IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13490

Training, Wages and a Missing School Graduation Cohort

Matthias Dorner
Katja Görlitz

JULY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13490

Training, Wages and a Missing School Graduation Cohort

Matthias Dorner
IAB Nürnberg

Katja Görlitz
HdBA, IZA and RWI

JULY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Training, Wages and a Missing School Graduation Cohort*

This study analyzes the effects of a missing high school graduation cohort on firms' training provision and trainees' wages. An exogenous school reform varying at the state and year level caused the missing cohort to occur. Using administrative social security data on all trainees and training firms, we show that firms provide less training by reducing their overall number of hired apprentices. We also show that the pool of firms that offer training in the year of the missing cohort shifts towards a higher share of low wage firms. After keeping firm characteristics constant, the findings indicate that the missing cohort increases training wages measured at the start of training. Further analyses shed light on the opposite case of dual cohorts, which we find to increase training provision and to decrease training wages. The evidence also shows that high and low wage firms differ in how they adjust training provision in response to a dual cohort.

JEL Classification: J21, J24, J31

Keywords: training wages, training provision, missing high school graduation cohort, high and low wage firms, dual high school graduation cohort

Corresponding author:
Katja Görlitz
Hochschule der Bundesagentur für Arbeit (HdBA)
Seckenheimer Landstr. 16
68163 Mannheim
Germany

E-mail: katja.goerlitz@hdba.de

* Financial support from the German Research Foundation (DFG) is gratefully acknowledged. The authors thank Ronald Bachmann, Uschi Backes-Gellner, Thomas K. Bauer, Stefan Bender, Francine Blau, Udo Brixy, Bernd Fitzenberger, Elke Jahn, Sandra McNally, Joachim Moeller, Jens Mohrenweiser, Uta Schoenberg, Marcus Tamm, Sevrin Waights and Till von Wachter for helpful comments and suggestions. We also acknowledge comments from participants at conferences and research seminars (Royal Economic Society 2012, Verein für Socialpolitik 2012, IAB 2012, EALE 2013, First RWI Research Network Conference 2014, IWAEE 2015, FU Berlin 2016, CVER London 2017, Max-Planck-Institute for Innovation and Competition 2017, University of Regensburg 2018, Committee for the Economics of Education 2019). We also want to express deep gratitude to Udo Brixy for his particular support of our project. Matthias Dorner was a research associate at the Institute for Employment Research (IAB) until 2019. All remaining errors are our own.

1. Introduction

Firms provide apprenticeship training in many countries including Australia, Austria, Canada, Denmark, Germany, Switzerland and the UK. Apprenticeship training is a successful vocational pathway for young school leavers to enter the labor market and to keep youth unemployment low (Lerman 2019). In the US, policy makers formulated the aim of expanding apprenticeship training starting with the Obama administration.[1] Analyzing the mechanisms and functioning of training markets is essential to understand why firms provide training and what determines wages of trainees. Gary Beckers' human capital theory distinguishes between general and firm-specific training as a key factor to explain why firms provide training (Becker 1962). While Becker modelled the training decision under perfect competition, another stream of the literature shows that market imperfections like information asymmetries can equip firms with monopsony power that allows them to recoup their training investments through their ability to compress wages (Katz and Ziderman 1990, Chang and Wang 1996, Acemoglu and Pischke 1998, 1999).

This study contributes to the training literature by analyzing the novel research question of whether the supply of trainees – a factor that the previous literature has largely neglected – is a determinant of firms' willingness to provide training and whether it has the potential to affect trainees' wages. We investigate the effects of a decrease in the supply of trainees, meaning the number of school graduates available for an apprenticeship training, that was caused by an exogenous schooling reform. The reform extended the years required to graduate from high school from twelve to thirteen years in two German states in the same year. This induced a missing high school graduation cohort, because the number of high school graduates dropped virtually to zero in the year after the last "12 years"-cohort had graduated and prior to the first "13 years"-cohort. The missing cohort decreased the number of potential trainees with high school degree who could apply for an apprenticeship in the two affected states. Germany is well suited for our analysis, because firms recruit apprentices on an annual basis mainly from the pool of current school graduates and two thirds of the German workforce have completed an apprenticeship program.[2] Apprenticeship training combines formal learning in state-funded vocational schools (for one to two days per week) with working at the training firm (for three to four days). Firms post vacancies for trainees offering a temporary apprenticeship contract that includes paying a training wage. If trainees sign an apprenticeship contract, firms employ them for two to three years at the training firm. While one out of six trainees hold a high school degree, the majority of trainees have acquired fewer years of schooling. Thus, the missing cohort induced an exogenous decrease in the supply of high school graduates available for an apprenticeship.

[1] For an overview of the objectives and initiatives under the former president Obama see https://obamawhitehouse.archives.gov/the-press-office/2016/04/21/fact-sheet-investing-90-million-through-apprenticeshipusa-expand-proven (accessed: 2020-06-16). Another example is President Trump's executive order that is issued on June 15, 2017 and available at https://www.whitehouse.gov/presidential-actions/3245/ (accessed: 2020-06-16).

[2] Focusing on the German apprenticeship system also follows Acemoglu and Pischke (1998, 1999) who investigate it to learn about training processes.

First, we analyze how firms' training provision developed in the year of the missing cohort, which we approximate by the number of newly hired trainees. Second, we investigate the effects of the missing cohort on training wages.[3] To identify these effects, we exploit exogenous variation in the occurrence of the missing cohort by state and year within a difference-in-difference model. The analysis uses data from administrative social security records available at the Institute for Employment Research (IAB) that provide accurate information on the universe of trainees and their training firms.

Analyzing employment and wage responses to a missing cohort also contributes to the literature explaining how labor markets respond to shifts in labor supply, which is essential to understand the fundamental question of how labor markets re-equilibrate.[4] Most of the previous literature is concerned with an immigration-induced positive labor supply shock (see e.g. Card 1990, 2001, Pischke and Velling 1997, Borjas 2003, 2006, Manacorda, Manning, and Wadsworth 2012, Ottaviano and Peri 2012, Glitz 2012, Dustmann et al. 2017).[5] Other studies address this topic investigating demographic shocks such as the size of the birth cohorts (Welch 1979, Berger 1985 and Korenman and Neumark 2000). The study most related to ours is Morin (2015) who uses micro data to investigate the wage effects of excess supply caused by a schooling reform that induced two high school cohorts to graduate in the same year. He shows that this dual cohort decreased weakly earnings significantly. In contrast to Morin (2015) who analyzes the Canadian labor market, German school graduates usually enter the labor market as trainees and rarely as unskilled workers. Our study answers the question whether the training market operates as predicted by the classical labor market model.

The previous theoretical literature describes training as an investment decision of firms who invest in the productivity of their future workforce and does not predict that supply should affect training provision (see Becker 1962, Acemoglu and Pischke 1998, 1999). However, trainees already work in a productive manner during their apprenticeship. Based on German data, Mohrenweiser and Zwick (2009) show that employing apprentices can increase profits in some firms, because trainees perform tasks that otherwise unskilled workers would have to conduct. Thus, firms do not only invest in their workers' human capital by providing training, but also demand productive tasks from trainees. Lerman (2019) reviews the international literature on the training costs of firms. He finds that firms in many countries already recoup much or all of the costs arising from apprenticeship training through the productive work of their trainees. For Germany, where many firms report to bear net costs of training, Mohrenweiser and Zwick (2009) provide evidence that some firms manage to recoup their training costs before the end

[3] Wage rigidities could prevent wage adjustments to happen, because wages are subject to collective wage agreements in Germany. However, firms only have to follow these agreements, if they are part of the employers' association that negotiates with unions over wages. In 2010, this was the case for 30 percent of the firms only (Federal Statistical Office 2010). Furthermore, firms can always pay wages exceeding the collective wage agreements, which is encouraged by the unions. Schnabel and Jung (2011) show that wage cushion is quite common across German firms. Mohrenweiser et al. (2015) show that training wages can differ within firms for trainees in the same occupation and year.

[4] Another strand of the literature analyzes shifts in labor demand e.g. caused by recessions. See von Wachter and Bender (2006) for an application exploring the German apprenticeship training system.

[5] For an overview, see Dustmann et al. (2016).

of the training period. Given that there is heterogeneity in the reasons why firms train, where some firms invest and others produce, our study is complementary to the previous literature.

Besides analyzing the consequence of the missing cohort, we also shed light on the reverse effect of excess supply in the training market. Shortly after the missing cohort has occurred, the German federal states decided to abolish the 13th grade. This reform increased the number of high school graduates to about twice the number of a regular cohort. We are aware of one study that analyzes the corresponding wage and employment effects in the training market. Based on aggregate data varying at the state and year level, Mühlemann et al. (2018) find that the dual cohorts have increased the number of trainees, while they find no evidence of wage adjustments. We will show that using micro data on trainees and their training firms and applying firm fixed effects is essential to uncover the mechanisms of adjustments in training wages.

This is because our results indicate that both the missing cohort as well as the dual cohort induced the average characteristics of training firms to differ from usual years. In particular, we show that high and low wage firms respond differently. This finding makes an additional contribution to the literature that is concerned with high and low wage firms and their role in the development of labor markets (Abowd et al. 1999, Card et al. 2013, Card et al. 2018 and Song et al. 2019). This literature documents that wage differentials for similar workers across these two groups of firms exist (Abowd et al. 1999, Card et al. 2013, Card et al. 2018 and Song et al. 2019). Song et al. (2019) show that much of the rise in earnings inequality derives from increased wage differentials between firms and not within firms. Little is known about how high and low wage firms differ in their hiring policies and whether they adjust wages differently to exogenous shifts in labor supply. In addition, this particular line of heterogeneity across firms is a novel topic in the context of the training market.

We find that the missing cohort reduced the number of newly hired trainees by at least ten percent and increased training wages by at least one percent. The latter finding challenges beliefs of rigid wages in Germany, at least upon first hiring as a trainee. We provide evidence that composition effects – that occur because trainees hired in the year of the missing cohort have acquired fewer years of schooling on average – do not influence our wage estimates. The effects are also robust to alternative calculations of the standard errors and to using a comprehensive set of model specifications considering time trends, regional covariates and alternative control states. Our main model and additional empirical analyses show that low wage firms continued employing trainees in the year of the missing cohort, while high wage firms stopped doing so. This could be because high wage firms abstain from hiring trainees, if their applicants do not satisfy the usual hiring criteria such as having a high school degree. In contrast, the dual cohorts decreased training wages and increased training provision. Our results further suggest that the dual cohort changed the sample of training firms towards a larger share of low wage firms. One explanation for this finding is that low wage firms took the unique opportunity to increase their share of trainees with high school degree, while high wage firms are always able to attract the desired amount and quality of trainees (unless there is a missing cohort). Overall, our findings suggest that shifts in labor supply cause employment and wages to adjust in accordance with the prediction of the classical labor market model.

The remainder of this study is divided into five sections. The following section briefly outlines the training system in Germany, describes the school reform that caused the missing high school graduation cohort to occur and introduces the data. The third section investigates the effects of the missing cohort on firms' training provision. Section four presents the wage results and discusses the role of firm effects. The fifth section presents the analyses and results of the dual cohort. The final section summarizes the results and draws conclusions.

2. Institutional background and data

2.1 The schooling and apprenticeship training system

The German schooling system is characterized by early tracking that separates students after primary school based on students' ability and school performance into three tracks of secondary education (high school, intermediate track and basic track). The track for the high ability students leads to a high school degree after 12 years of schooling in some states or after 13 years in others. This difference in the years of high schooling is due to the constitutional autonomy right of the 16 German federal states to set their own education policy. After school completion, the majority of high school graduates choose to enroll at university or to apply for an apprenticeship training. The intermediate track confers a 10th grade certificate and prepares graduates for an apprenticeship in white-collar occupations. Students from the basic track graduate after nine years of mostly vocationally oriented secondary schooling, which prepares them for an apprenticeship in blue-collar occupations. Entering the labor market directly as unskilled worker is a rare event in Germany for all school leavers. One important reason is that German law requires all adolescents to participate in the schooling or vocational training system until the age of 18.

