Monetary returns to upper secondary schooling, the evolution of unobserved heterogeneity and implications for employer learning

Anna Krumme
FernUniversität in Hagen
TU Dortmund University

Matthias Westphal
FernUniversität in Hagen,
RWI Essen, and
Leibniz Science Campus Ruhr

February 2024

Abstract

Using the massive opening of academic track schools throughout the German educational expansion (1960-1990), we analyze monetary returns to high school education across the life cycle. For the analysis, we use linked survey and administrative labor market data from Germany (NEPS-ADIAB) combined with a purpose-built data set on all academic track school openings for cohorts born between 1950 and 1985. We exploit local changes in the geographical access to academic track schools to estimate local average treatment effects (LATE) and marginal treatment effects (MTE). We find sizeable monetary returns to the highest secondary schooling degree with average returns of over 70% (14% per year of additional schooling) within the first 10 years of labor market experience. There is unobserved heterogeneity in the returns and clear evidence for selection into gains for higher experience levels. Positive returns and unobserved heterogeneity first arise two years after entering the labor market and increase with growing experience. We interpret increasing heterogeneity in the returns in light of employer learning – so far unconsidered in the literature.

Keywords: Returns to education, IV estimation, marginal treatment effect, unobserved heterogeneity

---

Anna Krumme: TU Dortmund University, Department of Business and Economics, 44221 Dortmund, Germany, E-mail: anna.krumme@tu-dortmund.de.

Matthias Westphal: FernUniversität in Hagen, Faculty of Business Administration and Economics, 48084 Hagen, Germany, E-mail: matthias.westphal@fernuni-hagen.de.

Funding by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged (Project number WE 6751/2-1). We further thank Dennis Schlamann and Moritz Pickhardt for their excellent research assistance and participants of the University of Hagen Workshop on Labor Economics, the Westphalian Research Workshop, and the LIfBi Workshop on Economics of Education in Bamberg for their valuable feedback.

1 Introduction

Evidence on monetary returns to schooling has blossomed over the past two decades, crystallizing into the fact that ”human capital explains a substantial share of the variation in labor earnings within and across countries” (Deming, 2022). Beyond the fact that education is a worthwhile investment on average, many studies have shown that returns are individual-specific and may correlate with individual preferences and abilities for education (Card, 2001; Carneiro et al., 2011; Nybom, 2017; Westphal et al., 2022). However, less is known about when this heterogeneity is shaped throughout the working career: Is it that individual differences in the returns to education are determined already at the beginning of the career, with little change in the ranks and the variance of the returns thereafter – or does this heterogeneity needs time to evolve? The answer to these questions may reveal important insights into how a labor market functions. First, it may be that human capital complements labor market experience, such that education has increasing effects on marginal productivity. Second, it could be that human capital at labor market entry is unobserved, and employers may learn about the true earnings potential of their employees over their careers, such that differential wage increases reflect this learning process (Altonji and Pierret, 2001; Lange, 2007; Aryal et al., 2022).

In this paper, we approach these questions step by step by estimating monetary returns to academic track education, a German elite school type, which 32 percent of a cohort attended during our study. The academic track lasts 13 instead of 10 years, qualifying graduates for university studies or more abstract white-collar professions. To eliminate the confounding effects of selection into this school type, we exploit the massive opening of these schools throughout the German educational expansion with an instrumental variables approach. After estimating average returns, we go beyond the average returns in two important dimensions. First, we follow individuals over time, relative to their labor market entry, to assess when the returns are formed. Second, we unveil heterogeneities along the distaste for education by estimating marginal treatment effects (MTEs). By combining these two dimensions, we are the first to assess how the marginal returns to schooling change after individuals enter the labor market.

We combine administrative data with survey data and can follow individuals from birth to their career prime in the labor market. While the survey data from the National Educational Panel Study (NEPS) allows us to analyze the residential history (such as the place of birth) and a rich set of background characteristics, the administrative labor market data from the IAB provides us with detailed yearly information on daily wage and employment from the individuals’ first appearance in the labor market onward. To this data, we merge self-collected information on all academic track schools in Germany. Overall, we have a data set on 3,471 male individuals.

Our paper contributes to several strands of the literature. First, we add evidence to the literature on monetary returns to secondary schooling. While many papers exist on the returns for compulsory schooling reforms (e.g., Bhuller et al., 2017), evidence on secondary schooling beyond the compulsory level is scarce because many secondary education systems are comprehensive, i.e., do not allow for general education choices until tertiary education starts. Exceptions include Clark and Del Bono (2016), who study labor market effects for elite schools, and Birkelund and Werfhorst (2022) and Matthewes and Ventura (2022), who study tracking decisions at the age of 16. We study long-run returns to tracking decisions at the age of 10 in Germany, thereby complementing Dustmann et al. (2017) who study the returns of initially and coincidentally attending a higher track. Second, we test for selection into gain, i.e., whether the unobserved propensity to attend a higher track correlates with future earnings returns, thereby adding to the evidence for college education (Kamhöfer et al., 2019; Carneiro et al., 2011; Nybom, 2017). Third, we investigate the development of this heterogeneity over time, which, to our knowledge, remains unexplored. As the evolution of this heterogeneity may be shaped by employers who learn about the true earnings potential of their employees over time, we are (besides Aryal et al., 2022) the second to rely on instrumental variable approaches in assessing employer learning and the first to propose using marginal treatment effects for this endeavor.

We find large positive monetary returns to academic tracks school education within the first 10 years in the labor market. Yet, these effects do not exist directly after labor market entry but rise after the first two years of working experience. Our results further suggest that the returns are heterogeneous, especially for higher experience levels, revealing clear selection into gains. High secondary schooling is not beneficial for everyone. It can actually go along with negative returns at later stages of the working careers for individuals with the lowest desire for academic track schooling. A further expansion of academic track education most likely affects only individuals with lower returns and, therefore, would probably not pay off. Our results additionally indicate that the heterogeneity with respect to characteristics unobservable yet valuable to employers increases with growing experience. We conclude that a combination of two productivity-related mechanisms drives this increase in heterogeneity. First, employers might learn about innate abilities as one component of productivity over time and adjust wages according to their beliefs on true productivity. Second, individuals with a higher resistance to high-track education might have lower productivity growth over experience than those who are more prone to choose the higher educational path.

This paper proceeds as follows. Section 2 presents the data set, the institutional background, and the baseline empirical strategy. We present the baseline results in Section 3. Section 4 introduces the marginal treatment effect estimation and the interpretation of potential heterogeneity in this context, before we present and discuss the results. Section 5 concludes.

2 Institutional Background, Data, and Baseline Identification Strategy

2.1 Institutional Setup

The (West-)German Secondary Schooling System

After four years of elementary school and at age 10, the elementary school teacher recommends children to one of three different secondary school types based on their (perceived) performance. Although elementary school teachers give recommendations for the track choice in all states, the parents are the final decision-makers in most of the West German states .[1] We refer to these tracks as the basic, intermediate, and academic.[2] The basic and intermediate track education aims at preparing students for apprenticeships in blue- or white-collar jobs and ends typically after 5 or 6 years, respectively.[3] The academic track education at a *Gymnasium* prepares students for tertiary education at colleges and universities and lasts nine years for all West German students (before the graduating class of 2007). After completing 13 years of schooling and passing final exams, the graduates achieve the *Abitur*, i.e., the highest schooling degree and university-entrance diploma.

The different school types also show content-related differences. Students in the academic track have more weekly teaching hours, and priorities are set on second or third foreign languages and natural sciences rather than social sciences, sports, or vocational preparation (Dustmann et al., 2017). In the core subjects of mathematics and German, the contents are typically more advanced the higher the track (for more detail on the differences between tracks, see Dustmann et al., 2017).

Academic Track School Openings during the Educational Expansion

In the 1950s, educational opportunities were scarce. For example, in 1952, only 12.4% of all grade 8 students attended the academic track, and not more than 3.8% of all school leavers were academic track graduates (Köhler and Lundgreen, 2014). These low numbers are caused by the supply rather than the demand side. In 1950, only 1,823 academic track

[1] Before 2010, in 2 of the 10 West German states (excluding Berlin), the recommendation of the elementary school was widely binding; deviations require an official procedure but are generally possible. For more information, see `https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_02_19-Uebergang_Grundschule-SI-Orientierungsstufe.pdf`.

[2] Since 1971, comprehensive schools have been founded that accommodate all students. However, they have played a minor role as only a small percentage of students attended this comprehensive school type. Until 1990, the share of students at a comprehensive school out of all students at general schools never exceeded the 10% limit, and less than 3 % of all graduates with Abitur received their degree at a comprehensive school (Köhler and Lundgreen, 2014).

[3] Compulsory schooling reforms between 1956 and 1969 increased the basic track duration from five to six years.

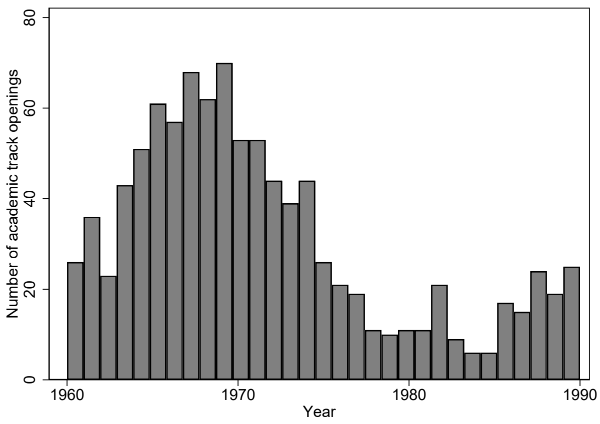
schools (Franzmann, 2006) and 33 universities (Kamhöfer and Westphal, 2019) existed in the entire territory of the Federal Republic of Germany. However, public and political opinion changed, and geographical or economic barriers should no longer restrict access to education (Becker, 2006). In the next decades, Germany's educational infrastructure changed substantially. We refer to this as the educational expansion, which we use for identification in this paper.

Economic and sociopolitical arguments drove the educational expansion in the early 60s. Picht (1964) proclaimed an education crisis that drastically impacted the country's economic situation and demanded public investments in higher education to ensure ongoing economic growth. At the same time, attention shifted more and more to equal educational opportunities, e.g., Dahrendorf (1965) requested "education as a civil right." Therefore, one of the main goals of educational reforms since the 1960s was to facilitate access to education. While the educational expansion also affected tertiary education,[4] the focus was on expanding higher education, as, for example, Picht (1964) called for doubling the number of graduates with Abitur.

