1995 and 2008, the time period covered by our analysis. Within this time period, the observed share of CBWs for the average firm increased from 2.73% to 3.87% – an increase of 1.14 percentage points (cf. table B.2).[25]

Everything else constant, our baseline 2SLS estimate thus informs us that this will translate into a decrease of 0.0164 apprentices by each employer. There were 212,081 firms in our sample in 1995, and 219,950 in 2008. Thus, approximately 3,500 apprenticeship positions were substituted by CBWs within this time period (-3,493 to -3,629, depending on whether we use the absolute number of firms in 1995 or in 2008) – corresponding to about 2% of the total of apprenticeship positions in relative terms (-2.498% to -1.891%).

[24]Note that controlling for cantonal fixed effects in column (3) partly accounts for institutional differences in the supply of apprentices because the cantonal administrations have autonomy in the design of their educational system.

[25]Note that the share of CBWs at the level of the firm differs from the share of CBWs in the workforce (compare the statistics from tables 1 and B.2). Apparently, larger firms tend to hire more CBWs than smaller firms.

## 5.2 Robustness

In a next step, we probe the robustness of our estimates.

### Alternative specifications

As a first check, table 4 presents a series of alternative specifications of our baseline model from column (5) of table 3 (note that the first column of table 4 simply replicates this specification for the ease of comparison). We estimate specifications that use a slightly different parameterization of the instrument. Instead of using $\ln(D_{j[i]})$ as in the baseline specification, we use either $D_{j[i]}$, a quadratic in $\ln(D_{j[i]})$ or a quadratic in $D_{j[i]}$ (columns (2) to (4) in table 4). Evidently, all these alternative specifications yield estimates of $\beta$ close to our baseline estimate (in these three specifications, $\widehat{\beta}$ varies between -1.637 and -1.086).

Table 4 around here

The remaining four columns use different parametrizations of either the endogenous variable or the dependent variable. Specifically, we use the absolute number of CBWs (instead of the share of CBWs) as endogenous variable in column (5). By construction, this yields a somewhat different estimate of $\widehat{\beta} = -0.068$; however, note that the associated elasticity is similar to that in column (1), i.e. -0.07 versus -0.062. Another variation, shown in column (6) of table 4, uses a slightly different definition of the share of CBWs. While we do count apprentices as employees in the baseline specification, we construct the share of CBWs in this specification as the ratio of the number of CBWs to the number of employees excluding apprentices. Again, this yields an almost identical estimate ($\widehat{\beta} = -1.392$) as our baseline specification. In column (7), we regress the share of apprentices on the share of CBWs. The resulting estimate of $\widehat{\beta} = -0.110$ is substantively smaller than that from the baseline specification. Note, however, that this is primarily due to the difference in the scaling of the dependent variable (also note that, again, the implied elasticity is virtually identical to that from the baseline specification). Finally, we construct a dummy being one for training firms and zero otherwise in column (8), which we then regress on the share of CBWs. The estimate ($\widehat{\beta} = -0.336$) yields the substitution effect between CBWs and apprentices on the extensive margin. Thus easier access to CBWs leads some firms to stop training apprentices altogether.

Overall, our estimate of the substitution between CBWs and apprentices appears robust to some obvious alternative parametrizations of the instrument as well as the endogenous and the dependent variable. We next turn to some more subtle issues.

**Unobservable firm characteristics**

In our baseline estimates (table 3) we include various firm specific controls in an effort to account for the fact that the potential to hire apprentices varies substantially among firms, e.g. of different size or across different industries. However, other firm characteristics interacting with firms' provision of apprenticeship positions might be unobservable.[26] To dispel concerns to that effect, table 5 applies our estimation approach to differenced data in a subsample of 212,730 firms surveyed at least twice within our sample.[27]

The regressions underlying the estimates shown in table 5 deviate slightly from the conventional IV setting using difference data because our instrument does not vary at all over time. We therefore instrument the change in the share of CBWs with a firm's distance to the border as well. This implies that we assume that the distance to the border not only predicts the share of CBWs in a given year, but also the change in the share of CBWs between two consecutive waves of the business census.[28] The corresponding first-stage F-statistics in table 5 displaying values between 45.28 and 52.60 confirm this.

Table 5 around here

The estimate in column (1) has about the same size as the baseline estimate from table 3, but is statistically insignificant due to its large associated standard error. In contrast, the estimates in columns (2) and (3) of table 5, respectively, suggest that an increase in the share of CBWs leads to a substitution away from apprentices. Taking into account that these estimates

[26] Moreover, Beerli *et al.* (2018) present evidence that the inflow of CBWs affected firms in many dimensions, e.g. increasing their size and productivity. Exploiting within-firm variation in the share of CBWs in this section also addresses concerns that such changes in firm characteristics threat our identification strategy.

[27] The total of 332,456 observations in table 5 consist of 20,246 observations from firms observed in 1995 and 2005, 72,758 observations from firms observed in 2005 and 2008, and 239,452 observations from 119,726 firms observed in all three waves.

[28] That is, we estimate regressions of the following form:

$$\Delta A_i = \beta \Delta CBW_i + \gamma \Delta x_i + \delta \Delta z_{j[i]} + \Delta \epsilon_i,$$

where $\Delta$ denotes that a variable has been differenced. As in the main analysis, we instrument $\Delta CBW_i$ with $\ln(D_{j[i]})$.

are relatively imprecisely estimated, they appear well in line with the our baseline estimate from section 5.1. This in turn suggests that our main finding remains robust when accounting for unobservable firm characteristics.

**Firms' age and endogenous locational choices by employers**

Endogenous location decisions by firms pose a potential threat to our identification strategy. We discuss this issue in some detail in appendix A and argue that the evidence points towards relatively large differences in the provision of apprenticeship positions between different types of firms (i.e. newly established firms, as well as firms changing their location, tend to train fewer apprentices than existing and non-moving firms).[29] At the same time, these differences in training behaviour also appear to correlate with locational choices of employers (even though the differences in terms of distance to the border are relatively small). In combination, these two findings may imply that endogenous location decisions could impact our 2SLS estimates.

Table 6 around here

Table 6 first tackles concerns to that effect by estimating our baseline specification within various subsamples of firms of similar age; consequently, the time at which and the duration since these firms made their initial location choice are comparable. In order to assign firms to any of the three subsamples of columns (2) to (4) in table 6, we approximate their age by exploiting the possibility to identify firms throughout our sample period. Concretely, we identify three types of firms. Firms included in all waves enter the subsample of column (2). These firms were already established in 1991 (first wave) and remained within our sample until 2008. The approximation of firms' age within this subsample remains relatively vague, since we can only state that these firms were present in our sample for the same duration but we are unable to estimate the duration of their existence prior to 1991. The column (3) subsample of newly established firms only considers observations of firms appearing in our sample for the first time; hence these firms are between zero and three years (the maximum number of years between two waves) old. In column (4), we construct a subsample of "middle aged" firms by restricting the sample to observations of the 2005 and 2008 wave, respectively, and only

[29]The finding that newly established firms train fewer apprentices is not new, however (see Müller and Schweri, 2012, for example).

include firms which appear for the first time in our sample either in 1995 or 1998. This limits the subsample in column (4) to firms of age seven to sixteen years.[30] All estimates in table 6 are statistically significant. In terms of effect size, the estimates in column (2) and (4) are fairly similar to our baseline specification in column (1), while the estimate in column (3) is somewhat smaller. Considering the lower number of apprentices across newly established firms in the subsample of column (3), this smaller estimate seems plausible and corresponds to a similar effect in relative terms.[31]

Finally, column (5) focuses on the small subsample of firms which are moving to another municipality between two consecutive wages of the business census (see also appendix A for additional details). We find $\widehat{\beta}$ to be of similar size than the baseline estimate when using only this subsample of moving firms. Overall, we conclude that the negative effect of CBW inflows on firms' training provision pertains across firms of various ages and is therefore neither biased by the establishment of new firms nor the moving behavior of already existing firms.

**Additional robustness checks**

Table 7 presents some additional robustness checks. In column (2) of table 7 (again, the first column of table 7 simply replicates the baseline specification column (5) of table 3), we first enlarge the sample by also including 492,278 observations from firms with 2 or fewer employees (cf. table B.2). The resulting 2SLS estimate of $\widehat{\beta} = -1.065$ becomes, less negative compared to our baseline specification because firms' average number of apprentices among this subsample is smaller. However, considering the statistical imprecision, the estimate in column (5) is again close to our baseline specification.