The apprenticeship system combines working in a firm (3-4 days per week) with publicly-financed vocational schooling (1-2 days).[6] Because it is governed at the federal level, there is no institutional heterogeneity across the 16 states. The curriculum, time schedules, exam requirements and the duration of training, which usually takes between two and three years, are constituted by law for each of the more than 400 officially recognized five-digit occupations (corresponding to 70 three-digit occupations used in this study). School graduates apply for an apprenticeship at the training firms that decide whom of the applicants to hire on a temporary training contract for the full duration of the apprenticeship. Training firms remunerate apprentices with a training wage. As was already mentioned are wages subject to collective

[6] Soskice (1994), Harhoff and Kane (1997) and Wolter and Ryan (2011) provide detailed outlines and an international comparison of the German apprenticeship system.

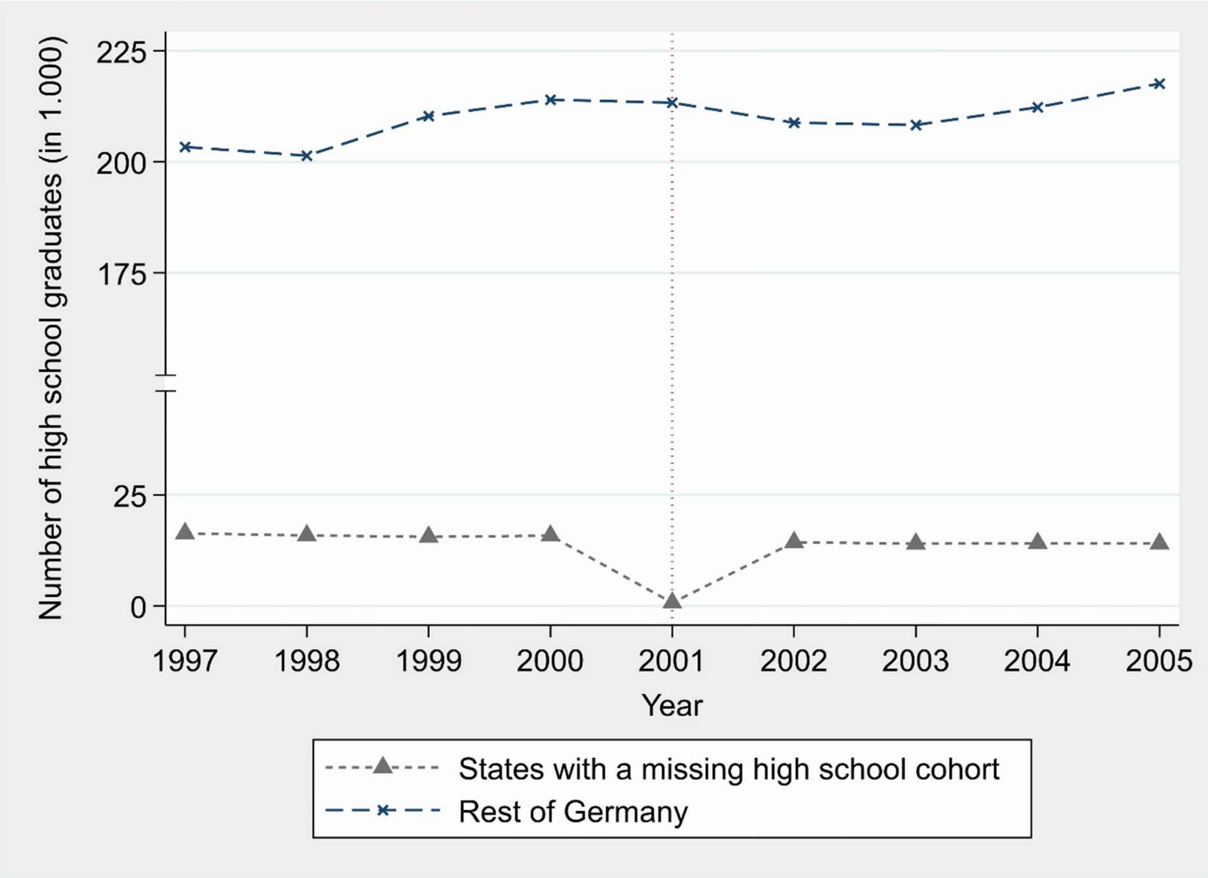
wage agreements, which firms have to follow, if they are part of the employers' association which was just the case for 30 percent of the firms in 2010 (Federal Statistical Office 2010).

Training wages vary significantly by occupation. For example, occupations that have a higher share of high school graduates also pay higher wages and provide better employment prospects measured e.g. in terms of lower unemployment rates and higher wages upon completion of the apprenticeship. At the end of the training period, firms are free to decide how many of their trainees they will retain by offering them a long-term employment contract. The average retention rate is approximately 60 percent (Franz and Zimmermann 2002, Euwals and Winkelmann 2004, von Wachter and Bender 2006, Göggel and Zwick 2012).

2.2. The missing high school graduation cohort

In 2001, there were no high school graduates available for the apprenticeship market in two East German states, Mecklenburg-Western Pomerania and Saxony-Anhalt. A major reform of the high school system caused the missing cohort to occur. In West Germany, attaining a high school degree uniformly took 13 years of schooling at this time, while these regulations varied in the five East German states. In the former German Democratic Republic (GDR), high school required only 12 years of schooling. Two East German states maintained the former requirements and another state switched to the West German standard shortly after German reunification in 1990. Mecklenburg-Western Pomerania and Saxony-Anhalt, extended the years until graduation from 12 to 13 years in the early 2000s. Both states experienced a missing high school cohort in 2001, because the last cohort graduating after 12 years left high school in 2000 and the first cohort graduating after 13 years left high school in 2002. Figure 1 documents that the reform caused the number of high school graduates to drop virtually to zero in 2001 in both states.

Figure 1: High school graduates by states over time



Source: Kultusministerkonferenz (2007).

Graduates from all school tracks are free to apply for an apprenticeship training in every occupation, but the occupation-specific composition of trainees by school degree varies tremendously. To illustrate the extent of this variation, we analyze data from the official statistics on training contracts.[7] Figure A-1 in the Appendix documents that the share of high school graduates among trainees differ greatly by occupation (Federal Statistical Office 1997, 2000). While many occupations exhibit only a low share of high school graduates of less than 10 percent (most of them being blue-collar jobs), few occupations in the service sector even have a share of more than 50 percent. This distribution is similar in 1997 and 2000 and, more generally considered as stable over time. Overall, about 16 percent of apprentices have previously obtained a high school degree in the period 1997 to 2000 (Federal Statistical Office 1997, 2000).

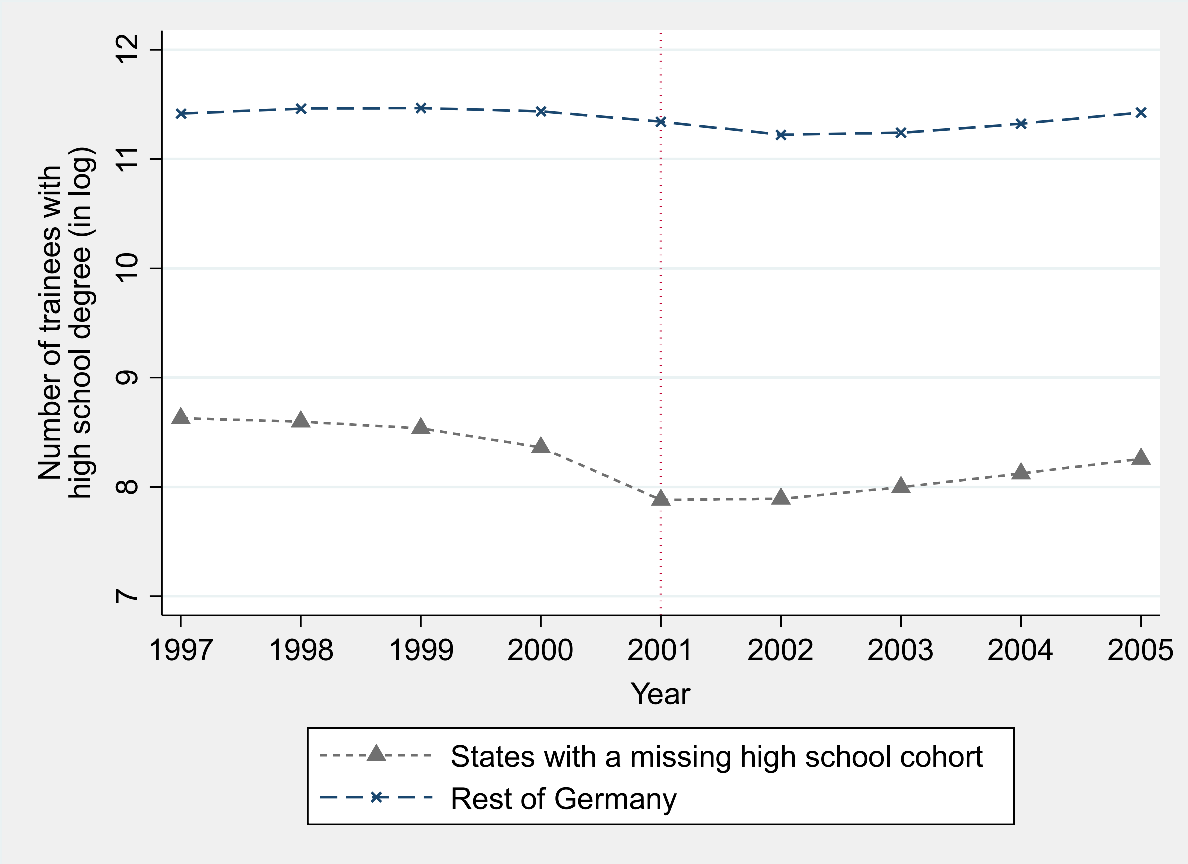
[7] Employers are obliged by law to report each training contract on December 31 to the chambers of industry and commerce. Because the data is not available at the micro level between 1997 and 2000, we use statistics on the aggregated number of new training contracts available by year, state and occupation cells that is provided by the Federal Statistical Office.

Figure 2 shows that the number of trainees with high school degree declined sharply when the two states experienced the missing high school cohort in 2001.[8] It can also be seen that this number did not recover to its pre-reform level. This is likely because graduating after 12 years of high school compared to 13 years affects educational decisions of students. Marcus and Zambre (2018) find that graduating after 12 years lowered the probability to enroll at university compared to graduating after 13 years. This effect occurs even though lengthening the years required for high school had no impact on the overall number of high school graduates (Huebener and Marcus 2017). For our main analysis of the missing cohort, we restrict our analysis to the period 1997 to 2001, which is a period in which high schooling required continuously 12 years. Our empirical model accounts for differences in the high schooling system across states.

The evidence from Figure 2, indicating that firms hired a lower number of trainees having graduated from high school, does not necessarily imply that firms provide less training in general. For instance, firms could anticipate the missing cohort and hire a larger number of trainees in the year prior to the missing cohort, attract high school graduates from unaffected states or substitute high school graduates with school leavers not having acquired a high school degree. Even though it is out of the scope of this study to analyze these adjustment strategies[9], we can show that the missing cohort actually decreased the number of trainees hired. Therefore, the empirical strategy will provide estimates of the causal effect of the missing cohort on training provision. This analysis relies on the data introduced in the next section. Analyzing additionally training provision based on data from the Federal Statistical Office (used in Figure 2) serves the purpose of robustness only. We choose to do so because our main data set allows us to analyze both training provision and wages, while the data from the Federal Statistical Office is restricted to the analysis of training provision only because of the lack of individual wages and firm level information.