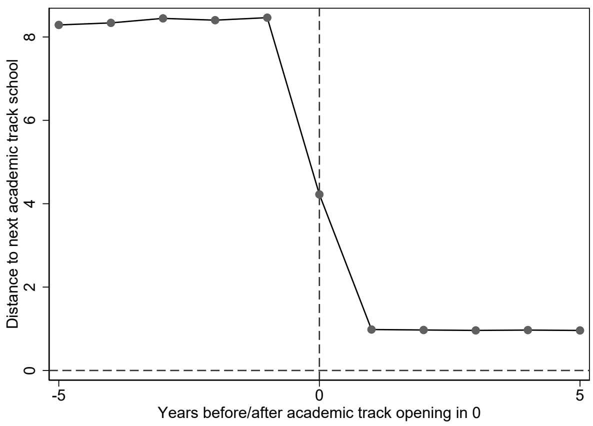
Between 1960 (the earliest secondary schooling entry cohort in our sample) and 1990, yearly public expenditures for academic track schools increased from 1,130 to 11,559 million Deutschmark. This increase led to the foundation of 796 academic track schools (relative to 1,396 existing schools), and the number of students nearly doubled from around 850 thousand to 1.6 million within this period (Franzmann, 2006). Figure 1 visualizes the distribution of academic track openings over time (Panel a) and space (c) and presents the effect of the openings on the distance to the closest school (b). We see at least six openings annually between 1960 and 1990 (our cohorts' main track decision years), with most openings, i.e., more than 40 per year, concentrated in the decade following the mid-sixties. This benefits our analysis because our panel is balanced for these cohorts. Panel (c) shows the spatial distribution of academic track schools. The black municipalities already had an academic track school in 1950. The red municipalities experienced the first opening of an academic track school between 1950 and 1990. This demonstrates that the political goal of the educational expansion seems to be fulfilled: more remote areas disproportionately benefit from improved educational opportunities. To visualize that this improvement also affected neighboring municipalities, we plot a 10km radius around the new academic track in light red. Finally, we quantify the increased accessibility of academic track schools in Panel (b), where we plot the average distance to the next academic track school relative to the opening year for municipalities with an opening. It documents that students had to travel about 8 km on average to the closest academic track school before one opened in their municipality (measured from municipality centroid to school location).

[4] Between 1950 and 1990, the number of colleges doubled in West Germany (see, for example, Kamhöfer et al., 2019 for information on college openings in Germany).

Figure 1: Distribution of academic track school openings and effects on the average distance



(a) Over time



(b) Relative to a school opening

(c) Across West Germany

Notes: Own illustration based on self-collected school information. Panel (b) is restricted to municipalities with a first opening. In opening year 0, typically half of the students (based on school year start and birth months) already have access to the newly opened school.

2.2 Data

Our main data set is the NEPS-SC6-ADIAB (Bachbauer et al., 2022), which links administrative records from the Institute for Employment Research (IAB) to survey data on educational paths from the Leibniz Institute for Educational Trajectories (LIfBi). This data covers adults born between 1944 and 1989 and living in Germany.

The NEPS-SC6 data includes 17,140 individuals whose educational trajectories we know and their approximate residence history at a very local level. The latter is the key information for our project since we need information on the residence before the individuals chose a secondary school track to correctly assign the data on geographical access to academic track schools. We use the municipality of residence in the last year of primary school whenever available and fill this information with the municipality of births else (around one-half of the cases). We include a dummy for the origin of the residence information in all regressions to control for systematic differences.

Most of the NEPS can be linked to the high-quality administrative labor market data provided by the IAB covering 1975 to 2019. 74% of the surveyed individuals consented to link their responses to and were identified in the longitudinal labor market data of the IAB. These individuals appear in the administrative data if they ever had insurable employment subject to social security or marginal part-time employment (since 1990), have registered as job seekers, or have participated in a labor market policy measure of the Federal Labor Office. Therefore, nearly complete labor market trajectories with employment information are included for most people. But generally, self-employed and civil servants are not included in the administrative data.[5]

The second data set we use is a purpose-built collection on academic track school openings. Starting with a list of all existing schools in 2010, we manually selected their opening years. With additional information on the addresses, we end up with the precise geolocation of all schools. We then built a panel with information on geographical access to an academic track school for individuals from a particular municipality and birth cohort. Since there were differences in the starting times of school years, we took the varying school entry cutoffs into account (for details, see Koebe and Marcus, 2022). Additionally, we distinguished between boys- and girls-only schools. Any distance measures rely on the geodetic distance from the center point of the municipality to a school because we have no access to residential information that is more local than the municipality. As long as the distribution of the residences around the centers is random, this does not lead to any problems, except that the access measures might be a little less precise. However, the level of instrument assignment is more local than in most previous analyses using geographical access to school education. The IAB and the LIfBi research data centers merged our dataset to the NEPS-SC6-ADIAB based on childhood residence municipality, gender, birth year and month.

We restrict our sample to males born between 1950 and 1985 in Germany and follow them in the labor market from 1975 to 2019. Hence, we have a balanced panel for individuals aged 25 to 34, i.e., the decisive wage-formation years after labor-market entry. We exclude females to rule out fertility decisions and avoid academic-track-induced selection into employment and hours worked, which may be pronounced for females in the studied cohorts (see Westphal et al., 2022 for a discussion about university education). To ensure comparability in the educational system, which was different in East Germany before the reunification, we restrict the data set further to individuals with West German munici-

[5]More detailed information on the linked data products are provided by Bachbauer and Wolf (2022).

palities of residence in childhood, excluding Berlin. We additionally exclude males with missing information on schooling degrees, childhood residence (municipality and district level), or geographical access to academic track schools.

The final yearly panel includes 3,471 males. We use this sample for our analysis but focus on labor market experience from 0 to 10 years. We use actual work experience taken from the labor market trajectories of the IAB data, i.e., the number of complete years worked in full- and part-time at the end of the year. Thus, an experience of zero indicates the year the individual entered the labor market for the first time. Excluded are any employment episodes before entering the labor market regularly, such as vocational training, student side job, or internship episodes. Since the experience only increases by one if one additional full year of employment is completed by the end of the observation year, individuals might be observed more than once at some experience levels. This is a result of any unemployment episode. Furthermore, not all individuals are observed when completing ten years of employment. Hence, our panel of 36,855 person-year observations is unbalanced across the experience.

Our treatment variable, "Abitur", indicates a university-entrance diploma (i.e., academic track graduation). Importantly, Abitur excludes the vocational baccalaureate diploma (*Fachabitur*), which restricts one to studies of certain subjects at universities of applied sciences (*Fachhochschulen*). With 1,125 of 3,471 individuals in the sample, around one-third have Abitur.

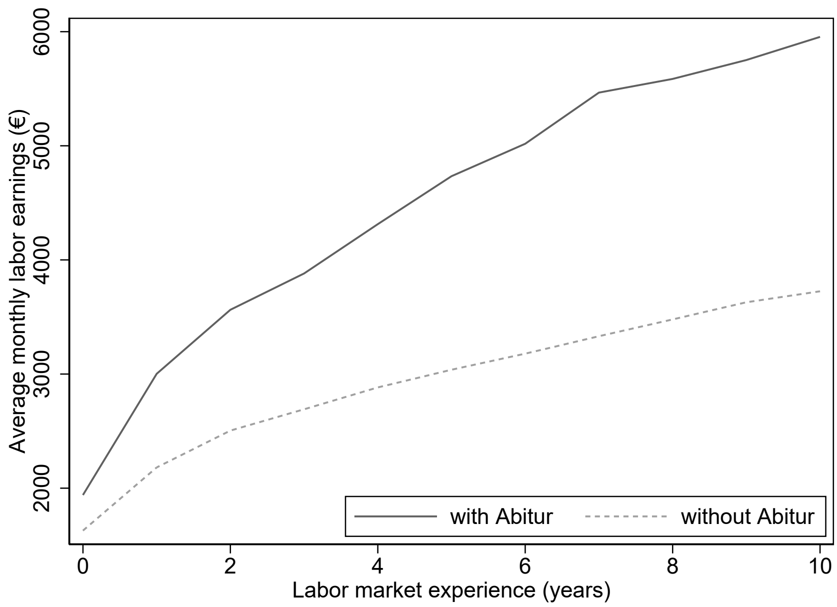
Dependent Variable

The most crucial information to our analysis is the gross daily wages for each job episode the employer obligatorily reports. We transform this information first into monthly labor earnings and then collapse it into a yearly panel using months with employment only. We further restrict our analysis to years with full-time employment.[6] We cannot use hourly wages because we do not have administrative information on hours worked. Nonetheless, because males' hours worked are inelastic in Germany (as they predominantly work full-time), we expect this measure to proxy hourly wages well. In preparing the earnings variable, we follow Dauth and Eppelsheimer (2020) and deflate daily wages with the consumer price index with the base year 2015. We also use the suggested procedure to identify and impute wages above an assessment ceiling. This is necessary because wages are only reported up to an upper limit for the statutory pension scheme, which varies with location and time. This step is particularly relevant for our paper as we want to analyze the whole range of heterogeneity in monetary returns.

[6]Full-time work in the IAB data depends on the ratio of own to the establishment-specific common working hours. Missing values in this variable are replaced with information from the NEPS survey data such that an individual is assumed to work full-time if the self-reported contractual working hours are greater or equal to 39.

Figure 2 reports the average monthly labor earnings per year (deflated to 2015 prices) for individuals with and without Abitur for experience levels from 0 to 10 years. After entering the labor market, the average monthly full-time earnings for individuals with Abitur amounts to 1,940.02€ . These earnings increase substantially in the first two years, with an annual growth rate of 55 and 19 percent, respectively. Afterward, these earnings still grow steadily (reaching 5,954.24€ after ten years), but the growth rate declines from 10 to about 4 percent. Albeit on a lower level, earnings without Abitur also increase markedly across experience (from 1,628.25€ to 3,725.19€). However, not only the absolute but also the relative gap increases. While the relative Abitur earnings differential is 19 percent without experience, the gap constantly widens to 64 percent with seven years of experience. Thereafter, the relative difference seems to remain at this level. In sum, this figure reveals the importance of the early labor market years for wage formation and the Abitur earnings premium, which we will explore in greater detail in the succeeding analyses.

Figure 2: Descriptive earnings trajectories over experience by academic track degree



Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure shows means for individuals with and without the academic track degree (Abitur) over labor market experience.

2.3 Baseline Empirical Strategy

Because individuals and their parents may choose the secondary school track, academic track education is endogenous, and a simple regression of earnings on Abitur (plus further observable characteristics) will not yield the causal effect of Abitur. We try to fix this problem using local changes in the academic track choice set as a quasi-experimental variation. Specifically, we use openings of academic track schools, measured by the variable Z_{it} (specified below), which we assume to be exogenous to individual decisions.

With such a variable, we can recover the causal effect through the following two-stage least squares (2SLS) regression model

$$\begin{aligned} D_i &= \pi_0 + \pi_1 Z_i + X_i'\gamma + \nu_i \\ Y_{it} &= \alpha_t + \beta_t \widehat{D}_i + X_i'\delta_t + \varepsilon_{it}. \end{aligned} \tag{1}$$

Because the outcome variable changes for different values of experience t, we estimate these two equations separately for each experience level. In the first regression (the first stage), we regress the Abitur indicator of individual i (D) on the instrument Z and a vector of control variables X. Under the assumptions provided below, π_1 captures the share of individuals who only attend an academic track school because of changes in Z (i.e., the opening of an academic track school in the surrounding), but not without. This is the group of compliers. For this group, the second stage estimates causal effects on monthly earnings Y by labor market experience t. To this end, we regress Y_{it} on the fitted values from the first stage $\widehat{D}_i$ plus the same set of controls. Our interest lies in the coefficient β_t, the absolute effect of Abitur on labor earnings with experience t. To test its statistical significance, we cluster our standard errors at the municipality level in all our regressions.