Table 7 around here

Column (3) of table 7 limits our sample to 256,040 firms located within 30 car driving minutes from the border (i.e. $D_{j[i]} \leq 30$ or $\ln(D_{j[i]}) \leq 3.40$, respectively). We argue that regions within this small bandwidth are comparable, e.g. in terms of geography or economic

[30] Among these firms, the possibly youngest firms appeared for the first time in the 1998 wave (founded in 1998) and are observed in 2005. The possibly oldest firms (founded in 1992) appear for the first time in the 1995 wave and are observed in 2008.

[31] The average number of apprentices per firm among the subsamples of columns (1) to (4) are 0.78, 1.05, 0.33, and 0.66. Multiplying the respective estimates with a supposed 10 percentage point increase in the share of CBWs yields a substitution of 18 to 20% of the total number of apprentices per firm across all subsamples of table 6.

structure. Reassuringly, the estimate of $\widehat{\beta} = -1.828$ applying for firms located close to the border is again of similar magnitude as in our baseline specification.

Next, we limit our sample to private profit-oriented firms in column (4). The coefficient of $\widehat{\beta} = -1.363$ in column (4) pertaining for profit-oriented firms is comparable to the baseline estimate in column (1). This assures that our results are not driven by state authorities, which arguably face different budget restrictions than privately-run firms.

In column (5) we control for a full set of local labor-markets fixed effects (using a set of 106 dummies) instead of including cantonal-level dummies. While cantons are Switzerland's most important jurisdictional entities, labor markets are often not bound to their borders and alternatively defined by the Federal Statistical Office based on commuting patterns. The resulting estimate of $\widehat{\beta} = -1.688$ in column (5) thus yields the effect of CBW inflows on the provision of apprenticeship positions of firms' that operate in the same market in terms of product demand and labor supply.

We finally check, in the last three columns of table 7, whether the negative effect of the share of CBWs on firms' training provision is persistent throughout all three waves of the Business Census covered by our sample. The three corresponding estimates of columns (6) to (8) vary little (between $\widehat{\beta} = -1.373$ to $\widehat{\beta} = -1.501$), with their confidence intervals widely overlapping, and are almost identical to the estimate in the baseline specification of column (1).

## 5.3 Heterogeneity

In the final part of the empirical analysis, we check whether we find heterogeneous effects conforming with prior expectations to further corroborate our main finding.

**Training costs**

In this subsection, we test whether varying apprenticeship costs among firms translate into heterogeneous substitution effects between CBWs and apprentices. A priori, we expect the substitution effect to increase with firms' apprenticeship costs for two reasons. First, roughly two thirds of all Swiss training firms profit financially from training apprentices.[32] Arguably, the

[32]This applies especially in occupations where training curricula require little instructions and little time for practicing. In these occupations the productivity gap between apprentices and regular workers is therefore relatively low. Another third of the Swiss training firms bear sometimes substantial net cost during an apprenticeship. This net costs can amount up to 137,000 Swiss Francs; see Schweri *et al.* (2003), Wolter and Strupler

incentive to substitute apprentices with CBWs is relatively low for firms that profit financially from their apprentices. Second, firms bearing net costs during an apprenticeship often pursue the goal to satisfy their future skill demand by retaining fully trained apprentices (Blatter *et al.*, 2015). Inflows of CBWs, which enlarge the supply of skilled workers and increase the possibility to recruit on the external labor market, lower this particular incentive of firms to provide apprenticeship positions.

We test this hypothesis by first estimating $\beta$ across subsamples defined by the average net cost (in 1,000 CHF) of an apprenticeship $\bar{c}_{k[i]}$ in industry $k$ firm $i$ operates in (see table 8, columns (1) to (3)).[33] While the first two columns aggregate firms operating in industries that display either strongly negative (column 1) or moderately negative (column 2) net costs (i.e. net profits in these cases) from apprenticeship training, respectively, column (3) aggregates industries in which firms on average bear positive net costs from offering apprenticeship positions.

Table 8 around here

The estimate of $\widehat{\beta} = -0.580$ in column (1) is relatively small and statistically insignificant. This suggests that firms which tend to profit financially from training apprentices are less likely to substitute apprentices with CBWs. In contrast, the estimate from column (2) suggests a negative association between the share of CBWs and the number of apprentices. Apparently, these firms partly substitute apprentices with CBWs – even though they also tend to make a profit from training. One potential explanation for this pattern might be that firms usually recruit teenagers with no previous work experience at all for open training positions. This considerably limits firms' ability to assess apprentices' productivity. Therefore, the expected costs and benefits are often very uncertain from the employer's point of view. The fact that about 20% of all apprenticeship contracts end prematurely further adds to this uncertainty. In contrast, CBWs with a formal degree and positive work experience perhaps represent a more predictable alternative to firms. Column (3) focuses on industries in which firms bear net costs

(2012), and section 2.2 for details.

[33] Ideally, we would like to calculate the apprenticeship net costs on firm- or at least occupational-level since training curricula, which partly determine firms' VET costs, are occupation specific. Unfortunately, we do not observe occupations in the Business Census. As a second best solution, we approximate firms' apprenticeship costs by the average net costs within 82 industries (at the NOGA 2-digit level) derived from the most recent data on the costs and benefits of apprenticeship training available (Gehret *et al.*, 2019).

from offering an apprenticeship on average. The estimate of $\widehat{\beta} = -3.344$ reveals that the rate of substitution is much higher in this particular subsample of firms.

Finally, we return to the full sample in column (4) but simply interact the share of CBWs, and consequently also the instrument $\ln(D_{j[i]})$, with the industry average costs of an apprenticeship $\overline{c}_{k[i]}$. Consistent with the estimates from columns (1) to (3), the negative coefficient associated with the interaction term implies that the effect of CBWs on firms' VET provision increases with respect to apprenticeship net costs of industry $k$ in which firm $i$ is operating in.

**Additional subsample results**

As a final robustness check we provide several additional subsample results in table 9 and compare them to the baseline specification displayed in column (1) for the ease of comparison.

Table 9 around here

First, columns (2) and (3) of table 9 present subsample estimates among small (50 or less employees) and large (more than 50 employees) firms, respectively. Both estimates are statistically significant on the 1%-level, but $\widehat{\beta} = -11.523$ in column (3) appears much larger than $\widehat{\beta} = -0.874$ in column (2). These differences arise largely due to different levels of the outcome variable across the two subsamples, however. In the baseline sample of column (1), firms employ an average of 0.78 apprentices, while this number amounts to 5.28 among firms with 50 or more employees considered in column (3). Consequently, $\widehat{\beta} = -1.445$ in column (1) and $\widehat{\beta} = -11.523$ in column (3) associates a ten percentage point increase in the share of CBWs with a decrease in $A_i$ of 17.5% among firms in the baseline sample and 21.8% among firms with 50 or more employees, respectively.

Another contrast of interest is based on the observation (cf. appendix figure B.2) that the majority of CBWs in German (French, Italian) language regions stem from Germany or Austria (France, Italy). These differences in the composition of CBWs' countries of origin combined with differences across the educational systems of Germany, Austria, France, and Italy likely translate into different skill compositions of CBWs across Swiss language regions.[34] One the

[34] According to Field *et al.* (2010) a dual VET system, where pupils spend most of their time within a firm, are especially known in German language countries including Switzerland, Germany, and Austria. In contrast, France and Italy have lower overall numbers of VET pupils at the upper secondary level and low shares of combined school and work-based VET programs.

one hand, Germany and Austria have VET systems similar to that in Switzerland, therefore, from a firm's perspective, hiring CBWs from these countries represent a relative close substitute for investments in VET. On the other hand, the Swiss vocational track differs both in size and type from the French or Italian vocational track, respectively. This limits the possibility to substitute VET investments by hiring CBWs for firms relying on skills provided specifically by a Swiss-type VET system and that are located in the French or Italian language regions of Switzerland.

Column (4) and (5) of table 9 test these hypotheses by applying the baseline specification (column 1) separately to a subsample of firms located in German language (column 4) and Romance language (column 5) regions, respectively.[35] Effectively, the point estimate of $\widehat{\beta} = -1.351$ (robust standard error of about 0.528) among German language firms exceeds the point estimate of $\widehat{\beta} = -0.883$ (robust standard error of about 0.256) among Romance language firms by roughly 50%. However, the confidence intervals of the two point estimates clearly overlap and indicate no statistically significant difference in the substitution effect between CBWs and apprentices across Swiss language regions.