[8] Figure A-2 in the Appendix documents that the results are similar when showing the corresponding trends separately for each of the two treated states.
[9] When presenting the robustness analyses, we show that there is no evidence of anticipation effects and we further investigate the role of interstate mobility.

Figure 2: The development of high school graduates among trainees by states



Source: Federal Statistical Office 1997, 2000

2.3 Data

The analysis exploits administrative micro data recording the universe of employees in Germany contributing to the social security system, including all trainees. For the purpose of labor market research, the Institute for Employment Research (IAB) store, process and anonymize the social security data. The data allows us to identify trainees based on the mandatory social security employment records that employers have to report to the social security authorities. From 1999 onwards, trainees can be uniquely distinguished from regular employees because of a major reform of the reporting system. Prior to that, there was no legally binding reporting scheme for trainees. Distinguishing trainees perfectly from interns, student workers or participants in further training is reliable since 1999 only. We address this issue and harmonize the data over the period 1997 to 2001 by implementing a set of heuristics that were originally proposed by von Wachter and Bender (2006).[10] The analysis of the missing cohort

[10] In particular, we exclude individuals who start their training at the age of 30 or older and whose training duration is shorter than 450 days. We further discard individuals for whom the social security data records regular employment prior to the first training observation.

exploits employment records at the reference date June 30 of the years 1997 to 2001. We assign all trainees being hired between July 1 and next years' June 30 to the same training cohort.[11]

The social security data include a comprehensive set of variables covering individuals' characteristics (e.g. gender, age, schooling degree and nationality), a 3-digit classification of occupations and, most importantly, precise wage information. Training firms report gross daily wages with high precision to the social security authorities.[12] We deflate wages to 2010 prices using the consumer price index of the Federal Statistical Office. Even though wages are only recorded up to a social security contribution limit and are top coded otherwise, this is not an issue for our study as these limits significantly exceed the wage levels of trainees. In fact, 99.998 percent of the trainees in our data appear to have non-censored wages. The individual records are complemented with firm characteristics based on unique firm identifiers. Firm characteristics originate from the IAB Establishment History Panel (BHP) and they were generated by aggregating the records of the firms' full workforce (Spengler 2008, Schmucker et al. 2018).

The firm characteristics cover firm size as measured by the number of employees (including trainees), the NACE industry classification, the median wage of full-time workers and the skill composition of the workforce. For the latter, we distinguish the three skill levels: high skilled employees who have graduated from college or university, medium skilled employees who have completed an apprenticeship and low skilled employees who have not attained any of these degrees. Furthermore, the data contain precise geographical identifiers for the location of the training firms at the level of the federal states (NUTS 1). Based on this information, we can assign trainees at the start of their training unambiguously to one of the 16 German states. The data also contains more than 400 counties (NUTS 3) nested within states which is exploited for running sensitivity checks. The data lacks additional information on the state where trainees have graduated from school, which prevents us from additionally analyzing school-to-training mobility across states.

Table A-1 in the Appendix presents summary statistics of the characteristics of trainees and training firms. While the share of high school graduates among apprentices is 16 percent using the administrative data from the Federal Statistical Office, Table A-1 shows that it is 13.3 percent in the social security data only. This confirms the previous literature suggesting that firms' reports of employees' education are not as reliable as the other information from the social security data. Employers tend to report more often the educational degree required by the average workers in the respective position (Fitzenberger et al., 2006). Since most trainees have not graduated from high school, the social security data underestimates the actual share of high school graduates slightly.

[11] More than 90 percent of all trainees start their apprenticeship in the second half of the year, of which 80 percent in the months August and September. This is the period where all school graduates have already finished schooling in Germany.

[12] Although the data contains establishments rather than firms, we refer to them as firms for ease of presentation.

3. The effects of the missing cohort on training provision

3.1 Empirical strategy

We approximate training provision by the number of newly hired trainees in each training cohort. To analyze whether the missing cohort affects training provision, we aggregate the micro data of the social security records at the state, year and three-digit occupation level. Each cell represents the number of newly hired trainees by state, year and occupation. Based on this data, the following OLS difference-in-difference model is estimated:

$$\text{Log}\ (trainees)_{sto} = \delta\ missing\_cohort_{sto} + \theta_s + \varphi_t + \pi_o + \varepsilon_{sto} \qquad (1)$$

where log $(trainees)$ indicates the log of the number of trainees of state s at year t (t=1997-2001) in occupation o.[13] The binary indicator variable $missing\_cohort$ takes the value of one in the two states that experience a missing cohort in the year 2001 (i.e. in Mecklenburg-Western Pomerania and Saxony-Anhalt), and zero otherwise. The remaining 14 German states constitute the control group taking economic or policy changes into account that are common to all German states. The regression includes state fixed effects (θ) to account for statewide difference in the high school system and in training provision. Year fixed effects (φ) absorb changes in economic conditions and occupation fixed effects (π) control for the occupation-specific variation of the share of high school graduates among trainees (see again Figure A-1). The indicator ε represents the idiosyncratic error term. To account for the significant differences in the size of the labor market by states, we weight the regressions by the state-specific number of trainees. The coefficient δ is the parameter of interest that identifies the lower bound of the missing cohort on training provision. One reason why we can only identify a lower bound is that our data does not observe high school graduates from other states to apply at the states with the missing cohort.[14]

Starting with Bertrand et al. (2004), there is a large debate on how to calculate standard errors in difference-in-difference applications when using micro data. When presenting wage estimates, we discuss this literature in detail. However, it is not of importance for our analysis of training provision because the analysis relies on aggregated and not on micro data. Inference is based on the model suggested by Donald and Lang (2007). As was already noted will we aggregate our main data from the social security system at the state, year and occupation level when estimating Equation (1). Additionally, we will estimate Equation (1) based on the data

[13] Table A-2 in the Appendix presents summary statistics of the log of the number of trainees by state and year in the first two columns.

[14] The difference-in-difference estimator can be calculated as (ignoring occupation, state and year fixed effects to keep it as simple as possible): $\hat{\delta} = [E(\log(trainees)_{sto} \mid missing\_cohort = 1, year_{2001} = 1) - E(\log(trainees)_{sto} \mid missing\_cohort = 1, year_{2001} = 0] - [E(E \log(trainees)_{sto} \mid missing\_cohort = 0, year_{2001} = 1) - E(E \log(trainees)_{sto} \mid missing\_cohort = 0, year_{2001} = 0)]$. We expect the sign of $\hat{\beta}$ to be negative. In this setting, mobility induces the first term to increase (as more school graduates apply in the treated states), while the second term decreases (as the school graduates leave from the control states). The effect of the missing cohort is attenuated towards zero.

provided by the Federal Statistical Office which is only available in aggregated form and which was already used in Figure 2. While both sources contain administrative information on the total population of trainees, they differ in the administrative process of reporting including the institutions where firms have to report to, the reference period and the exact unit of observation (number of hired trainees in the social security data versus training contracts).

3.2 Results on training provision

Using data from the social security records, Table 1 shows in the first column that firms train approximately ten percent fewer trainees in response to the missing cohort. The second column contains the sensitivity analysis estimating Equation (1) based on the data from the Federal Statistical Office. These results also show that firms have signed ten percent fewer training contracts.[15] This is an astonishing result given the great extent of differences in the reporting scheme of both data sets.

Table 1: The effect of the missing cohort on training provision

	Log of the number of trainees	
	(1)	(2)
The effect of the missing cohort	-0.102 ***	-0.104 ***
	(0.020)	(0.025)
Adj. R^2	0.939	0.945
Obervations	6,443	5,305

Notes: The dependent variable is the log of the number of trainees. Column (1) presents estimates of Equation (1) using the data from the social security system aggregated at the state, year and occupation level. Column (2) present estimates of Equation (1) using the aggregated data from the Federal Statistical Office. The standard errors are shown in parentheses. Statistical significance: p<0.1 *, p<0.05 **.

Source: Social security data in column (1) and data from the Federal Statistical Office (1997, 2000) in column (2).

[15] Because the social security data and the data from the Federal Statistical Office differ in the definition of the occupational codes, the number of observations differ between column (1) and (2).

4. The effects of the missing cohort on training wages

4.1 Empirical strategy

The analysis of training wages also employs a differences-in-differences design that uses exogenous variation in the missing cohort by state and year. We estimate the following linear regression model (that we henceforth refer to our baseline model):

$$\text{Log}(\text{w}age)_{ist} = \beta\ missing\_cohort_{st} + \rho_s + \gamma_t + \nu_{ist} \quad (2)$$

where log $(wage)$ indicates the log training wage of individual i starting training in state s at year t (t=1997-2001).[16] Again, the binary indicator variable $missing\_cohort$ is 1 in the two states with the missing cohort in the year 2001, and zero otherwise. As before, the analysis uses the rest of Germany to control for statewide economic or policy shocks. The regression includes state fixed effects (δ) that captures unobserved time-invariant state characteristics that could be correlated with training wages like differences in the economic environment or policy differences across states.[17] The year fixed effects (ρ) absorb contemporaneous events common to all states like the business cycle. The indicator ν represents the idiosyncratic error term. The coefficient β pools three potential mechanisms together that might have opposite effects on wages. This is because the missing cohort could affect wages through different channels.

First, it reduces the labor supply of trainees by which it should raise wages according to the classical labor market model (*supply effect*). This model also predicts that the missing cohort reduces the number of trainees, which we already documented in the previous section. Second, the composition of trainees has changed, because of a lower average schooling level. This *composition effect* could reduce average wages of the 2001 cohort. Third, the missing cohort could also mirror *firm effects*. Firm effects would matter, if the missing cohort reduces the number of trainees mainly in firms that pay higher or lower training wages. We consider this as likely because the missing cohort did not only decrease the size of the pool of training applicants, but it also decreased the average schooling quality of the applicants. If high wage firms offer training more often to high school graduates, the missing cohort would affect high wage firms more severely. If mostly these firms decided to offer less training in the year of the missing cohort, because they were unable to fill their open slots with the "usual" candidates, firm effects induce average wages to decrease in the year of the missing cohort.[18] In the opposite case of negative assortative matching of labor market entrants into training firms, low wage firms would have stopped hiring trainees more often.

[16] Table A-2 in the Appendix presents summary statistics of log wages at start of training by states and year.
[17] Introducing state fixed effects seems sufficient to consider state-specific wage differentials because Figure A-3 in the Appendix illustrates that these differentials mainly represent levels and not time trends. Further robustness tests controlling for time trend support this conclusion.
[18] This would assume positive sorting of workers into firms. The empirical literature has not yet reached consensus on whether positive or negative sorting exists (Abowd et al. 1999, Andrews et al. 2008, Eeckhout and Kircher 2011, Andrews et al. 2012, Card et al. 2013 and Ehrl 2019). The evidence on Germany is also mixed. While some previous studies conclude negative sorting to be apparent (Andrews et al. 2007, 2012), more recent studies by Card et al. (2013) and Ehrl (2019) provide evidence of positive assortative matching.

To disentangle the supply effect from the composition effect, we will proceed stepwise. To find out how much the composition effect contributes to the estimated wage results, we amend Equation (2) to control for trainees' characteristics, in particular, for the schooling degree, age, gender and nationality. Occupation fixed effects will be included in the regressions at the 3-digit level to account for the considerable occupational-specific heterogeneity regarding the share of high school graduates. These regressions only present suggestive evidence because changes in the individual characteristics directly stem from the reform itself. Therefore, we present findings from the literature and run an additional empirical analysis, supporting our conclusion that composition effects do not matter in our application.