We need to assume that, conditional on the controls X, the instrument is unrelated to unobserved factors that correlate with education (i.e., $cov(\nu, Z) = 0$) and wages $cov(\varepsilon, Z) = 0$). To make this assumption credible, we include a large set of fixed effects in X, which we specify below. We additionally assume that other municipality-level changes do not coincide with the timing of the academic track openings (exclusion restriction). This assumption seems plausible given a rich set of control variables that correctly absorbs general trends. However, it may well be that due to general equilibrium effects, the quality of newly opened schools and teachers differ. We attribute these to the potential mechanisms behind our results. Finally, we assume that the openings did not deter individuals from attending the academic track (monotonicity). While we cannot test this on the individual level, we gauge this as unlikely, given that the overall accessibility of academic track education clearly improves through the openings.

Control variables

The most important control variables are school cohort and district fixed effects to absorb differential trends and levels in wages, education, and academic track availability between districts. As the federal states may have implemented the educational expansion differently, we also account for federal-state-specific linear trends. Moreover, we include birth month fixed effects. Finally, we use a dummy variable indicating the source of the residence information (i.e., place of primary school or place of birth)[7]. Table A.1 in the Appendix

[7] All location-specific variables refer to the same location as the instrument.

provides information on all covariates of the main specification and variables used in any additional regression.

To make the independence assumption most credible in our setting, besides school entry cohort dummies, our ideal specification would include municipality-fixed effects to control for any time-constant differences that might affect school openings at the local level. Unfortunately, it is not feasible with our sample, which has fewer observations than municipalities. Hence, we are restricted to district-fixed effects, leading to the identifying assumption that the location and timing of academic track school openings are exogenous within districts. Since the educational expansion aimed at improving the comprehensive supply of (higher) schools to reduce social inequality, academic track schools should not only be opened preferably in municipalities with certain characteristics of inhabitants that also affect the later earnings of the children. The greatest expectable differences between municipalities are most likely driven by bigger cities that might be systematically distinct from smaller and more rural municipalities [8]. We mainly control for these differences by including district fixed effects, as urban districts are not subdivided into municipalities in Germany. This is also the case for the three German city-states. Furthermore, according to Henz and Maas (1995), already for the birth cohorts 1959-61, there were no urban-rural differences in the probability of attending a Gymnasium anymore. This might justify the assumption that municipality-specific inhabitant characteristics also affecting later wages did not promote academic track school openings in certain municipalities within districts. Therefore, in our specification, the location and timing of the openings might be random. However, the differences between municipalities that already had academic track schools before, those where academic track schools were opened, and those that never established any are problematic. We expect some municipalities with existing academic track schools to be large cities. Making use of the same argument as before, district-fixed effects thus partially control for the problem.

Instruments

We use two different measures as the instrument Z_i. The most intuitive measure for academic track school availability is a dummy variable indicating at least one corresponding school in the municipality. We use this binary instrument in our main specification for estimating the LATE.

We additionally use a continuous instrument and construct an academic track availability index like Kamhöfer et al. (2019) did for universities. We consider the distances to the three closest academic track schools, compute an inverse distance weight, and aggregate these weights for each municipality and year. Specifically, the index is defined as $Z^I_{m,\tau} =$

[8] For example, universities, labor market conditions, or any infrastructural advantages can attract especially higher educated families with children that might have better probabilities for having high wages in their later lives.

Table 1: Descriptive statistics of instruments and background information

	Statistics			
	Mean	SD	Min	Max
Background information				
Distance to 1st nearest school	3.427	4.339	0.1	35.3
Number of schools in municipality	5.2	10.283	0	65
Employed instruments:				
Academic track in municipality	0.633	0.482	0	1
Index (Z_I)	0.703	0.426	0	1.194

Notes: Own calculations based on NEPS-SC6-ADIAB data.

$\sum_{l=1}^{3} K\left(\frac{x_{m,l,\tau}}{5}\right)$ with $K(\cdot)$ denoting the Gaussian Kernel and x_{ji} being the distance in km from municipality m to its first, second and third ($l = 1, 2, 3$) closest academic track school im year τ. The denominator, or bandwidth, determines how quickly weights decay in the distance. We use a bandwidth of 5 km. For example, schools within a 1km distance enter the equation with 0.39, whereas schools 5 km away are valued lower with 0.24. Schools 10 or more km away enter the equation with only small values of 0.05 or less. Thus, individuals with more schools nearby have a greater index than those living farther away from the next three academic track schools.

Table 1 presents descriptive statistics for the instruments and background information on academic track school availability. The first two lines refer to background information underlying our employed instruments. For instance, the average distance from the school to the municipality centroid is 3.5 km, ranging from zero to 35. Likewise, the number of schools ranges from zero to 65 (in large cities like Hamburg and Munich), with a mean number of about 5. Regarding the employed instruments, nearly two-thirds of our sample have an academic track school in their municipality. The index additionally documents that some individuals in our sample are close to the maximum possible value (1.197), whereas the lowest value is zero (as implied by the distance measure). The mean value of the index is 0.7, which would be implied if the three nearest academic track schools were 5.15 km away.

3 Results I – OLS and Baseline 2SLS

Before exploring effect heterogeneity, we now present the baseline results in Table 2. The first column shows the first stage effect of an academic track school in the municipality on attaining the Abitur (academic track certificate) in our sample of 3,471 males. As expected, the sign of the coefficient is positive. The estimate indicates that having the corresponding school type upon completing elementary school significantly increases the probability of completing the track with the final degree by 8.25 percentage points (relative to an overall mean of 32 percent). An F-statistic exceeding 10 rules out weak instruments based on the Staiger and Stock (1997) rule of thumb and confirms that the relevance condition is satisfied for the full sample. The first stage estimates vary with the experience level since the panel is not balanced over experience. However, figure A.2 in the appendix shows that the F-statistic is above 10 for most experience levels.

Table 2: Regression results for first stage, OLS and IV estimations

	Abitur	Monthly labor earnings [after 10 years of experience]	
	First Stage	OLS	IV
Acad. track school in municipality	0.0825^{***}		
	(0.0240)		
Abitur		1354.52^{***}	2122.67^{***}
		(62.8445)	(809.67)
F-statistic (instrument)	11.84		
Baseline earnings w/o Abitur		2,865.86	2,865.86
Observations	3,471	3,471	3,471

Notes: Own calculations based on NEPS-SC6-ADIAB data. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. Standard errors in parentheses are clustered at the municipality level. Baseline: Average monthly earnings for individuals without Abitur. * ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$)

Next, we report OLS estimates of Abitur on earnings in column 2 of table 2 as a benchmark. Individuals with Abitur have, on average, about 1,355€ higher labor earnings per month after conditioning on covariates. Compared to a baseline of 2,867€ on average for people without Abitur, this results in a relative difference of 47%. But, due to the potential endogeneity, these OLS results are not causally interpretable. For the causal effect for the subgroup of compliers, we report the 2SLS estimates in column 3. Over the total years of experience considered here, i.e., from 0 to 10, having Abitur increases the monthly labor earnings on average by more than 2,000 € . Compared to people without Abitur, the higher degree increases earnings by around 74%. As common in the literature on returns to education, our IV estimates are larger than our OLS estimates. This, however, depends on the instrument and, accordingly, on the subgroup of compliers.

Table A.2 in the Appendix reveals that having Abitur goes along with a higher age when individuals enter the labor market, a higher probability of having a university degree and more completed years of education for the subpopulation of compliers. The huge effect of 74%, therefore, includes the effect of being older and potentially better at bargaining wages and the additional effect of having a university degree. The latter is included per definition because Abitur, as a university entrance diploma, is the precondition for being accepted.

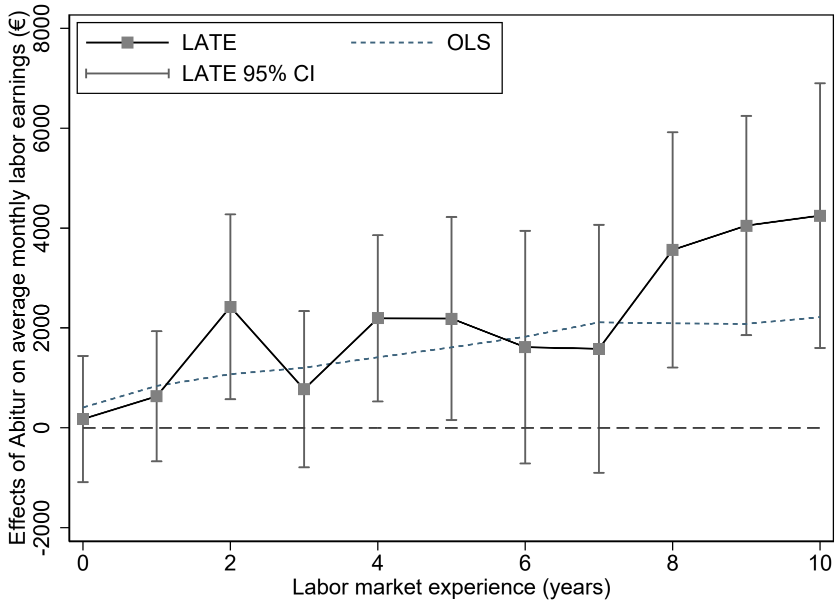
Typically, the Abitur is associated with three additional years of schooling. Surprisingly, it increases the total years of education, including vocational and tertiary education, by only around 4.2 years for our compliers. Thus, the annualized average monetary return to academic track education equals roughly 14% per additional year of education ($1.74^{-4.2}$).

Our general IV results align with previous studies finding positive effects of high secondary schooling. Our LATE results are similar to those from Carneiro et al. (2017) for upper secondary schooling in Indonesia. While the total effect of upper secondary schooling is slightly higher, the annualized returns are slightly lower than the effect we find for Germany. Matthewes and Ventura (2022) find negative effects of attending vocational education compared to academic education in England. These results also point in the same direction as ours but are much smaller in magnitude. For Scotland, Clark and Del Bono (2016) even find no effects of attending an elite school on male wages. One reason for smaller or no effects might be the big difference in the educational system, where there are no or small differences in years of schooling when attending the academic or elite track. However, Dustmann et al. (2017) also do not find any effects of initially attending (not graduating from) a more advanced track for marginal students on wages in Germany, leading to the conclusion that there must be big differences between attending and completing the track with the final degree. At least in Germany, this can be driven by up- and downgrading between tracks, which is the main argument of Dustmann et al. (2017) for their zero effect on wages. However, a priori, our compliers are comparable to marginal students from Dustmann et al. (2017) because our students are at the margin between two tracks without certainty that they will eventually graduate and receive the Abitur.