Finally, in the last two columns of table 9, we check that our results are not entirely driven by those regions with a disproportionate share of CBWs (i.e. the two half-cantons of Basel, as well as the cantons of Ticino and Geneva). Column (6) first shows the 2SLS estimate when these regions are excluded from the sample, yielding an estimate of $\widehat{\beta} = -1.412$, which is again very close to the baseline estimate (though less precisely estimated). Finally, we estimate the same specification using only firms from these regions, yielding again a similar estimate of $\widehat{\beta} = -1.342$ (note that this estimate remains statistically significant even though it uses a much smaller sample than the baseline specification).

# 6 Conclusions

In this paper, we study the potential substitution of employers between hiring CBWs and training apprentices of firms operating in the Swiss labor market. To account for potential biases and reverse causality issues related to employers' hiring decisions, we focus on CBWs

[35]The Romance language parts of Switzerland consist of the French, Italian, and Romansh language regions of Switzerland.

and apply an instrumental-variable approach, using a firm's distance to the national border as instrument for its share of CBWs. Within our sample period, we argue that firms' location choice is largely unrelated to the distance to the border and we thus argue that our instrument induces quasi-random variation in firms' share of CBWs. We apply this empirical strategy to three waves of the Business Census (1995, 2005, and 2008) covering a total of 645,137 observations stemming from 342,323 different firms.

Our preferred 2SLS estimate implies that the increase in firms' share of CBWs within our sample period from 1995 to 2008 of 1.14 percentage points led to a substitution of roughly 3,500 apprenticeship positions (about 2% of all apprentice positions in relativ terms). This effect is not only statistically, but also economically significant. Various robustness specifications show that this result is not sensitive to the exact specification neither of the dependent nor the exogenous variables, that it holds in different (sub)samples and for different sets of control variables, and IV estimates using differenced data further suggest our result to be robust to unobservable firm characteristics.

The present paper is, to the best of our knowledge, the first to analyze the effect of immigration on firms' provision of apprenticeship training. The potential substitution of firms between training apprentices and hiring CBWs is not only interesting from an economic point of view, but also bears political relevance because the apprenticeship system is a key pillar within the country's educational system. Importantly, however, note that our results do not imply that those young natives do not get any education or training because firms substitute away apprenticeship positions. In fact, our result are consistent with the finding from other studies that immigration pushes natives into other educational tracks and occupations (e.g. Foged and Peri, 2016; Peri and Sparber, 2009). Our results do, however, suggest that immigration undermines the incentives of firms to participate in firm-based system of education and training.

Moreover, our results also raise the more general question of substitution between firms training their own workforce versus hiring immigrant workers, whether crossborder or not. However, we believe that the answer to that question can not directly be inferred from our results. On the one hand, we suspect that the general immigrant population in the Swiss labor market differs with CBWs in their degree of substitutability. On the other hand, resident immigrants tend to make up for a much larger share of the workforce than CBWs. This applies especially for Switzerland, where roughly 25% of the working population are foreigners (while CBWs make up only about 5% of the workforce). We thus motivate further analyses focusing on immigration as a whole and not limit the analysis to CBWs as this paper does.

# References

Acemoglu, D. and Pischke, J.-S. (1999). Beyond Becker: Training in Imperfect Labour Markets. *Economic Journal*, **109**(453), 112–142.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Backes-Gellner, U. and Tuor, S. N. (2010). Avoiding labor shortages by employer signaling: on the importance of good work climate and labor relations. *Industrial & Labor Relations Review*, **63**(2), 271–286.

Bansak, C., Simpson, N. B., and Zavodny, M. (2015). *The economics of immigration*. Routledge.

Basten, C. and Siegenthaler, M. (2019). Do immigrants take or create residents' jobs? evidence from free movement of workers in Switzerland. *Scandinavian Journal of Economics*, **121**(3), 994–1019.

Beerli, A., Ruffner, J., Siegenthaler, M., and Peri, G. (2018). The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland. NBER Working Paper 25302.

Blatter, M., Muehlemann, S., and Schenker, S. (2012). The costs of hiring skilled workers. *European Economic Review*, **56**(1), 20–35.

Blatter, M., Muehlemann, S., Schenker, S., and Wolter, S. C. (2015). Hiring costs for skilled workers and the supply of firm-provided training. *Oxford Economic Papers*, **68**(1), 238–257.

Borjas, G. J. (2014). *Immigration Economics*. Harvard University Press.

Brunner, B. and Kuhn, A. (2018). Immigration, Cultural Distance and Natives' Attitudes Towards Immigrants: Evidence from Swiss Voting Results. *Kyklos*, **71**(1), 28–58.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, **50**(2), 317–372.

Dustmann, C., Schönberg, U., and Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. *Quarterly Journal of Economics*, **132**(1), 435–483.

Edo, A., Giesing, Y., Öztunc, J., and Poutvaara, P. (2019). Immigration and electoral support for the far-left and the far-right. *European Economic Review*, **115**, 99–143.

Favre, S., Lalive, R., and Zweimüller, J. (2013). Verdrängungseffekte des Freizügigkeitsabkommens Schweiz-EU auf dem Schweizer Arbeitsmarkt - Schlussbericht. Berne: SECO.

Field, S., Hoeckel, K., Kis, V., and Kuczera, M. (2010). Learning for Jobs. Synthesis Report of the OECD Reviews of Vocational Education and Training.

Foged, M. and Peri, G. (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, **8**(2), 1–34.

Gehret, A., Aepli, M., Kuhn, A., and Schweri, J. (2019). Lohnt sich die Lehrlingsausbildung für die Betriebe? Resultate der vierten Kosten-Nutzen-Erhebung. Zollikofen: Eidgenössisches Hochschulinstitut für Berufsbildung.

Gerfin, M. and Kaiser, B. (2010). The effects of immigration on wages: An application of the structural skill-cell approach. *Swiss Journal of Economics and Statistics*, **146**(4), 709–739.

Halla, M., Wagner, A. F., and Zweimüller, J. (2017). Immigration and voting for the far right. *Journal of the European Economic Association*, **15**(6), 1341–1385.

Hunt, J. and Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, **2**(2), 31–56.

Kerr, S. P., Kerr, W. R., and Lincoln, W. F. (2015). Skilled immigration and the employment structures of US firms. *Journal of Labor Economics*, **33**(S1), S147—-S186.

Kuhn, A., Schweri, J., and Wolter, S. C. (2019). Local Norms Describing the Role of the State and the Private Provision of Training. IZA Discussion Paper Nr. 12159.

Mohrenweiser, J. and Backes-Gellner, U. (2010). Apprenticeship training: for investment or substitution? *International Journal of Manpower*, **31**(5), 545–562.

Mohrenweiser, J. and Zwick, T. (2009). Why do firms train apprentices? The net cost puzzle reconsidered. *Labour Economics*, **16**(6), 631–637.

Mohrenweiser, J., Wydra-Somaggio, G., and Zwick, T. (2017). Information Advantages of Training Employers Despite Credible Training Certificates. University of Zurich, Department of Business Administration, Working Paper.

Moritz, M. (2011). The Impact of Czech Commuters on the German Labour Market. *Prague Economic Papers*, **2011**(1), 40–58.

Muehlemann, S. and Leiser, M. S. (2018). Hiring costs and labor market tightness. *Labour Economics*, **52**, 122–131.

Muehlemann, S. and Wolter, S. C. (2014). Return on investment of apprenticeship systems for enterprises: Evidence from cost-benefit analyses. *IZA Journal of Labor Policy*, **3**(1), 25.

Mühlemann, S., Fuhrer, M., Wüest, A., and Wolter, S. C. (2007). *Lehrlingsausbildung-ökonomisch betrachtet: Ergebnisse der zweiten Kosten-Nutzen-Studie*. Rüegger.

Müller, B. and Schweri, J. (2012). Die Betriebe in der dualen Berufsbildung: Entwicklungen 1985 bis 2008: eine Analyse der Betriebszählungsdaten. Bundesamt für Statistik.

Peri, G. (2012). The effect of immigration on productivity: Evidence from us states. *Review of Economics and Statistics*, **94**(1), 348–358.

Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, **1**(3), 135–169.