To find out how much observable firm characteristics matter, we proceed in the same manner as with our analysis of composition effects. Equation (2) additionally controls for time-varying characteristics such as firm size, the median of wages of full-time workers, the skill composition of the workforce and firms' industry affiliation (at the level of 17 NACE sections). Observing all training firms in Germany in our data allows us additionally to apply firm fixed effects to absorb all time-invariant firm characteristics that might influence training wages. Importantly, we will show that applying firm fixed effects absorb both time-invariant and time-varying firm effects.

The following model, henceforth referred to our main model, considers firm fixed effects:

$$\log(wage) = \eta\ missing\_cohort_{st} + \mu_t + \alpha_j + \omega_{ist}. \qquad (3)$$

The variables $\log(wage)$ and $missing\_cohort$ was already described when presenting the baseline model. μ_t represents the vector of year fixed effects and α_j represents the firm fixed effects where j indicates the training firm.[19] ω represents the idiosyncratic error term. The estimate of η displays the supply effect of the missing cohort on wages (given that we will show that composition effects are not an empirical issue and that firm fixed effects are sufficient to control for differences in firm characteristics). Again, η represents a lower bound because of the mobility-induced attenuation bias. Section 4.3 provides further empirical evidence on the extent of this bias by controlling in parts for mobility.

Comparison of the baseline estimates from Equation (2) and the main estimates from Equation (3) sheds light on the question whether the missing cohort affects high and low wage firms differently. If Equation (3) were the true model, the baseline estimate $\hat{\beta}$ would be calculated as (leaving out other controls for ease of exposition):

$$plim\ \hat{\beta} = \eta + \frac{cov(missing\_cohort,\ \alpha_j)}{var(missing\_cohort)}\ . \qquad (4)$$

[19] State fixed effects cannot be considered in addition because firms are almost entirely nested within states, leaving insufficient variation for identification.

If $\hat{\beta}$ and $\hat{\eta}$ were the same (for $cov(missing_{coho}\ \ , \alpha_j) = 0$), the missing cohort would reduce the number of trainees homogenously in all firms, meaning the missing cohort would be unrelated to firm characteristics. If $\hat{\beta} < \hat{\eta}$, the latter term of Equation (4) would become negative. This would happen, if the share of low (high) wage firms increases (decreases) among training firms in response to the missing cohort. If $\hat{\beta} > \hat{\eta}$, the opposite would be the case. To reinforce our conclusion drawn from comparing $\hat{\beta}$ and $\hat{\eta}$, further robustness analyses provide more direct evidence on the effects of the missing cohort on the characteristics of training firms and on the provision of training by high and low wage firms.

Clustering of standard errors and inference

In their seminal paper, Bertrand et al. (2004) emphasized the importance of the choice of method of estimating standard errors in differences-in-differences applications. Standard errors are likely downward biased when the errors are serially correlated over time or within units (Moulton 1990). In general, Abadie et al. (2017) highlight that the choice of the clustering unit for the standard errors should be aligned with the experimental design of each study. Following this literature, we cluster standard errors at the state level in our main model analyzing training wages. Another issue hotly debated is how to proceed in settings with only few clusters as in our case where the highest level of aggregation allows clustering at only 16 German states (see Donald and Lang 2007, Cameron et al. 2008, Conley and Taber 2011). Donald and Lang (2007) as well as Cameron et al. (2008) suggest applying alternative critical *t*-values for inference that can at least reduce the bias in samples with few clusters, which we implement additionally.[20] Furthermore, adjusting standard errors by combinations of the unit structure and time dimensions, i.e. units by years or pre-/post-period, represents another multi-way cluster-robust approach practiced in the literature (Donald and Lang 2007). For reason of sensitivity, we present a variety of alternative calculations of the cluster-robust standard errors.

As an alternative to the cluster-robust inference, Cameron et al. (2008) recommend applying the wild cluster bootstrap to eliminate bias. This cluster-robust variance estimator allows for unrestricted intra-group correlation in differences-in-differences settings and is heteroscedasticity robust. We will implement this approach as a test of robustness for our main model. However, in settings with few treated states as in our main analysis where we only have two treated out of 16 states, MacKinnon and Webb (2017) document that the desirable properties of the approach do not hold and instead lead to unreliable statistical inference. Precisely, the wild cluster bootstrap will produce inference that overrejects the null hypothesis. To find out whether overrejection is apparent in empirical applications, Roodman et al. (2019) suggest implementing a version of the wild cluster bootstrap with restricted and unrestricted heteroscedasticity and, then, compare the conformity of the two estimates. As a rule of thumb, the wild cluster bootstrapped standard errors would be problematic, if the inference from the

[20] They suggest *t(G-1)* or *t(G-2)*, with G denoting the number of clusters in the data.

restricted and the unrestricted model differ from another. In these cases, MacKinnon and Webb (2018) suggest the subcluster wild bootstrap to improve reliability of inference. This approach draws on wild bootstraps that are performed at levels of subunits in nested data. As we observe counties in which training firms are located in our data set that are unambiguously nested in the 16 federal states, we perform the suggested approach at the level of the 401 counties.

4.2 Results on training wages

Table 2 shows the baseline results without applying firm fixed effects using the log training wage at start of training as dependent variable. The estimate is significantly negative, suggesting the missing cohort to decrease wages by four percent. Including trainee characteristics does only modestly alter the estimate as can be seen from column (2). This result is astonishing given that the missing cohort reduced the average schooling level of trainees. We suggest that the most likely reason for this finding is that training wages are only slightly higher for high school graduates compared to school leavers without having acquired a high school degree.[21]

Column (3) indicates that considering occupation fixed effects only slightly decreases the estimate of the missing cohort, but leaves its sign and significance unchanged. In contrast, controlling for log firm size, the median of wages, the skill composition of the firms' workforce and industry identifiers alters the estimate noticeably. The estimate decreases by the factor ten and is no longer statistically distinguishable from zero. This suggests that firm characteristics need to be held constant when analyzing the effect of the missing cohort on wages of trainees.

[21] Pischke and von Wachter (2008) find that increasing schooling by one more year has zero wage returns in Germany. They explain their result by the German schooling system in which basic vocational skills - that are essential for successful completion of an apprenticeship - are already learned by grade 8 regardless of school track. Even though high school graduates attend school longer than their counterparts in the lower tracks, their additional skills and knowledge do prepare them for subsequent academic education, but might not provide large productivity advantages in the training period. To shed more light on this suggestion, we regress the log starting wage on a dummy being 1 for high school graduates and 0 for all other school degrees. These results show that high school graduates earn only three percent higher wages at the start of training compared to other school graduates.

Table 2: Baseline results of the effect of the missing cohort on training wages

	Log training wage			
	(1)	(2)	(3)	(4)
The effect of the missing cohort	-0.040 *** (0.013)	-0.037 *** (0.011)	-0.029 ** (0.013)	-0.003 (0.007)
Trainee characteristics	No	Yes	Yes	Yes
Occupation fixed effects	No	No	Yes	Yes
Firm characteristics	No	No	No	Yes
Adj. R^2	0.264	0.319	0.538	0.711
Obervations	2,151,726	2,151,726	2,151,726	2,151,726

Notes: OLS results of our baseline specification that incorporate a dummy for the missing cohort in addition to state and year fixed effects (see Equation (2)). Trainee characteristics cover gender, age, German nationality and high school degree (y/n). Occupation fixed effects are introduced at the 3-digit level. Firm characteristics indicate log firm size, median of wages of full-time employees, the skill composition of the workforce (high, medium and low skilled worker shares) and 17 NACE industry sections. Standard errors shown in parentheses are clustered at the state level. Significance levels: *** 1%, ** 5%.

Source: Social security data

Table 3 illustrates our main results applying firm fixed effects where the first column again shows the baseline results from column 1 of Table 2 for reason of comparison. Column (2) documents that the lower bound effect of the missing cohort on wages of trainees is one percent once taking firm fixed effects into account. This result remains the same after controlling for trainee characteristics, occupation fixed effects or time-varying firm characteristics (see Column (3) and (4)). This suggests that the firm fixed effects already cover all factors that correlate with training wages and the missing cohort. Comparing the results from the baseline with the fixed effects model shows that $\hat{\beta} < \hat{\eta}$, which is evidence that the missing cohort induced a negative selection of training firms. This could be either because firms with inferior characteristics and low wages kept hiring trainees, while firms with superior characteristics and higher wages stopped to provide training. Alternatively, it could also be that both high and low wage firms decreased training provision, but the former reduced it to a greater extent. The following two paragraphs answer this question and provide further sensitivity analyses supporting our conclusion that the missing cohort affects the quality of the average training firm.

Table 3: Firm fixed effects results of the effect of the missing cohort on training wages

	Log training wage			
	(1)	(2)	(3)	(4)
The effect of the missing cohort	-0.040 *** (0.013)	0.010 ** (0.004)	0.010 *** (0.003)	0.010 *** (0.003)
Firm fixed effects	No	Yes	Yes	Yes
Trainee characteristics	No	No	Yes	Yes
Occupation fixed effects	No	No	Yes	Yes
Firm characteristics	No	No	No	Yes
Adj. R^2	0.264	0.907	0.919	0.919
Obervations	2,151,726	2,151,726	2,151,726	2,151,726

Notes: Column (1) reports again the baseline results from Table 2. Column (2) documents the firm fixed effects estimates that incorporate a dummy for the missing cohort in addition to year and firm fixed effects (see Equation (3)). Columns (3) and (4) stepwise include further controls. Trainee characteristics cover gender, age, German nationality and high school degree (y/n). Occupation fixed effects are introduced at the 3-digit level. Firm characteristics indicate log firm size, median of wages of full-time employees, the skill composition of the workforce (high, medium and low skilled worker shares) and 17 NACE industry sections. Standard errors shown in parentheses are clustered at the state level. Significance levels: *** 1%, ** 5%.

Source: Social security data

First, we document which firms kept hiring trainees in the year of the missing cohort by regressing several firm characteristics on a the treatment dummy in addition to state and year fixed effects. The dependent variables the log firm size, the log median wage and the shares of high, medium and low skilled workers are investigated with a one-year lag in order to describe changes in firm characteristics.[22] Table 4 documents that training firms that usually employ a larger share of low skilled workers and a lower share of medium and high skilled workers were more likely to provide training in the year of the missing cohort.

[22] Investigating firm characteristics in 2001 would answer another question that is out of the scope of this study, meaning e.g. whether firms have hired a larger number of low or medium skilled workers to compensate for the missing high school graduation cohort.

Table 4: Characteristics of training firms and the missing cohort

	Log firm size	Log median wage	Share of low skilled workers	Share of medium skilled workers	Share of high skilled workers
	(1)	(2)	(3)	(4)	(5)
The effect of the missing cohort	-0.017 (0.036)	-0.013 (0.009)	0.025 *** (0.003)	-0.017 *** (0.005)	-0.007 ** (0.003)
Adj. R^2	0.022	0.158	0.082	0.103	0.022
Obervations	1,188,146	1,188,146	1,188,146	1,188,146	1,188,146

Notes: The dependent variables that are measured with a one-year lag are shown in the first row. The number of observations differs from Table 3, because the regression is conditional on the training firm being active in the previous year. All training firms in 1997 had to be deleted additionally because there is no pre-year information. Low skilled workers have not obtained a vocational degree, medium skilled workers have successfully completed an apprenticeship and high skilled workers have graduated from university or college. Standard errors shown in parentheses are clustered at the state level. Significance levels: *** 1%, ** 5%.