To go a step further, we also report the evolution of the LATEs on earnings over labor market experience in Figure 3. Again, OLS estimates are included as a benchmark and depicted by the dashed line, whereas the squares represent the experience-specific LATE estimates. As before, the OLS estimates are mostly smaller than the IV estimates. Shortly after entering, Abitur's earnings premium is small and statistically insignificant. A potential explanation is that the untreated group includes individuals with vocational education within a firm. After completing their vocational training, they often officially enter the labor market with more practical experience than those with Abitur. At the beginning of the working careers, the advantage of practical experience might offset the signaling effect of higher education. Generally, the returns to education tend to increase over experience, even though the LATE estimate is not statistically significant for every experience level. Comparing the effects to the average labor market earnings for males without Abitur displayed in figure 2, we get relative effects between 10 and 114%.

The findings on monetary returns over experience are mainly comparable to results on the evolution of returns for different ages. Bhuller et al. (2017) use Norwegian data and

Figure 3: Local average treatment effects of Abitur on earnings over experience



Notes: Own illustration based on NEPS-SC6-ADIAB data. The graph reports regression results for OLS and IV estimations of Abitur on monthly labor earnings over labor market experience. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. The vertical bars denote the 95% confidence interval based on standard errors clustered at the municipality level.

find that additional years of schooling increase earnings as soon as individuals reach the age of 25. This is most likely when higher-educated individuals (for example, people with high-track schooling) enter the labor market. After that point, monetary returns increase sharply, at least until the age of 45. This matches our generally increasing returns to Abitur in the first 10 years of labor market experience.

4 The evolution of unobserved heterogeneity and its interpretation

4.1 Marginal Treatment Effects and its evolution

We now want to explore the unobserved heterogeneity in the Abitur earnings premium and how it evolves for individuals who are indifferent between Abitur and the next best track by estimating marginal treatment effects (MTEs, see Heckman and Vytlacil, 2005). We therefore introduce a potential outcome framework, where we model these potential outcomes as

$$Y_{it}^{j} = \eta_t^{j} + X_i' \mu_t^{j} + U_{it}^{j}, \quad j \in \{0, 1\}, \tag{2}$$

where X_i is a vector of demeaned (time-invariant) controls, such that η^j gives the mean earnings in the treatment state j. Equation 2 implies that the individual treatment effect is given by $Y_{it}^1 - Y_{it}^0 = (\eta_t^1 - \eta_t^0) + X_i'(\mu_t^1 - \mu_t^1) + U_{it}^1 - U_{it}^0$ and consists of observable (X) and unobservable factors (U^1, U^0). Both can potentially drive heterogeneity in the returns.

To generally allow for a correlation between the individual returns with the Abitur propensity underlying D_i, we set up a Roy model, which derives U_D, a value on the unit interval, which defines a hypothetical value, at which the individual is indifferent about Abitur. As forward-looking individuals should base their education decision on lifetime earnings, we use $\widetilde{Y}_i^j = \sum_{t=0}^{T} \frac{Y_{it}^j}{1+r}$ to denote discounted lifetime earnings (with interest rate r) accruing across the labor market career from 0 to experience level T. With these considerations, we can be more explicit about the underlying tracking decision. We start by relating the lifetime Abitur benefit $\widetilde{Y}_i^1 - \widetilde{Y}_i^0$ to potential costs $C(X, Z, U^C)$, which we assume to be additively separable in the instrument Z, the covariates X, and the unobserved costs U^C:

$$
\begin{aligned}
D_i &= \mathbb{1}\left\{ (\widetilde{Y}_i^1 - \widetilde{Y}_i^0) \geq C(X_i, Z_i, U_i^C) \right\} \\
&= \mathbb{1}\left\{ \pi_0 + \pi_1 Z_i + X_i'\gamma \geq V_i \right\} \\
&= \mathbb{1}\left\{ F_V(\pi_0 + \pi_1 Z_i + X_i'\gamma) \geq F_V(V_i) \right\} \\
&= \mathbb{1}\left\{ P(Z_i, X_i) \geq U_i^D \right\}
\end{aligned} \tag{3}
$$

Using the definition of lifetime earnings and the individual treatment effect from equation 2, the second step separates all the observable components on the left- and all the unobserved components ($V := U^C - (\widetilde{U}^1 - \widetilde{U}^0)$) on the right-hand side of the inequality. Note that coefficients equal those from the first stage if both are estimated by a probit model. To clarify this, we apply the cumulative distribution function on V (a rank-preserving monotonic transformation) to both sides. The last step defines the notation of the resulting quantities. First, the propensity score $P(Z, X)$ maps the observables to the unit interval by assigning a probability to receive Abitur based on Z and X. Second, the unobserved value U_D – being the quantiles of V – can be interpreted as the unobserved resistance to Abitur. Thus, the MTE is now defined as $E(Y^1 - Y^0 | U_D = u_D)$, i.e., the causal effect of D at different quantiles u_D of V. For every higher value of the observables induced by Z (holding X constant), more people select an academic track education. Individuals who change D due to a marginal change in Z are indifferent, and $P(Z, X)$ equals U_D. For these additional marginal individuals being at the u_D-th quantile of the distribution ($P(Z_i, X_i) = u_D$), we can evaluate the marginal treatment effect at any observed value of the propensity score or, equivalently, any quantile U_D. This is generally not restricted to stronger assumptions than those required for the LATE estimation, as shown by Vytlacil (2002).

The latent index U_D and its implications for employer learning

Assessing the evolution of MTEs over time can additionally give insights into wage development since U_D typically negatively correlates with earnings returns (see, for example Carneiro et al., 2011, Carneiro et al., 2017 or Kamhöfer et al., 2019). At the same time, although potentially important for wage setting, U_D, is unknown to employers. Generally, the unobserved component U^j_{it} may consist of private information about individual abilities A_i known by the individual, their parents, or their elementary teacher at the time of the tracking decision and other unobserved characteristics that may partially materialize after the tracking decision, such that $U^j_{it} = \theta^j_t A_i + \omega^j_{it}$. Then, the unobserved heterogeneity underlying the MTE may be decomposed as follows:

$$U^1_{it} - U^0_{it} = (\theta^1_t - \theta^0_t)A_i + (\omega^1_{it} - \omega^0_{it})$$

As long as $\theta^1 > \theta^0$ (which is true if initial abilities complement with human capital investments), this model implies negative relationships between (i) A_i and U_D (i.e.,$\frac{\partial E(A_i|U_D=p)}{\partial p} <$ 0) and (ii) and A_i and $E(Y^1 - Y^0|U_D)$. Hence, information on U_D is relevant for predicting the earnings potential. The interesting question is what causes $\theta^1_t - \theta^0_t$ to increase along experience. We can think of two mechanisms:

1. Differential productivity growth: Individuals with high A_i learn faster on the job with academic track education than without.
2. Employer learning: Time in the labor market reveals private information to employers, which is important for wage setting.

We now detail the second mechanism and defer the discussion of the first one to the end. The employer learning theory states that employers learn about time-invariant productivity factors (unobserved innate abilities A_i) over time and adopt wages according to the expected productivity adjusted in every period based on job performance (Ablay and Lange, 2023).[9] Over time, the information asymmetry on productivity decreases, and the signal of education as an easily observable signal becomes less determining for wages. This is only true if the MTE slopes become steeper with more labor market experience. Individuals at lower quantiles of the V_i distribution are most likely those with unobserved characteristics that lower individuals' higher education costs and increase productivity in employment, i.e., high innate abilities A_i. The opposite applies to individuals with high resistance to treatment located at the right margin of U_D. If employers learn about the unobserved component of their employee's productivity with every additional time period and adjust wages accordingly, the returns to education must follow this process

[9]The common employer learning model introduced by Farber and Gibbons (1996) and further developed by Altonji and Pierret (2001) assumes competitive labor markets with asymmetric information on productivity between employers and employees. Additionally, all employers can access the same information (symmetric employer learning).

by being more heterogeneous along U_D with more labor market experience. If one additionally assumes that the productivity effect of education over experience is unrelated to unobserved ability, increasing heterogeneity detected by MTEs can even prove common employer learning under the remaining assumptions of the standard employer learning model.

4.2 MTE estimation

For estimating the MTE, we primarily need the conditional expectation $E[Y_i|X_i = x, P(Z_i, X_i) = p]$. Its derivative with respect to the propensity score results in the MTE:

$$MTE(X_i = x, U_{Di} = p) = \frac{\partial E[Y_{it}|X_i = x, P(Z_i, X_i) = p]}{\partial p} \tag{4}$$

It is the effect of an increase in the propensity score on the outcome. An estimable expression of the conditional expectation $E[Y_i|X_i = x, P(Z_i) = p]$ is given by

$$E[Y_{it}|X_i = x, P(Z_i) = p] = \eta_0 + X_i\mu_0 + (\eta_1 - \eta_0)p + X_i(\mu_1 - \mu_0)p + K(p) \tag{5}$$

as shown by Heckman and Vytlacil (2005) $K(p)$ is a function of the propensity score and is generally not further specified. Hence, we must first estimate the propensity score to estimate equation 5.[10]

In our main specification, we estimate a linear MTE, i.e. $K(p) = \lambda p^2$. This is rather inflexible but eases the comparison of heterogeneity between different MTEs, since we have only one coefficient on the slope. Furthermore, the extent of heterogeneity is the same as long as the maxima and minima of true MTE are never higher or lower than the linear MTE at the outer margins. However, we check whether MTEs based on more flexible specifications deviate from the linear MTE of our main specification. For estimation, we only use observations within common support, i.e. the overlap of $P(Z)$ for $D = 1$ and $D = 0$. We report standard errors clustered at the municipality level based on bootstrapping using 200 repetitions.

[10]To forego further assumptions, we follow the approach of Kamhöfer et al. (2019) and estimate MTEs that only vary over the unobservables. Hence, we restrict μ_1 to equal μ_0 in equation 5 , such that

$$E[Y_{it}|X_i = x, P(Z_i, X_i) = p] = X_i\mu_0 + (\eta_1 - \eta_0)p + K(p). \tag{6}$$

Thus, we restrict the covariates to have the same effect in both potential outcome equations. This might seem a harsh restriction but it is equivalent to common IV approaches (such as LATE) where the treatment indicator is typically not interacted with other covariates. Even if the true effects vary over X, the restriction only affects the intercept of the MTE as the derivative of equation 5 concerning the p is constant in X. Since we are mainly interested in the heterogeneity driven by differences in unobservables for the whole population, any of the restrictions' shortcomings are irrelevant.

In the case of homogenous effects, the MTE does not vary with different quantiles along U_D. In that case, the treatment effect is equal for all subpopulations. However, regarding monetary returns, it is rather unlikely to be the case. If individuals self-select into higher education based on expected gains, the MTE has a negative slope. The higher the propensity of taking the treatment, the lower the expected gains need to be for indifferent individuals to take the treatment, and vice versa. Therefore, the MTE slope might reveal a pattern called selection into gains.