Ruedin, D., Alberti, C., and D'Amato, G. (2015). Immigration and integration policy in Switzerland, 1848 to 2014. *Swiss Political Science Review*, **1**(21), 5–22.

Ruffner, J. and Siegenthaler, M. (2016). From labor to cash flow? The abolition of immigration restrictions and the performance of Swiss firms. KOF Working Papers No. 424.

Schweri, J., Mühlemann, S., Pescio, Y., Walther, B., Wolter, S. C., and Zürcher, L. (2003). Kosten und Nutzen der Lehrlingsausbildung aus der Sicht Schweizer Betriebe.

SECO (2019). 15. Bericht des Observatoriums zum Freizügigkeitsabkommen Schweiz - EU. Bern: Staatssekretariat für Wirtschaft SECO.

SERI (2014). Vocational and Professional Education and Training in Switzerland. Facts and Figures 2014. State Secretariat for Education, Research and Innovation SERI.

Stöhr, T. (2015). The returns to occupational foreign language use: Evidence from Germany. *Labour Economics*, **32**, 86–98.

Wettstein, E., Schmid, E., and Gonon, P. (2014). *Berufsbildung in der Schweiz: Formen, Strukturen, Akteure*. hep Verlag Bern.

Wolter, S. and Strupler, M. (2012). *Die duale Lehre: Eine Erfolgsgeschichte - auch für die Betriebe*. Rüegger Verlag.

Wolter, S. C. and Ryan, P. (2011). Apprenticeship. In *Handbook of the Economics of Education*, volume 3, pages 521–576. Elsevier.

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

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Table 1: Descriptive statistics

| | Sample Mean | Standard deviation | Minimum | Maximum | Level of aggregation |
|---|---|---|---|---|---|
| **(a) Key variables** | | | | | |
| Number of apprentices | 0.775 | 3.692 | 0 | 623 | firm |
| Number of CBWs | 0.795 | 10.592 | 0 | 2,503 | firm |
| Share of apprentices | 0.059 | 0.118 | 0 | 1 | firm |
| Share of CBWs | 0.034 | 0.116 | 0 | 1 | firm |
| Any apprentices (yes = 1) | 0.286 | 0.452 | 0 | 1 | firm |
| Any CBWs (yes = 1) | 0.131 | 0.337 | 0 | 1 | firm |
| **(b) Firm-level controls** | | | | | |
| Total number of employees | 16.42 | 66.76 | 3 | 10,789 | firm |
| 3 - 10 employees | 0.710 | 0.454 | 0 | 1 | firm |
| 11 - 50 employees | 0.239 | 0.426 | 0 | 1 | firm |
| 51 - 100 employees | 0.030 | 0.170 | 0 | 1 | firm |
| > 100 employees | 0.022 | 0.148 | 0 | 1 | firm |
| Share resident immigrant workers | 0.170 | 0.246 | 0 | 1 | firm |
| Public firm (yes = 1) | 0.116 | 0.320 | 0 | 1 | firm |
| Foreign owned (yes = 1) | 0.004 | 0.065 | 0 | 1 | firm |
| Sector III (yes = 1) | 0.782 | 0.413 | 0 | 1 | firm |
| **(c) Location-specific controls** | | | | | |
| Log inhabitants | 9.35 | 1.68 | 3.09 | 12.64 | municipality |
| Population density | 1,906 | 2,548 | 0.740 | 12,214 | municipality |
| Log firms | 8.58 | 0.902 | 5.68 | 10.17 | LM-region |
| Log firms same industry | 5.14 | 1.42 | 0 | 8.20 | LM-region |
| % of population working | 62.39 | 2.10 | 55.61 | 67.48 | LM-region |
| % Employed in sector III | 71.00 | 10.74 | 27.08 | 91.45 | LM-region |
| Median income (CHF) | 7,139 | 691.0 | 5,500 | 9,000 | LM-region |
| % Romance | 0.290 | 0.454 | 0 | 1 | municipality |
| Social norm | 0.196 | 0.075 | 0.010 | 0.956 | municipality |
| **(d) Distance to schools in km** | | | | | |
| High school | 6.60 | 9.24 | 0 | 134.15 | municipality |
| VET school | 4.45 | 6.50 | 0 | 70.35 | municipality |
| Full-time VET school | 8.78 | 10.71 | 0 | 136.83 | municipality |
| **(e) Distance to border** | | | | | |
| Travel time by car in minutes | 35.38 | 18.90 | 3.52 | 106.26 | municipality |
| Log travel time by car in minutes | 3.394 | 0.634 | 1.258 | 4.666 | municipality |
| Observations (year-firm) | 645,137 | - | - | - | - |
| Observations (firms) | 342,323 | - | - | - | - |
| Observations (municipalities) | 2,331[a] | - | - | - | - |
| Observations (LM-region) | 106 | - | - | - | - |

Notes: [a]In 2014 there were 2,352 Swiss municipalities. However, in 21 municipalities we do not observe a firm within our sample. Sources: Business Census 1995, 2005, 2008; SLFS 2014; Census 2000; Federal Office of Topography 2014; own calculations.

Table 2: Main results (OLS estimates)

| | Number of apprentices | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Share of CBWs | -0.220 | -0.250 | -1.039*** | -0.710*** | -0.709*** |
| | (0.171) | (0.173) | (0.114) | (0.133) | (0.133) |
| | [-0.010] | [-0.011] | [-0.045] | [-0.031] | [-0.031] |
| Census year dummies | *No* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *No* | *No* | *Yes* | *Yes* | *Yes* |
| Location controls | *No* | *No* | *No* | *Yes* | *Yes* |
| Distance to schools | *No* | *No* | *No* | *No* | *Yes* |
| R-squared | 0.000 | 0.001 | 0.298 | 0.300 | 0.300 |
| Observations | 645,137 | 645,137 | 645,137 | 645,137 | 645,137 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table 3: Main results (2SLS estimates)

| | Number of apprentices | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Share of CBWs | -1.778*** | -1.790*** | -1.958*** | -1.319*** | -1.445*** |
| | (0.437) | (0.438) | (0.301) | (0.291) | (0.272) |
| | [-0.077] | [-0.077] | [-0.085] | [-0.057] | [-0.062] |
| Census year dummies | *No* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *No* | *No* | *Yes* | *Yes* | *Yes* |
| Location controls | *No* | *No* | *No* | *Yes* | *Yes* |
| Distance to schools | *No* | *No* | *No* | *No* | *Yes* |
| R-squared | 0.000 | 0.000 | 0.298 | 0.300 | 0.300 |
| Observations | 645,137 | 645,137 | 645,137 | 645,137 | 645,137 |
| F-statistic (first-stage) | 201.6 | 201.8 | 174.4 | 170.7 | 183.0 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table 4: Alternative specifications (2SLS estimates)

| | Number of apprentices | | | | | | Share of apprentices | Training dummy |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Share of CBWs | -1.445*** | -1.637*** | -1.086*** | -1.456*** | | | -0.110*** | -0.336*** |
| | (0.272) | (0.450) | (0.288) | (0.247) | | | (0.014) | (0.051) |
| | [-0.062] | [-0.071] | [-0.047] | [-0.063] | | | [-0.063] | [-0.039] |
| Number of CBWs | | | | | -0.068*** | | | |
| | | | | | (0.013) | | | |
| | | | | | [-0.070] | | | |
| Share of CBWs[a] | | | | | | -1.392*** | | |
| | | | | | | (0.263) | | |
| | | | | | | [-0.063] | | |
| Census-year dummies | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Location controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Distance to schools | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| R-squared | 0.300 | 0.300 | 0.300 | 0.300 | 0.277 | 0.299 | 0.088 | 0.117 |
| Observations | 645,137 | 64,5137 | 645,137 | 645,137 | 645,137 | 645,137 | 645,137 | 645,137 |
| F-statistic (first stage) | 183.0 | 104.2 | 89.1 | 92.6 | 81.9 | 188.9 | 183.0 | 183.0 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities. [a]Refers to a firm's ratio of CBWs to total employees excluding apprentices.