Source: Social security data

Second, we analyze how high and low wage firms hired trainees in the year of the missing cohort. To do so, we calculate the residuals from a firm level regression of the log training wage on log firm size, the skill composition of the workforce, industry, year and states dummies for the pooled period 1997 to 2000 for every firm with at least one training record during this period. Taking the mean of the residuals for every firm allows us to observe each firms' position in the training wage distribution adjusted for firm and workforce characteristics. We define high and low wage firms based on quartiles (but also on tertiles for reason of robustness). High wage firms represent the top quartile (tertile) and low wage firms represent the lower quartile (tertile). To analyze how high and low wage firms provide training, we follow our previous analysis of training provision and calculate the overall number of trainees hired by state, year and occupation cells as well as by high and low wage firms.[23] Using this data, the log of the number of trainees is regressed on a binary indicator of the missing cohort, state, year and occupation fixed effects in separate regressions distinguishing high and low wage firms. As before, the regressions consider weights to account for differences in the size of the state's labor market. Column (1) and (3) of Table 5 illustrates that low wage firms did not modify the number of hired trainees in response to the missing cohort. In contrast, columns (2) and (4) documents that high wage firms reduced hiring trainees by about 17 percent (exp(($-$0.192)-1)*100) to 20 percent (exp(($-$0.232)-1)*100). Taken all the evidence from Table 3 to 5 together, we conclude

[23] Because of aggregating the data, the number of observations differ slightly between using quartiles or tertiles.

that particularly firms with superior characteristics decided not to hire as many trainees as usually when facing a missing cohort.

Table 5: Training provision by high and low wage firms

	Log of the number of trainees			
	Low wage firms (quartiles)	High wage firms (quartiles)	Low wage firms (tertiles)	High wage firms (tertiles)
	(1)	(2)	(3)	(4)
The effect of the missing cohort	-0.006	-0.192 **	-0.017	-0.232 ***
	(0.063)	(0.074)	(0.062)	(0.074)
Adj. R^2	0.855	0.851	0.878	0.878
Obervations	5,555	5,312	5,766	5,611

Notes: The dependent variable is the log of the number of trainees. See more information on the empirical proceeding in the text. Standard errors shown in parentheses are clustered at the state level. Significance levels: *** 1%, ** 5%.

Source: Social security data

4.3 Robustness analyses of the main wage results applying firm fixed effects

Clustering of the standard errors

To assess whether clustering of the standard errors affects the statistical inference of our main effects, we vary the level and type of clustering. Table 6 summarizes the results. Panel A shows again the main results in parentheses where inference applies cluster-robust standard errors at the state level adjusted for a small number of state clusters. This seems important given that the results from Panel B point out that the standard errors are underestimated without clustering at the state level. Before presenting the results of the wild cluster bootstrap (shown in brackets), we compare our main results with the subcluster-robust inference that are indicated in parentheses. Panel C to E shows results clustering at the county level (401 clusters), at the states × year level (80 clusters) and at the level of states × pre-/post-indicator (32 clusters), respectively. These findings confirm the statistically significance of our main results.

Next, we discuss the results when applying the wild cluster bootstrap (WCB). Panel A shows that the *p*-values of the restricted and the unrestricted WCB model lead to ambiguous conclusions about the estimates' statistical significance. Such an ambiguous finding suggests that the WCB approach is not valid in cases with few treated clusters only as apparent in our analyses (Roodman et al. 2019). In such cases, MacKinnon and Webb (2018) recommend instead estimating the WCB clustered at a finer level of aggregation. Clustering at the level of 401 counties, which leads to 26 treated units, brings the inference from the restricted and the unrestricted WCB in accordance (see Panel C) and shows our findings to differ significantly from zero. Subclustering wild bootstraps at the level of combinations between state and year as in Panel D and E is not superior to clustering at the county level because the restricted and the unrestricted model are not in line with each other. One likely reason for this is that these models also consider only an insufficiently small number of treated units.

These findings suggest that the restricted WCB model clustered at the state level (Panel A) overrejects the null hypothesis in our setting. To shed further light on this issue, section 5 presents the results from the dual cohorts, i.e. high school cohorts that are twice as large as usually occurring in as much as 13 states between 2007 and 2016. The analysis finds statistically significantly results of both the restricted and the unrestricted WCB irrespectively of clustering at the state, county or any combinations of the state and year level. This supports our conclusion that the WCB does not work well when analyzing the effects of the missing cohort because of too few treated units.

Table 6: Firm fixed effects results with alternative calculations of the standard errors

	Log training wage	
	(1)	(2)
Panel A		
Firm fixed effects results (shown in Table 3) clustered at the state level (G=16)	0.010 **	0.010 ***
	(0.004)	(0.003)
p -value, wild cluster bootstrap (restricted)	[0.188]	[0.168]
p -value, wild cluster bootstrap (unrestricted)	[0.092]	[0.031]
Panel B		
Unclustered standard errors	0.010 ***	0.010 ***
	(0.001)	(0.001)
Panel C		
Standard errors clustered at the county level (G=401)	0.010 ***	0.010 ***
	(0.004)	(0.004)
p -value, wild cluster bootstrap (restricted)	[0.008]	[0.002]
p -value, wild cluster bootstrap (unrestricted)	[0.005]	[0.001]
Panel D		
Standard errors clustered at the level of state x year (G=80)	0.010 ***	0.010 ***
	(0.004)	(0.003)
p -value, wild cluster bootstrap (restricted)	[0.186]	[0.170]
p -value, wild cluster bootstrap (unrestricted)	[0.028]	[0.014]
Panel E		
Standard errors clustered at the level of states x pre-/post-reform year (G = 32)	0.010 ***	0.010 ***
	(0.003)	(0.002)
p -value, wild cluster bootstrap (restricted)	[0.252]	[0.250]
p -value, wild cluster bootstrap (unrestricted)	[0.061]	[0.018]
Firm fixed effects	Yes	Yes
Trainee characteristics	No	Yes
Occupation fixed effects	No	Yes
Firm characteristics	No	Yes
Obervations	2,151,726	2,151,726

Note: The standard errors are shown in parentheses. The *p*-values reported in brackets were obtained using regressions with wild cluster bootstrapped standard errors (999 iterations) clustered at the level as indicated in the label. Significance level: p<0.1 *, p<0.05 **.

Source: Social security data

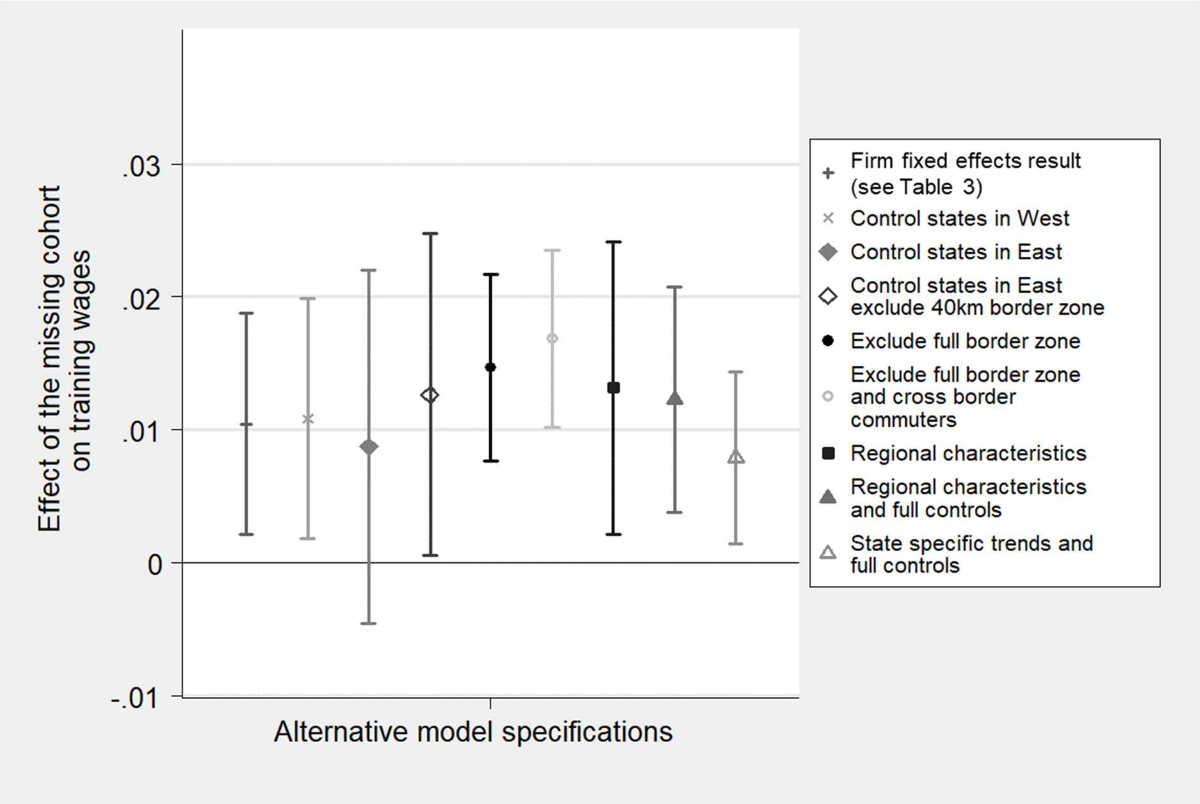
Alternative definitions of the control group and model specifications

It is an open question which states should constitute the control group. To proof the robustness of our main findings, we alternatively use either the rest of West Germany or the rest of East Germany as control states. Figure A-3 in the Appendix shows that wages follow parallel trends regardless of using the rest of West or East Germany. Figure 3 illustrates that the main conclusions remain the same when only West German states are used in the control group. The result is very similar to our main estimator both in terms of magnitude and in terms of statistical precision. Considering only East German decreases the estimate modestly to 0.9 percent, but it turns statistically insignificant (p-value: 0.12). As was already noted suffers the estimator from attenuation bias due to mobility. This bias is likely more pronounced when using only the rest of East Germany as control. This is because labor market mobility such as daily commuting as well as the frequency of relocations are much higher in the East compared to the West (Bogai et al. 2008; Jost et al. 2019) and mobility is mostly directed from the East to the West (Decressin 1994; Suedekum 2004; Hunt 2000, 2006). Figure A-4 in the Appendix shows that the greater part of the border of the states with the missing cohort are located adjacent to the rest of East Germany (which covers the federal states Berlin, Brandenburg, Saxony and Thuringia).

To account for the mobility-induced attenuation, we show results that delete all regions in the states with the missing cohort that are located less than 40 km from the state border. This reduces the bias arising from permanent residential relocations or daily commuting to these counties from adjacent states. Running the main results on this geographically restricted sample with the rest of East Germany as control group shows that the estimate increases to 1.3 percent and it turns statistically significant ($p<0.05$). In addition to deleting the 40 km border zone, we also delete the counties of the unaffected states that are located adjacent to either of the two treated states. Because the adjacent counties are located both in the West and in the East, we use the rest of Germany as control states. This estimate even increases to 1.5 percent ($p<0.05$). As of 1999, we have also access to information on the place of residence of trainees, which enables us to identify commuting trainees by comparing the state of residence with the state of the workplace. Dropping the full border zone in addition to all interstate commuters yields an estimate of 1.68 percent ($p<0.01$). These results suggest that mobility indeed attenuates the estimates of the missing cohort.

Further robustness analyses keep time-varying regional characteristics constant. The state-specific unemployment rate accounts for business cycle fluctuations. Considering the state-specific share of individuals being 17 years old varies by state because of yearly fluctuations in the size of the birth cohorts. Figure 3 shows that including these regional covariates leaves the main conclusion unchanged, both in the main specification as well as in specification considering full controls. Furthermore, we also consider state-specific trends in the analysis to find out whether differential trends in the development of wages can bias our main results. The previous analysis assumes that the development of training wages over time is similar across German states. The results are robust to controlling for state-specific time trends, albeit the estimate of the effect size decreases slightly to 0.8 percent.

Figure 3: Sensitivity analyses considering alternative control states, regional characteristics and time trends



Notes: The graph shows the estimates together with the 95% confidence intervals. Regional characteristics comprise the state-specific unemployment rate and the share of the population being age 17.