4.3 Results II: Marginal Treatment Effects

Our analysis proceeds in two steps. We first present evidence on selection into gains (an effect gradient along U_D with larger effects for the lowest values) in a pooled sample for all experience values. This demonstrates that U_D reveals important information about the earnings potential of employees. Subsequently, we explore when this gradient evolves, which we finally interpret in light of the employer learning literature.

Detecting a pattern in unobserved heterogeneity: Selection into gains
Table 3 includes first stage results and MTE results for the entire sample, i.e. average effects for experience levels from 0 to 10 years. Column 1 displays the average marginal effects of the instrument on the probability of having Abitur from a probit estimation of the selection equation in equation 3. As expected, the effect is significantly positive for the continuous instrument since a higher index indicates that academic track schools are relatively close. The table also shows the χ^2 for a test of the significance of the excluded instrument, which rejects that the index on geographical access does not determine the probability of having Abitur. Nevertheless, one could worry about weak instruments, especially because the test statistics vary with experience, so some experience levels have even lower values. Therefore, we also use the binary instrument (also used for estimating the LATE) to estimate the linear MTE as a robustness check. The corresponding χ^2-statistics of the first stage selection models over experience are shown in the Appendix in figure A.3. It reveals higher values for the binary instrument, indicating a lower risk of a weak instrument.

The propensity scores of the first stage estimation provide a large common support. Figure A.4 in the appendix shows the estimated propensity score density for individuals with and without Abitur, and the overlap reveals that common support ranges from 0.04 to 0.86. The variation in the propensity scores is generated by the instrument and all covariates. We estimate the MTE as described above for observations within the common support.

Estimating the outcome equation in 5 with $K(p) = \lambda p^2$ and taking the derivative with respect to the propensity score provides the parameters for the linear MTE. Column 2 in

Table 3: Regression results for the selection equation and a linear MTE

	First Stage	**MTE**
	Abitur	Monthly labor earnings [in 10 years experience]
Index geographical access	0.1196***	
	(0.0396)	
Intercept		2755.83
		(1917.05)
Slope		-2834.66
		(2871.86)
χ^2-statistic (instrument)	9.05	
Observations	3,261	3,261

Notes: Own calculations based on NEPS-SC6-ADIAB data. Column 1 reports average marginal effects from a probit selection model. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. Standard errors in parentheses are clustered at the municipality level and in column 2, bootstrapped with 200 repetitions. * $(p < 0.1)$, ** $(p < 0.05)$, *** $(p < 0.01)$

table 3 shows the relevant parameters of the linear MTE. $(\hat{\eta}_1 - \hat{\eta}_0)$ is the intercept, while $2\hat{\lambda}$ specifies the slope parameter of the linear MTE. Even though the slope is negative, the effect is not significantly different from zero. Thus, for all experience levels jointly, we find no clear evidence for heterogeneity in unobservables for the returns to Abitur. The negative slope, however, points towards selection into gains: Individuals with the highest returns to Abitur are the first to attend the newly opened academic track schools. This suggests that individuals (or their parents and teachers) choose their educational path based on expected gains.

The marginal effects along U_D in figure A.5 in the Appendix show the results of the linear MTE graphically. For individuals with low quantiles of U_D, i.e., with low resistance to treatment, we find the highest monetary returns to Abitur. For people with low resistance to academic track education, already low values of $P(Z)$ exceed the U_D, making them very likely to choose academic track education. 90% confidence intervals given by the dashed lines indicate positive yet insignificant returns up to a U_D close to one. For the highest U_D, in other words, for individuals with the highest resistance, the returns are close to zero.

Our linear MTE is similar to the MTE curves estimated by Carneiro et al. (2017) for Indonesia. The main difference is that we do not find relevantly negative returns for any part of the U_D distribution. Relating to the following results on heterogeneity along working experience, this difference might be driven by observations at the beginning of the working career. In this overall sample with experiences ranging from 0 to 10 years, the lower experience levels might be overrepresented. MTE results on the returns to college education in Germany from Kamhöfer and Westphal (2019) also align with our finding of a negative slope, suggesting selection into gains for monetary returns to education in general.

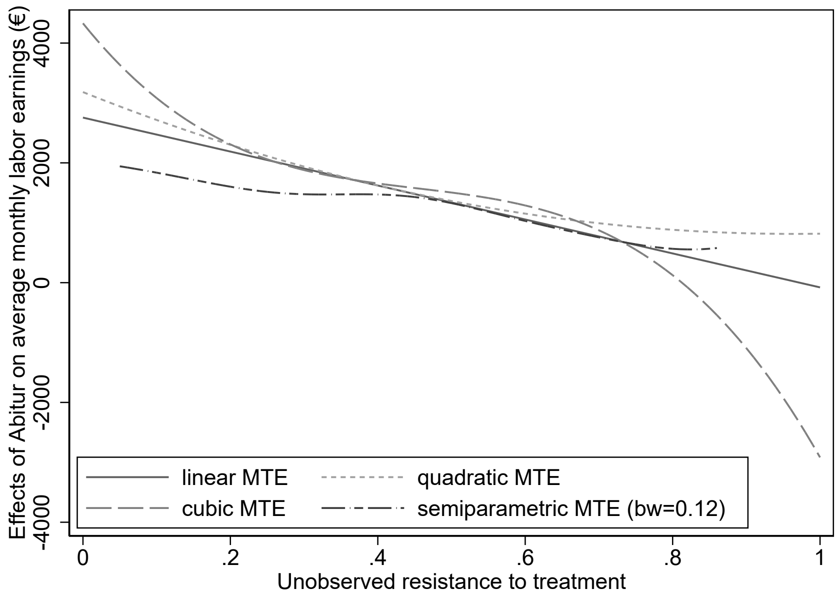
We carry out two types of robustness checks. First, we use more flexible approaches for estimating $K(p)$ with the same instrument. Figure 4 shows the MTEs of parametric approaches with higher order polynomials and a semiparametric estimation method suggested by Robinson (1988). The MTE curves are all similar to the linear one for most parts of the U_D distribution. This is also true for the semiparametric estimation, which allows the most flexible pattern. Thus, the assumption of a linear MTE does not seem to be a restriction that is too strong in our setting, especially when the extent of heterogeneity is of interest. Second, we use a binary instrument to estimate the linear MTE. As argued before (and shown in figure A.3 for different experience levels), the binary instrument generates a stronger first stage and might serve better as an instrument. We still use the continuous instrument for our main results because it does not allow estimating a non-linear MTE for comparison without further assumptions. Generally, estimating an MTE with a binary instrument requires the additional assumption of additive separability as shown by Brinch et al. (2017). The comparison in the Appendix in figure A.6 shows that both instruments lead to very similar linear MTEs. We omit the discussion of the benefits and drawbacks of the two approaches because of this comparability in results. Overall, these checks suggest that assuming linearity for the MTE is appropriate, and the results are robust to the use of a different instrument.

The evolution of selection into gains

Once we have shown that there might be heterogeneous returns for individuals with different unobserved resistance to treatment, we look at the evolution of the unobserved heterogeneity over working experience. As previously argued, estimating linear MTEs seems to be suitable for detecting the degree of heterogeneity with respect to the unobserved resistance to treatment. Moreover, it provides a single slope parameter per experience level that can be easily compared to others. Figure 5 shows the slope parameters of the linear MTEs per year of working experience.

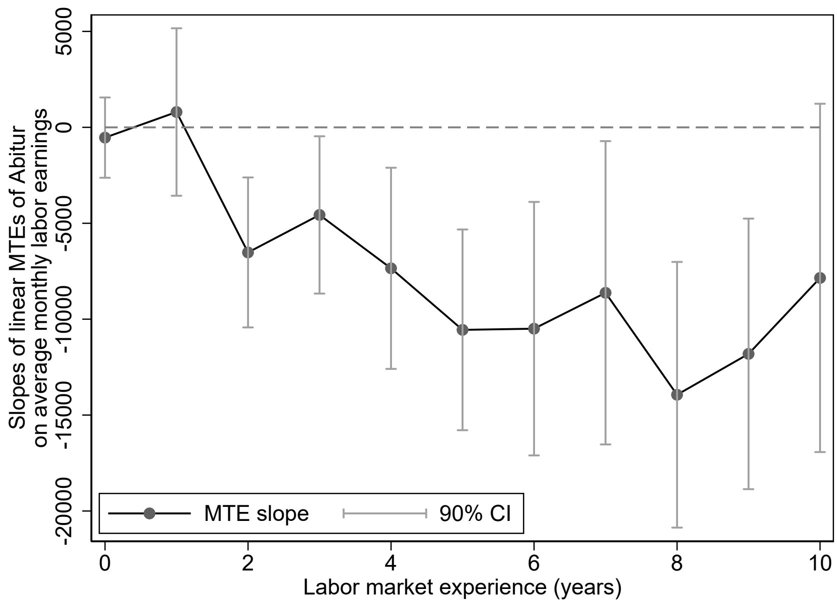
In the year of entering the labor market and the following year, the slope is close to zero. The 90% confidence intervals show that the heterogeneity in unobservables is also not statistically significant at that time. But from experience year two onwards, we observe significantly negative slopes revealing a clear pattern of selection into gains. The generally increasing slope parameters also detect increasing heterogeneity in monetary returns regarding resistance to treatment. The longer the individuals are in the labor market, the greater the differences in returns between individuals from different quantiles of U_D. Figure 6 illustrates these differences. The intercept of the linear MTE gives the monetary

Figure 4: Comparison of MTE results with different functional form assumptions



Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure shows the regression results of MTE estimations of Abitur on monthly labor earnings under different functional form assumptions. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. The semiparametric MTE is based on a local linear regression with a bandwidth of 0.12 using a semiparametric Robinson estimator (Robinson, 1988); it is only estimated within the common support.

Figure 5: Slope parameters of linear MTEs over experience

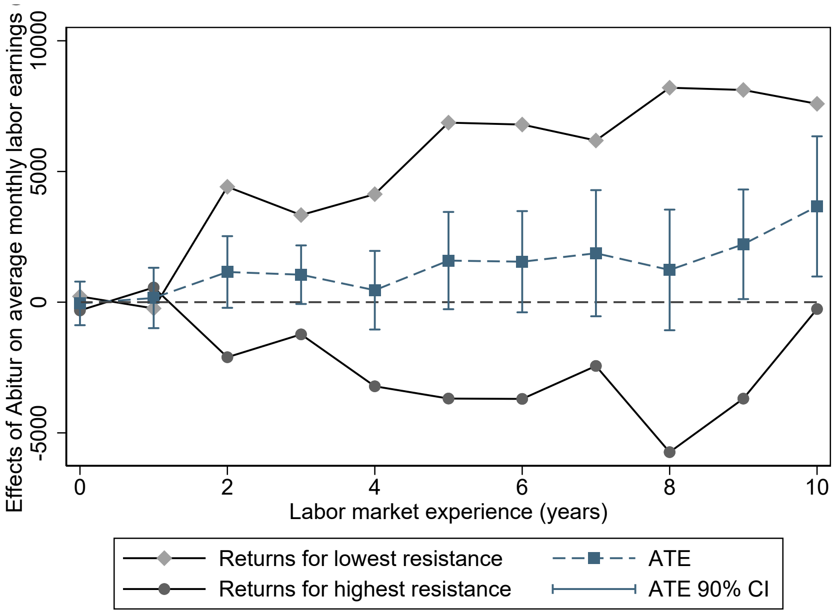


Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure presents the estimated slope parameters of linear MTE estimations of Abitur on monthly labor earnings over labor market experience. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. 90% confidence intervals are based on bootstrapped standard errors with 200 repetitions clustered at the municipality level.

returns to Abitur for those with the lowest resistance. The returns for the individuals with the highest resistances are calculated by adding the negative slope parameter to the intercept.[11] Thus, the slope parameters in figure 5 determine the distance between the returns for low- and high-resistance individuals in figure 6. Noticeably, with the inserting heterogeneity in year 2, the returns for the individuals with high resistance to academic track education become negative. Negative monetary returns to education for a part of the distribution have also been found in the previous literature (Carneiro et al., 2011, 2017; Nybom, 2017).