Table 5: First-differenced data (2SLS estimates)

| | Δ Number of apprentices | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Δ Share of CBWs | -1.572 | -3.000** | -3.006** |
| | (1.215) | (1.189) | (1.195) |
| | [-0.045] | [-0.086] | [-0.088] |
| Δ Firm controls | *No* | *Yes* | *Yes* |
| Δ Location controls | *No* | *No* | *Yes* |
| R-squared | 0.000 | 0.066 | 0.066 |
| Observations | 332,456 | 332,456 | 332,456 |
| F-statistic (first stage) | 45.28 | 51.18 | 52.56 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table 6: Heterogeneity by firms' age (2SLS estimates)

| | Number of apprentices | | | | |
|---|---|---|---|---|---|
| | (1) Baseline (all firms) | (2) Included in all waves | (3) Newly established | (4) "Middle aged" firms | (5) Moving firms only |
| Share of CBWs | -1.445*** | -1.921*** | -0.611** | -1.312** | -1.204* |
| | (0.272) | (0.408) | (0.269) | (0.516) | (0.701) |
| | [-0.062] | [-0.052] | [-0.069] | [-0.071] | [-0.145] |
| Census year dummies | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Location controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Distance to schools | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| R-squared | 0.300 | 0.398 | 0.090 | 0.156 | 0.091 |
| Observations | 645,137 | 292,838 | 90,238 | 56,194 | 24,943 |
| F-statistic (first stage) | 183.0 | 202.4 | 127.5 | 123.4 | 102.8 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table 7: Additional robustness checks (2SLS estimates)

| | Number of apprentices | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) Baseline | (2) Including micro firms | (3) $D_{j[i]}$ < 30 | (4) Profit-oriented firms | (5) Within MS-region | (6) Wave 1995 | (7) Wave 2005 | (8) Wave 2008 |
| Share of CBWs | -1.445*** | -1.065*** | -1.828*** | -1.363*** | -1.688*** | -1.373*** | -1.496*** | -1.501*** |
| | (0.272) | (0.227) | (0.364) | (0.251) | (0.478) | (0.266) | (0.318) | (0.316) |
| | [-0.062] | [-0.059] | [-0.198] | [-0.065] | [-0.073] | [-0.057] | [-0.064] | [-0.068] |
| Census year dummies | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Location controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Distance to schools | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| R-squared | 0.300 | 0.306 | 0.261 | 0.292 | 0.300 | 0.303 | 0.283 | 0.314 |
| Observations | 645,137 | 1,136,188 | 256,040 | 570,606 | 645,137 | 212,081 | 213,106 | 219,950 |
| F-statistic (first stage) | 183.0 | 158.4 | 95.5 | 185.6 | 99.7 | 217.7 | 177.0 | 155.6 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table 8: Heterogeneity by costs of apprenticeship training (2SLS estimates)

| | Number of apprentices | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\overline{c}_{k[i]}$ | $< p25$ | $\in [p25, p75]$ | $> p75$ | Full sample |
| Mean of $\overline{c}_{k[i]}$ | -18.88 | -6.74 | 6.86 | -7.38 |
| Share of CBWs | -0.580 | -1.257*** | -3.344*** | -1.724*** |
| | (0.393) | (0.303) | (0.726) | (0.285) |
| | [-0.022] | [-0.057] | [-0.150] | [-0.074] |
| Share of CBWs$\times\overline{c}_{k[i]}$ | | | | -0.042*** |
| | | | | (0.010) |
| $\overline{c}_{k[i]}$ | | | | -0.003*** |
| | | | | (0.001) |
| Census year dummies | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *Yes* | *Yes* | *Yes* | *Yes* |
| Location controls | *Yes* | *Yes* | *Yes* | *Yes* |
| Distance to schools | *Yes* | *Yes* | *Yes* | *Yes* |
| R-squared | 0.333 | 0.243 | 0.462 | 0.304 |
| Observations[a] | 159,274 | 373,078 | 111,988 | 644,340 |
| F-statistic (first stage) for: | | | | |
| Share of CBWs | 180.0 | 144.2 | 210.5 | 95.16 |
| Share of CBWs $\times\overline{c}_{k[i]}$ | | | | 131.15 |

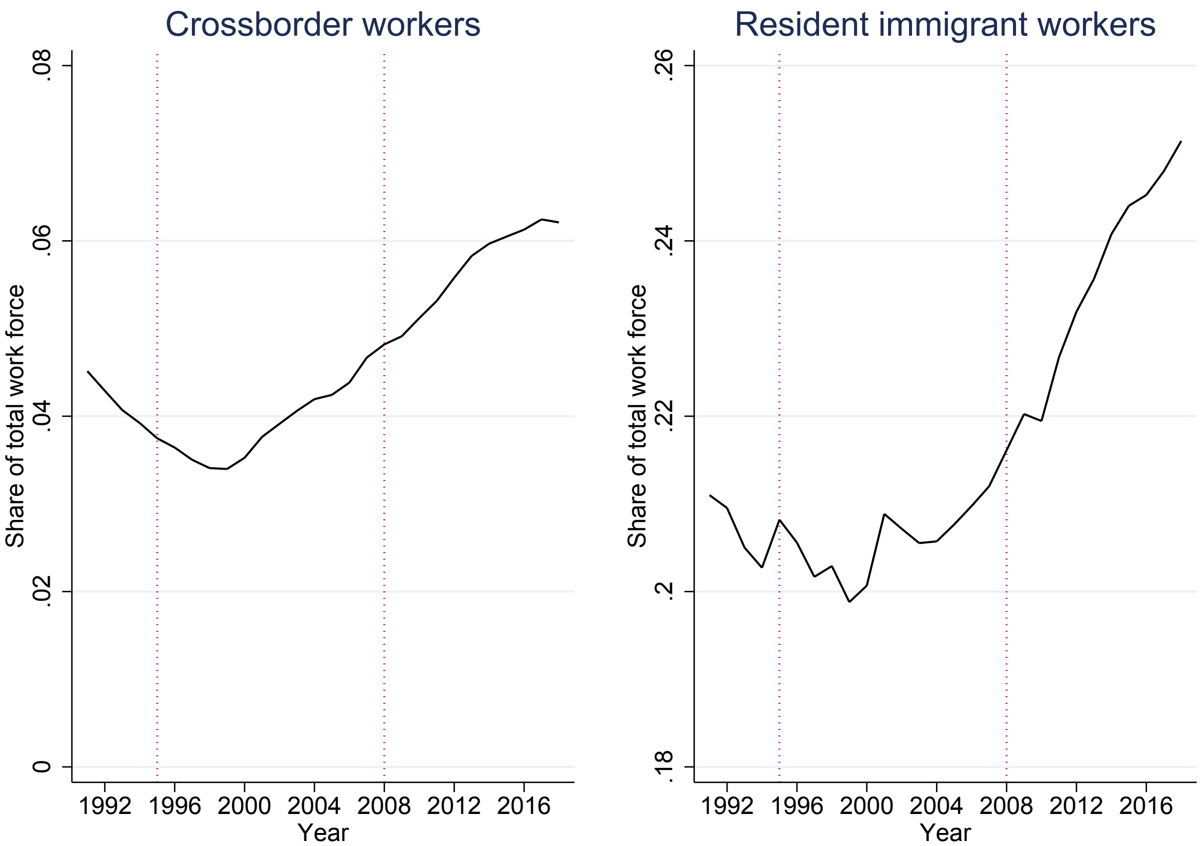
Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities. $\overline{c}_{k[i]}$ denotes the average costs of an apprenticeship (in 1,000 Swiss Francs) in NOGA-2-digit category $k$ firm $i$ operates in. $p25$ and $p75$ denote, respectively, the 25th (75th) percentile of the distribution of $\overline{c}_{k[i]}$. [a]Because we do not observe at least one apprenticeship in every NOGA-2-digit industry in the cost/benefit data, we lose some observations compared to our baseline sample.