Source: Social security data

Choice of reference period and placebo tests

The previous results have pooled all years before 2001 together in the reference period. To prove whether these results are robust to the choice of the reference period, we show estimates using the earliest year possible, i.e. 1997, as base category. The estimated model is similar to our main model, but it differs by including additional interaction terms between the states experiencing the missing cohort in 2001 and the years 1998, 1999 and 2000, respectively.[24] Note that these interaction terms have a meaningful interpretation itself, because they indicate whether the missing cohort already had an impact on training wages in 1998, 1999 or 2000.

[24] The following model is estimated: $y_{ist} = \delta_{2001}\, states\ with\ missing\ cohort\ \times 2001_{st} + \delta_{2000}\, states\ with\ missing\ cohort\ \times 2000_{st} + \delta_{1999}\, states\ with\ missing\ cohort\ \times 1999_{st} + \delta_{1998}\, states\ with\ missing\ cohort\ \times 1998_{st} + \mu_t + \alpha_j + \nu_{ist}$ where δ_{2001} indicates the effect of the missing cohort in 2001 when 1997 serves as the base category. The coefficients δ_{2000}, δ_{1999} and δ_{1998} reveal whether the treatment states' development of training wages already deviates from those of the non-affected states prior to the occurrence of the missing cohort.

Further note that the interpretation of these terms is similar to interpreting results from an analysis of placebo effects. If any of the effects prior to 2001 were statistically significant, this would hint at effects arising from firms anticipating the reform or at contemporaneous shocks that occur at the level of the affected states. Both would render our empirical strategy invalid.

Table 7: Sensitivity analyses varying the choice of the reference year

	Log training wage			
	(1)	(2)	(3)	(4)
The effect of the missing cohort (in 2001)	0.010 ** (0.004)	0.014 *** (0.004)	0.012 *** (0.004)	0.012 *** (0.004)
States with missing cohort interacted with year 1999	not included	0.004 (0.007)	0.003 (0.007)	0.003 (0.007)
States with missing cohort interacted with year 1998	not included	0.002 (0.009)	0.0003 (0.008)	0.0002 (0.008)
States with missing cohort interacted with year 1997	not included	0.007 (0.005)	0.006 (0.005)	0.006 (0.005)
Firm fixed effects	Yes	Yes	Yes	Yes
Trainee characteristics	No	No	Yes	Yes
Occupation fixed effects	No	No	Yes	Yes
Firm characteristics	No	No	No	Yes
Adj. R^2	0.907	0.907	0.919	0.919
Obervations	2,151,726	2,151,726	2,151,726	2,151,726

Notes: The first column repeats the results from our main fixed effects model. Column (2) to (4) shows firm fixed effects results that use 1997 as baseline year and consider additional interaction terms, which indicate whether wage effects already occurred in the states with the missing cohort before 2001. Standard errors shown in parentheses are clustered at the state level. Significance levels: *** 1%, ** 5%.

Source: Social security data

The first column of Table 7 displays again the results of our main firm fixed effects model. Columns (2) to (4) contain the sensitivity results after stepwise inclusion of covariates. They support our conclusion of a statistically significantly positive effect of the missing cohort on training wages. The main results are, therefore, robust to using 1997 as reference period. Table 7 also shows that there are neither anticipation effects nor contemporaneous shocks in previous years because each of the additional interaction terms is statistically indistinguishable from zero and much smaller in magnitude.

5. The training and wage effects of dual high school graduation cohorts

Thirteen out of the sixteen German states decided to abolish the 13th grade of high school between 2007 and 2016, including the two states extending the years of schooling shortly beforehand. Abolishing one year of high school leads to dual high school graduation cohorts because the last "13 years" cohort graduates jointly with the first "12 years" cohort. These dual high school graduation cohorts occur across the different states at different points in time (see Table A-3 in the Appendix).

Covering the period 2001 to 2015, the analysis relies on data from the social security system. The effect of the dual cohort on training provision and wages follows the previous differences-in-differences models described in Equation (1) and (3), respectively, with the important difference of analyzing dual cohorts instead of the missing cohort.[25] Table 8 documents the corresponding results. We find a statistically significantly positive effect on the number of trainees of 5.7 percent and a statistically significantly negative wage effect of -0.6 percent. Applying the wild cluster bootstrap at the state level unambiguously supports inference of our findings. Table A-4 in the Appendix contains detailed sensitivity results of alternative calculations of the standard errors of which none contradicts our conclusions.[26]

[25] As was already mentioned can wages adjust downwards in Germany because many firms do not commit themselves to pay according to the collective agreements. However, even firms for whom the collective agreements are binding can reduce wages up to the union wage if they usually remunerate above the level of the agreements. Jung and Schnabel (2011) show that 40 per cent of establishments covered by collective agreements pay wages above the level stipulated in the agreement.

[26] These results reinforce our previous conclusion from the missing cohort that the reason for issues with the WCB inference were due to the low number of treated states. While the analysis of the missing cohort only considers two treated states, the dual cohort emerges for 13 German states.

Table 8: Effect of the dual cohort on training provision and wages

	Log of the number of trainees	Log training wage		
	(1)	(2)	(3)	(4)
Effect of the dual high school graduation cohort	0.057 *	-0.006 **	-0.006 **	-0.006 **
	(0.030)	(0.002)	(0.002)	(0.002)
p -value, WCB (restricted)		[0.019]	[0.014]	[0.018]
p -value, WCB (unrestricted)		[0.002]	[0.001]	[0.001]
Firm fixed effects	-	Yes	Yes	Yes
Trainee characteristics	-	No	Yes	Yes
Occupation fixed effects	Yes	No	Yes	Yes
Firm characteristics	-	No	No	Yes
Adj. R^2	0.954	0.904	0.908	0.908
Obervations	17,153	5,218,430	5,218,430	5,218,430

Notes: Dependent variables are indicated in the first row. Standard errors shown in parentheses are clustered at the state level. The *p*-values reported in brackets were obtained from OLS regressions with wild cluster bootstrapped standard errors (999 iterations) at the state level. Significance levels: *** 1%, ** 5%, * 10%.

Source: Social security data

Comparing the firm fixed effects results to our baseline specification without applying firm fixed effects reveals how the composition of training firms changed in response to the dual cohort. The first column of Table 9 contains the baseline results showing that the estimate becomes even more negative than our main fixed effects estimate. This suggests that the dual cohort induced the sample of training firms to shift towards a larger share of low wage firms. Put differently, low wage firms increased their training provision to a larger extent than high wage firms did. One explanation for this finding could be that high wage firms can always fill their open slots with high quality workers through their ability to pay higher wages, while low wage firms take the opportunity of a dual cohort to increase their number of trainees with high school degree.

Table 9: Baseline results for the dual cohort

	Dependent variable: Log wage at start of training		
	(1)	(2)	(3)
Effect of the dual high school graduation cohort	-0.022 *	-0.024 **	-0.014 **
	(0.012)	(0.010)	(0.005)
Training firm fixed effects	No	No	No
Trainee characteristics	No	Yes	Yes
Occupation fixed effects	No	Yes	Yes
Firm characteristics	No	No	Yes
Adj. R^2	0.215	0.473	0.657
Obervations	5,218,430	5,218,430	5,218,430

Notes: The results were estimated from OLS regressions. Standard errors shown in parentheses are clustered at the state level. Significance levels: *** 1%, ** 5%, * 10%.

Source: Social security data

6. Conclusion

The previous literature has modelled training as an investment of firms into their workforces' skills. We provide novel evidence that gives a more nuanced view on the economics of the training system by showing that a missing cohort, which exogenously decreases the supply of trainees, reduces the number of trainees hired and increased at the same time trainees' wages. Analyzing the opposite case of dual cohorts, leading to excess supply of trainees, increases the number of trainees and reduces training wages. These results reveal that fundamental labor market mechanisms are at work when the training market is confronted with supply shifts. These results also confirm that wages are not completely rigid in Germany. They can adjust up- and downward in response to shifts in supply, at least upon first hiring.

The missing cohort did not only change the general supply of trainees, but rather the supply of high school graduates in particular. In this setting, we find that high and low wage firms respond differently to the missing cohort. While high wage firms abstain from hiring new trainees, we cannot observe adjustments in training provision by low wage firms. In the case of excess supply of high school graduates, our results suggest that the share of low wage firms increases within the population of training firms. This finding illustrates that using a data set covering firm characteristics and a unique firm identifier is essential to eliminate estimation biases by applying firm fixed effects. It is an open question whether our findings on the importance of firm effects also occur in an analysis where shifts in labor supply hit the labor market in general. Or whether it is a particularity of the apprenticeship system where some firms invest in the future skills of their workforce, while others employ apprentices mainly to perform productive work during the training period.

From a policy perspective, we find that scarce supply of apprentices having acquired a high school degree prevents high wage firms from providing training. If these firms provide training of higher quality, lowering their training activities would also decrease the overall human capital accumulated in apprenticeship training. Thus, demographic change could have adverse effects on human capital formation. To attenuate the effects of negative supply shifts in low-mobility countries like Germany, responsible authorities should implement policies like information campaigns to increase interstate mobility of school graduates to compensate for a missing cohort in one state. An alternative option involves avoiding reforms to occur in adjacent states because the joint appearance of missing cohorts reduces the supply of school leavers from other states.

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J., 2017. When Should You Adjust Standard Errors for Clustering? NBER Working Paper No. 24003.

Abowd, J.M., Kramarz, F., Margolis, D.N., 1999. High Wage Workers and High Wage Firms. Econometrica 67, 251-333.

Acemoglu, D., Pischke, J.-S., 1998. Why Do Firms Train? Theory and Evidence. The Quarterly Journal of Economics 113, 79-119.

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

Andrews, M.J., Gill, L., Schank, T., Upward, R., 2008. High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias? Journal of the Royal Statistical Society: Series A (Statistics in Society) 171, 673-697.

Andrews, M.J., Gill, L., Schank, T., Upward, R., 2012. High Wage Workers Match with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias. Economics Letters 117, 824-827.

Becker, G.S., 1962. Investment in Human Capital: A Theoretical Analysis. Journal of Political Economy 70, 9-49.

Berger, M.C., 1985. The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence. Journal of Political Economy 93, 561-573.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How Much Should We Trust Differences-in-Differences Estimates? The Quarterly Journal of Economics 119, 249-275.

Bogai, D., Seibert, H., Wiethölter, D., 2008. Duale Ausbildung in Deutschland: Die Suche nach Lehrstellen macht junge Menschen mobil. IAB-Kurzbericht, 09/2008.

Borjas, G.J., 2003. The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. The Quarterly Journal of Economics 118, 1335-1374.

Borjas, G.J., 2006. Native Internal Migration and the Labor Market Impact of Immigration. Journal of Human Resources 41, 221-258.

Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. Review of Economics and Statistics 90, 414-427.

Card, D., 1990. The Impact of the Mariel Boatlift on the Miami Labor Market. Industrial and Labor Relations Review 43, 245-257.

Card, D., 2001. Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. Journal of Labor Economics 19, 22-64.

Card, D., Cardoso, A.R., Heining, J., Kline, P., 2018. Firms and Labor Market Inequality: Evidence and Some Theory. Journal of Labor Economics 36, 13-70.

Card, D., Heining, J., Kline, P., 2013. Workplace Heterogeneity and the Rise of West German Wage Inequality. The Quarterly Journal of Economics 128, 967-1015.

Chang, C., Wang, Y., 1996. Human Capital Investment under Asymmetric Information: The Pigovian Conjecture Revisited. Journal of Labor Economics 14, 505-519.

Conley, T.G., Taber, C.R., 2011. Inference with 'Difference in Differences' with a Small Number of Policy Changes. Review of Economics and Statistics 93, 113-125.

Decressin, J.W., 1994. Internal Migration in West Germany and Implications for East-West Salary Convergence. Weltwirtschaftliches Archiv 130, 231-257.