In the middle of the monetary returns to Abitur for individuals at the maximum and minimum of U_D, the average treatment effects (ATEs) are shown in blue. For most experience levels, the ATE is below the LATE, emphasizing the importance of effect heterogeneity between different subpopulations. Even though both effects are positive after one year in the labor market, contrary to the LATEs, the ATEs are only statistically significant from zero after nine years of working experience.

Figure 6: Effects of Abitur on earnings over experience by resistance to treatment



Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure shows monetary returns to academic track education for individuals with different resistances to treatment. Individuals with the highest resistance to academic track education have a propensity score of 0, and those with the lowest resistance have a propensity score of 1. Regressions include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. Average treatment effects derived from MTEs are marked in blue. 90% confidence intervals are based on bootstrapped standard errors with 200 repetitions clustered at the municipality level.

Again, as a robustness check, we compare our main results of the slope parameters over experience to MTE estimates with the binary instrument indicating whether an academic track school was in the residence municipality. We do not see any statistically significant differences between the slope parameters, and they are, except for one experience level,

[11]Note that the slope parameter is $2\hat{\gamma}$ and since U_D and p are bounded within the 0/1-interval, the marginal individual with the highest potential resistance to treatment have a propensity score of 1. Accordingly, the returns for the highest resistance to treatment is given by $(\hat{\eta}_1 - \hat{\eta}_0) + 2\hat{\lambda} \cdot 1$.

quite comparable in magnitude. The results of this comparison are included in figure A.7 in the Appendix. It allows for the conclusion that the previously presented results on the evolution of unobserved heterogeneity are robust to a change in the instrument.

4.4 Interpreting the evolution of unobserved heterogeneity in light of employer learning

We begin with the interpretation under the assumptions of the common employer learning model introduced by Farber and Gibbons (1996) before shortly discussing their validity and implications for our results. The central assumptions are: (i) one component of the individuals' productivity is time-invariant, (ii) employers set wages equal to the expected marginal productivity at any time, and (iii) employers learn about the time-invariant productivity component with the time individuals spend in the labor market. We refer to the time-invariant component as innate ability, denoted by A_i. This unobserved component affects selection into higher education and later wages (via higher productivity). Applying this to our MTE setting implies that innate abilities A_i strongly correlate with the unobserved characteristics included in V_i. To ease the interpretation in this chapter, we talk about individuals with low (high) abilities instead of individuals with high (low) resistance to treatment, i.e. individuals at a high (low) quantile of the V distribution. The original version of the model additionally implies that the experience component of productivity is common to all, thus unrelated to education and innate abilities.

Rejecting the existence of any significant monetary returns to Abitur already in the first two years after the labor market entry suggests that Abitur has no relevant signaling effect. However, this does not necessarily mean that there are no information asymmetries between employers and employees concerning the individuals' productivity. It only states that information on Abitur as a schooling degree does not help overcome these frictions. This result aligns with the diploma effects analyzed by Clark and Martorell (2014), finding signaling effects close to zero.

Employer learning can take place even without a signaling effect of the schooling degree. Employers may pay people with and without Abitur similarly at the beginning of their careers. Still, they might learn about the employee's unobserved abilities over time as they observe job performance and adjust their wages accordingly. Suppose there is such a process, it is straightforward to imply that it is reflected not only in direct wages but also in the returns to education. Our results show that the monetary returns to Abitur of high- and low-ability individuals tend to diverge more with increasing working experience. This increasing heterogeneity in the effects suggests that employers actually learn about the time-invariant productivity component as long as productivity growth is the same for high- and low-ability individuals. Of course, this implication is only valid for heterogeneous effects driven by selection into gains, i.e., if the detected heterogeneity is based on MTEs with a negative slope. The majority of the existing empirical literature uses an ability measure unobservable to employers for analyzing employer learning and provides similar results. Evidence for employer learning was found for the US and several other countries, at least for certain subgroups (see for example Altonji and Pierret, 2001; Lange, 2007; Broecke, 2015). [12] Bauer and Haisken-DeNew (2001) also build on the same approach but can not conclude that employer learning generally occurs in Germany. Their results even suggest that the assumption of productivity growth over experience, unrelated to schooling and innate abilities, does not hold.

Let us still rely on this assumption once more. Taking a closer look at the evolution of unobserved heterogeneity reveals not only that employers learn but that they learn relatively fast. Already after two years of labor market experience, there is a clear and sharp decrease in the slope parameter of the MTE. It decreases further but at a much smaller rate. Our results indicating fast learning, again, match the previous literature analyzing the speed of employer learning (Lange, 2007; Aryal et al., 2022).

Now, we want to discuss the implications for our interpretation if any of the assumptions are not fulfilled. Most doubtful is the assumption that the productivity growth with experience is the same for all individuals. Therefore, for example, Aryal et al. (2022) explicitly allow the experience component of productivity to vary for different skill levels, while others critically discuss their implications for the case the assumption does not hold. It is more likely that individuals with more education or higher innate abilities benefit more from working experience. Higher probabilities for on-the-job training is one potential driver. In our setting, a correlation between productivity growth and Abitur should not be a problem as long as it is the same along with innate abilities. The additional productivity growth (for people with compared to people without Abitur) is then simply captured in the returns to Abitur over experience. However, if individuals at higher quantiles of V have stronger productivity growth over working experience, the increasing heterogeneity could be driven solely by the increasing productivity differences between high- and low-ability individuals. Even under perfect information on productivity, the returns to education would diverge the longer the employees are in the labor market. Most likely, the increase in effect heterogeneity regarding unobserved abilities is, in fact, driven by a combination of both, employer learning about time-invariant abilities and productivity growth varying with those abilities. Nevertheless, without increasing unobserved heterogeneity, we could reject the employer learning hypotheses as there are no hints for any countervailing effects.

Our interpretation is based on the common or symmetric employer learning model, which means that every employer has the same information and sets wages equal to expected productivity at any time. First, this rules out any long-term contracts and collective

[12] For an overview of the empirical literature on employer learning, see Ablay and Lange (2023).

payment agreements. In Germany, most employees have unlimited contracts, typically including a payment scheme negotiated ex-ante. A collective bargaining system also regulates wages according to pay scales bargained, for example, by trade unions for whole industries. The assignment of employees to these pay scales is typically not directly based on any productivity measures. Taken together, this substantially hampers setting wages equal to expected productivity. However, productivity can still affect promotions into higher positions, which allows for wage adjustments based on productivity. Observing increasing heterogeneity in monetary returns along innate abilities is already a clear sign that there are any productivity-related payment mechanisms. Thus, the German circumstances do not rule out employer learning in a broader sense and do not change the interpretation of our results.

Second, learning is assumed to be symmetric. It might be reasonable to assume that current employers have better information than future employers. There has partly been empirical support for asymmetric employer learning, especially for higher-educated employees (for example, in Schönberg, 2007; Kahn, 2013; Ge et al., 2021). But job references or information about previous employers, tasks, and wages still allow information to flow across employers. Thus, asymmetric learning does not rule out common learning, but without the possibility of job changes, symmetric employer learning could be even faster. Our interpretation is still valid for symmetric learning as long as there are no strategic decisions of employers and employees based on asymmetric learning.

Last, the common model assumes competitive labor markets. If, contrarily, the employers have any market power, wages might fall below marginal productivity. As long as potential mark-downs are proportionally the same for all employees, it does not change our interpretation. Differences in mark-downs might result from varying elasticities of labor supply. Aryal et al. (2022) argue that education supports specialization and the supply of specialized labor might be less elastic. Potential differences in mark-downs driven by education, however, do not restrict the general interpretation of our findings. If the deviation between wages and productivity is smaller for individuals with than without Abitur, but this deviation is constant over time, the increase in heterogeneity still results in the same interpretation. Certainly, this is not true for interpreting the overall extent of heterogeneity.

5 Conclusion

In this paper, we estimate monetary returns to academic track education by exploiting variations in the availability of those schools induced by the educational expansion in West Germany. We use a high-quality data set linking representative survey data to administrative labor market data (NEPS-ADIAB). Combining this with a purpose-built data set on academic track school openings enables us to use measures of geographical access to higher secondary schooling to instrument the choice of academic track education. We analyze how returns evolve with increasing labor market experience and uncover heterogeneities in the effects regarding the resistance to education. Therefore, our paper is the first to analyze the evolution of unobserved heterogeneity in returns to education after individuals enter the labor market.

We find average returns to the highest German schooling degree (Abitur) of over 70% within the first 10 years after labor market entry, corresponding to returns of 14% per additional year of education. These positive returns first appear after two years of working experience. Lacking positive effects directly after labor market entry reveals a weak signaling value of this schooling degree.

We also document substantial heterogeneity in the returns to Abitur with respect to unobserved characteristics from year two onwards. Clear selection into gains shows that high secondary schooling education does not pay off for everybody. Individuals who most likely select academic track education benefit the most. With increasing experience, we even find negative returns for a part of the distribution. Although average effects are positive (yet mainly insignificant), further incentives for academic track education are less likely to pay off as they will attract people with lower returns.

The evolution of heterogeneity with respect to characteristics unobservable to employers gives further insights into employers' wage decision-making. With growing experience, the returns for individuals with the highest and lowest resistances to academic track education get more heterogeneous. We assume that employers adjust wages over time depending on productivity expectations. Employers might learn about innate abilities as a time-constant component of productivity. People with unobserved characteristics reducing resistance to high schooling might simultaneously have higher productivity returns over experience. We conclude that a combination of both most likely drives our findings on increasing heterogeneity. All in all, we do not need to reject the hypothesis of employer learning, albeit we do not clearly disentangle its impact.

References

Ablay, M. and Lange, F. (2023). Approaches to learn about employer learning. *Canadian Journal of Economics/Revue canadienne d'économique*, 56(2):343–356.

Altonji, J. G. and Pierret, C. R. (2001). Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics*, 116(1):313–350.

Aryal, G., Bhuller, M., and Lange, F. (2022). Signaling and Employer Learning with Instruments. *American Economic Review*, 112(5):1669–1702.