Table 9: Subsample results (2SLS estimates)

| | Number of apprentices | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) Baseline | (2) Small firms | (3) Large firms | (4) German language | (5) Romance language | (6) BL, BS, GE, TI excluded | (7) BL, BS, GE, TI only |
| Share of CBWs | -1.445*** | -0.874*** | -11.523*** | -1.351** | -0.883*** | -1.412** | -1.342*** |
| | (0.272) | (0.130) | (2.364) | (0.528) | (0.256) | (0.406) | (0.443) |
| | [-0.062] | [-0.053] | [-0.119] | [-0.026] | [-0.122] | [-0.023] | [-0.338] |
| Census year dummies | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Location controls | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Distance to schools | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| R-squared | 0.300 | 0.186 | 0.271 | 0.331 | 0.210 | 0.291 | 0.402 |
| Observations | 645,137 | 611,748 | 33,389 | 457,768 | 184,637 | 537,630 | 107,507 |
| F-statistic (first stage) | 183.0 | 181.6 | 120.5 | 183.6 | 75.6 | 276.8 | 28.7 |

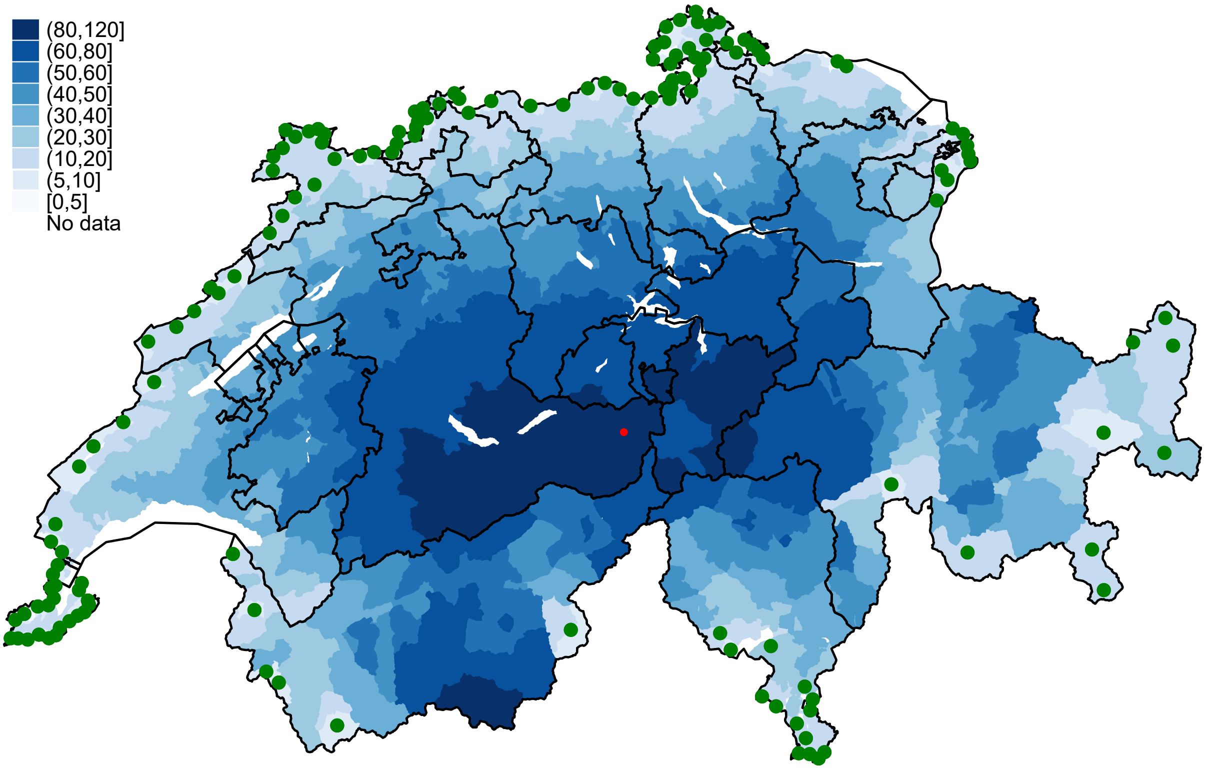
Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities. Column (6) excludes observations from the two half-cantons of Basel (BS and BL), as well as from the cantons of Geneva (GE) and Ticino (TI).

Figure 1: Share of CBWs and resident immigrant workers



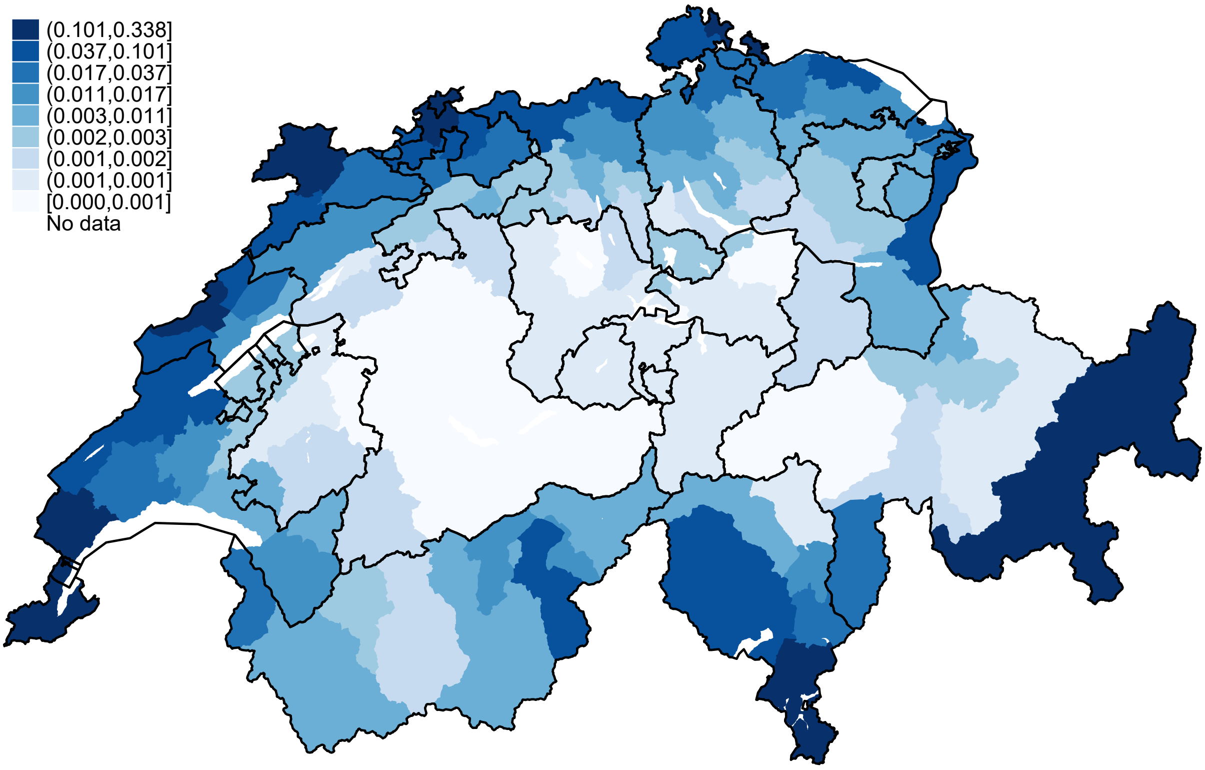

Notes: The two dotted vertical lines delineate our analysis period (1995 to 2008).
Source: Federal Statistical Office 2017; own calculations.

Figure 2: Travelling distance to the national border




Notes: The figure shows the minimum travelling distance from each municipality to the national border. The green dots indicate the center of a bordering municipality; the red dot indicates the center of *Innertkirchen*, the municipality most far away from the national border.

Figure 3: Share of CBWs




Notes: The figure shows the spatial distribution of the share of CBWs (i.e. the ratio of the number of CBWs to the total work force) at the district level (there are 148 distinct districts).

# A Assessing instrument validity

In this appendix we discuss the validity of our instrument, i.e. a firm's minimum distance to the national border. While we can use the first-stage estimates to assess the partial correlation between our instrument and the endogenous variable, the other two requirements for an instrument – i.e. the assumption concerning the (quasi-)random assignment of the instrument and the exclusion restriction – cannot be tested empirically (e.g. Angrist and Pischke, 2008; Wooldridge, 2010).

## A.1 Endogenous location and training decisions by employers

Quite obviously, a firm's distance to the border is not random. Rather, firms themselves decide where to locate. Within our instrumental-variable approach, firms deliberately choosing their location with respect to the distance to the border may violate the assumption of the instrument being quasi-randomly assigned. In case these firms differ systematically from other firms in terms of their provision of apprenticeship positions, our 2SLS estimates will be biased. We thus first try to assess whether newly established and/or moving firms are any different in their training behavior. To that end, we estimate regressions of the following form:

$$A_i = \gamma_0 + \gamma_1 N_i + \gamma_2 M_i + \gamma_3 x_i + \gamma_4 z_{j[i]} + \psi_{t[i]} + \epsilon_i, \tag{A.1}$$

where the dependent variable $A_i$ denotes, as in the main text, the number of apprentices trained by firm $i$. $N_i$ and $M_i$ denote, respectively, whether firm $i$ has been newly established or whether firm $i$ has changed its location between two consecutive waves of the business census.[36] The controls are otherwise the same as those used in the regressions discussed in the main text (i.e. we include both firm- and regional-level controls).