Donald, S.G., Lang, K., 2007. Inference with Difference-in-Differences and Other Panel Data. Review of Economics and Statistics 89, 221-233.

Dustmann, C., Schonberg, U., Stuhler, J., 2017. Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment. The Quarterly Journal of Economics 132, 435-483.

Eeckhout, J., Kircher, P., 2011. Identifying Sorting - In Theory. Review of Economic Studies 78, 872-906.

Ehrl, P., 2019. On The Use Of Firm Fixed Effects As A Productivity Measure For Analyzing Labor Market Matching. Bulletin of Economic Research 71, 195-208.

Euwals, R., Winkelmann, R., 2004. Training intensity and first labor market outcomes of apprenticeship graduates. International Journal of Manpower 25, 447-462.

Federal Statistical Office/ Statistisches Bundesamt, 1997. Bildung und Kultur. Berufliche Bildung. Berichtszeitraum 1997. Wiesbaden.

Federal Statistical Office/ Statistisches Bundesamt, 2000. Bildung und Kultur. Berufliche Bildung. Berichtszeitraum 2000. Wiesbaden.

Federal Statistical Office/ Statistisches Bundesamt, 2010. Verdienste und Arbeitskosten. Tarifbindung in Deutschland. Wiesbaden.

Fitzenberger, B., Osikominu, A., Volter, R., 2006. Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. Schmollers Jahrbuch: Journal of Applied Social Science Studies 126, 405-436.

Franz, W., Zimmermann, V., 2002. The transition from apprenticeship training to work. International Journal of Manpower 23, 411-425.

Glitz, A., 2012. The Labor Market Impact of Immigration: A Quasi-experiment Exploiting Immigrant Location Rules in Germany. Journal of Labor Economics 30, 175-213.

Göggel, K., Zwick, T., 2012. Heterogeneous wage effects of apprenticeship training. The Scandinavian Journal of Economics 114, 756-779.

Harhoff, D., Kane, T.J., 1997. Is the German Apprenticeship System a Panacea for the U.S. Labor Market? Journal of Population Economics 10, 171-196.

Huebener, M., Marcus, J., 2017. Compressing instruction time into fewer years of schooling and the impact on student performance. Economics of Education Review 58, 1-18.

Hunt, J., 2006. Staunching Emigration from East Germany: Age and the Determinants of Migration. Journal of the European Economic Association 4, 1014-1037.

Jost, O., Seibert, H., Wiethölter, D., 2019. Regionale Mobilität von Lehrlingen: Auszubildende in MINT-Berufen pendeln besonders häufig. IAB-Kurzbericht, 02/2019.

Jung, S., Schnabel, C., 2011. Paying More Than Necessary? The Wage Cushion in Germany. Labour 25, 182-197.

Katz, E., Ziderman, A., 1990. Investment in General Training: The Role of Information and Labour Mobility. Economic Journal 100, 1147-1158.

Korenman, S., Neumark, D., 2000. Cohort Crowding and Youth Labor Markets (A Cross-National Analysis), Youth Employment and Joblessness in Advanced Countries. National Bureau of Economic Research, pp. 57-106.

Kultusministerkonferenz, 2007. Schüler, Klassen, Lehrer und Absolventen der Schulen 1997 bis 2006.

Lerman, R., 2019. Do firms benefit from apprenticeship investments? IZA World of Labor 2019: 55.

MacKinnon, J.G., Webb, M.D., 2017. Wild Bootstrap Inference for Wildly Different Cluster Sizes. Journal of Applied Econometrics 32, 233-254.

MacKinnon, J.G., Webb, M.D., 2018. The Wild Bootstrap for Few (Treated) Clusters. Econometrics Journal 21, 114-135.

Manacorda, M., Manning, A., Wadsworth, J., 2012. The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain. Journal of the European Economic Association 10, 120-151.

Marcus, J., Zambre, V., 2019. The Effect of Increasing Education Efficiency on University Enrollment: Evidence from Administrative Data and an Unusual Schooling Reform in Germany. Journal of Human Resources 54, 468-502.

Mohrenweiser, J., Wydra-Sommaggio, G., Zwick, T., 2015. Work-related Ability As Source of Information Advantages of Training Employers. ZEW Discussion Paper No. 15-057.

Mohrenweiser, J., Zwick, T., 2009. Why Do Firms Train Apprentices? The Net Cost Puzzle Reconsidered. Labour Economics 16, 631-637.

Morin, L.-P., 2015. Cohort Size and Youth Earnings: Evidence from a Quasi-experiment. Labour Economics 32, 99-111.

Moulton, B.R., 1990. An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit. Review of Economics and Statistics 72, 334-338.

Mühlemann, S., Pfann, G.A., Pfeifer, H., Dietrich, H., 2018. The Effects of Supply Shocks in the Market for Apprenticeships: Evidence from a German High School Reform, IZA Discussion Paper 11264.

Ottaviano, G.I.P., Peri, G., 2012. Rethinking the Effect of Immigration on Wages. Journal of the European Economic Association 10, 152-197.

Pischke, J.-S., Velling, J., 1997. Employment Effects of Immigration to Germany: An Analysis Based on Local Labor Markets. Review of Economics and Statistics 79, 594-604.

Roodman, D., MacKinnon, J.G., Webb, M.D., Nielsen, M.A., 2019. Fast And Wild: Bootstrap Inference In Stata Using Boottest. The Stata Journal 19, 4-60.

Song, J., Price, D., Guvenen, F., Bloom, N., von Wachter, T., 2019. Firming Up Inequality. The Quarterly Journal of Economics 134, 1-50.

Soskice, D., 1994. Reconciling Markets and Institutions: The German Apprenticeship System. National Bureau of Economic Research.

Suedekum, J., 2004. Selective Migration, Union Wage Setting and Unemployment Disparities in West Germany. International Economic Journal 18, 33-48.

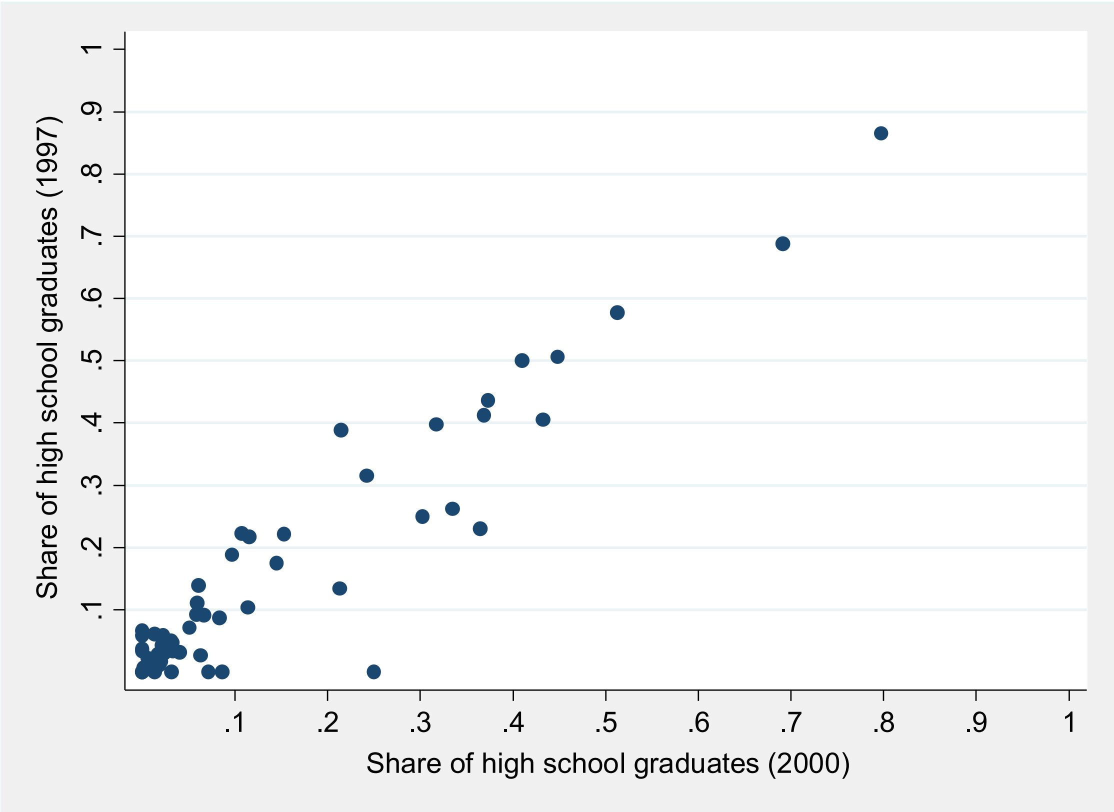
von Wachter, T., Bender, S., 2006. In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers. American Economic Review 96, 1679-1705.

Welch, F., 1979. Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust. Journal of Political Economy 87, 65-97.

Wolter, S.C., Ryan, P., 2011. Apprenticeship, in: Hanushek, E., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education. Elsevier, pp. 521-576.

Appendix

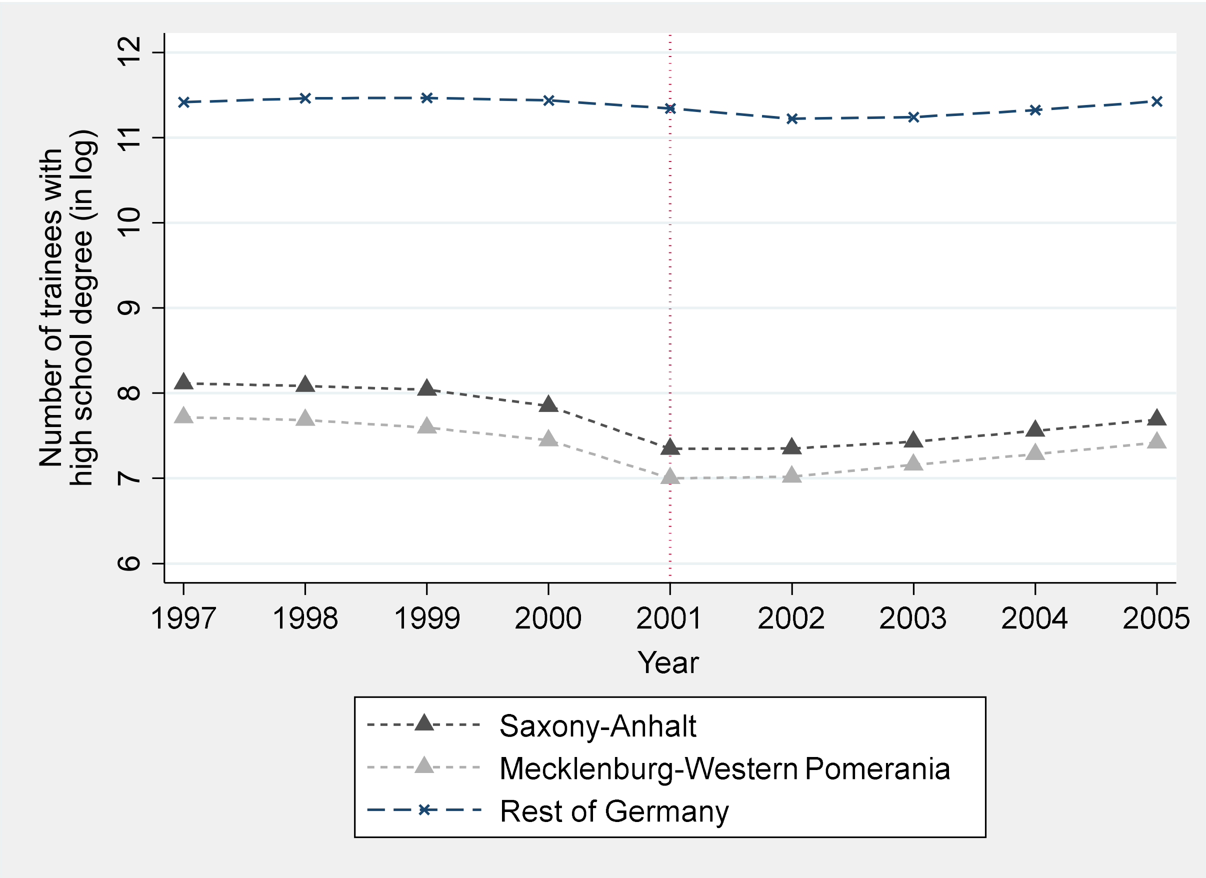
Figure A-1: Share of high school graduates among trainees by occupations



Notes: Each dot presents one out of 70 three-digit training occupations.