Bachbauer, N., Wolf, C., Graf, T., Grießemer, S., Kaimer, S., Köhler, M., Lehnert, C., Oertel, M., Schneider, A., and NEPS-Netzwerk (LIfBi) (2022). National Educational Panel Study (NEPS), Starting Cohort 6 (SC6) survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB) – Version 7520 v1. Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Bachbauer, N. and Wolf, C. B. (2022). NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB). FDZ-Datenreport, 01/2022 (en), Nuremberg.

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

Becker, R. (2006). *Dauerhafte Bildungsungleichheiten als unerwartete Folge der Bildungsexpansion?*, pages 27–61. VS Verlag für Sozialwissenschaften, Wiesbaden.

Bhuller, M., Mogstad, M., and Salvanes, K. G. (2017). Life-Cycle Earnings, Education Premiums, and Internal Rates of Return. *Journal of Labor Economics*, 35(4):993–1030.

Birkelund, J. F. and Werfhorst, H. G. v. d. (2022). Long-term labor market returns to upper secondary school track choice: Leveraging idiosyncratic variation in peers' choices. *Social Science Research*, 102:102629.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond late with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.

Broecke, S. (2015). Experience and the returns to education and skill in OECD countries.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69(5):1127–1160.

Carneiro, P., Heckman, J. J., and Vytlacil, E. (2011). Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–2781.

Carneiro, P., Lokshin, M., and Umapathi, N. (2017). Average and marginal returns to upper secondary schooling in indonesia. *Journal of Applied Econometrics*, 32(1):16–36.

Clark, D. and Del Bono, E. (2016). The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom. *American Economic Journal: Applied Economics*, 8(1):150–176.

Clark, D. and Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):282–318.

Dahrendorf, R. (1965). *Bildung ist Bürgerrecht*. Wegener, Hamburg.

Dauth, W. and Eppelsheimer, J. (2020). Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide. *Journal for Labour Market Research*, 54(1):1–14. Publisher: SpringerOpen.

Deming, D. J. (2022). Four Facts about Human Capital. *Journal of Economic Perspectives*, 36(3):75–102.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2017). The Long–Term Effects of Early Track Choice. *The Economic Journal*, 127(603):1348–1380.

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

Franzmann, G. (2006). The Development of the Educational System in Germany: Comprehensive Schools 1960–2000Bildungsentwicklung in Deutschland: Allgemeinbildende

Schulen 1960 bis 2000.

Ge, S., Moro, A., and Zhu, B. (2021). Testing for asymmetric employer learning and statistical discrimination. *Applied Economics*, 53(12):1361–1377.

Heckman, J. J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation1. *Econometrica*, 73(3):669–738.

Henz, U. and Maas, I. (1995). Chancengleichheit durch die bildungsexpansion. In Sahner, H. and Schwendtner, S., editors, *27. Kongreß der Deutschen Gesellschaft für Soziologie - Gesellschaften im Umbruch: Sektionen und Arbeitsgruppen*, pages 389–393. Westdt. Verl., Opladen.

Kahn, L. B. (2013). Asymmetric Information between Employers. *American Economic Journal: Applied Economics*, 5(4):165–205.

Kamhöfer, D. A., Schmitz, H., and Westphal, M. (2019). Heterogeneity in Marginal Non-Monetary Returns to Higher Education. *Journal of the European Economic Association*, 17(1):205–244.

Kamhöfer, D. A. and Westphal, M. (2019). *Fertility effects of college education: Evidence from the German educational expansion*. Number 316. DICE Discussion Paper.

Koebe, J. and Marcus, J. (2022). The Length of Schooling and the Timing of Family Formation. *CESifo Economic Studies*, 68(1):1–45.

Köhler, H. and Lundgreen, P. (2014). *Allgemeinbildende Schulen in der Bundesrepublik Deutschland: 1949-2010: ZA8570 Datenfile Version 1.0.0*. GESIS, Köln.

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

Matthewes, S. H. and Ventura, G. (2022). On track to success? CEPA Discussion Papers (58).

Nybom, M. (2017). The Distribution of Lifetime Earnings Returns to College. *Journal of Labor Economics*, 35(4):903–952.

Picht, G. (1964). *Die deutsche Bildungskatastrophe: Analyse und Dokumentation*. Walter.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Schönberg, U. (2007). Testing for Asymmetric Employer Learning. *Journal of Labor Economics*, 25(4):651–691.

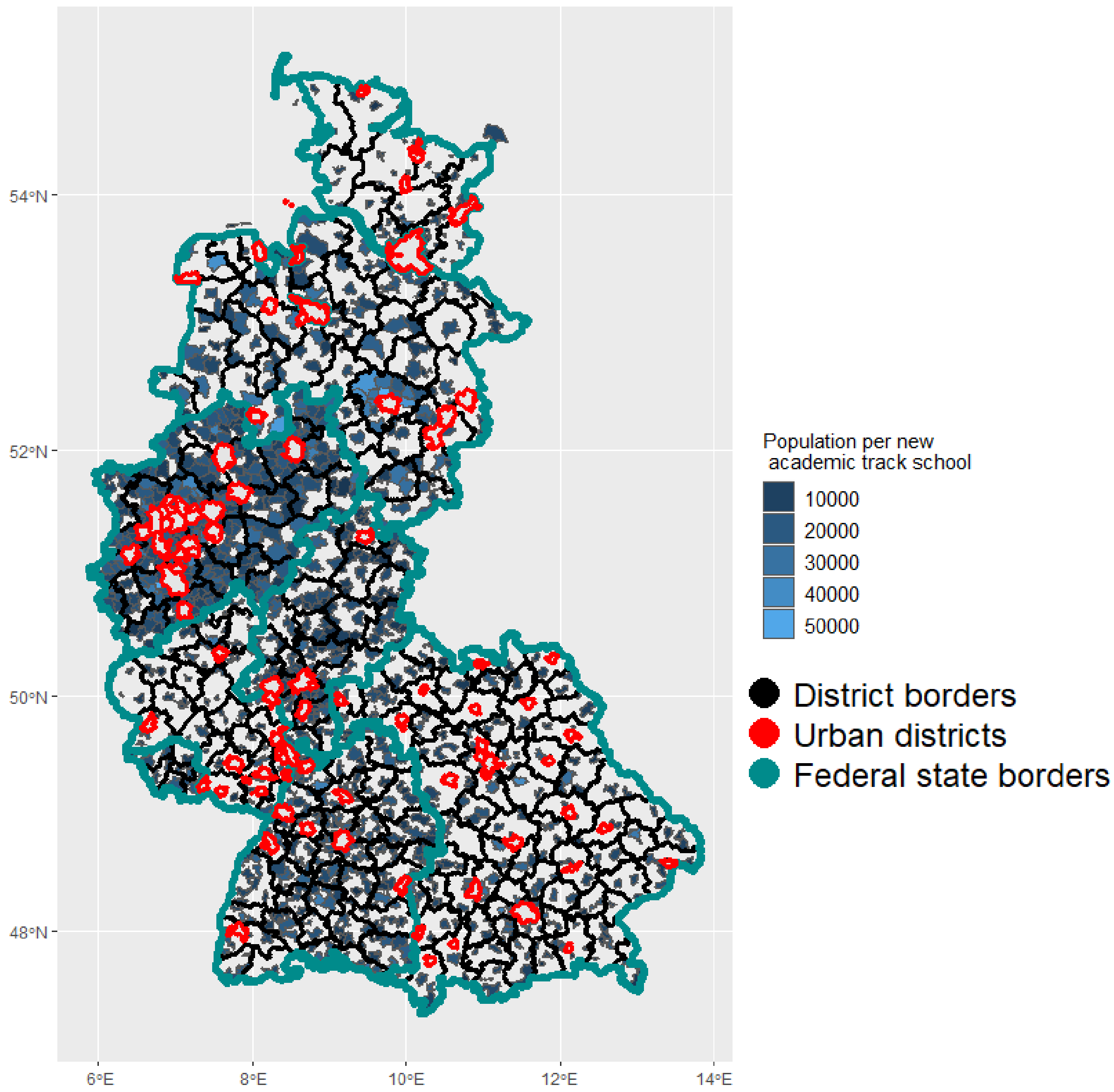
Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586. Publisher: Wiley, Econometric Society.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341.

Westphal, M., Kamhöfer, D. A., and Schmitz, H. (2022). Marginal College Wage Premiums Under Selection Into Employment. *The Economic Journal*, 132(646):2231–2272.

Appendix

Figure A.1: Academic tracks schools across West Germany



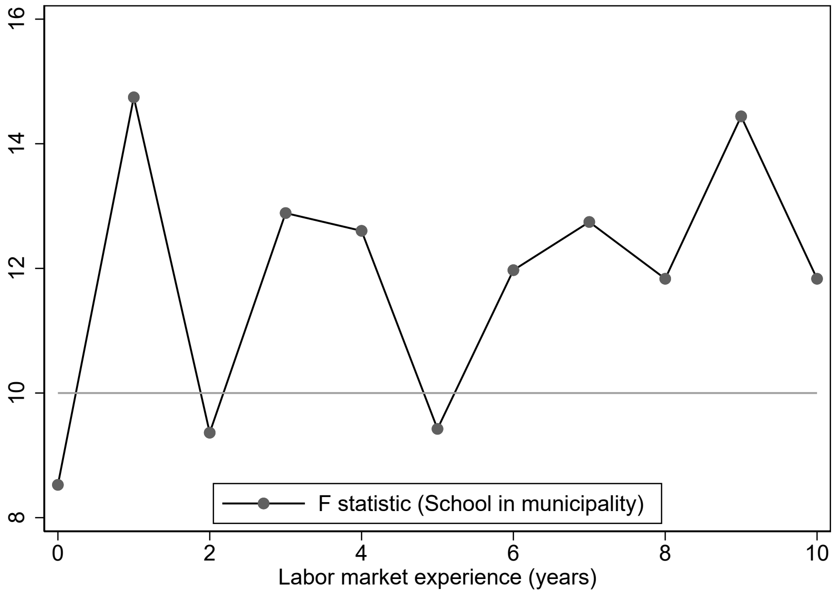
Notes: Own illustrations based on self-collected school information. The blue areas indicate the population in municipalities with an academic track school opening between 1955 and 1990. Borders of the federal states are marked in turquoise, usual district borders in black, and borders of the urban districts are highlighted in red.