Table A.1: Number of apprentices and sample appearance

| | Number of apprentices | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Newly established firm (yes = 1) | -0.525*** | -0.426*** | -0.398*** |
| | (0.018) | (0.024) | (0.025) |
| Moving firm (yes = 1) | -0.292*** | -0.300*** | -0.271*** |
| | (0.020) | (0.020) | (0.015) |
| Constant | 0.877*** | 0.013 | 1.651*** |
| | (0.024) | (0.062) | (0.368) |
| Census year dummies | *No* | *Yes* | *Yes* |
| Firm controls | *No* | *Yes* | *Yes* |
| Location controls | *No* | *No* | *Yes* |
| R-squared | 0.003 | 0.298 | 0.300 |
| Observations | 645,137 | 645,137 | 645,137 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

[36] That is, the dummy variable $N_i$ equals 1 if firm $i$ is observed in wave $t$, but not in wave $t-1$. Similarly, $M_i$ is equal to 1 if the location (i.e. municipality) of firm $i$ in wave $t$ differs from the same firm's location in wave $t-1$.

The resulting estimates of $\gamma_1$ and $\gamma_2$ are shown in table A.1. Column (1) indicates that newly established as well as moving firms train significantly less apprentices than already existing and non-moving firms. These substantial differences in the number of apprentices remain when including control variables, as evident from columns (2) and (3). Altogether, table A.1 suggests a substantially different training behavior among newly established or moving firms. This finding threatens our identification strategy in case firms' location choice also differs between those types of firms.

We thus next analyze the location choice of newly established and moving firms with respect to our instrument by running several regressions of the following form:

$$\ln(D_{j[i]}) = \delta_0 + \delta_1 N_i + \delta_2 M_i + \delta_3 x_i + \psi_{t[i]} + \epsilon_i, \tag{A.2}$$

where the dependent variable now is the log distance of a firm to the border, $\ln(D_{j[i]})$. The regressors of main interest in the above equation are again the two dummies $N_i$ and $M_i$, and the controls are also the same as in equation (A.1), except that we do not include regional-level controls here (because these are, at least in part, determined by a firm's choice of $\ln(D_{j[i]})$, i.e. the dependent variable in equation (A.2)).

Table A.2: Distance to border of newly established and moving firms

| | $\ln(D_{j[i]})$ | |
|---|---|---|
| | (1) | (2) |
| Newly established firm (yes = 1) | -0.046*** | -0.040*** |
| | (0.013) | (0.011) |
| Moving firm (yes = 1) | -0.051** | -0.041* |
| | (0.025) | (0.024) |
| Constant | 3.406*** | 3.554*** |
| | (0.047) | (0.043) |
| Census year dummies | *No* | *Yes* |
| Firm controls | *No* | *Yes* |
| Location controls | *No* | *No* |
| R-squared | 0.001 | 0.013 |
| Observations | 645,137 | 645,137 |

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table A.2 shows the resulting estimates of the two parameters of interest. Column (1), where we do not include any control variables, reveals that newly established firms are located, in terms of driving time, about 4.6% closer to the border than already existing and non-moving firms. Similarily, moving firms are located 5.1% closer to the border than non-moving and already existing firms. In absolute terms, however, these differences are relatively small and amount to about 1.6 and 1.8 car driving minutes only. It seems questionable whether such small differences in commuting durations affect the supply of CBWs for establishing or moving firms. In short, we thus conclude that firms neither tend to establish themselves nor to move closer to the border compared to existing, immobile firms.

However, because table A.1 documents large differences in the number of apprentices, the fact that firms choosing their location within our sample tend to train less apprentices deserves greater attention nonetheless. We address concerns to that effect in three steps. First, we control for various observable firm characteristics, e.g. industry dummies to account for industrial clusters which may have evolved dependent on the distance to the border (cf. section 3.1). Second, table A.1 shows that firms' number of apprentices is especially low at the beginning of their lifecycle. This motivates us to consider firms' age as an additional firm specific control variable (as discussed in section 3.1). Third, we estimate our baseline specification within various sub-samples based on firms' approximated age in section 5.2 in the main text.

## A.2 Other threats to our identification strategy

Besides firm behavior, other channels potentially pose a threat to our identification strategy by violating the exclusion restriction. For instance, municipalities close to the border may systematically differ from regions further away. Suppose, therefore, that urbanization is higher in border regions. Presumably, this might affect the supply of apprentices due to the fact that general education facilities (e.g. high schools or universities) are mostly located in or close to cities; the instrument $\ln(D_{j[i]})$ might therefore be correlated with the error term $\epsilon_i$. To mitigate these concerns, we control for various municipality characteristics, such as the size and the population density. Moreover, the distance between firm $j$'s municipality and the nearest municipality hosting any of the three educational facilities at the upper secondary level (high school, dual VET school, full-time VET school) accounts for exogenous apprentice supply shifts. All these location specific controls are described in section 3.2 in the main text (descriptives are shown in table 1).

Another factor potentially interfering with the distance to the border are the number of resident immigrant workers. On the one hand, immigrants from neighboring countries might choose to settle close to the border and thus to their country of origin. On the other hand, the labor supply of CBWs might crowd out resident immigrant workers and encourage them to move to more central regions of Switzerland. Assuming complementarity between resident immigrant workers and CBWs, both these associations would bias our results in a violation of the exclusion restriction. To mitigate concerns to that effect, we add firms' share of resident immigrant workers to the firm specific controls (for details see section 3.1).

A final threat to the exclusion restriction we consider stems from cross-border apprentices. Any positive correlation between the supply of cross-border apprentices and the distance to the border would weaken the reduced form association between firms' apprentices positions and their distance to the border (shown in table B.4) and would bias the 2SLS estimate downwards. However, data of the Federal Statistical Office indicates that foreign apprentices commuting into Switzerland are extremely rare, almost non-existing.[37]

[37] In the year 2012 (2013; 2014) only 37 (50; 42) out of 311,070 (313,853; 320,883) apprentices commuted from outside of Switzerland to a firm within Switzerland (LVA data 2016).

# B Additional tables and figures

Table B.1: Educational attainment by population groups

| | Swiss | Immigrants | CBWs |
|---|---|---|---|
| University degree | 11.80 | 12.40 | 13.65 |
| Tertiary-B degree, teaching diploma | 10.97 | 5.12 | 8.79 |
| High school degree | 2.21 | 1.74 | 3.17 |
| Vocational qualification (basic level) | 63.45 | 41.12 | 53.11 |
| Compulsory schooling | 9.05 | 29.49 | 12.37 |
| Remaining category | 2.53 | 10.14 | 8.90 |
| Total | 100.00 | 100.00 | 100.00 |

Notes: Immigrants refers to resident immigrant workers (i.e. non-natives working and living in Switzerland). There are slight discrepancies between these figures and those derived from the data used in the analysis below (due to, for example, different sample definitions).
Source: Swiss Earnings Structure Survey (2004).

Table B.2: Descriptive statistics of individuals

| | Full population[a] | | | Sample[b] | | |
|---|---|---|---|---|---|---|
| Wave | 1995 | 2005 | 2008 | 1995 | 2005 | 2008 |
| Number of employees | 3,556,506 | 3,713,028 | 4,013,453 | 3,328,516 | 3,486,617 | 3,780,068 |
| Numer of apprentices | 141,981 | 173,971 | 192,085 | 139,746 | 171,265 | 189,168 |
| Number of CBWs | 140,644 | 170,785 | 211,663 | 138,254 | 167,499 | 207,317 |
| Number of immigrant workers | 857,899 | 895,797 | 1,033,993 | 830,670 | 861,083 | 993,737 |
| Share of apprentice | 0.040 | 0.047 | 0.048 | 0.042 | 0.049 | 0.050 |
| Share of CBWs | 0.040 | 0.046 | 0.053 | 0.042 | 0.048 | 0.055 |
| Share of immigrant workers | 0.241 | 0.241 | 0.258 | 0.250 | 0.247 | 0.263 |
| Share public sector | 0.183 | 0.166 | 0.157 | 0.190 | 0.173 | 0.163 |
| Share third sector | 0.694 | 0.735 | 0.736 | 0.686 | 0.730 | 0.731 |
| Number of firms | 373,472 | 375,392 | 388,324 | 212,081 | 213,106 | 219,950 |

Notes: Immigrant workers refers to the number (share) of resident immigrant workers (i.e. non-natives living and working in Switzerland). [a]All firms operating in the second or third sector. [b]All firms operating in the second or third sector with three or more employees.