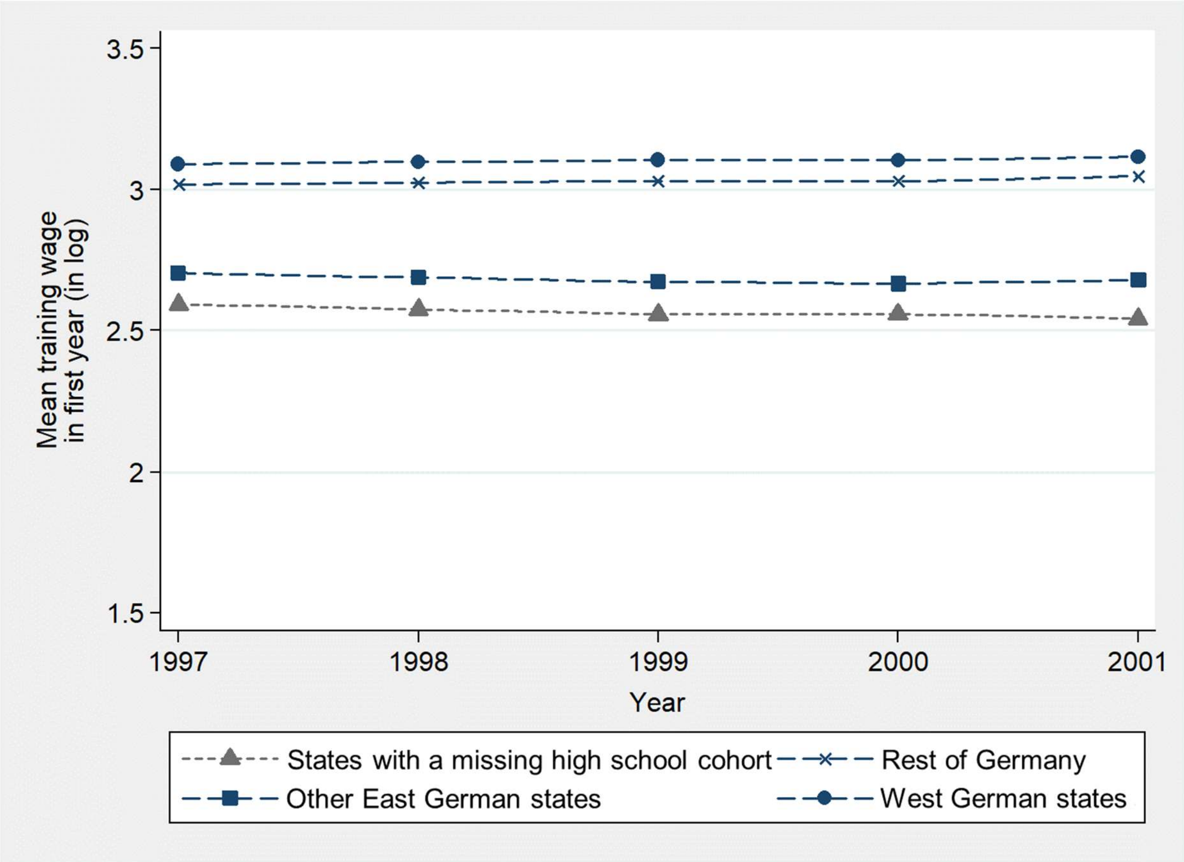
Source: Federal Statistical Office (1997, 2000)

Figure A-2: High school graduates among trainees by state and over time separately for the two treatment states



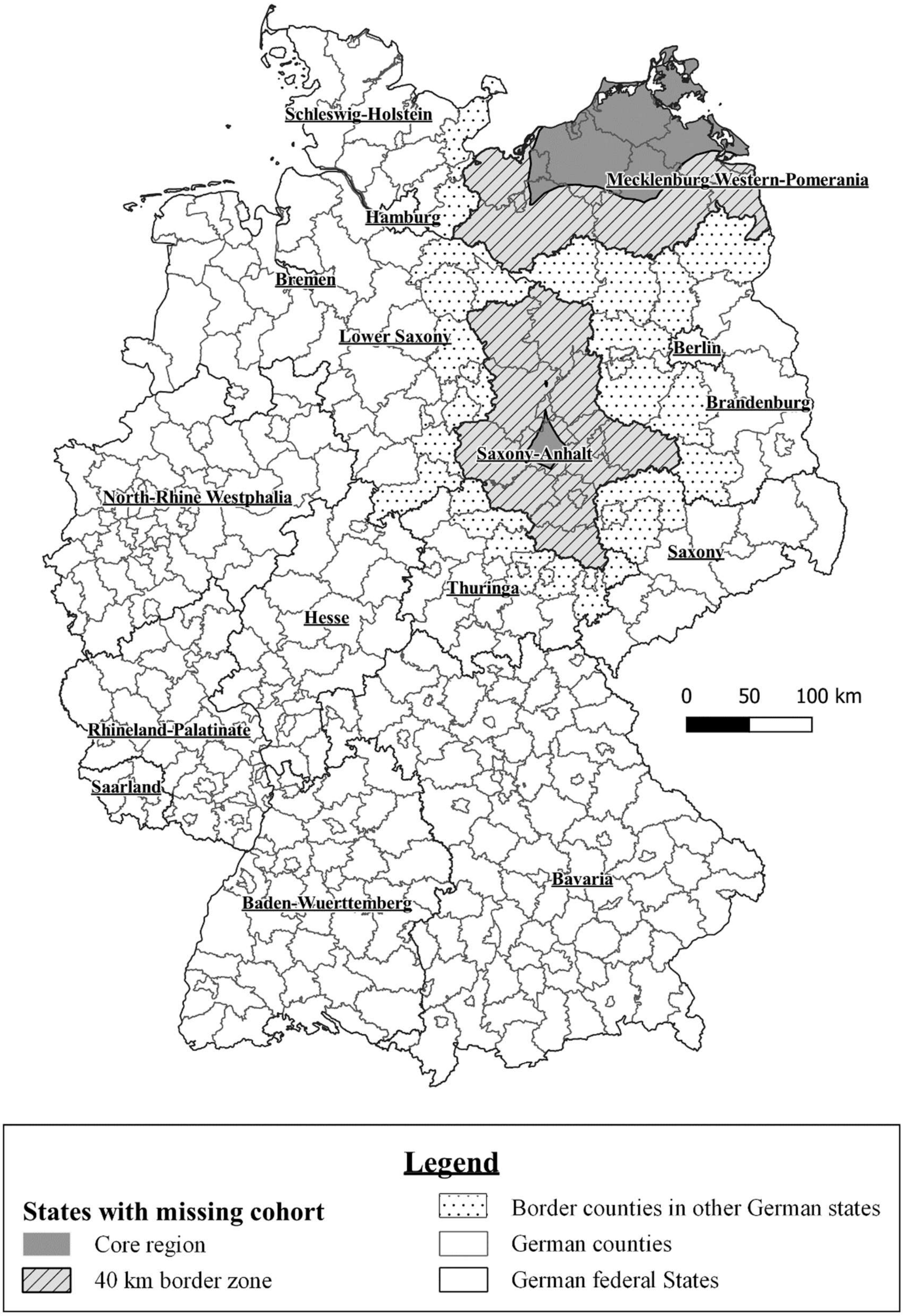
Source: Federal Statistical Office (1997, 2000)

Figure A-3: Log wage at start of training over time



Source: Social security data

Figure A-4: Map of Germany



Source: Own representation (Geometry: GeoBasis-DE / BKG 2015)

Table A-1: Summary statistics of individual and firm level characteristics

	Mean	Std. dev.
Log of training wage (EUR)	3.000	0.392
Trainee characteristics		
Female (0/1)	0.438	0.496
Age (years)	18.634	2.145
German (0/1)	0.95	0.218
High school degree (0/1)	0.133	0.340
Training firm characteristics		
Log of the number of employees	4.174	2.077
Log of the median wage of full time employees (EUR)	4.400	0.352
Share of high skilled workers	0.075	0.115
Share of medium skilled workers	0.693	0.213
Share of low skilled workers	0.219	0.197
NACE industry section manufacturing (0/1)	0.230	0.421

Notes: The number of observations is 2,151,726.

Source: Social security data

Table A-2: Descriptive statistics of training provision and wages

	Log of the number of trainees		Log training wage	
	1997-2000	2001	1997-2000	2001
States with a missing cohort	12.05	10.56	2.573	2.544
			(0.448)	(0.455)
Rest of Germany	14.65	13.26	3.026	3.046
			(0.371)	(0.364)

Notes: The table contains means and the corresponding standard deviation in parentheses.

Source: Social security data

Table A-3: Dual high school graduation cohorts in the German federal states

State	Dual cohort in graduation year
Saxony-Anhalt	2007
Mecklenburg-Western Pomerania	2008
Saarland	2009
Hamburg	2010
Bavaria	2011
Lower Saxony	2011
Baden-Wuerttemberg	2012
Berlin	2012
Brandenburg	2012
Bremen	2012
Hesse	2012/ 2013/ 2014
North Rhine-Westphalia	2013
Schleswig Holstein	2016
Rhineland-Palatinate	-
Saxony	-
Thuringia	-

Source: Information based on Marcus and Zambre (2018).

Table A-4: Alternative calculations of the standard errors of the effect of dual cohorts on training wages

	Log training wage	
	(1)	(2)
Panel A		
Firm fixed effects results (shown in Table 8)	-0.006 **	-0.006 **
cluster at the state level (G=16)	(0.002)	(0.002)
p -value, wild cluster bootstrap (restricted)	[0.019]	[0.018]
p -value, wild cluster bootstrap (unrestricted)	[0.002]	[0.001]
Panel B		
Baseline specification, unclustered std. errors	-0.006 ***	-0.006 ***
	(0.0003)	(0.0003)
Panel C		
Standard errors clustered at the level of	-0.006 ***	-0.006 ***
counties (G=401)	(0.001)	(0.001)
p -value, wild cluster bootstrap (restricted)	[0.000]	[0.000]
p -value, wild cluster bootstrap (unrestricted)	[0.000]	[0.000]
Panel D		
Standard errors clustered at the level of	-0.006 **	-0.006 **
states x years (G=219)	(0.003)	(0.003)
p -value, wild cluster bootstrap (restricted)	[0.090]	[0.091]
p -value, wild cluster bootstrap (unrestricted)	[0.048]	[0.038]
Panel E		
Standard errors clustered at the level of	-0.006 *	-0.006 *
states x pre/reform year (G = 26)	(0.003)	(0.003)
p -value, wild cluster bootstrap (restricted)	[0.094]	[0.082]
p -value, wild cluster bootstrap (unrestricted)	[0.039]	[0.027]
Firm fixed effect	Yes	Yes
Trainee characteristics	No	Yes
Occupation fixed effects	No	Yes
Firm characteristics	No	Yes
Obervations	5,218,430	5,218,430

Note: The standard errors are shown in parentheses. The *p*-values reported in brackets were obtained using regressions with wild cluster bootstrapped standard errors (999 iterations) at the level indicated in the label. Significance level: *** 1%, ** 5%, * 10%.

Source: Social security data