Table A.1: Variables and means by academic track degree

	Definition	Without Abitur	With Abitur
Secondary school entry	Year individual entered secondary school	1975	1978
Month of birth	Month the individual was born	6	6
Residence information	=1 if childhood residence is based on primary school instead of birth location	0.506	0.396
Father born in Germany	=1 if father was born in Germany	0.923	0.916
Mother born in Germany	=1 if mother was born in Germany	0.944	0.940
Raised by single parent	=1 if raised by a single parent (from birth to age 15)	0.065	0.044
Raised by patchwork family	=1 if raised in a patchwork family (from birth to age 15)	0.052	0.021
Nr siblings	Number of siblings	1.995	1.458
Nr older siblings	Number of older siblings	1.430	1.178
Firstborn	=1 if individual was the firstborn child in the family	0.296	0.342
Age start working	Age (years) when individual first entered the labor market (without vocational training)	20.798	24.396
University degree	=1 if individual has a degree from a university	0.018	0.489
Years of education		13.103	16.535
Repeated year in primary school	=1 if individual repeated at least one year in primary school	0.010	.
Number of observations		2346	1125

Notes: Own calculations based on NEPS-SC6-ADIAB data. The table shows means of the variables for individuals with and without the academic track degree (Abitur). Values might be missing due to data protection rules.

Figure A.2: First stage F-statistic over experience



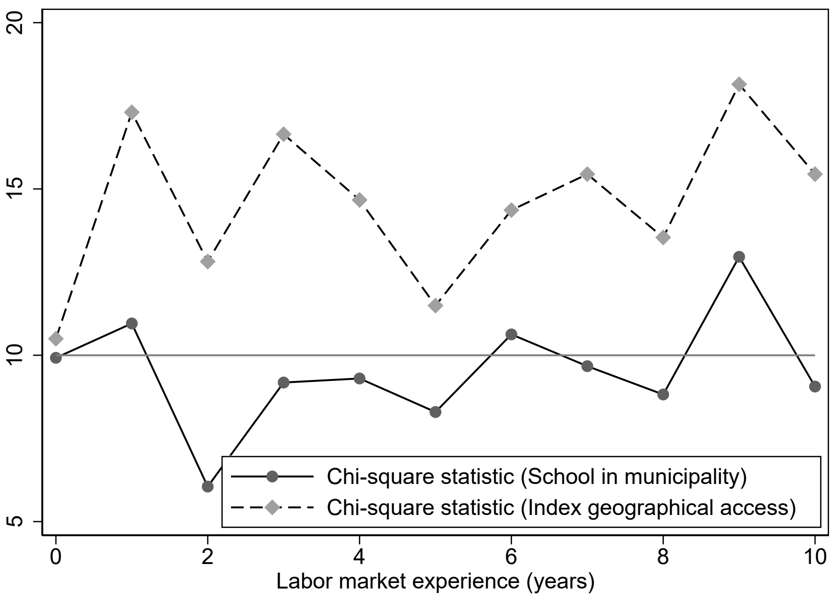
Notes: Own illustrations based on NEPS-SC6-ADIAB data. The figure reports the F-statistic from a linear first stage regression of the instrument on the treatment (Abitur) over labor market experience. The regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information, and state-specific trends.

Table A.2: Regression results for IV estimations with other dependent variables

	Outcome		
	Age start working	University degree	Years of education
IV			
Abitur	5.81***	0.55***	4.17***
	(2.01)	(0.18)	(0.99)
Baseline outcome w/o Abitur	20.31	0.00	13.10
N	3471	3469	3448

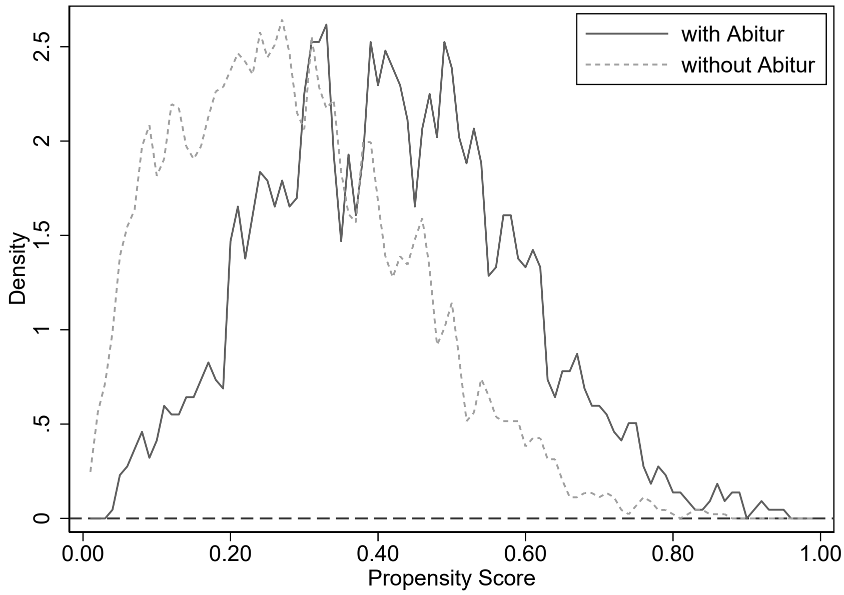
Notes: Own calculations based on NEPS-SC6-ADIAB data. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. Standard errors in parentheses are clustered on the municipality level. Baseline: Average outcome for individuals without Abitur. * ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.01$)

Figure A.3: Chi-squared statistics of a probit selection model over experience



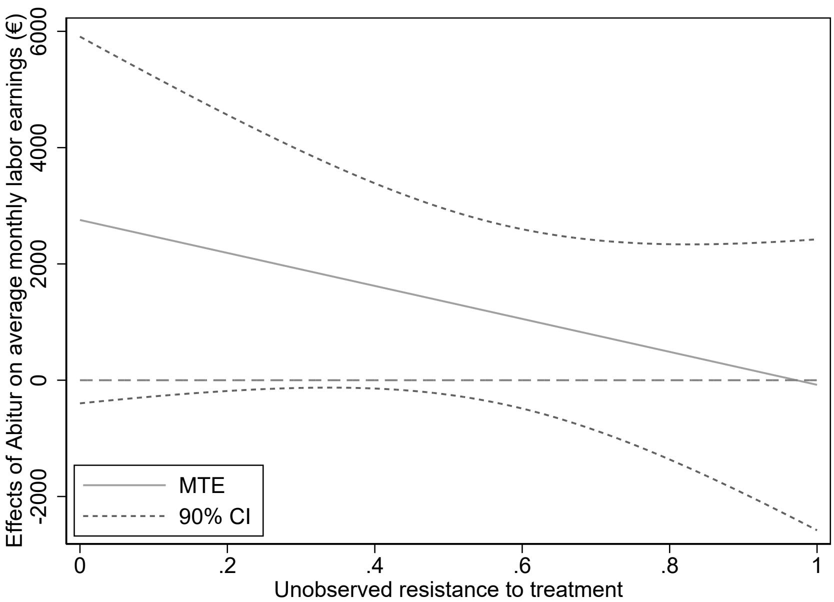
Notes: Own illustrations based on NEPS-SC6-ADIAB data. The figure reports the chi-squared statistics from a probit regression of the instrument on the treatment (Abitur) over labor market experience. The regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information, and state-specific trends.

Figure A.4: Support by treatment status



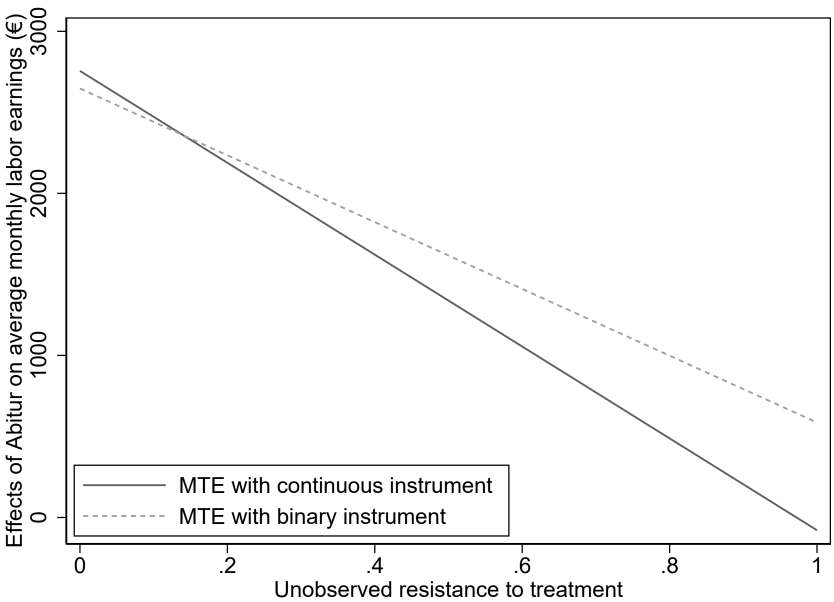
Notes: Own illustration based on NEPS-SC6-ADIAB data. This graph depicts the estimated density of the propensity score separately for individuals with and without Abitur.

Figure A.5: Linear MTE of Abitur on earnings for 0-10 years of working experience



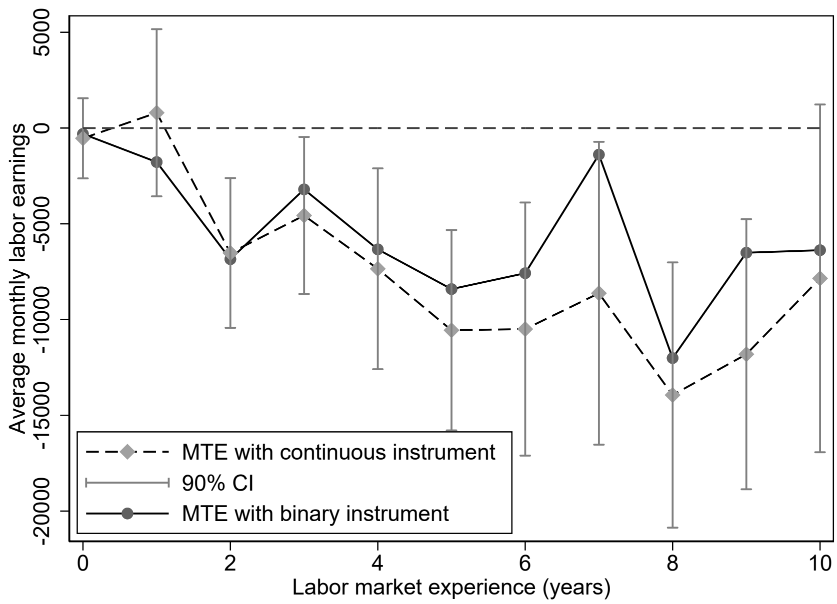
Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure presents the regression results of a linear MTE estimation of Abitur on monthly labor earnings. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends. The dashed lines give the 90% confidence interval based on bootstrapped standard errors with 200 repetitions clustered at the municipality level.

Figure A.6: Comparison of linear MTE results with different instruments



Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure presents the regression results of linear MTE estimations of Abitur on monthly labor earnings. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information and state-specific trends.

Figure A.7: Comparison of slope parameters of linear MTEs over working experience with different instruments



Notes: Own illustration based on NEPS-SC6-ADIAB data. The figure presents the estimated slope parameters of linear MTE estimations of Abitur on monthly labor earnings over labor market experience. Regressions also include district and entry cohort fixed-effects, birth month dummies, origin of residence information, and state-specific trends. 90% confidence intervals are based on bootstrapped standard errors with 200 repetitions clustered at the municipality level.