Table B.3: First-stage estimates

| | Share of CBWs | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $\ln(D_{j[i]})$ | -0.067*** | -0.067*** | -0.068*** | -0.054*** | -0.056*** |
| | (0.005) | (0.005) | (0.005) | (0.004) | (0.004) |
| Census year dummies | *No* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *No* | *No* | *Yes* | *Yes* | *Yes* |
| Location controls | *No* | *No* | *No* | *Yes* | *Yes* |
| Distance to schools | *No* | *No* | *No* | *No* | *Yes* |
| R-squared | 0.132 | 0.134 | 0.159 | 0.242 | 0.244 |
| Observations | 645,137 | 645,137 | 645,137 | 645,137 | 645,137 |
| F-statistic (first stage) | 201.6 | 201.8 | 174.4 | 170.7 | 183.0 |

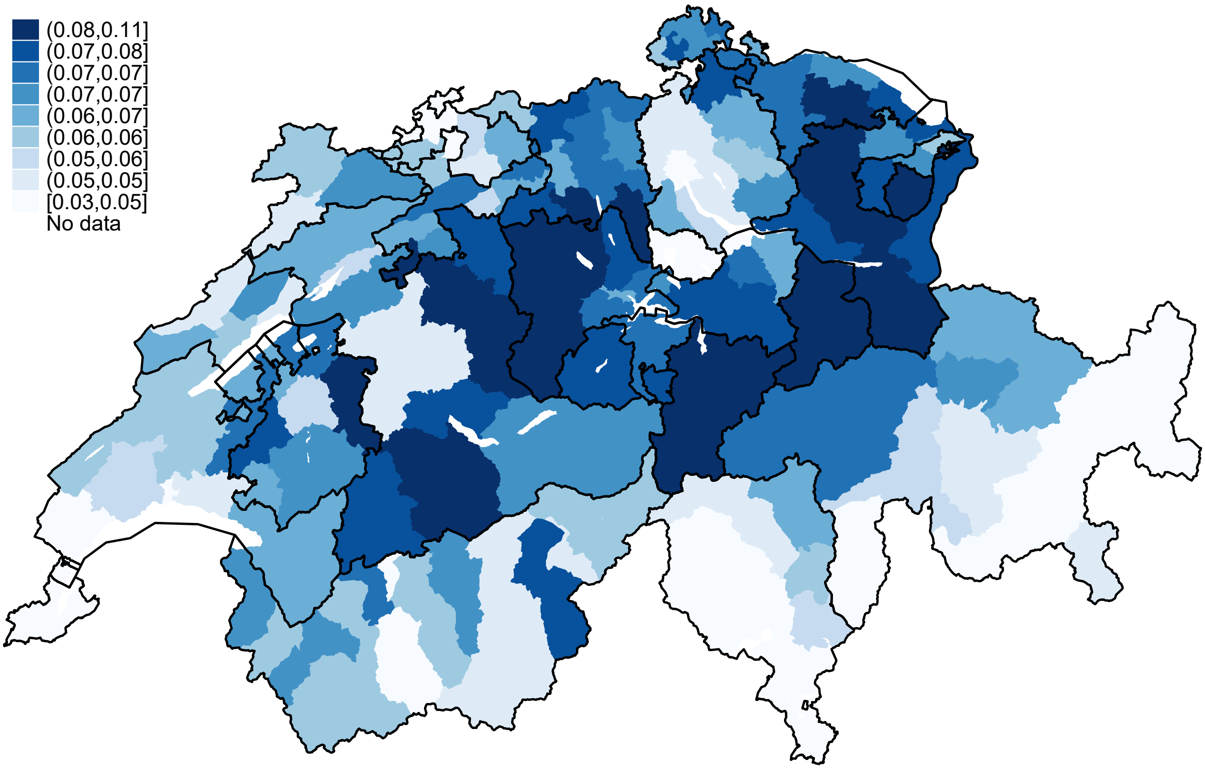
Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table B.4: Reduced-form estimates

| | Number of apprentices | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $\ln(D_{j[i]})$ | $0.118^{***}$ | $0.119^{***}$ | $0.133^{***}$ | $0.072^{***}$ | $0.080^{***}$ |
| | (0.032) | (0.032) | (0.026) | (0.017) | (0.016) |
| Census year dummies | *No* | *Yes* | *Yes* | *Yes* | *Yes* |
| Firm controls | *No* | *No* | *Yes* | *Yes* | *Yes* |
| Location controls | *No* | *No* | *No* | *Yes* | *Yes* |
| Distance to schools | *No* | *No* | *No* | *No* | *Yes* |
| R-squared | 0.000 | 0.001 | 0.298 | 0.300 | 0.300 |
| Observations | 645,137 | 645,137 | 645,137 | 645,137 | 645,137 |

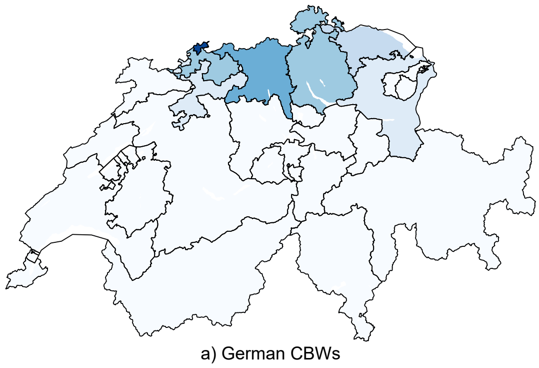
Notes: $^{***}$, $^{**}$, and $^{*}$ denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Figure B.1: Share of apprentices

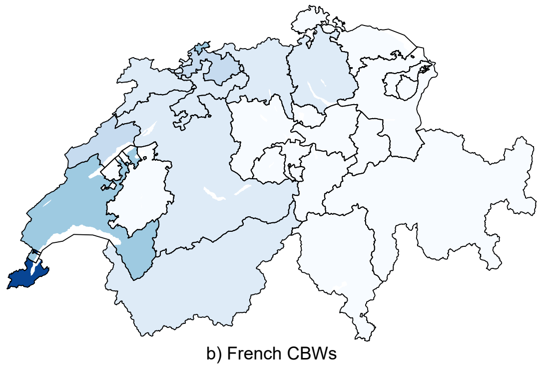



Notes: The map shows the spatial distribution of the ratio of apprentices to total work force on district level (there are 148 unique districts).

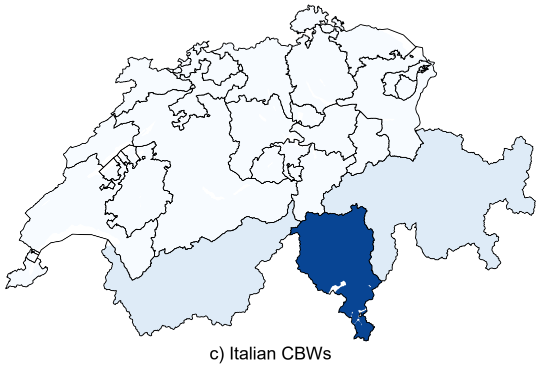
Figure B.2: Cantonal numbers of CBW by country of residence



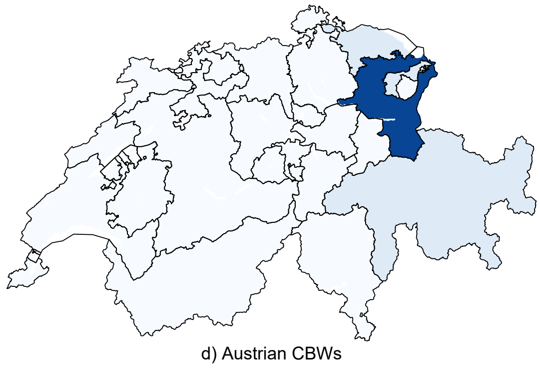

(a) German CBWs




(b) French CBWs




(c) Italian CBWs




(d) Austrian CBWs

Notes: The figure shows cantonal-level means of the absolute number of CBWs in the years 2002 to 2008, by their country of residency.
Source: Official Statistic on Crossborder Commuters by the Federal Statistical Office.